**Computer Science & Information Technology 127**

Dhinaharan Nagamalai
David C. Wyld (Eds)

# Computer Science & Information Technology

7th International Conference on Computer Science, Engineering and Information Technology (CSEIT 2020)
September 26 ~ 27, 2020, Copenhagen, Denmark

AIRCC Publishing Corporation

## Volume Editors

Dhinaharan Nagamalai,
Wireilla Net Solutions, Australia
E-mail: dhinthia@yahoo.com

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

ISSN: 2231 - 5403
ISBN: 978-1-925953-25-1
DOI: 10.5121/csit.2020.101101- 10.5121/csit.2020.101125

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

# Preface

The 7th International Conference on Computer Science, Engineering and Information Technology (CSEIT 2020) September 26 ~ 27, 2020, Copenhagen, Denmark, 12th International Conference on Wireless & Mobile Networks (WiMoNe 2020), 12th International Conference on Network and Communications Security (NCS 2020), 2nd International Conference on Internet of Things (CIoT 2020), 2nd International Conference on Machine Learning & Applications (CMLA 2020), International Conference on Data Mining and Software Engineering (DMSE 2020), International Conference on NLP & Big Data (NLPD 2020), 7th International Conference on Signal, Image Processing and Multimedia (SPM 2020) was collocated with 7th International Conference on Computer Science, Engineering and Information Technology (CSEIT 2020). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The CSEIT 2020, WiMoNe 2020, NCS 2020, CIoT 2020, CMLA 2020, DMSE 2020, NLPD 2020 and SPM 2020 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically.

In closing, CSEIT 2020, WiMoNe 2020, NCS 2020, CIoT 2020, CMLA 2020, DMSE 2020, NLPD 2020 and SPM 2020 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the CSEIT 2020, WiMoNe 2020, NCS 2020, CIoT 2020, CMLA 2020, DMSE 2020, NLPD 2020 and SPM 2020.

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

Dhinaharan Nagamalai
David C. Wyld (Eds)

| General Chair | Organization |
|---|---|
| Dhinaharan Nagamalai, | Wireilla Net Solutions, Australia |
| David C. Wyld, | Southeastern Louisiana University, USA |

### Program Committee Members

| | |
|---|---|
| Abdel-Badeeh M. Salem, | Ain Shams University, Egypt |
| Abdelhak Merizig, | University of Biskra, Algeria |
| Abdellatif BOUZID-DAHO, | Paris-Est Creteil University, Algeria |
| Abdellatif I. Moustafa, | Umm AL-Qura University, Saudi Arabia |
| Abdelouahab MOUSSAOUI, | Ferhat Abbas University, Algeria |
| Abderrahim Siam, | University of Khenchela, Algeria |
| Abhijit Bhowmick, | Vellore Institute of Technology, India |
| Abhishek Shukla, | R.D. Engineering College Technical Campus, India |
| Ahmad A. Saifan, | Yarmouk university, Jordan |
| Ahmed Farouk AbdelGawad, | Zagazig University, Egypt |
| Ajay Anil Gurjar, | Sipna College of Engineering & Technology, India |
| Akhil Gupta, | Lovely Professional University, India |
| Alexander Gelbukh, | Instituto Politécnico Nacional, Mexico |
| Ali Abdrhman Mohammed Ukasha, | Sebha University, Libya |
| Ali Adil Yassin, | Basra University, Iraq |
| Ali Asghar Aanvary Rostamy, | Tarbiat Modares University, Iran |
| Ali Kashif Bashir, | Manchester Metropolitan University, UK |
| Allel HADJALI, | LIAS/ENSMA, France |
| Amari Houda, | Networking & Telecom Engineering, Tunisia |
| Amizah Malip, | University of Malaya, Malaysia |
| Anand Nayyar, | Duy Tan University, Vietnam |
| Anita Dixit, | SDM college of Engineering Technology, India |
| Antonio Carlos Bento, | Anhembi Morumbi University, Brazil |
| Apu Kumar Saha, | National Institute of Technology Agartala, India |
| Armir Bujaria, | Studi di Padova, Italy |
| Ashutosh Kumar Dubey, | Chitkara University, India |
| Assem Mousa, | E Commerce Tech Support Systems Manager, Egypt |
| Ayush Singhal, | Contata Solutions, USA |
| Azeddine Wahbi, | Hassan II University, Morocco |
| B.K Verma, | Chandigarh Engineering College Landran, India |
| B.K.Tripathy, | VIT University, India |
| Baihua Li, | Loughborough University, UK |
| Benyamin Ahmadnia, | UC Davis, USA |
| Benyettou Mohamed, | University of Sciences and technology, Algeria |
| Bouchra Marzak, | Hassan II University, Morocco |
| Boukari Nassim, | Skikda University, Algeria |
| Chaman Lal Sabharwal, | Missouri University of Science and Technology, USA |
| Cherkaoui LEGHRIS, | University Hassan II of Casablanca, Morocco |
| Chin-Chih Chang, | Chung Hua University, Taiwan |
| Chiranjiv Chingangbam, | Manipur Technical University, India |
| Christian Mancas, | Ovidius University, Romania |

| | |
|---|---|
| M. Mirrashid, | Semnan University, Iran |
| M.A.Jabbar, | Vardhaman College of Engineering, India |
| M.Anuradha, | St.Joseph's College of Engineering, India |
| M.Suresh, | Kongu Engineering College, India |
| Mabroukah Amarif, | Sebha University, Libya |
| Mahendra B. Gawali, | Sanjivani College of Engineering, India |
| Malka N. Halgamuge, | The University of Melbourne, Australia |
| Mamoun Alazab, | Charles Darwin University, Australia |
| Mamun Bin Ibne Reaz, | Universiti Kebangsaan, Malaysia |
| Mario Henrique Souza Pardo, | University Of Sao Paulo, Brazil |
| Mayank Dave, | National Institute of Technology, India |
| MERIAH sidi Mohammed, | Univerity of Tlemcen, Algeria |
| Mihai Carabas, | University POLITEHNICA of Bucharest, Romania |
| Mirsaeid Hosseini Shirvani, | Islamic Azad University, Iran |
| Mo Sha, | National University of Singapore, Singapore |
| Mohamed Ismail Roushdy, | Ain Shams University, Egypt |
| Mohamed Yacoab, | University of Madras, India |
| Mohammad A. Alodat, | Sur University College, Oman |
| Mohammad Jafarabad, | Qom University, Iran |
| Mohammed B. M. Kamel, | Eotvos Lorand Univresity (ELTE), Hungary |
| Morris Riedel, | University of Iceland, Iceland |
| Morteza Alinia Ahandani, | University of Tabriz, Iran |
| Mourad Chabane Oussalah, | University of Nantes, France |
| Muhammad Hafeez Javed, | Southwest Jiaotong University, China |
| Muhammad Sajjad Ashfaq, | Eastern Mediterranean University, Cyprus |
| Muhammad Sajjadur Rahim, | University of Rajshahi, Bangladesh |
| Murali Manohar.B, | Vellore Institute of Technology, India |
| N.Syed Siraj Ahmed, | VIT University, India |
| Nadia Abd-Alsabour, | Cairo university, Egypt |
| Nahlah Shatnawi, | Yarmouk University, Jordan |
| Nidal Turab, | Al-ahliyya Amman University, Jordan |
| Nidhi Lal, | IIIT Nagpur, India |
| Nihar Athreyas, | CTO & Founding Member at Spero Devices, USA |
| Nikola Ivković, | University of Zagreb, Croatia |
| Nikolai Prokopyev, | Kazan Federal University, Russia |
| Nirmalya Thakur, | University of Cincinnati, USA |
| Nishant Doshi, | Pandit Deendayal Petroleum University, India |
| O. Ayhan ERDEM, | Gazi University, Turkey |
| Omid Mahdi Ebadati, | Kharazmi University, Iran |
| Osama Rababah, | The University of Jordan, Jordan |
| Osman Toker, | Yildiz Technical University, Turkey |
| P.Subashini, | Avinashilingam University, India |
| Parameshachari B D, | GSSSIET for Women, India |
| Pranita Mahajan, | Mumbai University, India |
| Punnoose A K, | Flare Speech Systems, India |
| Purnendu Pandey, | Bml Munjal University, India |
| R Senthil, | Shinas college of technology, Oman |
| R.Arthi, | SRM Institute of Science and Technology, India |
| R.Muthukkumar, | National Engineering College, India |
| Rabhat Mahanti, | University of New Brunswick, Canada |
| Ragab El Sehiemy, | Kafrelsheikh University, Egypt |
| Rahim Messaoud, | Yahia Farès University of Médéa, Algeria |

## Technically Sponsored by

**Computer Science & Information Technology Community (CSITC)**

**Artificial Intelligence Community (AIC)**

**Soft Computing Community (SCC)**

**Digital Signal & Image Processing Community (DSIPC)**

## Organized By

**Academy & Industry Research Collaboration Center (AIRCC)**

## TABLE OF CONTENTS

## 7th International Conference on Computer Science, Engineering and Information Technology (CSEIT 2020)

## 12th International Conference on Wireless & Mobile Networks (WiMoNe 2020)

## 12th International Conference on Network and Communications Security (NCS 2020)

## 2nd International Conference on Internet of Things (CIoT 2020)

## 2nd International Conference on Machine Learning & Applications (CMLA 2020)

## International Conference on Data Mining and Software Engineering (DMSE 2020)

## International Conference on NLP & Big Data (NLPD 2020)

## 7th International Conference on Signal, Image Processing and Multimedia (SPM 2020)

# Evaluating the impact of different types of crossover and selection methods on the convergence of 0/1 Knapsack using Genetic Algorithm

Waleed Bin Owais, Iyad W. J. Alkhazendar, and Dr.Mohammad Saleh

Department of Computer Science and Engineering,
Qatar University,Doha

**Abstract.** Genetic Algorithm is an evolutionary algorithm and a metaheuristic that was introduced to overcome the failure of gradient based method in solving the optimization and search problems. The purpose of this paper is to evaluate the impact on the convergence of Genetic Algorithm vis-a'-vis 0/1 knapsack. By keeping the number of generations and the initial population fixed, different crossover methods like one point crossover and two-point crossover were evaluated and juxtaposed with each other. In addition to this, the impact of different selection methods like rank-selection, roulette wheel and tournament selection were evaluated and compared. Our results indicate that convergence rate of combination of one point crossover with tournament selection, with respect to 0/1 knapsack problem that we considered, is the highest and thereby most efficient in solving 0/1 knapsack.

**Keywords:** Genetic, Crossover, Selection, Knapsack, Roulette, Tournament, Rank, Single Point, Two Point, Convergence

## 1 Introduction

A genetic algorithm can be defined as a search heuristic algorithm that is motivated by Darwin's theory of natural evolution. As mentioned, this is motivated by Darwinian Natural Section and then applies them to soft computing. In Darwinian Natural Section there are three key principles that need to be in place for evolution.[13]

- Hereditary: A procedure by which children obtain the characteristics of their parents.
- Variation: Their should be an element of variety in a population, that is, it should not be homogeneous throughout.
- Selection: A mechanism by which some members of the population have the opportunity to be the parents and pass down their genetic information and some do not. Also known as the survival of the fittest [1]. There are five phases associated with a genetic algorithm 1. Initial population 2. Fitness function 3. Selection 4. Crossover 5. Mutation [2]

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 01-16, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101101

## 2 DEFINITIONS

The paper in later section's uses some of the technical keywords that are related to the Genetic Algorithm so it is imperative to define such technical keywords.

- Initial Population: As illustrated in the Figure 1, in a genetic algorithm, the set of genes of an individual is characterized using a string of 1s and 0s is used. This can be referred to as coding. Genes can be defined as a single element . The genes then join together into a string to form a Chromosome (solution)[18].
- Fitness Function: The fitness function can be defined as a function that helps in determination of how fit an individual is and gives an inference about its ability to compete with other individuals. The fitness function usually assigns a fitness score to each individual [14].
- Crossover: Crossover is the most significant stage in a genetic algorithm. Two chromosomes are mated by choosing a point of crossover. The crossover in Genetic Algorithm creates new generation as the same is done in the process of natural mutation.. The resulting chromosomes are known as offspring's [3]
- Mutation: In some of the new offspring made, some of the genes are subjected to a mutation. In soft computing the purpose of mutation is to ensure that the solution is not stuck at local optima and the solution explores the entire search space in the pursuit of finding the global maxima or global minima [17].

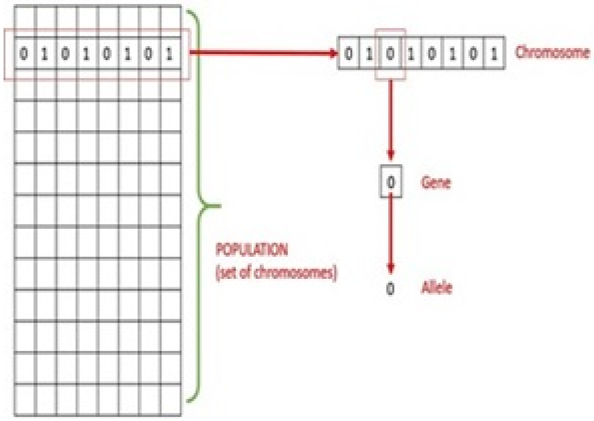


**Fig. 1.** Various Parameters

This paper will compare the result of various crossover techniques and selection methods in 0-1 knapsack. The knapsack problem deals with the idea of filling the knapsack with different items to maximize the profit without exceeding the net capacity of the knapsack. In other words, it is maximizing the profit while minimizing the cost. We will use a combination of most widely used crossover methods and selection methods and observe the results by doing a one to many mapping between the crossover methods and selection techniques.

## 3 BACKGROUND AND RELATED WORK

In the field of combinatorial optimization, knapsack problem can be defined as the process of finding an optimal solution from a finite set. The aim is to maximize the profit by including the items(each having a given weight and value) in the knapsack without exceeding the total capacity of the knapsack. 0/1 knapsack means that there is an additional constraint of either selecting the item in its entirety(1) or not selecting the item at all(0).
This can be explained as follows:
A set of finite items from 1 to n, each having a value of v and weight of w, and x being the number of copies of each item, whilst the total weight of knapsack being W, then based on 0/1 knapsack we have:

$$\sum_{i=1}^{n} wi * xi \tag{1}$$

Subjected to the following constraints:

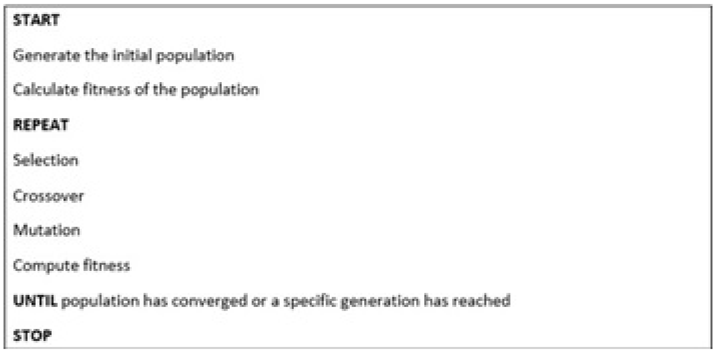
$$\sum_{i=1}^{n} wi * xi \leq W, xi \in 0, 1 \tag{2}$$

Knapsack problem has found a lot of applications in real world for example in making decisions related to investment banking, selection of project, and vote trading problem. In the research related to comparing the results of 0/1knapsack much of the research has been done in comparing the results of Genetic Algorithm with Greedy Approach , Branch and Bound and dynamic algorithm [3]-[5]. In another comparative study titled “ Comparative study of meta-heuristic algorithms using Knapsack Problem” [6] the authors have compared various meta-heuristic techniques to solve knapsack problem. Also [7] used chaotic crossover operator on Genetic Algorithm which produced improved results. In the research by [8] job scheduling problem(an application of 0/1 knapsack) was solved using Genetic Algorithm performance of various crossover techniques was presented. Similarly in [9] the research evaluated the impact of various crossover techniques on a web classifier. In [10] the authors have compared six crossover techniques and evaluated their performance on 0/1 knapsack. The experiments that followed, two point crossover showed the best results. In [11] and [12] the authors have evaluated the

performance of various selection techniques. In [15] the authors evaluated various algorithmic techniques used in optimization of 0/1 knapsack.
To the best of the knowledge of author's, none of the research hitherto has combined various selection and crossover methods and evaluated their performance on convergence of 0/1 knapsack using the Genetic Algorithm.

## 4 WORKING OF GENETIC ALGORITHM

This section describes the working of Genetic Algorithm
Step 1: Generating the initial population randomly.
Step 2: Calculating the fitness of the population.
Step 3: Selecting the fittest individuals based on fitness.
Step 4: Producing offspring's by crossover of selected chromosomes.
Step 5: Applying mutation.
Step 6: Go back to step 2 until termination condition is satisfied.



```
START
Generate the initial population
Calculate fitness of the population
REPEAT
Selection
Crossover
Mutation
Compute fitness
UNTIL population has converged or a specific generation has reached
STOP
```

**Fig. 2.** Psuedocode of Genetc Algorithm

## 5 CROSSOVER METHODS

As described crossover, also known as recombination is the most imperative process in the stages of Genetic Algorithm. It is in this phase that two parents exchange genetic information by choosing a single or multi point of crossover. Crossover operators divides a pair of selected chromosomes into two or more parts. After that the combination of chromosomes takes place to produce a new offspring (child). There are two types of crossover's that we have used:

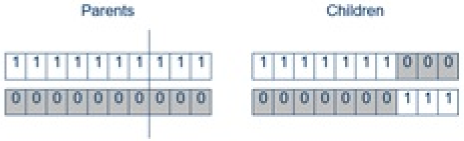
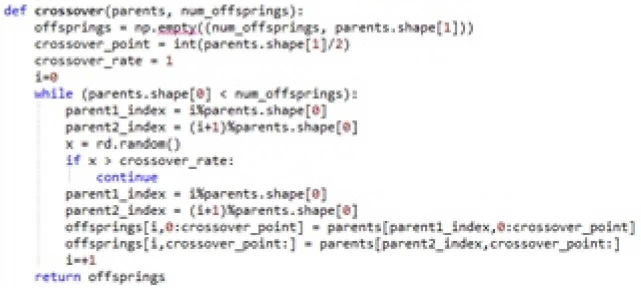


**Fig. 3.** Diagrammatic representation of Single Point Crossover

### 5.1 Single Point Crossover

In a single point crossover only one point is designated as a crossover location.After a random point is chosen, the parents slip at the crossover point and offspring's are created by exchanging tails. In the



```
def crossover(parents, num_offsprings):
    offsprings = np.empty((num_offsprings, parents.shape[1]))
    crossover_point = int(parents.shape[1]/2)
    crossover_rate = 1
    i=0
    while (parents.shape[0] < num_offsprings):
        parent1_index = i%parents.shape[0]
        parent2_index = (i+1)%parents.shape[0]
        x = rd.random()
        if x > crossover_rate:
            continue
        parent1_index = i%parents.shape[0]
        parent2_index = (i+1)%parents.shape[0]
        offsprings[i,0:crossover_point] = parents[parent1_index,0:crossover_point]
        offsprings[i,crossover_point:] = parents[parent2_index,crossover_point:]
        i=+1
    return offsprings
```

**Fig. 4.** Code Snippet of Single Point Crossover

Figure 4 as can be seen the 8th point is chosen as crossover point and the bits to the right side of the crossover point (111 and 000) are exchanged inter alia.

### 5.2 Two Point Crossover

In the two point crossover, the there are two points wherein the exchange of information takes place. The information between these two points is exchanged between parents to form offspring's.

As can be seen in Figure 5, the 2nd point and the 7th point are designated as the two points for the crossover. The genes between them are swapped (11111 and 00000) to create two new offspring's. Figure 6 is a snapshot of coding scheme used in Two Point crossover.

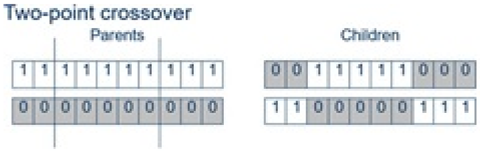
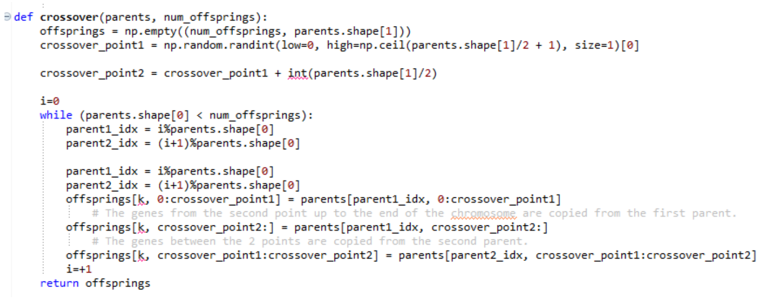


**Fig. 5.** Diagrammatic representation of Two Point Crossover



```
def crossover(parents, num_offsprings):
    offsprings = np.empty((num_offsprings, parents.shape[1]))
    crossover_point1 = np.random.randint(low=0, high=np.ceil(parents.shape[1]/2 + 1), size=1)[0]

    crossover_point2 = crossover_point1 + int(parents.shape[1]/2)

    i=0
    while (parents.shape[0] < num_offsprings):
        parent1_idx = i%parents.shape[0]
        parent2_idx = (i+1)%parents.shape[0]

        parent1_idx = i%parents.shape[0]
        parent2_idx = (i+1)%parents.shape[0]
        offsprings[k, 0:crossover_point1] = parents[parent1_idx, 0:crossover_point1]
            # The genes from the second point up to the end of the chromosome are copied from the first parent.
        offsprings[k, crossover_point2:] = parents[parent1_idx, crossover_point2:]
            # The genes between the 2 points are copied from the second parent.
        offsprings[k, crossover_point1:crossover_point2] = parents[parent2_idx, crossover_point1:crossover_point2]
        i=+1
    return offsprings
```

**Fig. 6.** Code Snippet of Two Point Crossover

## 6 SELECTION METHODS

In each generation, based on certain pre-determined criteria, only some of the chromosomes of the population are chosen to take part in the mating process, which is the crossover and mutation. This filtering of the population based on the fitness function is known as selection. The purpose of selection is to choose only those chromosomes who satisfy a specific criteria with regards to the fitness of the chromosome.

Such filtering of the population ensures that only healthy individuals are promoted to the next generation and the unhealthy ones (who do not satisfy a criteria) are left behind.

For instance, consider the solution set of a 0/1 knapsack. Any subset of population that has Formula(1), will not be considered for selection, because it doesn't satisfy the fitness function, because the weight exceeds the total weight of the knapsack. Fitness function is the criteria that is used to filter chromosomes in the selection step. Not only does the selection do the filtering of the fit chromosomes from the unfit chromosomes,it also helps in arranging the chromosomes based on their fitness.

For instance in the max one problem, the chromosome's are arranged in a hierarchy with the most fit( the chromosome having maximum 1's )at the top. It should be noted that in the process of selection, filtering based on the fitness function doesn't always take place, as can be seen in the case of maxone problem, because all the chromosomes satisfy the fitness function. In such cases the ordering of chromosome's based on their fitness takes place. There are a lot of selection methods that are used in the Genetic Algorithm. We have used the following :

### 6.1 Rank Selection

In the method of rank selection, after assignation of fitness to each individual, they are arranged in decreasing order of rank, in other words, the most fit individual gets rank one and so on. After each individual is assigned a rank, the chromosomes have a chance to get selected.The probability of an individual getting selected is given by the formula:

$$\rho(i) = rank(i)/n * (n-1) \tag{3}$$

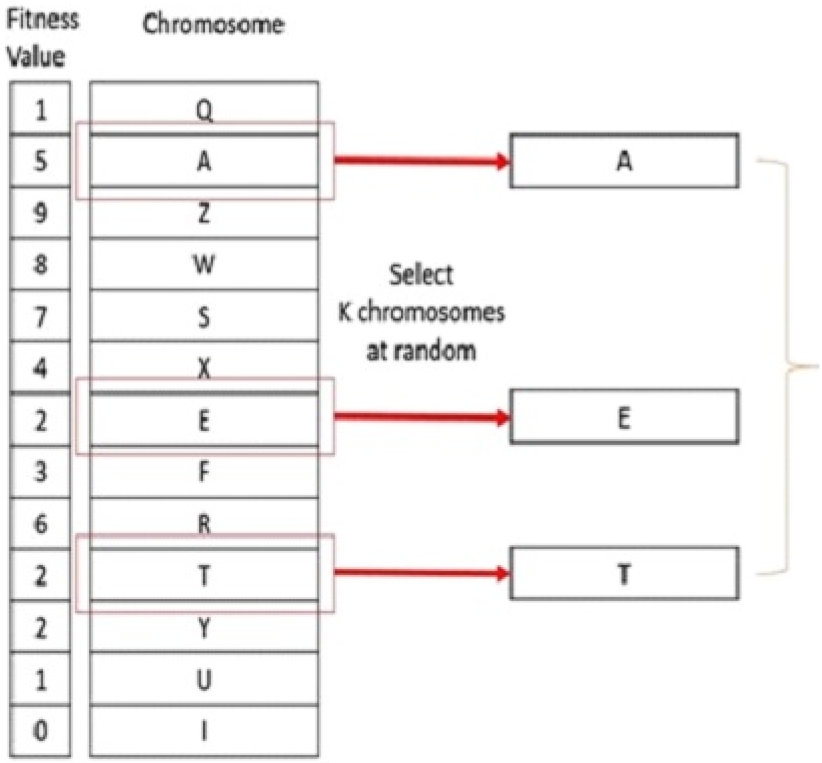
where p is the probability of individual i and n is the total number of individuals competing.
For instance if chromosome's 1 through 5 have a fitness of 37,6,36,30, and 28 respectively, then on the basis of rank selection, the individuals will be ranked as 1,3,4,5,2 based on their fitness. Chromosome 1 has the first rank and chromosome 2 has the last rank.

### 6.2 Roulette Wheel Selection

Roulette wheel selection, also known as Fitness Proportionate Selection, is another widely used selection method in Genetic Algorithm. After each individual is assigned the fitness score via fitness function, the roulette wheel determines the selection. The higher the fitness of an individual, the higher the chances will be to get selected. The process in roulette selection does a linear search through a roulette wheel where each individual gets a share in the roulette wheel. That is, higher the fitness, higher will be the share in the wheel, and thereby higher chance of to be selected when the wheel spins. Weaker individuals, having less share in the wheel, have very less probability of getting selected.
The probability of an individual to get selected via roulette wheel selection is given by the formula:

$$\rho(i) = f(i)/\sum_{j=1}^{n} f(j) \tag{4}$$

where p is the probability of individual i, f is the fitness of individual i and f(j) is the total fitness of the population. In regards to the simplicity and easiness of implementation, the roulette wheel selection is the most preferred method of selection.



| Fitness Value | Chromosome |
|---|---|
| 1 | Q |
| 5 | A |
| 9 | Z |
| 8 | W |
| 7 | S |
| 4 | X |
| 2 | E |
| 3 | F |
| 6 | R |
| 2 | T |
| 2 | Y |
| 1 | U |
| 0 | I |

**Fig. 7.** Diagrammatic representation of Tournament Selection

### 6.3 Tournament Selection

In this selection method,shown in Figure 7, the individuals contend against each other. The one with the highest fitness apparently wins the tournament and is selected for the subsequent generation. Weak individuals(one having low fitness) have less chances to be selected. The number of the chromosomes that contend against each other is termed as tournament size. The default tournament size is 3. It should be noted that in tournament selection, every chromosome is given an equal chance to compete.

## 7 Results

In the implementation of various crossover techniques and selection methods, we used Python 3.7.4, in the PyDev module and implemented Genetic Algorithm on the following:
0/1 knapsack which has the following characteristics:
Weight = [2, 3, 6, 7, 5, 9, 4, 5, 2, 3, 4, 1, 7, 8, 4, 5, 3]
Value = [6, 5, 8, 9, 6, 7, 3, 7, 4, 2, 5, 8, 3, 1, 5, 2, 8]
Knapsack threshold = 29
There are a total of 17 items that can be chosen as 1 or left over as 0, and the respective weights and values are given. The aim is to maximize the profit of the knapsack without breaking the knapsack, that is the net weight should be less than or equal to Knapsack threshold. The genetic algorithm ran 20 times for each scenario that will be discussed fore with. The following parameters were constant throughout the experiment.
Number of generations= 50
Solutions per population =8(No. of chromosome's within each population )
Mutation Rate=0.4
Type of mutation=Bit flip mutation
Crossover Rate=0.8
The Genetic Algorithm was applied in the following scenarios. Although attributed to the 'No Free Lunch Theorem' [16], there is no best crossover or selection technique but the authors chose the following selection methods and crossover techniques as they have been most widely used in research done thus far and shown interesting results.

### 7.1 Scenario 1

As shown in figure 8 ,in this scenario the type of the crossover is one point whilst the selection method is rank selection. The remaining parameters defined above remain constant. The results show a spike towards the optimal solution from the 15th generation onwards. The optima was stuck at 43 for about 9 generations.

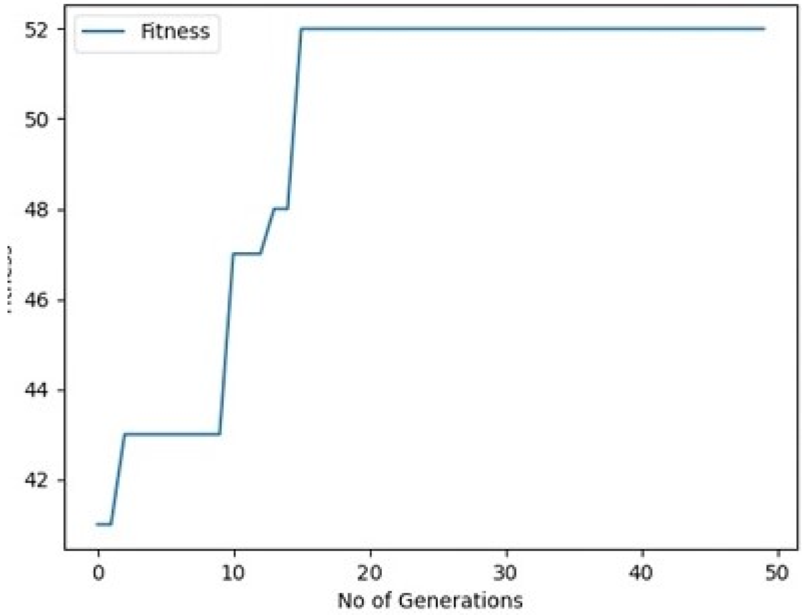


**Fig. 8.** Combination of OnePoint Crossover with Rank Selection

### 7.2 Scenario 2

As shown in figure 9, In this scenario the type of the crossover is one point whilst the selection method is roulette wheel selection. The remaining parameters defined above remain constant.

As is evident from the figure, this combination is quite slow as it does not converge to optimal solution in the fixed 50 generations. There can also be seen a trend of fitness increasing and decreasing till the 20th generation, and then it gets stuck at local optima of 43 for about 25 generations. The results indicate that the combination of One point crossover with roulette wheel selection is slow to converge to optimal solution.

### 7.3 Scenario 3

As shown in figure 10,in this scenario, the type of the crossover is one point whilst the selection method is tournament selection. The remaining parameters defined above remain constant.

The results indicate that the combination of one point crossover with tournament

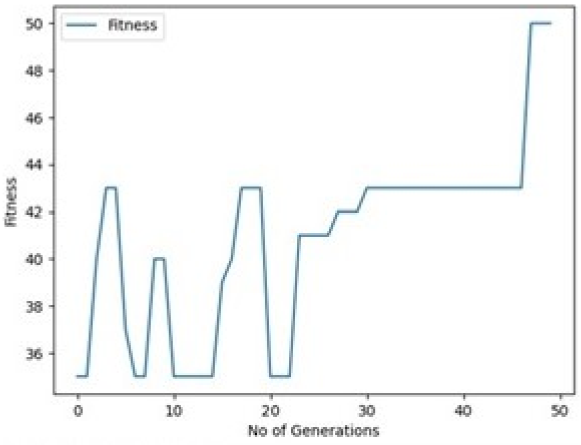


**Fig. 9.** Combination of OnePoint Crossover with Roulette Wheel Selection

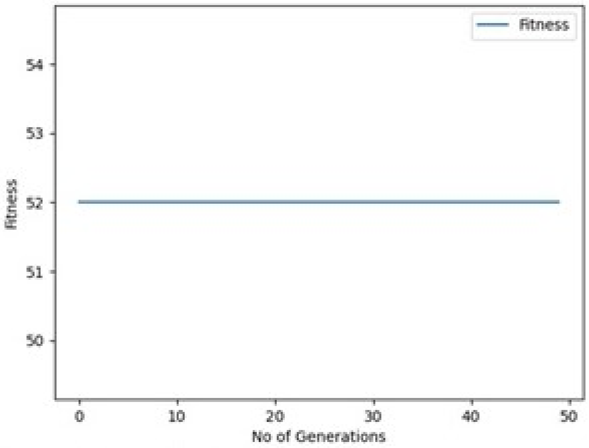


**Fig. 10.** Combination of OnePoint Crossover with Tournament Selection

selection converges as quickly as 1st generation and remains constant till the end of the 50th generation.

### 7.4 Scenario 4

As shown in figure 11,in this scenario we have combined two point crossover with the rank selection method.The remaining parameters defined above remain constant. As is evident from the figure the solution is stuck at local optima of 40 for about 8 generations, where it shows a steep increase to 48,and after getting stuck at 48,it gives the optimal solution from the 15th generation. No much variation is seen, the optima changes just two values before converging at global optima.

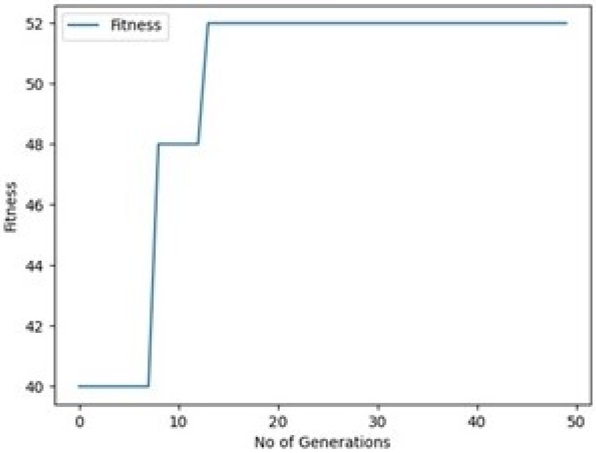


**Fig. 11.** Combination of TwoPoint Crossover with Rank Selection

### 7.5 Scenario 5

As shown in figure 12,in this we have combined two point crossover with roulette wheel selection method. The remaining parameters defined above remain constant. The results do not show a healthy trend, first it doesn't converge to the optimal solution of 52 in the fixed 50 generations, and, as is evident from the figure after getting stuck at local optima of 45 there is a decrease in fitness from 25th generation onwards where the fitness decreases to 39 and again increases to 45 subsequently. Results indicate that convergence of this combination is very slow.

### 7.6 Scenario 6

As shown in figure 13, in this we have combined two point crossover with tourna-

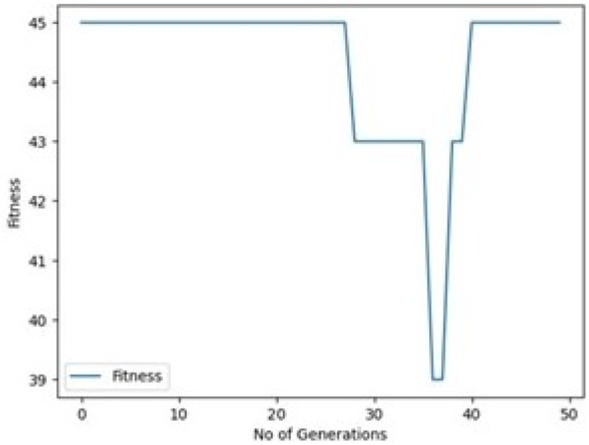


**Fig. 12.** Combination of TwoPoint Crossover with Roulette Selection

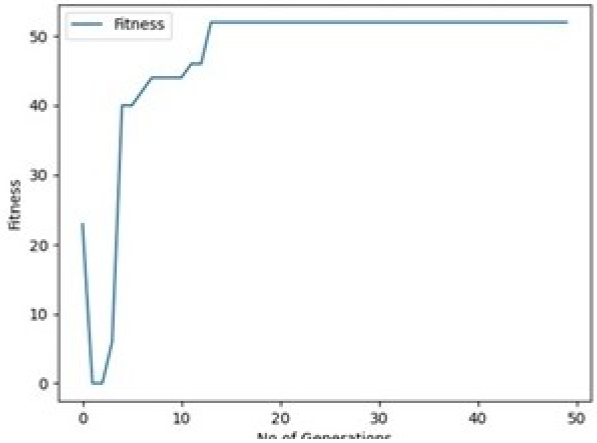


**Fig. 13.** Combination of TwoPoint Crossover with Tournament Selection

ment selection method. The remaining parameters defined above remain constant. The results indicate that optimal solution converged to 52 at the beginning of the 20th generation. Initially the optimum showed a downward trend, where the fitness even fell to 0.

## 8 Discussion

In the six combinations that we tested, all the combinations converged except for the combination of two point crossover method with the roulette wheel selection method. Although this combination will converge too if we increase the number of generations. Since we used 50 generations as baseline in all of the experiments, we can say that the convergence of this combination is slow. Also the combination of one point crossover with roulette wheel selection doesn't converge to optimal solution in 50 generations, but it is faster than the combination of two point crossover method with the roulette wheel selection method, as it converges to 50 as against to 45 of latter.

On the contrary, the results show that the convergence rate of combination of one point crossover with tournament selection converges as quickly as 1st generation and remains constant until the end of the 50th generation. The optimal solution of 52 is achieved in the first-generation itself. In the six combinations, the combination of one point crossover with tournament selection remains the fastest with respect to the 0/1 knapsack problem that we considered. The other combination methods converge to the optimal solution of 52, but we conclude that convergence rate of combination of one point crossover with tournament selection, with respect to 0/1 knapsack problem that we considered, is the highest and thereby most efficient in solving 0/1 knapsack.

In future endeavors we would like to test this on a randomly chosen sample space of weight and values and adjust the knapsack threshold accordingly.It would also be interesting to combine Genetic Algorithm with other meta heuristic algorithms and gauge the impact on the convergence of 0/1 knapsack thereof and test whether it has any substantial impact on the reduction of number of iterations and time of convergence. Another possible future work would be evaluating the impact of various types of mutation vis-a-vis the scenarios aforementioned.

The properties of the computer used in experimentation are Intel(R) Core(TM) i7-4600U CPU @ 2.10 GHz 2.69 GHz, and 8GB RAM with x64 based processor.The algorithm is written in Python 3.7 in the PyDev module of Eclipse.

## 9 Conclusion

The knapsack problems have a wide variety of applications in real world like cargo loading, budgeting, project management et.al. In our paper we combined different types of crossover methods with different selection techniques and evaluated their rate of convergence. We concluded that the combination of one point crossover with tournament selection is the most efficient. We also have discussed some of the future directions that the authors would like to work on.

## References

1. A. Gad, “Introduction to optimization with genetic algorithm,” 2018.
2. D. R. Bhattacharjya, “Introduction to genetic algorithms,” 2013.
3. X. Pan and T. Zhang, “Comparison and analysis of algorithms for the 0/1 knapsack problem,” in Journal of Physics: Conference Series, vol. 1069, no. 1. IOP Publishing, 2018, p. 012024
4. A. Shaheen and A. Sleit, “Comparing between different approaches to solve the 0/1 knapsack problem,” International Journal of Computer Science and Network Security (IJCSNS), vol. 16, no. 7, p. 1, 2016
5. R. K. Yadav, H. Gupta, H. Jhingran, A. Agarwal, and A. Gupta, “An enhanced genetic algorithm to solve 0/1 knapsack problem,” Interna- tional Journal of Computer Science and Information Security (IJCSIS), vol. 15, no. 5, 2017.
6. D. Sapra, R. Sharma, and A. P. Agarwal, “Comparative study of metaheuristic algorithms using knapsack problem,” in 2017 7th Interna- tional Conference on Cloud Computing, Data Science Engineering- Confluence. IEEE, 2017, pp. 134–137.
7. H. Demirci, A. Ozcerit, H. Ekiz, and A. Kutlu, “Chaotic crossover operator on genetic algorithm,” in Proceedings of 2nd International Conference on Information Technology, 2015.
8. T. Kellegz, B. Toklu, and J. Wilson, “Comparing efficiencies of genetic crossover operators for one machine total weighted tardiness problem,” Applied Mathematics and Computation, vol. 199, no. 2, pp. 590–598, 2008.
9. W. Chinnasri, S. Krootjohn, and N. Sureerattanan, “Performance com- parison of genetic algorithm’s crossover operators on university course timetabling problem,” in 2012 8th International Conference on Comput- ing Technology and Information Management (NCM and ICNIT), vol. 2. IEEE, 2012, pp. 781–786.
10. G. A. E.-N. A. Said, A. M. Mahmoud, and E.-S. M. El-Horbaty, “A comparative study of metaheuristic algorithms for solving quadratic assignment problem,” arXiv preprint arXiv:1407.4863, 2014.
11. Thada, Vikas, and Shivali Dhaka. ”Genetic Algorithm based Approach to Solve Non Fractional (0/1) Knapsack Optimization Problem.” International Journal of Computer Applications 100.15 (2014): 21-26.
12. Guler, A., M. E. Berberler, and U. G. Nuriyev. ”A new genetic algorithm for the 0-1 knapsack problem.” Acad Platf J Eng Sci 4.3 (2016): 9-14.
13. Changdar, Chiranjit, G. S. Mahapatra, and Rajat Kumar Pal. ”An improved genetic algorithm based approach to solve constrained knapsack problem in fuzzy environment.” Expert Systems with Applications 42.4 (2015): 2276-2286.
14. McCall, John. ”Genetic algorithms for modelling and optimisation.” Journal of computational and Applied Mathematics 184.1 (2005): 205-222.
15. Ezugwu, Absalom E., et al. ”A Comparative study of meta-heuristic optimization algorithms for 0–1 knapsack problem: Some initial results.” IEEE Access 7 (2019): 43979-44001.
16. Ho, Yu-Chi, and David L. Pepyne. ”Simple explanation of the no-free-lunch theorem and its implications.” Journal of optimization theory and applications 115.3 (2002): 549-570.
17. De Falco, Ivan, Antonio Della Cioppa, and Ernesto Tarantino. ”Mutation-based genetic algorithm: performance evaluation.” Applied Soft Computing 1.4 (2002): 285-299.
18. Maaranen, Heikki, Kaisa Miettinen, and Marko M. Mäkelä. ”Quasi-random initial population for genetic algorithms.” Computers Mathematics with Applications 47.12 (2004): 1885-1895.

## Authors

**Waleed Bin Owais** received his Bachelor's of Technology from BGSBU, India, and is currently pursuing his MS at Qatar University, Doha. His research interests include Virtual Reality, Empathy, Nuerofeedback and Algorithms.

**Iyad W. J. Alkhazendar** received Bachelor's degree in computer science from Al-Azhar University, Palestine and is currently pursuing his MS at Qatar University, Doha. His research interests include Network security, communication, Industrial IOT, SCADA , and Algorithms.

**Dr. Mohammad Saleh** received his PhD, Computer Science(Computer Networking) from University Putra Malaysia (UPM) and is currently Associate Professor in Department of Computer Science and Engineering, College of Engineering, Qatar University. His research interests include Simulations and modeling, OO Metrics.

# A NOVEL MOBILE ECG SENSOR WITH WIRELESS POWER TRANSMISSION FOR REMOTE HEALTH MONITORING

Jin-Chul Heo, Eun-Bin Park, Chan-Il Kim,
Hee-Joon Park and Jong-Ha Lee

Department of Biomedical Engineering, School of Medicine,
Keimyung University, Daegu, Korea

### *ABSTRACT*

*For electromagnetic induction wireless power transmission using an elliptical receiving coil, we investigated changes in magnetic field distribution and power transmission efficiency due to changes in the position of the transmitting and receiving coils. The simulation results using the high-frequency structure simulator were compared with the actual measurement results. It has been shown that even if the alignment between the transmitting coil and the receiving coil is changed to some extent, the transmission efficiency on the simulator can be maintained relatively stable. The transmission efficiency showed the maximum when the center of the receiving coil was perfectly aligned with the center of the transmitting coil. Although the reduction in efficiency was small when the center of the receiving coil was within ± 10 mm from the center of the transmitting coil, it was found that the efficiency was greatly reduced when the receiving coil deviated by more than 10 mm. Accordingly, it has been found that even if the perfect alignment is not maintained, the performance of the wireless power transmission system is not significantly reduced. When the center of the receiving coil is perfectly aligned with the center of the transmitting coil, the transmission efficiency is maximum, and even if the alignment is slightly changed, the performance of wireless power transmission maintains a certain level. This result proposes a standardized wireless transmission application method in the use of wireless power for implantable sensors.*

### *KEYWORDS*

*ECG, Implantable sensors, Simulation, Power transmission efficiency, Wireless power transmission*

## 1. INTRODUCTION

Wireless power transmission devices have been applied to many medical devices that can be inserted into the human body Inductively coupled wireless power transmission using electromagnetic induction is increasingly being applied to medical electronic devices[1-3]. There is an increasing demand for small electronic devices such as microneural stimulators, cochlear implants, and pacemakers[4-6]. As the size of the device decreases, there is a need to minimize the volume by reducing the size of the coil for power transmission. Therefore, realizing an efficient coil structure within a limited area is a very important task in wireless power transmission[7, 8].

Although various wireless power transmission technologies have been developed so far, they have not yet been commercialized except for some non-contact induction coupling methods. In

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 17-27, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101102

the past, some studies have been made to use microwaves, such as 5.8 GHz, to transmit large powers of several tens of watts or more, but they are not actively commercialized due to the influence on the human body and the directivity due to the use of high efficiency antennas[9-11]. Particularly, the wireless transmission system that can be used for human body has little progress in development due to the problem of safety due to electromagnetic waves of wireless transmission. Wireless charging using magnetic induction phenomenon is performed in a short distance of a few millimeters, and it is possible to use a small-sized device of 3 W or less, and it is known that it can be applied to a human body by use of relatively low energy. However, the charging efficiency is extremely low due to a short receiving distance and a large amount of heat[12, 13].

The purpose of this study is to develop a wireless sensor that can be used in human body and applied it to wireless electrocardiogram. The wireless ECG sensor is a power supply method by the wireless power transmission technology through electromagnetic induction between a pair of coils. The transmitting coil and the receiving coil form a pair, and the electromotive force induced in the receiving coil by the magnetic field generated when the current is supplied to the transmitting coil supplies the DC power to the circuit connected to the receiving coil. In this process, we will evaluate the theoretical considerations of wireless power transmission system and the power transmission characteristics through simulation[14].

In this study, the theoretical considerations of the wireless power transmission system and the power transmission characteristics were evaluated through simulation.

## 2. MATERIAL AND METHOD

### 2.1. System mode

The magnetic field analysis is determined by the size and shape of the transmitting and receiving coils. Induction current and voltage are evaluated according to the gap between the transmitting coil and the receiving coil. Computer analysis simulations were performed using the finite element method (FEM).

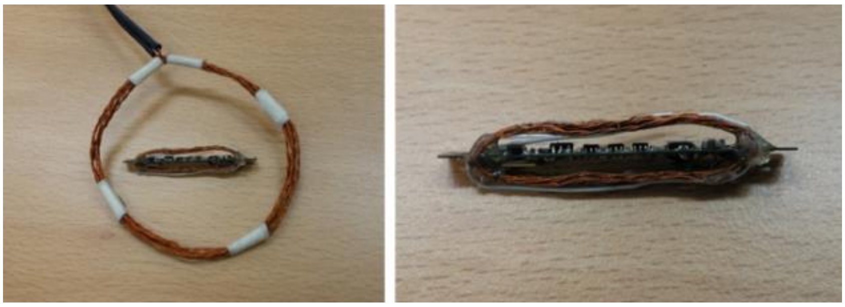


Figure 1. Round transmitting coil (Tx coil, left) and elliptical receiving coil (Rx coil, right).

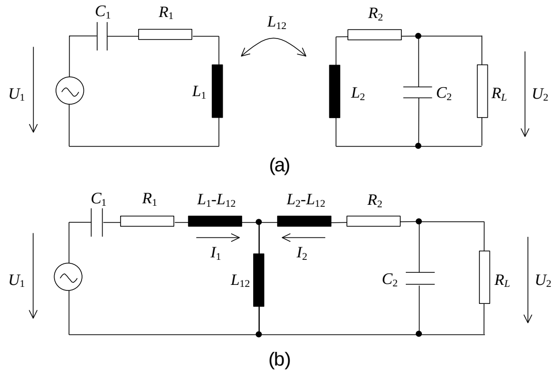


Figure 2. Wireless power transmission system. The circuit in which the transmitting coil (L1) and the receiving coil (L2) are combined is the equivalent circuit of the transmitting and receiving coils (L1-L2).

Transmitting coil (L1) and receiving coil (L2) use capacitance for L-C resonance to simulate electromagnetic induction in wireless power transmission. Capacitance C1 was connected in series to L1, and capacitance C2 was connected in parallel to L2. The total resistance of the transmit and receive coils and circuits is R1 and R2 (Figure 1 and 2). The two circuits that are magnetically coupled appear as a coupling factor that normalizes the mutual inductance or the mutual inductance for each coil. The delivered power can be calculated by the power delivery efficiency and the load of the receiving circuit.

$$k = \frac{L_{12}}{\sqrt{L_1 L_2}}$$

In general, the power transfer efficiency that is widely used when expressing the power transfer performance is defined as follows.

$$\eta = \frac{P_L}{P_S}$$

At this time, the power PL transmitted to the load of the receiving circuit can be obtained as follows.

$$P_L = \frac{V_{pk}^2}{2R_L}$$

## 2.2. FEM modelling

ANSYS Company's High Frequency Structure Simulator (HFSS) was used for finite element analysis of electromagnetic fields. HFSS can be used to simulate electromagnetic fields and electronic circuits in all frequency domains. Therefore, the electromagnetic field generated by the coil can be simulated even if the coil for wireless power transmission is connected to the L-C resonant circuit. The simulation shows the phenomenon that the transmitting and receiving coils resonate at 13.56 MHz, and the degree of activity of the magnetic field formed inside the transmitting and receiving coils. In addition, the evaluation of the voltage and power of the receiving coil is confirmed by the distance between the transmitting coil and the receiving coil, and the alignment between the transmitting coil and the receiving coil.

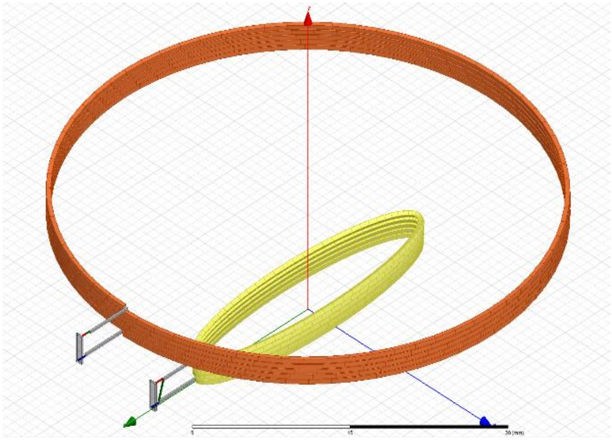


Figure 3. Modeling the transmitting coil (orange) and receiving coil (yellow).

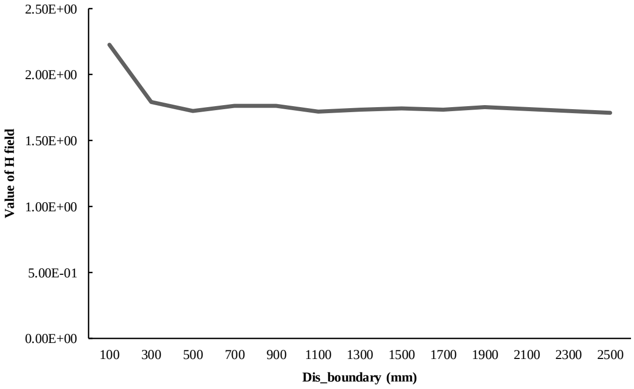


Figure 4. Convergence test for magnetic field strength as the size of the system simulation increases.

Wires with a circular cross section diameter of 0.5 mm require a long time to generate mesh and simulation of FEM. Therefore, you can convert a circular cross section into a rectangular cross section of the same area to model the wire as a rectangle with a length of 0.44 mm.

The transmitting coil (Tx coil) is a 5 cm round coil, modeled in the form of a 5th winding, and the receiving coil (Rx coil) is modeled as a 5th winding with an elliptical coil with a long axis of 3 cm and a short axis of 0.7 cm. Orange indicates the transmitting coil and yellow indicates the receiving coil (Figure 3).

While determining the size of the space for simulation, by increasing the boundary between the magnetic field strength and the inductance value at a specific location, even if the boundary increases, the value of the magnetic field strength does not change. Inductance convergence is set as the optimal boundary that satisfies both efficiency and accuracy. The size suitable for simulation was determined to be 500 mm (Figure 4).

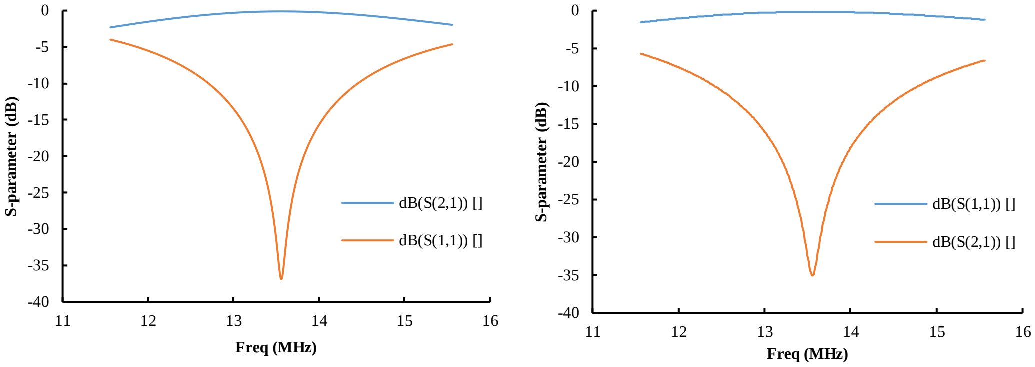


Figure 5. 13.56MHz resonance check for transmit (orange) coil (top) and receive (yellow) coil (bottom).

In order to attach a capacitor for resonating at 13.56 MHz to each of the transmitting coil and the receiving coil, a two-port model was selected among the options available in HFSS. After calculating the resonant capacitance from $f = \frac{1}{2\pi\sqrt{LC}}$ to $C = \frac{1}{(2\pi f)^2 L}$, based on the calculated value, the capacitance value was changed little by little to find the correct capacitance value at which resonance occurred at the desired resonant frequency (Figure 5).

## 2.3. Simulation and verification of ECG sensor

The distribution of magnetic field (H-field) is as follows when 13.56 MHz AC current is applied to the transmitting coil and the receiving coil is located at a certain distance on the Z axis. In addition, the power transmission efficiency was determined according to the relative position of the transmit / receive coils as well as the formed magnetic field(Figure 6).

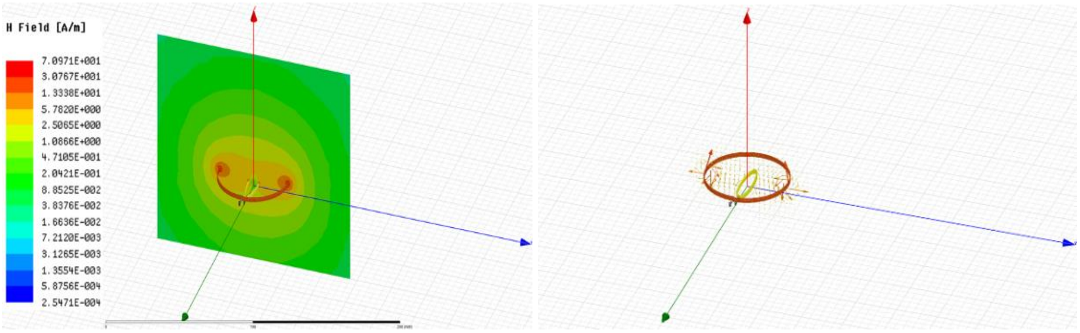


Figure 6. Distribution of magnetic field formed when 13.56 MHz alternating current flows through the transmitting coil: (top) contour plot, (bottom) vector plot.

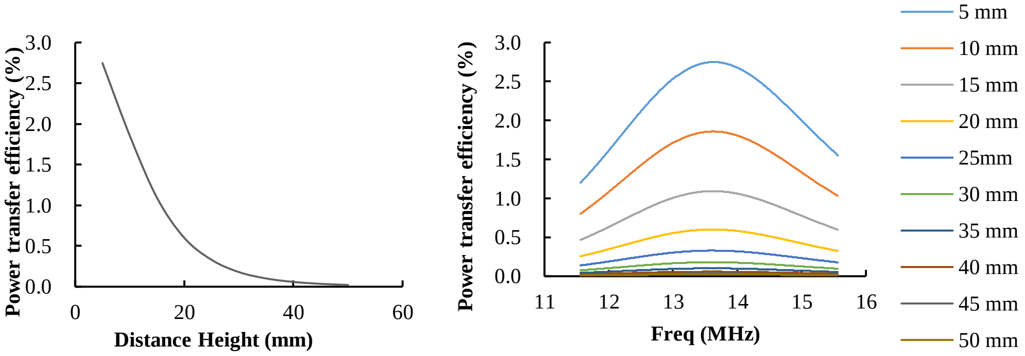


Figure 7. Changes in power transmission efficiency as the distance between the transmitting coil and the receiving coil increases (top) and frequency (bottom)

The simulation results show that the power transmission efficiency at each distance. When the distance between coils is 5 mm (minimum distance in the graph below), the transmission efficiency is about 2.74%. The power transmission efficiency according to the frequency change around 13.56 MHz. It has the maximum transmission efficiency at the resonance frequency of the transmitting / receiving coil, and the transmission efficiency is decreased at a frequency other than the resonance frequency (Figure 7).

## 2.4. Variations due to misalignment when coil-to-coil misalignment

The change in the magnitude of the magnetic field in the case where the center of the receiving coil and the transmitting coil are perfectly aligned (misalignment is zero), the transmission coil is shifted 5 mm in the major axis direction of the ellipse until it becomes 25 mm. The change of the power transmission efficiency when the receiving coil is moved in the major axis direction and the alignment turn aside. At this time, the distance between the transmitting and receiving coils in the Z-axis direction was fixed to 5 mm. In a perfectly aligned state, the transmission efficiency is about 1.85%, and the transmission efficiency is decreased as the degree of misalignment increases. However, there is no significant difference in transmission efficiency until the alignment is changed by about 10 mm (Figure 8).

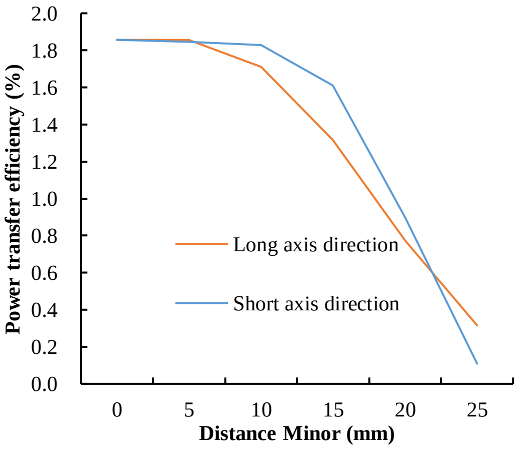


Figure 8. When the alignment of the transmitting coil and the receiving coil is different in the long/short axis direction of the receiving coil.

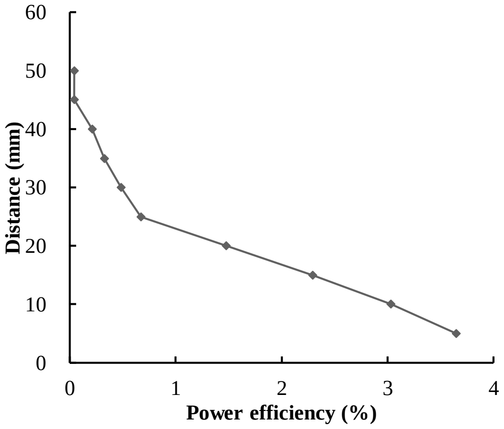


Figure 9. Power transmission efficiency.

The change of the magnetic field by the alignment was changed moving the transmission coil by 0 mm to 25 mm in the minor axis direction of the ellipse. The change in power transfer efficiency when the receiving coil was moved in the uniaxial direction and the alignment is wrong. The distance in the Z-axis direction between the transmitting and receiving coils is fixed at 5 mm, and the transmission efficiency is the highest in a perfectly aligned state. As the degree of misalignment increases, the transmission efficiency decreases. However, it can be seen that there is no significant difference in transmission efficiency (Figure 9).

## 2.5. Verification through experiments

In order to verify the wireless power transmission through the simulation, we performed actual experiments. For the experiment, a transmission coil with a circular shape with a diameter of 5 cm and an elliptical shape with a long axis/short axis of 30 mm/7 mm were fabricated. Both the transmitting and receiving coils were coated with copper wire having a diameter of 0.5 mm, and the number of turns of the coils was set to 5 times. The inductance of the fabricated coil was measured to be 2.8 µH for the transmit coil and 0.6 µH for the receive coil. The resonant capacitance used to resonate the transmit and receive coils at 13.56 MHz was 37 pF and 228 pF, respectively.

This is an experiment in which the voltage value obtained at the receiving coil is measured while the distance between the coils is changed while the input to the transmitting coil is constant using the transmitting/receiving coil. And the result of measuring the voltage and power obtained from the receiving coil according to the distance between the coils. The power transmission efficiency was calculated as described.

Transmission efficiency: η(%) = (PL / PS) * 100
The power at the load of the receiving coil: PL = (Vpk)2 / 2RL
Power in the transmit coil: Ps = Vs * Is

Table 1. Power transmission efficiency.

| Distance (mm) | Vp (V) | VR_Tx (V) | Is (A) | Ps (W) | PL(W) | Power efficiency (%) |
|---|---|---|---|---|---|---|
| 5 | 7.3 | 1.46 | 0.146 | 0.73 | 0.026645 | 3.65 |
| 10 | 6.7 | 1.48 | 0.148 | 0.74 | 0.022445 | 3.033108108 |
| 15 | 5.9 | 1.52 | 0.152 | 0.76 | 0.017405 | 2.290131579 |
| 20 | 4.8 | 1.56 | 0.156 | 0.78 | 0.01152 | 1.476923077 |
| 25 | 3.28 | 1.6 | 0.16 | 0.8 | 0.005379 | 0.6724 |
| 30 | 2.8 | 1.62 | 0.162 | 0.81 | 0.00392 | 0.483950617 |
| 35 | 2.3 | 1.62 | 0.162 | 0.81 | 0.002645 | 0.32654321 |
| 40 | 1.84 | 1.6 | 0.16 | 0.8 | 0.001693 | 0.2116 |
| 45 | 0.82 | 1.62 | 0.162 | 0.81 | 0.000336 | 0.041506173 |
| 50 | 0.85 | 1.62 | 0.162 | 0.81 | 0.000361 | 0.044598765 |

As a result, when the distance between the transmitting and receiving coils is 5 mm, the voltage of the receiving coil (using 1 kΩ) was 7.3 V, the power was 26.6 mW, and the power transmission efficiency was 3.6%. When the distance between the coils is 20 mm, the voltage, power, and transmission efficiency at the rod of the receiving coil were 4.8 V, 11.5 mW, and 1.5%, respectively. If the actual load (IC chip of various amplifiers) to which the receiving coil will transmit power requires 3.3 Vdc operating power, if a full-wave rectifier is used to convert the induced AC voltage to DC voltage, When the input exceeds 4.8V, the 3.3V DC voltage can be generated sufficiently. In other words, if you use the same coils as those used in this measurement, you can conclude that the transmit and receive coils can supply enough voltage and power even when they are 2 cm apart.

## 2.6. In vivo model validation

Fiteen healthy adult male SD rat (Sprague–Dawley rats) weighing 350–390 g were implanted with ECG sensors, and activated. The sensor between the peritoneal epithelium and the skin tissue, and approximately four weeks were allowed for the wounds at the surgical site to heal.

The wireless power system received signals from the sensor containing the ECG electric power supply, to transmit to an external monitor via Bluetooth. The display system consisted of an ECG signal output using a smartphone. The experiments were performed in accordance with the guidelines outlined in the Declaration of Helsinki, and were approved by the Ethic Committee of Keimyung University (Approval number, KM-2015-20R1). The transplanted ECG sensor showed a normal signal and confirmed that the experimental animal showed a pattern similar, less than 5% difference, to the simulation result of the wireless transmission in the fixed state (Figure 10).

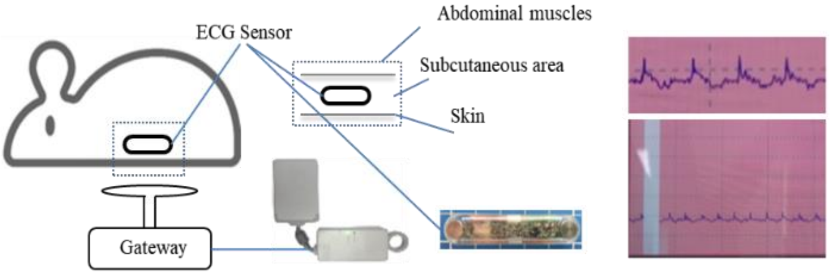


Figure 10. ECG sensor verification using in vivo model and wireless signal transmission.

## 3. DISCUSSION

As the population ages and the welfare increases, research on implantable medical devices is being actively conducted, and products that provide more diverse functions are being released. On the other hand, the power consumption of implantable medical devices is increasing due to various functions, it is not possible to supply sufficient power with only the primary battery. In this study, the efficiency of a wireless power transmission system using a magnetic induction method was investigated. Computer modeling was used to verify the effectiveness of ECG detector power transmission, and to determine the applicability of the human body using a magnetic induction wireless power transmission system, we applied it to animal models using mice. As the required functions of implantable devices are diversified, power usage time is shortened and the restart cycle for power replacement is shortened, which increases the additional cost and physical and psychological burden. The development of a sustainable power module that reduces the cost and psychological burden of these additional surgeries is an essential technology enabling functional implementation of medical devices for human implantation[15-17].

The market for implantable medical devices is growing rapidly. Implantable medical devices are increasingly being applied to various disease treatment fields to help human weak functions, and power modules are increasingly needed for active treatment through devices. As research on wireless power transmission technology has been actively conducted, many attempts have been made to apply a method using electromagnetic waves, such as an electromagnetic induction technique or a magnetic resonance method, which is commercialized in electronic products to human insertion devices. However, the low-frequency method has a low absorption rate of the human body, a short transmission distance, and the high-frequency method has a high absorption rate of the human body, causing a temperature increase in skin tissue[18, 19].

The magnetic induction wireless transmission system used in this study receives electric induction between the transmitter coil and the receiver coil. When a magnetic field is generated in the transmitter coil, the receiver coil receives the magnetic field to induce electric power. This method has a high transmission efficiency of 90% or more, but the transmission distance is very short in several millimeters, and if the centers of the coils are not aligned with each other, the transmission efficiency is greatly reduced. However, it is known to be most suitable for application in the medical field due to its high safety and efficiency compared to the magnetic resonance method and electromagnetic wave method[20]. The results of this study can be used as a model of a power transmission system such as a human implant sensor using magnetic induction. The limitations of this study did not reflect animal movement for in vivo transplantation. In order to effectively apply the results of this study, it is necessary to verify the wireless power transmission efficiency for the movement of the detector.

IEC TC106 discusses the evaluation and measurement methods of electromagnetic human body exposure to radio power, and implantable medical devices have become a big issue in terms of safety[21-23]. Various wireless charging techniques may fail to reach commercialization because of possible hazards to the human body. High frequency and high power can induce human injury in implant application, too low output causes problems in operation of implant. In order to solve such a problem, it is possible to operate at low power and an electromagnetic wave environment within the human body protection standard is required. The development of miniaturization with lower power and the development of low-power-based implants is needed. In addition, strong electric fields and magnetic fields can lead to malfunctions of active implant devices such as artificial heart pacemakers, artificial eyes, and artificial heart, and problems must be solved[24-26].

In this study, we investigated the distribution of the magnetic field and the power transmission efficiency according to the relative position between the transmitting and receiving coils when the electromagnetic induction type wireless power transmission is implemented using the elliptical receiving coil through the simulation study. If the coil is wound five times with the coil, the voltage that the coil can transfer to the load is more than 4.8V, which is enough to drive the IC with 3.3V operating voltage. And the transmission efficiency at this time is about 1.5%.

It is necessary to develop a variety of implants that can operate at low voltage, and to study the resistance to near field of active implant devices such as pacemakers and artificial eyes that should be mounted in the human body and various studies on low frequency band interference.

## 4. CONCLUSION

Through the simulation study, we have investigated how the distribution of magnetic field and the power transmission efficiency of the magnetic field formed according to the relative position between the transmit / receive coils change when the electromagnetic induction type wireless power transmission is implemented using a receive coil having an elliptical shape. The simulation using HFSS is relatively consistent with the actual measurement results, and can be usefully used for wireless power transmission simulation. When the center of the receiving coil is perfectly aligned with the center of the transmitting coil, the transmission efficiency is maximum, and even if the alignment is slightly changed, the performance of wireless power transmission maintains a certain level.

## ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1D1A1B04031182 and 2017R1D1A1B03031747), and this research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C0575010018)and this research was supported by a grant of Innopolis Foundation (Bio-signal authentication and abnormal sign detection system through multi-object detection).

## REFERENCES

[1] K. Kim, B. Kim, and C. H. Lee, “Printing Flexible and Hybrid Electronics for Human Skin and Eye-Interfaced Health Monitoring Systems,” Advanced Materials, vol. 32, no. 15, 2020.

[2] Y. Chen, Y. Cheng, Y. Jie et al., “Energy harvesting and wireless power transmission by a hybridized electromagnetic-triboelectric nanogenerator,” Energy and Environmental Science, vol. 12, no. 9, pp. 2678-2684, 2019.

[3] Q. S. Abdullahi, R. Joshi, S. K. Podilchak et al., “Design of a wireless power transfer system for assisted living applications,” Wireless Power Transfer, vol. 6, no. 1, pp. 41-56, 2019.

[4] Y. Ben Fadhel, S. Ktata, S. Rahmani et al., "Near-field Wireless Power Transfer is a promising approach to Power-up Active Implants." pp. 399-404, 2020.

[5] K. Agarwal, R. Jegadeesan, Y. X. Guo et al., “Wireless Power Transfer Strategies for Implantable Bioelectronics,” IEEE Reviews in Biomedical Engineering, vol. 10, pp. 136-161, 2017.

[6] T. Campi, S. Cruciani, F. Maradei et al., “Pacemaker Lead Coupling with an Automotive Wireless Power Transfer System,” IEEE Transactions on Electromagnetic Compatibility, vol. 61, no. 6, pp. 1935-1943, 2019.

[7] R. Chávez-Santiago, J. Wang, and I. Balasingham, "The ultra wideband capsule endoscope." pp. 72-78, 2013.

[8] G. Ahmed Zeeshan, R. Sundaraguru, and F. Naaz, “Wearable wireless sensor system with RF remote activation for industrial applications,” International Journal of Recent Technology and Engineering, vol. 8, no. 3, pp. 4716-4720, 2019.

[9] M. Wang, J. Chen, X. Cui et al., "Design and Fabrication of 5.8GHz RF Energy Harvesting Rectifier." Wireless Power Transfer, vol. 6, on. 1, pp. 41-56, 2019.

[10] S. Suganuma, D. Hung Nguyen, Y. Nishioka et al., "The Logistics System by Rotary Wing Unmanned Aerial Vehicle with 28GHz Microwave Power Transmission." pp. 413-416, 2019.

[11] S. Almorabeti, M. Hanaoui, M. Rifi et al., "Microstrip patch antennas at 5.8GHz for wireless power transfer system to a MAV." ACM International Conference Proceeding Series, 2017.

[12] N. Khan, H. Matsumoto, and O. Trescases, “Wireless Electric Vehicle Charger with Electromagnetic Coil-Based Position Correction Using Impedance and Resonant Frequency Detection,” IEEE Transactions on Power Electronics, vol. 35, no. 8, pp. 7873-7883, 2020.

[13] S. A. Sis, and H. Akca, “Maximizing the efficiency of wireless power transfer systems with an optimal duty cycle operation,” AEU - International Journal of Electronics and Communications, vol. 116, 2020.

[14] K. Wang, L. Sang, Y. Zhang et al., "Optimization Design for Wireless Power Transfer System under the Variation of Load and Coupling Coefficients." pp. 867-872, 2019.

[15] T. Ghomian, and S. Mehraeen, “Survey of energy scavenging for wearable and implantable devices,” Energy, vol. 178, pp. 33-49, 2019.

[16] A. Koruprolu, S. Nag, R. Erfani et al., "Capacitive Wireless Power and Data Transfer for Implantable Medical Devices." 2018 IEEE Biomedical Circuits and Systems Conference, BioCAS 2018 - Proceedings, 2018.

[17] E. Katz, "Implantable biofuel cells operating in vivo: Providing sustainable power for bioelectronic devices: From biofuel cells to cyborgs." Proceedings - 2015 6th IEEE International Workshop on Advances in Sensors and Interfaces, IWASI, pp. 2-13, 2015

[18] A. W. S. Putra, M. Tanizawa, and T. Maruyama, “Optical wireless power transmission using si photovoltaic through air, water, and skin,” IEEE Photonics Technology Letters, vol. 31, no. 2, pp. 157-160, 2019.

[19] K. Takahashi, T. Yamada, and Y. Takemura, “Circuit parameters of a receiver coil using a wiegand sensor for wireless power transmission,” Sensors (Switzerland), vol. 19, no. 12, 2019.

[20] K. Zhang, C. Liu, Z. H. Jiang et al., “Near-Field Wireless Power Transfer to Deep-Tissue Implants for Biomedical Applications,” IEEE Transactions on Antennas and Propagation, vol. 68, no. 2, pp. 1098-1106, 2020.

[21] C. Xiao, K. Wei, D. Cheng et al., “Wireless charging system considering eddy current in cardiac pacemaker shell: Theoretical modeling, experiments, and safety simulations,” IEEE Transactions on Industrial Electronics, vol. 64, no. 5, pp. 3978-3988, 2017.

[22] B. Cheng, Y. Chatzinoff, D. Szczepanski et al., “Remote acoustic sensing as a safety mechanism during exposure of metal implants to alternating magnetic fields,” PLoS ONE, vol. 13, no. 5, 2018.

[23] A. C. Özen, B. Silemek, T. Lottner et al., “MR safety watchdog for active catheters: Wireless impedance control with real-time feedback,” Magnetic Resonance in Medicine, vol. 84, no. 2, pp. 1048-1060, 2020.

[24] M. V. Tholl, A. Haeberlin, B. Meier et al., “An intracardiac flow based electromagnetic energy harvesting mechanism for cardiac pacing,” IEEE Transactions on Biomedical Engineering, vol. 66, no. 2, pp. 530-538, 2019.

[25] D. B. Ahire, and V. J. Gond, "Wireless power transfer system for biomedical application: A review." ICEI 2017, pp. 135-140, 2018

[26] C. Kasia, A. Appannagari, A. Joshi et al., “Safety of wireless capsule endoscopy in patients with implantable cardiac devices,” JGH Open, vol. 4, no. 2, pp. 241-244, 2020.

# LEARNING FOR E-LEARNING

Carsten Lecon[1] and Marc Hermann[2]

[1]Department of Computer Science, Media Computer Science,
Aalen University, Germany
[2]Department of Computer Science, User Experience,
Aalen University, Germany

***ABSTRACT***

*In response to the heterogeneity of previous knowledge of the students when beginning their studies, we present a solution, where undergraduate students as well as advanced students (hopefully) will benefit from 'AdLeR' (Additive Learning Resources): A tool for the rapid generation of small e-learning courses. The undergraduates can catch up lack of knowledge by our mini courses (self-regulated). The advanced students are involved in the development of our tool or in the creation process of learning material, which is suited for self-regulated learning. When implementing the tool, the students have to deal with various aspects of computer science domains for example, which consolidates their knowledge and their competences.*

***KEYWORDS***

*E-Learning, self-regulated learning, learning by teaching, XML, learning path, search functionality.*

## 1. INTRODUCTION

In this paper, we address two challenges of higher education, especially for computer science. First: Due to the increasing heterogeneity of the students (for example in Germany, now also technicians are allowed to study), there exist different levels of knowledge at the beginning of their studies. By this, there exist different levels of knowledge at the beginning of their studies. With the Bologna-Reform put into practice, the teacher nowadays cannot respond flexible to the individual needs of the individuals due to the rigid curriculum. Although tutorials can help, these are not sufficient. An appropriate solution is to offer individual e-learning courses in the sense of microteaching. These are learning units, which cover specific learning matters. Learning paths offer a multimodal access to the learning material, which supports the self-regulated learning. Second: the students themselves (students in higher semesters) can compile just these learning objects (teaching units, exercises, etc.). In order to create 'mini courses', we use an easy-to-use tool: 'AdLeR' (*Additive Learning Resources)*. The development process of this tool covers many disciplines of the curriculum: Software Engineering/ Programming (the tool has to be implemented), human-computer-interaction (in order to allow a good-looking, motivating, as well as functional screen), formal languages (a subfield of the big topic 'theoretical computer science') for a flexible search functionality. All these subjects are handled in computer science lectures. By implementing our tool, these subjects can be deepened by the practical application, which will also lead to better and sustainable competence in the respective areas.

Up to now, most of the university teaching is organized like depicted in figure 1 (left side): the teachers present the learning matter in (presence) lectures, which take place in lecture halls, seminar rooms or labs. The students have access to the learning material by – for example – a

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 29-43, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101103

learning management system (LMS). Because this system is available online, the students use these data outside of the campus of the university as well. Furthermore, especially in companies, virtual worlds (virtual reality, VR) are used for training and learning. Martín-Gutiérrez, Mora, Añorbe-Díaz and González-Marrero give an overview of such learning scenarios [11], Lueckemeyer presents a concrete VR learning setting for teaching programming languages [10].

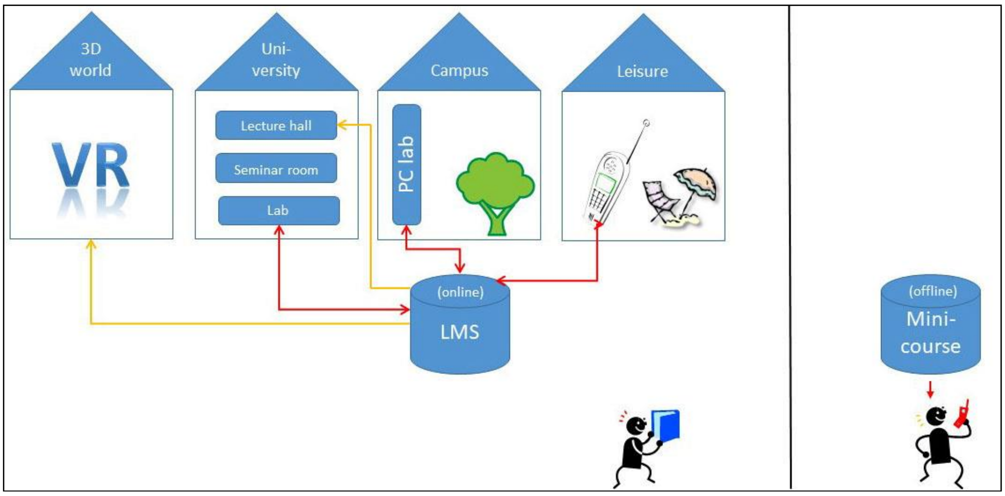


Figure 1: Learning places

Reading (real) books seems to be somewhat outdated since the availability of web sources, all needed information seems to be found in the internet (with the risk of incomprehension by ‘fast food learning’ – see section 2). Therefore, we propose (electronic) mini courses for the above mentioned self-regulated learning, which offer self-contained learning material. By this, learning outside the university infrastructure is possible. However, because in this way the safe environment of the university is left (see figure 1, right side), we have to consider some didactic issues in order to ensure a positive learning effect as much as possible.

The rest of the paper is organized as follows: First, we outline the didactic concepts, which are used in this work (section 2). Then we describe our tool ‘AdLeR’ (section 3).In this chapter, we also describe one important feature of this tool: learning paths (sometimes named ‘learning trail’, subsection 3.2) and aspects of human-computer-interaction (subsection 3.4). In our tool not only a classical full text search is available, but also a flexible search functionality, which will be described in chapter 4. The paper ends with as short summary and an outlook (section 5).

## 2. Didactic Concept and Related Work

It is a challenging approach to offer a kind of learning environment for students which are used to visit presence lectures at the university. This particularly makes sense, if the students get additional learning material just when there is a need: In this case the most intrinsic motivation of the students can be expected. Generally, in our approach several didactic concepts play a role and should be considered when developing learning material for the mini courses and when implementing the tool for generating the course.

In our paper, we consider the following didactic concepts: self-regulated learning, microteaching, blended learning, inverted/ flipped classroom, learning by teaching. In this project, one can often find these concepts in combination.

With the mini courses, we implement some important aspects of e-learning: learning at any place and at any location. This requires self-regulated learning of the students [13], although this term still is ambivalent ([5], [14]). In our context, we expect from the learners (students) to organize themselves and to be aware of their learning behaviour. This sometimes could be a serious problem for the students, which we try to address by the following steps (extract):

- The learning matter of the mini course fits the lacks (for example missing previous knowledge)
- The length of the courses is short
- The course offers possibilities of self-evaluation by quizzes
- A multimodal access to the learning units is possible, for example by learning paths (see subsection 3.3)

The above mentioned length of the course leads to the didactic concept *microteaching* [8], [2]. In [2] one can find an appropriate definition:

> ‘Microteaching is a teaching situation which is scaled down in terms of time and number of students. […] The lesson is scaled down to reduce some of the complexities of the teaching act, thus allowing the teacher to focus on selected aspects of teaching. […]’

This definition reflects our intention: students can learn the subjects in small portions and just when a specific knowledge is needed. This can be done at any place, for example using a mobile phone when commuting to the university. However, a challenge is to avoid ‘fast food learning’ (quickly consuming learning along the way); that means that some topics cannot be treated without previous knowledge. For example, it would be difficult to understand the JPEG compression method using the DCT (discrete cosines transformation) without a fundamental knowledge about trigonometry. The elements of a microteaching learning material can be: texts, short videos, animations, pictures as well as hyperlinks to opportune sites in the internet. In addition, the learners can review their acquired knowledge by performing quizzes. Therefore, every mini course should begin with the required previous knowledge (eventually measured by a short initial test).

Another more general concept is the *Blended Learning* (see for example [7]): The students visit lessons at the university, and also learn by electronic learning objects (self-regulated) – if necessary. In contrast to the usual variant of blended learning, here most of the study matter is presented in presence lessons whereas the electronic material is an (optional) enrichment.

The aim to support all students is also the idea of the didactic concept ‘inverted classroom’/ ‘flipped classroom’ [4]: Here, the learning is transferred completely to electronic material, whilst the presence phases are used for repeating and applying the learning matter. The motivation for this is similar to our observation, that ‘not all students learn in the same way at the same pace’ [6]. This concept is promising, but is rarely used nowadays, because a great effort is necessary, for example, the learning materials have to be prepared for self-study considering various kinds of learning behaviour. If the learners are not motivated, the worst case could be, that they participate in the presence phase without having learnt anything.

The didactic concepts described so far, refer to consuming (additional) learning material. From another point of view, the persons – in our case students – who create learning contents, apply the didactic concept ‘*learning by teaching*’ (see for example [3], [18]). This is like a tutorial, but electronically and without face-to-face communication. Therefore, this is a special challenge for the students, because the teaching content has to be prepared for self-regulated learning by themselves.

# 3. THE TOOL 'ADLER'

In this section, we present our tool for generating mini courses. First, we give a motivation and an overview of our tool (subsection 3.1). The structure and the content of such courses is specified by an XML specification (subsection 3.2). The significance and realization of learning paths will be described in subsection 3.3. Some aspects of the human-computer-interaction, which is relevant for the screen design and for the user interface of our tool, are addressed in subsection 3.4.

## 3.1. Introduction to 'AdLeR'

For the above mentioned reasons, we have decided to offer additional (or deepening) learning material. This should not be a complete e-learning course, but a summary of a specific learning matter. We name this sequence of HTML pages 'mini course'. Because this is an addition to the curriculum, we named our tool *Additional Learning Resources* – shortened '*AdLeR*' (the German word for *eagle*).

The aim of this tool is not necessarily the creation of new learning material. Instead, our tool is able to use existing assets in order to combine this media to a mini course. Therefore, the assets can be used in multiple ways.

For this, we have implemented a prototype for the generation of 'mini courses' in the sense of microteaching. The structure (chapter, subsection, hyperlinks, etc.) and the content (assets like text, image, video, quiz, etc.) is specified in an XML document (see subsection 3.2).
The general mechanism of the tool is depicted in figure 2.

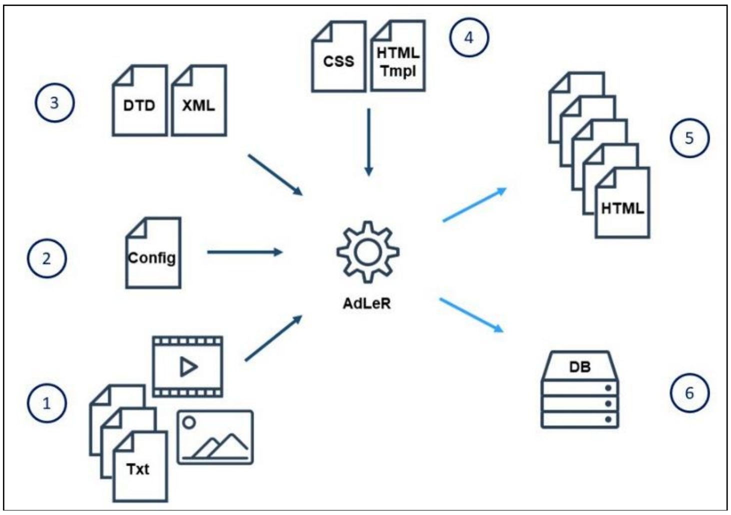
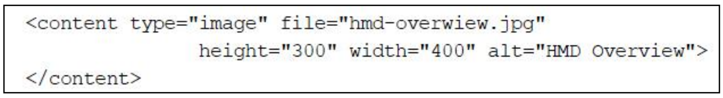


Figure 2: General mechanism of the tool 'AdLeR'

The programming language for this tool is Java. The graphical user interface for the end user is written in C++.

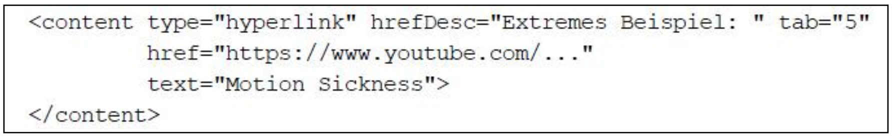
A generated HTML page (5) of the generated course consists of media assets (1) pages, which are specified in an XML document (3). For example, the syntax for a picture looks like this:



```
<content type="image" file="hmd-overwiew.jpg"
               height="300" width="400" alt="HMD Overview">
</content>
```

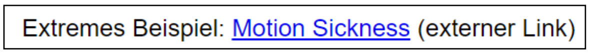
In this case, the picture file 'hmd-overwiew.jpg' will be included in the HTML page. If the size (height and width – because of space saving not denoted in the XML document in the next section) is given, the picture will be scaled accordingly; when clicking on the picture, it will be shown in original size – in another browser tab or window. The 'alt' attributed is used for the figure caption (and appears by 'mouse over').

In order to refer to external learning resources in the internet, hyperlinks are also possible. The syntax looks like this:



```
<content type="hyperlink" hrefDesc="Extremes Beispiel: " tab="5"
         href="https://www.youtube.com/..."
         text="Motion Sickness">
</content>
```

The attribute 'hrefDesc' is the caption of the hyperlink, the attribute 'text' describes the clickable text, when clicking on this text,the specified hyperlink (attribute href') will be opened.
In the generated HTML page, this content description looks like this:



Extremes Beispiel: Motion Sickness (externer Link)

'(externer Link)' symbolizes a hyperlink to a resource outside the actual course ('externer' means 'external'). This hyperlink will be opened in a new browser window. Besides, the attribute 'tab' signs an indenting (for formatting). Hyperlinks to other pages in the current course are also possible. In this case, the hyperlink would refer to a page id (see DTD in section 3.2). Links between different mini courses are not provided: that would presuppose that a course is online available – at a certain location.

Currently, the following media types are supported by the tool:

- text: text written directly in the XML file
- textfile: content of a text file
- image: picture file
- video: video file; with HTML5 a video player can easily be realized
- audio: audio file; with HTML5 an audio player can easily be realized
- hyperlink: link to an external resource in the internet (see above)
  (Internal links to pages inside the course also are possible by using the prefix '#' and the id of the referred page)
- A special content type is 'quiz': This is a hyperlink to webpage generated by an external tool (with hyperlinks to pages of the mini course, where the actual content is described, on the other hand). Quizzes are specified in the XML document by an own element 'quiz' (see section 3.2).

The external media (picture, video, audio) is copied into the target directory of the mini course.
The layout of every page should look similarly. Therefore, we use HTML template files (4 in figure 2; also see section 3.4), which can be adapted to special needs of the teachers by rewriting the HTML code of the standard template or by adjusting the CSS file.

Structure information and meta data are extracted for building a kind of database (6). In our approach, no traditional database management system (DBMS) is used, because the mini course should be used also online, and an installation of a DBMS on one's own device would be too wasteful. Instead, the data are stored as text files or serialized Java classes, respectively.

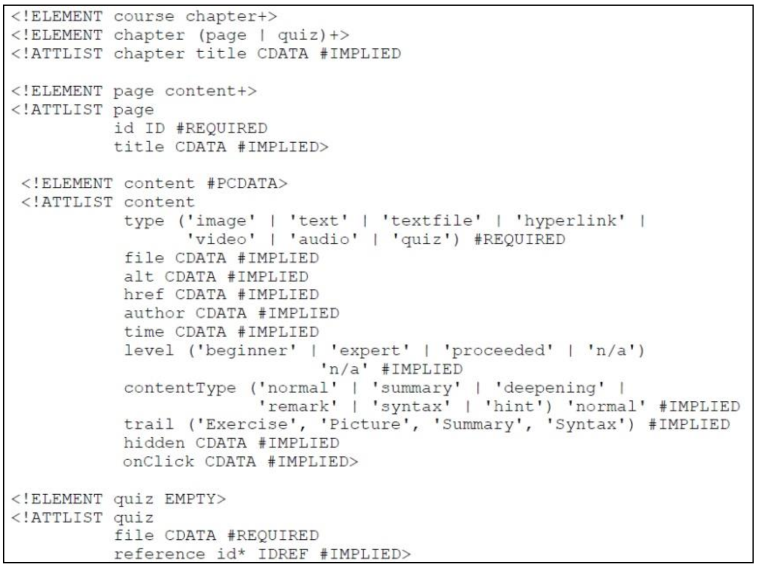
Several parameters of the tool can be adjusted by a configuration file (2 in figure 2; if such a file exists, these values are treated – otherwise default values). Beside the specification of filenames and directories (name of the template document, target directory of the generated files, location of the XML and the template file, etc.), for example – among others – these parameters can be set:

- IMPRESSUM: Imprint (there exits an appropriate placeholder in the template file)
- NUMBERING: Indication, if the chapter and sections should be numbered
- WRITE_AUTOR: Indication, if the authors name and the generation time should be included in the page
- TRAIL: Name of learning trails, which will be generated automatically based on the occurrence of this word in the full text (also see section 3.3)

## 3.2. XML: Structure and Meta Data

In principle, the structure of a mini course consists of chapters (and subsections if needed) which consist of pages. One single page is composed of assets. Furthermore, hyperlinks to other pages or to external learning sources in the WWW can be integrated in a page. The default navigation consists of 'page up'/ 'page down' or 'chapter up'/ 'chapter down', respectively. In addition, so called learning paths are possible (see subsection 3.3).

The structure, as well as meta data are specified in an XML document. A part of this specification is depicted as an XML-DTD (*XML Document Type Definition*) in figure 3.



```
<!ELEMENT course chapter+>
<!ELEMENT chapter (page | quiz)+>
<!ATTLIST chapter title CDATA #IMPLIED

<!ELEMENT page content+>
<!ATTLIST page
          id ID #REQUIRED
          title CDATA #IMPLIED>

 <!ELEMENT content #PCDATA>
 <!ATTLIST content
           type ('image' | 'text' | 'textfile' | 'hyperlink' |
                 'video' | 'audio' | 'quiz') #REQUIRED
           file CDATA #IMPLIED
           alt CDATA #IMPLIED
           href CDATA #IMPLIED
           author CDATA #IMPLIED
           time CDATA #IMPLIED
           level ('beginner' | 'expert' | 'proceeded' | 'n/a')
                               'n/a' #IMPLIED
           contentType ('normal' | 'summary' | 'deepening' |
                        'remark' | 'syntax' | 'hint') 'normal' #IMPLIED
           trail ('Exercise', 'Picture', 'Summary', 'Syntax') #IMPLIED
           hidden CDATA #IMPLIED
           onClick CDATA #IMPLIED>

<!ELEMENT quiz EMPTY>
<!ATTLIST quiz
          file CDATA #REQUIRED
          reference id* IDREF #IMPLIED>
```

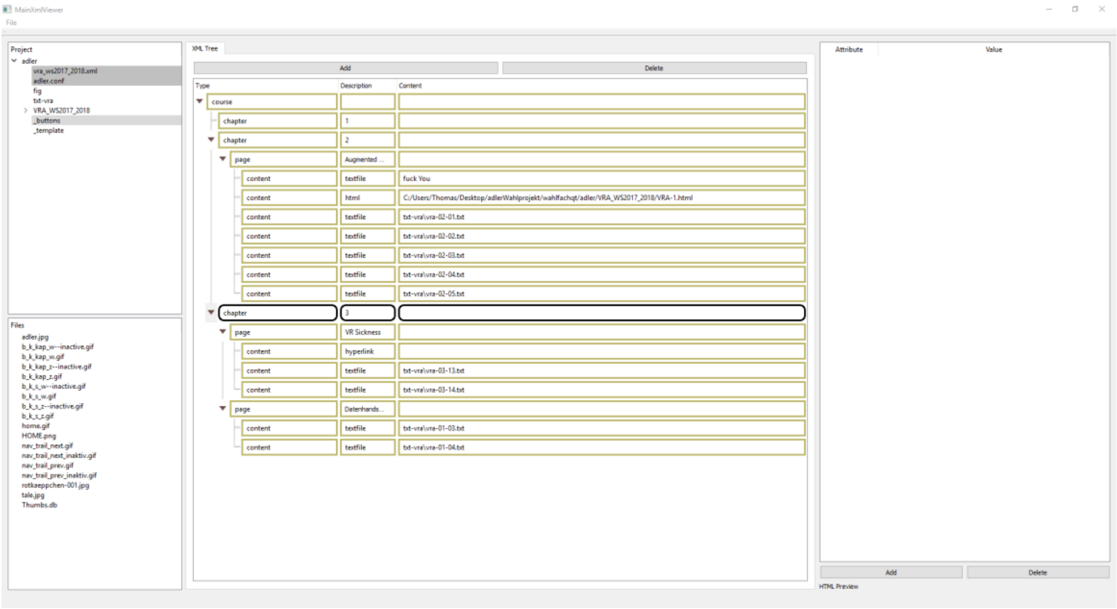
Figure 3: XML-DTD: Structure specification of a mini course (extract)

In this specification, a `course` consists of `chapters`; each chapter consists of subchapters or `page`s (which are transferred to HTML pages). Each page consists of assets (element `content`). For a better understanding, we explain the meaning of some attributes:

- *author*: This is an informative data. It will only appear on the generated HTML page, if the appropriate placeholder exists in the template (see section 3.4 User Interface).
- *level*: This optional attribute indicates the difficulty of the appropriate topic. It is not visible on the page, but can be used for generating individual learning trails (section 3.3) and for the search functionality (section 4).
- *trail*: In the DTD default learning trails are specified. Appropriate statements in the configuration document (see above) can overwrite these.
- *hidden*: This text is hidden for the user, but is accessible for the search functionality (section 4); furthermore, this text can be used in order to generate learning trails (see section 3.3).
- *onClick*: If this attribute exists, the text will appear – with a link to a first hidden text specified by the 'ref' attribute (not listed in the above DTD). In this way, for example, a kind of self-evaluation can be realized: the 'onClick' text shows a question, the answer is hidden and appears only when clicking on the question text.

Even attributes, which are not visible on the generated HTML page, are accessible for the search functionality (see section 4).

With regard to a pleasant use, when building a new course or editing an existing course,students have implemented a graphical editor (figure 4) for the easy-to-use composition of mini courses.



Project
adler
vra_ws2017_2018.xml
adler.conf
fig
txt-vra
VRA_WS2017_2018
_buttons
_template

| Type | Description | Content |
|---|---|---|
| course | | |
| chapter | 1 | |
| chapter | 2 | |
| page | Augmented ... | |
| content | textfile | fuck You |
| content | html | C:/Users/Thomas/Desktop/adlerWahlprojekt/wahlfachqt/adler/VRA_WS2017_2018/VRA-1.html |
| content | textfile | txt-vra\vra-02-01.txt |
| content | textfile | txt-vra\vra-02-02.txt |
| content | textfile | txt-vra\vra-02-03.txt |
| content | textfile | txt-vra\vra-02-04.txt |
| content | textfile | txt-vra\vra-02-05.txt |
| chapter | 3 | |
| page | VR Sickness | |
| content | hyperlink | |
| content | textfile | txt-vra\vra-03-13.txt |
| content | textfile | txt-vra\vra-03-14.txt |
| page | Datenhands... | |
| content | textfile | txt-vra\vra-01-03.txt |
| content | textfile | txt-vra\vra-01-04.txt |

Files
adler.jpg
b_k_kap_w--inactive.gif
b_k_kap_w.gif
b_k_kap_z--inactive.gif
b_k_kap_z.gif
b_k_s_w--inactive.gif
b_k_s_w.gif
b_k_s_z--inactive.gif
b_k_s_z.gif
home.gif
HOME.png
nav_trail_next.gif
nav_trail_next_inaktiv.gif
nav_trail_prev.gif
nav_trail_prev_inaktiv.gif
rotkaeppchen-001.jpg
tale.jpg
Thumbs.db

Figure 4: User interface for *AdLeR* (prototype)

After clicking on a button, a mini course will be generated as a sequence of HTML documents – which can be used offline on any device (or online on a server) –at all times and locations.

## 3.3. Learning paths

The standard navigation path – page by page – serves as a 'guided tour' through the course. This gives the learner a sense of safety. The creator of this course orders and selects appropriate content and media. However, e-learning allows to address the individual needs of the learner. For example, this concerns the selection of media (learning by text, audio, pictures, videos, etc.) as well as the scope and the order of the learning material. With expanding the standard navigation by so-called learning paths, more flexibility is possible during the learning process.

As an example, we consider some use cases (following [16]):

- Modal structure: The students can learn best with pictures, a navigation path could lead through all pages, which contain at least one picture.
- Selective structure: A course about object-oriented programming consists of the introduction of object-oriented programming in general and the concrete realization with the programming language Java. If a learner is familiar with the concepts of object oriented programming, but wishes to learn how this is done with Java, he/ she can skip the path ‘concept’ and can only use a path ’syntax’.
- Repetitive structure: By a path ‘summary’ and/ or ‘exercises’ the learner can prepare oneself for an exam, whereas the navigation path leads through pages, which include summaries or exercises, respectively.
- Didactic structure: By this, it is possibility to react to the heterogeneity of the previous knowledge of the learners, for example offering paths on a beginner level, which includes more details than the standard learning path.

The specification of learning paths in our tool can be done via two ways:

- Automatically: Based on the meta data in the XML document (attributes ‘level’, ‘type’, ‘contentType’, etc.) specific navigation paths will be generated.
- Manually: The attribute ‘trail’ marks the belonging to a specific learning path.

The belonging to learning path(s) is recognizable on every page. This could be helpful for the orientation. On figure 5 an example of learning paths of a mini course about selected themes about augmented reality (AR) and virtual reality (VR) is depicted (meanwhile, the design of the page has been revised; see next section).

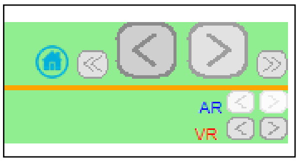


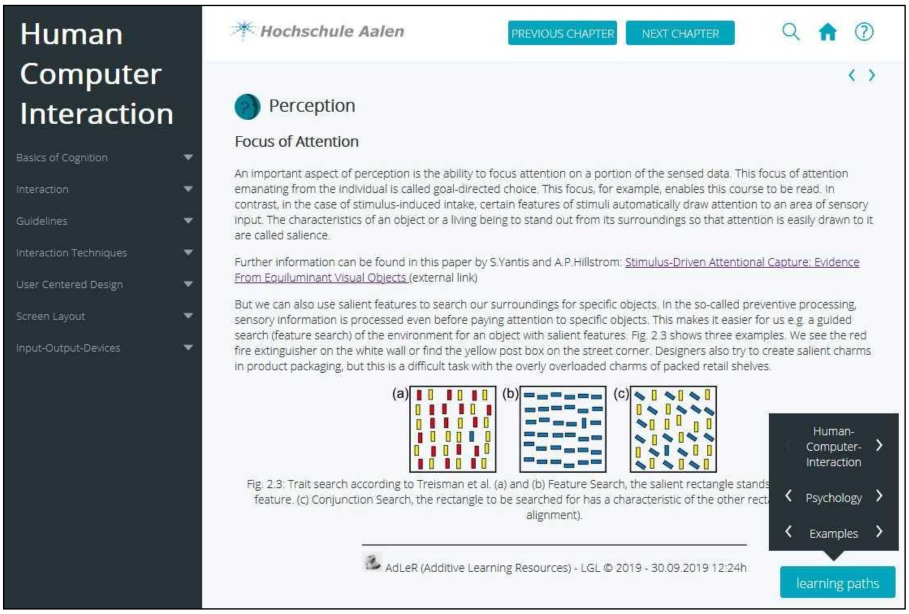
Figure 5: Learning Paths

One can see that this page deals with augmented reality (AR) as well as with virtual reality (VR). This page is the only one, which includes the topic AR (white – inactive – navigation buttons) whereas the topic VR will be treated on the previous and the next page.

## 3.4. User Interface

In modification to ‘you eat with your eyes’, one can say ‘you learn with your eyes’; that means, the screen design plays an important role at learning on the screen (monitor screen, mobile phone screen, etc.). Thus, we consider the appearance when displaying the learning objects – which also is a student´s project in one of our lessons.

As a student´s project, templates for HTML pages – using HTML5 and CSS3 –were designed and evaluated. The design of the template can easily be adapted by the external CSS3 file to special needs. The template contains placeholders for elements/ functionality, which are filled at the generation process: title, content blocks, navigation, footer text, etc. for example ‘navP’ for page-navigation, ‘$navT’ for path-navigation (‘T’ stands for ‘trail’), ‘$foot’ for the logo, copyright, and date. Actually, this project is in progress and some drafts have already been evaluated. In the proposal at figure 6, one can see a typical distribution of the single elements: On the left, there is a table of content, at the top there is a chapter up/ chapter down navigation, at the subjacent level a page up/ page down navigation. When clicking on the button at the bottom right, the list of (clickable) available learning paths appears.



Perception

Focus of Attention

An important aspect of perception is the ability to focus attention on a portion of the sensed data. This focus of attention emanating from the individual is called goal-directed choice. This focus, for example, enables this course to be read. In contrast, in the case of stimulus-induced intake, certain features of stimuli automatically draw attention to an area of sensory input. The characteristics of an object or a living being to stand out from its surroundings so that attention is easily drawn to it are called salience.

Further information can be found in this paper by S.Yantis and A.P.Hillstrom: Stimulus-Driven Attentional Capture: Evidence From Equiluminant Visual Objects (external link)

But we can also use salient features to search our surroundings for specific objects. In the so-called preventive processing, sensory information is processed even before paying attention to specific objects. This makes it easier for us e.g. a guided search (feature search) of the environment for an object with salient features. Fig. 2.3 shows three examples. We see the red fire extinguisher on the white wall or find the yellow post box on the street corner. Designers also try to create salient charms in product packaging, but this is a difficult task with the overly overloaded charms of packed retail shelves.

Fig. 2.3: Trait search according to Treisman et al. (a) and (b) Feature Search, the salient rectangle stands feature. (c) Conjunction Search, the rectangle to be searched for has a characteristic of the other rect alignment).

AdLeR (Additive Learning Resources) - LGL © 2019 - 30.09.2019 12:24h

Figure 6: Layout for an HTML page of the mini course

In addition, designing the user interface for our application was done considering the results of the human-computer-interaction research (a student´s project as well).

## 4. Search Functionality

Even though the mini courses are mostly small, it could be useful to offer a search functionality. For example, to decide, if a mini course is meaningful for the current need, and to find easily the actual relevant topics which are part of the mini course. In our context, we have interesting possibilities: The underlying XML document offers rich structure and metadata information, which can be exploited for a powerful search functionality. In addition, because an XML structure bases on a (simplified) formal grammar, aspects of the theoretical computer science can be integrated in the data modeling (for the search function) and offers an alternative application of formal languages to the students –in contrast to the traditional dealing with this topic in lessons. Thus, we can present a flexible search functionality, which goes much further than a simple full text search.

It is possible to search for all attributes, which are specified in the XML document: Apart from *title*, *date*, *size*…, one can also look for prerequisites, etc.

In addition, structure information is available: the hierarchy of the XML elements and the linking between elements. The structure (hierarchically order, sequences) exists implicitly in the underlying document(XML document). There is a semi structured data view to the data, which has already been described in [1], adapted in [9].

This idea allows some structure-oriented queries, for example by using:

*Structure-describing attributes:* The (recursive) structure of the learning object can be described by object-valued attributes. For example, the titles of all objects of a hierarchy can be described as a character string like o6.title$_{path}$='Databases/Languages/SQL'. In this manner, also regular expressions can be used, for example in order to look for all subsections with the title 'Data Models' or 'SQL': */('Data Models'|'SQL')/*'.

*Derived attributes:* The theoretical concept of attributed context free grammars (for example used at compiler construction), can be adapted to hierarchically structured data: The deriving (inferring) of attributes and attribute values of the XML elements. This can be done alongside the hierarchy (parent/ children relationship) or alongside the links or paths.

As an example, consider a part of a lesson about extended reality (XR) (see figure 7), which is organized in a hierarchical structure (see figure 7, left side): At the top, the title of the chapter of the lesson is described (title='XR'). This chapter consists of three subsections, each with a subtitle and the estimated time to complete this topic (*dur*: duration). The duration can be inherited upwards, whereby the sum of all times is calculated, so the top element (chapter) offers the total estimated time to complete the topic 'XR'. On the other hand, the title of the top element can be inherited to the subjacent lessons downwards. In figure 7, the derived attributes are written in *italic* (based on [10]).

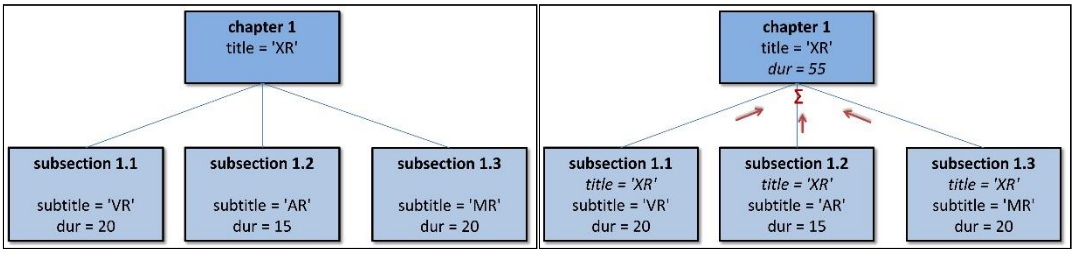


Figure 7: Inheriting of attributes and attribute values

This means, searching for objects with the title 'XR' results not only in the top element (chapter with the title 'XR'), but also in the dependent lesson objects 'VR', 'AR' and 'MR'.

In order to distinct origin and derived attributes, the derived values can be weighted when performing the query processing. This will be used to rank the search results.

In addition, pages, which are connected by learning paths, can be considered as a hierarchical structure. The linked pages will be decreasingly weighted.

*Attributed context free grammar approach*

In order to process such structure-oriented queries, we need an adequate internal representation of the data. The underlying data model can be described like an attributed context free grammar (foundations see [15], prior working concerning data modeling see [19]). Referring to the above example, we extend the specification as follows:

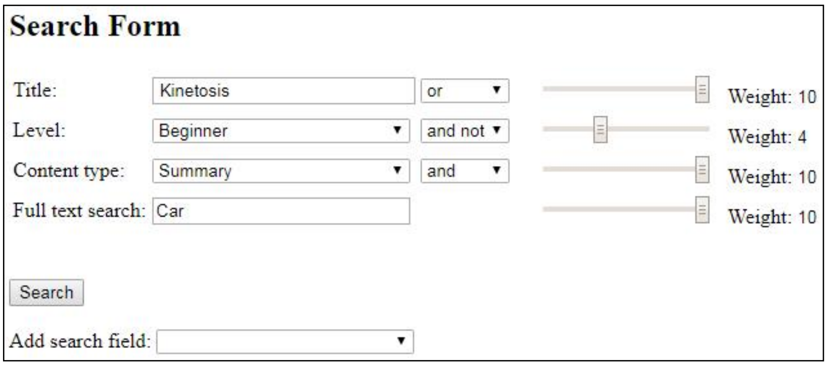
- A course consists optionally of slides (described by a heading) and of at least one chapter(described by a title).
- A chapter consists of at least one subsection (described by a title and a duration - dur)
- A subsection consists of inner subsections (optional) or at least one page.
- A page represents the generated (HTML) page and consists of at least one asset (text, picture, etc.)

As an addition to the actual course, slides are possible, for example, in order to present some topics for classroom teaching. An attributed context free grammar for this scenario can look like this (highly simplified; words with capital letters are non-terminal symbols, words with lowercase letters are terminal symbols, words with italic letters are attributes):

| | | |
|---|---|---|
| COURSE | → | SLIDE*CHAPTER* |
| | | *title* := SLIDE.*heading*∪ (∪ CHAPTER.*title*) |
| SLIDE | → | asset |
| CHAPTER | → | SUBSECTION+ |
| | | *dur* := ∑(SUBSECTION.*dur*) |
| | | *title* := ∪(SUBSECTION.*subtitle*) |
| SUBSECTION | → | SUBSECTION* \|PAGE+ |
| PAGE | → | CONTENT+ |
| CONTENT | → | asset |

Remark: The XML-DTD above (figure 4, as well as the following example in figure 8) does not include subsection elements: it uses a simplified specification, which in many cases is sufficient. For more complex learning units, the XML-DTD can easily be customized; the generator tool ‘AdLeR’ is prepared to deal with subsections.

This data model allows expressive possibilities of the search functionality. The query processing itself is invisible (deriving/ inheriting and calculation of attributes and attribute values), but some extended parameters of the search process can be controlled by user input directly using an appropriate search form like in figure 8.



**Search Form**

| | | | |
|---|---|---|---|
| Title: | Kinetosis | or ▾ | Weight: 10 |
| Level: | Beginner ▾ | and not ▾ | Weight: 4 |
| Content type: | Summary ▾ | and ▾ | Weight: 10 |
| Full text search: | Car | | Weight: 10 |

Add search field: ▾

Figure 8: Search form

The data are extracted from the XML document and are stored as a file, because the mini course is used for offline learning, so that an online database system is not appropriate.

# 5. CONCLUSION AND FURTHER WORK

## 5.1. Summary

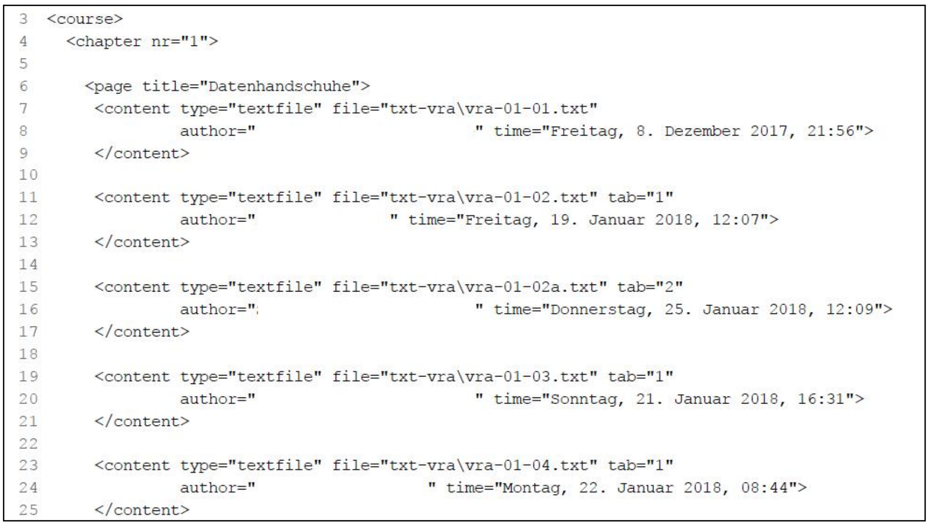
Considering the observed gaps in previous learning, we offer mini courses in the sense of microteaching, so that students can catch up on or repeat the required knowledge.

Undergraduate Students can benefit from advanced students, who might explain the learning matter in another way than the teacher by using mini courses for self-regulated learning. Advanced students themselves benefit from a better understanding of the learning matter by putting the theory of the visited lessons into practice (*learning by teaching*).

Concretely, students are involved in the development of our tool in the following part disciplines of computer science:

- Software engineering/ programming: The students have to understand existing software, and have to specify and implement new features – including testing and debugging.
- Human-computer-interaction: The learned concepts of user experience have to be adapted to the screen design of the generated mini courses, as well as to the design of the user interface for the generation tool *AdLeR*.
- Theoretical computer science & database theory: The concept of formal languages (theoretical computer science) has to be adapted to the data modelling using context-free grammars inclusive ranking and query processing (database theory).

The mentioned benefit for undergraduate students (learning missed teaching content) can be adapted for advanced students as well: Every now and then interesting topics or deeper analysis cannot be handled in the lessons because of the tight schedule. Mini courses can fill this gap. For example, in our lecture ‘Virtual Reality and Animation’, the students picked out interesting topics. The results were combined to a mini course. The appropriate XML file is depicted in figure 9 (reduced and extracted; the names of the students are disguised).



```
<course>
  <chapter nr="1">

    <page title="Datenhandschuhe">
     <content type="textfile" file="txt-vra\vra-01-01.txt"
              author="                      " time="Freitag, 8. Dezember 2017, 21:56">
     </content>

     <content type="textfile" file="txt-vra\vra-01-02.txt" tab="1"
              author="              " time="Freitag, 19. Januar 2018, 12:07">
     </content>

     <content type="textfile" file="txt-vra\vra-01-02a.txt" tab="2"
              author="                      " time="Donnerstag, 25. Januar 2018, 12:09">
     </content>

     <content type="textfile" file="txt-vra\vra-01-03.txt" tab="1"
              author="                      " time="Sonntag, 21. Januar 2018, 16:31">
     </content>

     <content type="textfile" file="txt-vra\vra-01-04.txt" tab="1"
              author="                  " time="Montag, 22. Januar 2018, 08:44">
     </content>
```

Figure 9: XML document for a mini course (extract)

## 5.2. Future Work

Our future steps are divided into two parts: the tool itself and using the tool.

First, we will evaluate and improve the user interface of the generation tool and we will improve the implementation of the search functionality and other features (graphical sitemap, etc.). In addition, an HTML preview is planned.

Because the particular learning objects (assets) are separated from the kind of presentation, a multimodal publication of the learning matter is possible: besides HTML web pages, the output could be a PDF file or slides for a presence presentation. XML tags can mark this. When considering a PDF output, the intermediate step LaTeX is also interesting: The XML structure is translated into LaTeX, which can easily be transformed to a PDF file; furthermore, the LaTeX document can be adapted to the special needs after compilation of the mini course.

Furthermore, we observe an increased use of mobile phones for (self-regulated) learning. Thus, we intend to use augmented reality learning apps like the use of so-called ‘Learning Factories’, in which the reality is augmented with additional information [12]. One-step further is the visualization of non-visible processes or abstract issues, which could improve the motivation and the learning success [20].

Secondly, we will motivate students to create new assets and reuse existing learning material in order to generate more mini courses. This brings the students to recapitulate their knowledge and force them to structure the learning subjects. The advanced students are nearer to undergraduate students, they know about their own problems and are able to present the learning matters in another way as the teachers. The tool AdLeR allows concentrating on the intended learning goal, the transfer to a course is done automatically. An intensive evaluation is planned in order to confirm or disprove the effects to the learning students – regarding the comprehension (found by tests) and the motivation (found by questionnaires).

## ACKNOWLEDGEMENTS

The program 'Hochschuldidaktische Projekte' (Baden-Wuerttemberg, Germany, 2017) funded previous parts of this work. The authors would like to thank the students Paula Schwenger, Florian Russl, Ufuk Karaali, and Thomas Dennighoff for the implementation of parts of the tool 'AdLeR'.

## REFERENCES

[1] Abiteboul,Serge (1999) "On Views and XML". Proc. of ACM Symposium on Principles of Database Systems, pp 1-9. 1999

[2] Allen,Dwight William & Cooper,James Michael & Poliakoff,Lorraine (1972) "Microteaching", U.S. Department of Health, Education, and Welfare, Office of Education, National Center for Educational Communication. 1972

[3] Aslan,Safiye (2015) "Is Learning by Teaching Effective in Gaining 21st Century Skills? The Views of Pre-Service Science Teacher", Educational Sciences: Theory and Practice, 2015

[4] Bergmann, Jonathan & Sam,Aaron (2012) „Flip your classroom – Reach every Student in Every Class Every Day", International Society for Teaching in Education, 2012

[5] Boekarts,Monique (1999) "Self-regulated learning: where we are today", International Journal of Educational Research, Volume 31, Issue 6, pages 445-457, 1999

[6] Campos-Sánchez, Antonio &del Carmen Sánchez-Quevedo,María &Crespo-Ferrer,Pascual Vicente & García-López,José Manuel & Alaminos,Miguel (2013) "Microteaching as a self-learning tool. Students' perceptions in the preparation and exposition of a microlesson in a tissue engineering course", Journal of Technology and Science Education, Vol. 3, No. 2, 2013

[7] Garrision,D. Randy & Heather,Kanuka (2004) "Blended Learning: Uncovering its transformative potential in higher education", The Internet and Higher Education. Volume 7, Issue 2, pp 95-105. 2004

[8] Hug, Theo (Ed) (2007) "Didactics of Microlearning", Waxmann. 2007

[9] Lecon, Carsten & Seehusen,Silke (2002) "Combining Structure Search and Content Search for Online Courses", 13th International Workshop on Database and Expert Systems Applications (DEXA), 2-6 September 2002, Aix-en-Provence (France)

[10] Lecon,Carsten &Hermann,Marc (2019) "Flexible Search Function for Online Courses in the Sense of Attribute Grammars",15th International Conference on Information Technology & Computer Science. 20-23 May 2019, Athens (Greece)

[10] Lueckemeyer,Gero (2015) "Virtual blended learning enriched by gamification and social aspects in programming education", 10th International Conference on Computer Science & Education (ICCSE), 22-24 July 2015, Cambridge (UK)

[11] Martín-Gutiérrez,Jorge & Mora,Carlos Efrén & Añorbe-Díaz,Beatriz & González-Marrero.Antonio (2017) "Virtual Technologies Trends in Education", EURASIA Journal of Mathematics Science and Technology Education, 2017 13(2):469-486

[12] Mueller-Frommeyer,Lena C. et. al. (2017) "Introducing competency models as a tool for holistic competency development in learning factories: Challenges, example and future application", 7th Conference on Learning Factories (CLF 2017), 2017

[13] Pintrich,Paul R.& V. de Groot, Elisabeth (1990) "Motivational and self-regulated learning components of classroom academic performance", Journal of Educational Psychology, 82(1), pp 33-40. 1990

[14] Pintrich,Paul R. (2000) "The Role of Goal Orientation in Self-Regulated Learning", Handbook of Self-Regulation, chapter 14, Academic Press, pages 451-502, 2000

[15] Reghizi, Stefano Crespi & Breveglieri, Luca & Morzenti, Angelo (2019) „Formal Languages and Compilation", 3rd edition. Text in Computer Sciences, Springer, 2019

[16] Seehusen, Silke & Lecon, Carsten & Kaben, Cay (2000)„Specification of learning trails in virtual courses", Frontiers in Education Conference. FIE 2000, Vol. 2. 2000

[17] Stansburry,Meris (2010) "Teachers turn up learning upside down", eSchool News. December 22nd 2010. https://www.eschoolnews.com/2010/12/22/teachers-turn-learning-upside-down/?all (read on 02/16/2020)

[18] Stollhans. Sascha (2016) “Learning by teaching: developing transferable skills”, in E. Corrandi & K. Borthwick & A. Gallagher-Brett (eds), Employability for languages: a handbook (pp 161-164), Dublin: Research-publishing.net. 2016

[19] Stuschka, Ulrike & Linnemann,Volker(1995) “Attributierte Grammatiken als Werkzeug zur Datenmodellierung“. Datenbanken in Büro, Technik und Wissenschaft (BTW), Dresden (Germany), 22-24 March 1995 (in German)

[20] Yuen,Steve Chi-Yin & Yaoyuneyong,Gallayanee & Johnson,Eric (2011) “Augmented Reality. An Overview and Five Directions for AR in Education”, Journal of Educational Technology Development and Exchange (JETDE), Vol. 4, Iss. 1, Article 11, 2011

## AUTHORS

**Prof. Dr. Carsten Lecon**

Short Bio

- Study of computer science
  (Technical University Braunschweig, Germany)
- Software Quality Assurance
  (Siemens AG, Braunschweig)
- Database systems, Media archives
  (University Luebeck, Germany)
- Virtual University of Applied Sciences
  (FH Luebeck, Germany)
- Since 04/2004 Professor for media computer science
  (Aalen University for Applied Sciences, Germany)

**Dr. Marc Hermann**

Short Bio

- Study of Computer Science
  (Ulm University, Germany)
- Certificate of Higher Education Pedagogy
  (Baden-Württemberg Certificate)
- Software Developer
  (Inneo Solutions GmbH, Germany)
- Senior Developer
  (Veroo Consulting GmbH, Germany)
- Since 04/2009 Lecturer for several courses
  like C, C++ and Java Programming, Software Engineering, Human Computer Interaction
  (Aalen University for Applied Sciences, Germany)

# USABILITY EVALUATION TO IMPROVE OPERATION INTERFACE OF WIRELESS DEVICE: PRESSURE RANGE OF TOUCH SENSOR

Sangwoo Cho[1] and Jong-Ha Lee[2]

[1]The Center for Advanced Technical Usability and Technologies,
Keimyung University, Daegu, South Korea
[2]Department of Biomedical Engineering, School of Medicine,
Keimyung University, Daegu, South Korea

### *ABSTRACT*

*Usability evaluation of wireless device can find improvement about user convenience. This study investigated natural finger pressure range when presses touch sensor. Fifteen adults (Male: 10, Female: 5, Age: 26.13 ± 3.98 years) were recruited in this experiment. Subjects carried out a usability evaluation about wireless device operation. The usability evaluation measured finger pressure on touch sensor operation of wireless device using finger pressure sensor. Subjects performed 1.76±0.95 times until pressing the touch sensor to complete task (t = 3.091, p = 0.008). In comparisons between natural movement and the movement to complete task, more finger pressure value was decreased in natural movement than the movement to complete task (t = -2.277, p = 0.039). This study found a finger pressure values to improve effectiveness of wireless device operation interface. Finger pressure value was presented to induce natural movement for the use of touch sensor.*

### *KEYWORDS*

*Usability Evaluation, Operation Interface, Finger Pressure Range, Wireless Device*

## 1. INTRODUCTION

Wireless devices are being widely disseminated in various fields with the development of Bluetooth technology. In particular, Bluetooth technology had been utilized in IoT-based home appliances and healthcare filed. Manufacturers are increasingly interested in usability evaluations to improve the efficiency of their user interface.

Many usability studies have been conducted on the user interface of Wireless devices [1, 2, 6]. Traditional usability evaluations used interview-based subjective scoring methods to validate product-use interface [3, 4]. According to the development of measuring devices (motion, cognition and sensation) that can analyse human factors, it is possible to investigate the user interface that can induce natural behavior in usability evaluations [2, 7]. Chang et al. developed a usability evaluation method that evaluates the efficiency of product-use behavior by motion analysis technology [2]. Human factor analysis can provide an objective result of user interface from usability evaluation [5, 8].

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 45-49, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101104

Usability evaluation of wireless device can find improvement about user interface. To improve effectiveness of wireless device, it is necessary to reduce the joint load in product use. This study investigated finger pressure range for the use of touch sensors in wireless devices.

# 2. METHODS

## 2.1. Subjects

We recruited 15 healthy adults with no history of neurological disorders (F=5, M=10, 26.13 ± 3.98 yrs) as shown in Table 1. All subjects that consented to participate in this study were informed about the experimental protocol. All subject had the use experiment of wireless devices in daily life.

Table 1. Demographics of subjects

| **Subject** | **Gender** | **Age** | **Occupation** | **Wireless device Use experiment* (O/X)** |
|---|---|---|---|---|
| S1 | Male | 27 | Office Worker | O |
| S2 | Female | 26 | Office Worker | O |
| S3 | Male | 26 | Undergraduates | O |
| S4 | Male | 26 | Office Worker | O |
| S5 | Male | 26 | Undergraduates | O |
| S6 | Male | 26 | Office Worker | O |
| S7 | Female | 21 | Undergraduates | O |
| S8 | Female | 22 | Undergraduates | O |
| S9 | Male | 24 | Undergraduates | O |
| S10 | Male | 20 | Undergraduates | O |
| S11 | Male | 24 | Undergraduates | O |
| S12 | Male | 36 | Office Worker | O |
| S13 | Male | 28 | Office Worker | O |
| S14 | Female | 30 | Office Worker | O |
| S15 | Female | 30 | Office Worker | O |

* O: Subject has experience controlling touch sensors on wireless devices such as Smart phone.

## 2.2. Experiment protocol

Wireless device has two functions both standby mode and operating mode. The standby mode was deactivated touch sensor of wireless device. The operating mode can control a volume size and play/stop of audio using touch sensor of wireless device. The operating mode is activated when the user presses the touch sensor harder than the reference value. The usability evaluation requires finger pressure to be applied to the touch sensor to move from standby mode to operating mode. All subjects performed the usability evaluation three times. Natural finger movement is that exclude finger joint load and subject was pressed touch sensor without uncomfortable of operation interface. The first attempt was performed with natural finger movement. If the subject fails to move to the operating mode in the first attempt, increase the pressure on the fingers until moving to the operating mode. The usability evaluation measured finger pressure on touch sensor of wireless device using finger pressure sensor (pliance®, novel.de).

### 2.3. Data analysis

The One sample t-test has been used in order to investigate the effectiveness of touch sensor operation using the SPSSWIN 20.0 software package. In addition, the paired t-test has been used to compare the finger pressure value between natural finger movement and finger movement to complete task.

## 3. RESULTS AND DISCUSSION

### 3.1. Problem of touch sensor operation interface

If the touch sensor operation interface is efficient, the number of touch sensor presses will be close to 1. Five subjects out of subjects were completed the task that presses touch sensor in natural finger movement. All subjects performed 1.76 ± 0.95 times until pressing the touch sensor to complete task (*t = 3.091, p = 0.008*) as shown in Figure 1. Subjects repeatedly pressed the touch sensor until operation mode activated. This result suggests that subjects had difficulty pressing the touch sensor of wireless device. To improve effectiveness of touch sensor operation, finger pressure value should be measured at natural finger movement.

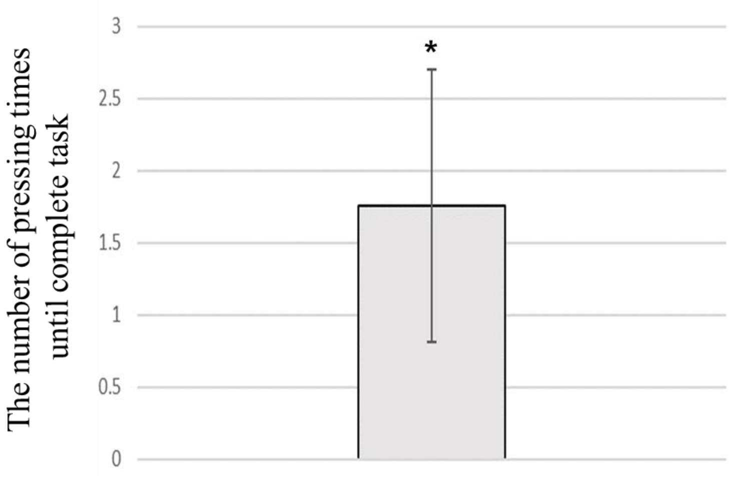


Figure 1. The number of pressing times until complete task (*: p < 0.01)

### 3.2. Change of finger pressure by finger joint load

The One sample t-test has been used in order to compare the training effects between tasks. All inserts, figures, diagrams, photographs and tables must be centre-aligned, clear and appropriate for black/white or greyscale reproduction. The mean finger pressure value of natural movement was 9.31 ± 2.64 kPa. The mean finger pressure value to complete task was 11.00 ± 3.71 kPa. In comparisons between natural movement and the movement to complete task, more finger pressure value was decreased in natural movement than movement to complete task. (*t = -2.277, p = 0.039*)

Table 2. Comparison of finger pressure value between natural movement and the movement to complete task

| | **Natural finger movement** | **finger movement to complete task** | **p-value** |
|---|---|---|---|
| Finger pressure value (mean ± sem(kPa)) | 9.31± 2.64 | 11.00±3.71 | 0.039 |

In the results of finger movement to complete task, finger pressure value was shown that finger joint load of subjects is increasing. On the other hand, finger pressure range was decreased in natural finger movement that excluded finger joint load. This result suggests that the response range of the touch sensor should be adjusted to improve the user interface of the wireless device.

## 4. CONCLUSIONS

This study investigated finger pressure range to improve operation interface of touch sensors in wireless devices. The limitations of this study were that subject group did not include elderly group and teenager group. But, considering the fact that wireless devices have become more popular in the last 10 years, young adults can buy more wireless devices than elderly group and teenager group. Finger pressure sensor was used to evaluate an effectiveness of touch sensor interface. Finger pressure range could show natural finger movement that excluded joint load on touch sensor operation. This study found that the response range of the touch sensor should be adjusted to improve the user interface of the wireless device.

### ACKNOWLEDGEMENTS

The Foundation Assist Project of Future Advanced User Convenience Service" through the Ministry of Trade, Industry and Energy (MOTIE) (R0004840, 2020) and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1D1A1B04031182 / NRF-2020R1I1A1A01072101) and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI17C2594).

### REFERENCES

[1] Choi, Jiho. Lee, Seongil. & Cho, Joo Eun, (2011) “The Usability Evaluation of Mobile Phone Interfaces Designed for the Elderly”, *Journal of the Ergonomics Society of Korea*, Vol. 30, No. 1, pp265-273.

[2] Chang, Joonho. Jung, Kihyo. Lee, Wonsup & You, Heecheon (2017) “Development of usability evaluation method using natural product-use motion”, *Applied Ergonomics*, Vol. 60, pp171-182.

[3] Ahmad, Naseer. Boota, M Waqas & Masoom, A Hye, (2004) “Smart phone Application Evaluation with Usability Testing Approach”, *Journal of Software Engineering and Applications*, Vol. 7, pp1045-1054.

[4] Lee, S.heum, (1999) “Usability Testing for Developing Effective Interactive multimedia software: Concepts, dimensions, and procedures”, *Journal of Educational Technology & society*, Vol. 2, No. 2, pp1-12.

[5] Strawderman, Lesley & Koubek, Rick, (2008) “Human factors and usability in service quality measurement”, *Human Factors and Ergonomics in Manufacturing & service Industries*, Vol. 18, No. 4, pp454-463.

[6] Chang, Hsien-Tsung. Tsai, Tsai-Hsuan. Chang, Ya-Ching. & Chang, Yi-Min (2014) “Touch Panel usability of elderly and children”, *Computers in Human Behavior*, Vol. 37, pp258-269.

[7] Ocak, Nihan & Cagiltay, Kursat (2017) “Comparison of Cognitive Modeling and User Performance Analysis for Touch Screen Mobile Interface Design”, *International Journal of Human-Computer Interaction*, Vol. 33, No. 8, pp33-641.

[8] Chang, Youli. L’Yi, Sehi. Koh Kyle & Seo, jinwook (2015) “Understanding Users’ Touch Behavior on Large Mobile Touch-Screens and Assisted Targeting by Tilting Gesture”, *CHI’15:Proceedings of The 33$^{rd}$ Annual ACM Conference on Human Factors in Computing Systems*, pp1499-1508.

## AUTHORS

**Sangwoo Cho** He received the B.A. degree in computer engineering from Kyungnam University, Changwon, South Korea, in 2006, and the Ph.D. degree in biomedical engineering from Hanyang University, Seoul, South Korea, in 2013. He is currently senior research engineer of the Center for Advanced Technical and Usability and Technologies, Keimyung University, Daegu, South Korea. His interests include virtual reality technology in rehabilitation therapy, usability test, and human computer interaction.

**Jong-Ha Lee** He received the B.A. degree in electronic engineering from Inha University, Incheon, South Korea, in 2000, the M.S. degree in electrical and computer engineering from New York University, New York, USA, in 2005, and the Ph.D. degree in electrical and computer engineering from Temple University, Philadelphia, USA, in 2011. He is currently an associate professor of department of biomedical engineering, School of Medicine, Keimyung University, Daegu, South Korea. He is currently director of the Center for Advanced Technical and usability and Technologies, Keimyung University, Daegu, South Korea. His interests include artificial intelligence algorism in medical science, usability test, and image processing.

# COVID CT NET: A TRANSFER LEARNING APPROACH FOR IDENTIFYING CORONA VIRUS FROM CT SCANS

Smaranjit Ghose and Suhrid Datta

SRM Institute of Science and Technology, India

## *ABSTRACT*

*The pandemic of COVID-19 has been rapidly spreading across the globe since it first surfaced in the Wuhan province of China. Several governments are forced to have nationwide lockdowns due to the progressive increase in a daily number of cases. The hospitals and other medical facilities are facing difficulties to cope with the overwhelming number of patients they can provide support due to the shortage in the number of required medical professionals and resources for meeting this demand. While the vaccine to cure this disease is still on the way, early diagnosis of patients and putting them in quarantine has become a cumbersome task too. In this study, we propose to build an artificial intelligence-based system for classifying patients as COVID-19 positive or negative within a few seconds by using their chest CT Scans. We use a transfer learning approach to build our classifier model using a dataset obtained from openly available sources. This work is meant to assist medical professionals in saving hours of their time for the diagnosis of the Coronavirus using chest radiographs and not intended to be the sole way of diagnosis.*

## *KEYWORDS*

*COVID-19, Deep Learning, CT Scans, Deep Convolutional Neural Networks, computer tomography scans.*

## 1. INTRODUCTION

The sars-cov2 virus responsible for COVID-19 is rich in a cell surface receptors called angiotensin converting enzyme 2 [3].The virus after entering to the body attaches to the host cells and creates copies of itself. Our immune system forms antibodies to fight the virus but it is unable to keep up with the myriad copies of the virus, it spreads to the other organs of the bodies. The lung gets affected in this process as it suffers from acute respiratory disease, which is seen in the radiographs as a white space [15].The primary process for diagnosing the presence for sars-cov2 is by doing a polymerase chain reaction testing. It is a laboratory procedure in which the ribonucleic(RNA) and deoxyribonucleic acid(DNA) are used for finding the exact volume of ribonucleic acids by the help of fluorescence. The samples for this procedure is collected by inserting swab into the nasal area for collecting the secretions. The process is very long and complex and there is shortage of test kits in some countries.

A possible substitute to the PCR testing is the use of chest radiographs like computer tomography scans [2] [6] and x-rays for detecting the presence of the virus. However, doing it manually is cumbersome and requires a specific skill set. In order to automate this process, we propose to make use of deep learn-ing [8].As deep convolutional neural networks are known to work on images, we use CNN to classify between chest x-rays as healthy or infected with COVID

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 51-57, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101105

[7] [10].A Convulational network [9] is used to extract meaningful features from images and differentiate amongst them by using certain learning parameters. The layers in it are grouped as convolutional layers, pooling layers, fully connected layers and at last an activation function is used in case of a classification task. Due to the small size of the dataset and lower quality of images ,we used transfer Learning. Transfer learning uses a model trained on large datasets like imagenet [5] and MS coco[4]and the pretrained weights collected are taken and layers are added over it to train the model on the new dataset.

## 2. DATASET DESCRIPTION

The dataset [16] that we used for this study has been collected from a public repository that consists of CT scan images amongst them 349 images are CT scan images collected from 216 COVID patients and 397 images that were CT scan images of healthy patients. The dataset is culmination of different CT scan images collected from websites such as medRxiv, bioRxiv, NEJM, JAMA, Lancet. The dataset also included the age, gender of the patient and the location from where the CT scans were taken.

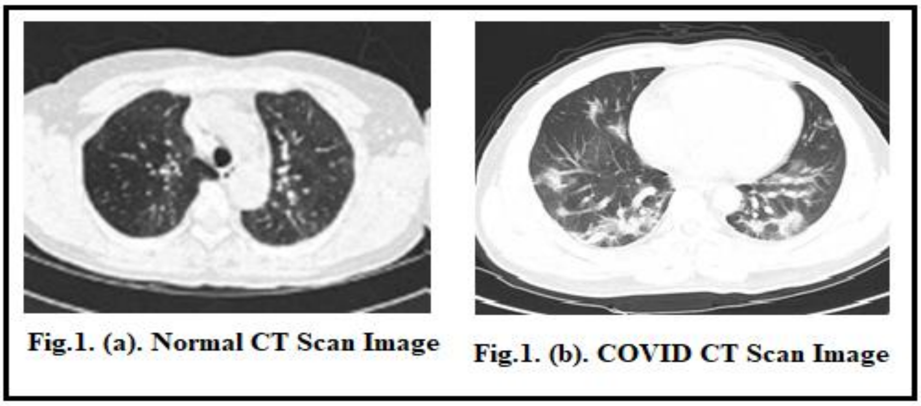


Fig.1. (a). Normal CT Scan Image

Fig.1. (b). COVID CT Scan Image

**Fig.1.** Normal and COVID CT Scan Images

The above figure.1. (b) is an example of a CT-scan belonging to a COVID patient and the figure1. (a) shows an example image orfl CT-scan image belonging to a healthy patient taken from the dataset.

## 3. PREPROCESSING

We performed the following pre-processing steps for facilitating better feature extraction during our model training.

### 3.1. Cleaning

The dataset comprised certain images which contained markings or additional labelling as they were taken from journals and books. We choose to remove all the images.

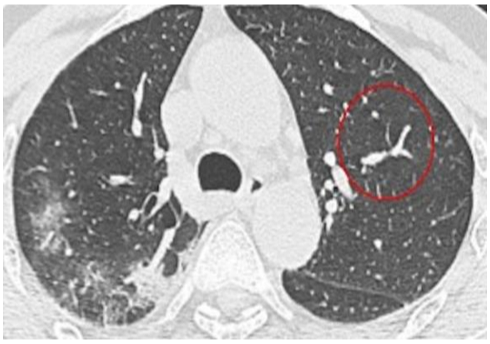


**Fig.2.** COVID CT Scan Image.

As seen in figure 2 it contains a red circle, these kinds of images were removed and a custom dataset was constructed from the dataset.

## 3.2. Data Augmentation

Owing to the limited size of the dataset we generated a custom dataset by performing data augmentation, data augmentation is a method used for increasing the size of the dataset by generating different iterations of the samples in the dataset. we subjected the images to different methods which included rotating the image to different perspectives, width and height shifting of the image, flipping the image and also padding the image. This method also helps us to address the class imbalance problem, reduction of over fitting and also improve the convergence rate of the model to yield a better accuracy. **Table.1** shows the increase in number of images after preprocessing.

**Table.1.** number of images in each class after pre-processing

| Classes | No.of images |
|---|---|
| COVID-Negative | 5000 |
| COVID-positive | 5000 |

## 3.3. Converting the pixel values to Hounsfield Units

In this method we converted all the pixels in a CT scan image to Hounsfield Units, this gives the relative radio density that is used for measuring CT scan images [12].This was done because the dataset the images used were of low quality that lead to the loss of hounsfield units.

## 3.4. Normalization

The pixel values in the image consists of integers, the lower the value the lower will be the learning time of the neural network, so that is why we normalized images in which the max bound was set to 400.0 and the minimum bound was set to -1000.0.

### 3.5. Zero Centring

This refers to a pre-processing method in which the mean value is subtracted from each data point, in our case the zero centred value was found to 0.25.

### 3.6. Model Architecture

In our study we developed a deep convolutional neural network for classifying between two classes of CT-scans. We had to collect the data from various public sources, there was limited number of images available and we had to compromise with the image quality as well hence we used transfer learning [11].Transfer learning takes advantage of previous knowledge of extraction that was obtained from training the network in datasets like imagenet [5].The pre-trained model obtained after training a model on a large dataset can be used for a image classification task.The main intuition behind the use of transfer learning is if any model is trained on a large dataset, we can make use of the feature map without requiring to train the model again from scratch on a large dataset.

We make use of Efficient B3 [13] for our experiment, the architecture of Efficient Net B3 optimizes flops by using a multi neural search architecture. The Convolution layers of efficient net are divided into two parts point wise convolution and depth wise convolution, this helps in reducing calculation time while having minimum loss in accuracy. The MBconv block in Efficient Net first extends to the channels of the images and then compresses them which results in lesser number of skipped connections unlike other architectures resnet [14].Efficient Net also utilises compound scaling, in compound scaling the length breadth and width of the network is increased with respect to the baseline architecture of Efficient Net as seen in table 2.

**Table.2.** Model Architecture

| stage | Operator $F^{\wedge}_{i}$ | Resolution $H_i$ * $W_i$ | # Channels $C_i$ | # Layers $L_i$ |
|---|---|---|---|---|
| 1 | Conv 3x3 | 224x224 | 32 | 1 |
| 2 | MBConv1, k3x3 | 112x112 | 16 | 1 |
| 3 | MBConv, k3x3 | 112x112 | 24 | 2 |
| 4 | MBConv, k5x5 | 56x56 | 40 | 2 |
| 5 | MBConv, k3x3 | 28x28 | 80 | 3 |
| 6 | MBConv,k5x5 | 14x14 | 112 | 3 |
| 7 | MBConv, k5x5 | 14x14 | 192 | 4 |
| 8 | MBConv, k3x3 | 7x7 | 320 | 1 |
| 9 | Conc 1x1 & pooling & FC | 7x7 | 1280 | 1 |

The layers in Efficient Net have been increased by keeping a fixed constant ratio, this helps in boosting accuracy of the model. In order to make a sequential model we used Efficient Net as the the head model and added layers. The output obtained from the last layer is fed to an average pooling layer where the input is down sampled by a kernel of size 4x4.The output is fed to a flatten layer for converting the matrix of features into vectors which are then fed to a dense layer. Relu activation function is used to introduce non-linearity. In the last layer sigmoid activation is used for classifying the images into 2 classes..The operator column in Table 2 shows the exact orientation of blocks, the resolution represents the input resolution that is gonna be utilised by the blocks and the channels represents the number of output channels of the blocks and the layers represents the number of times the blocks were repeated

## 3. RESULT

The dataset was split into a ratio of 8:2 that is 80% for train and 20% for test. The Learning rate of our model is set to 5e-5 with a batch size of 16 on the tensorflow2.0 [1] framework. The model was then trained for a cycle of 100 epochs. The figure 3 shows the accuracy curve obtained after training the model, a validation and test accuracy of 100% was obtained.

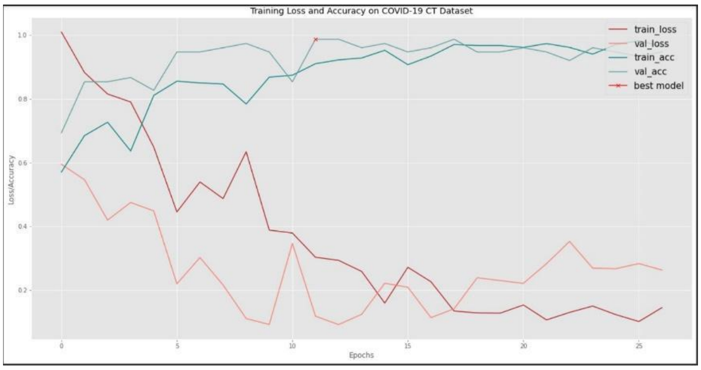


Fig.3. Training Accuracy.

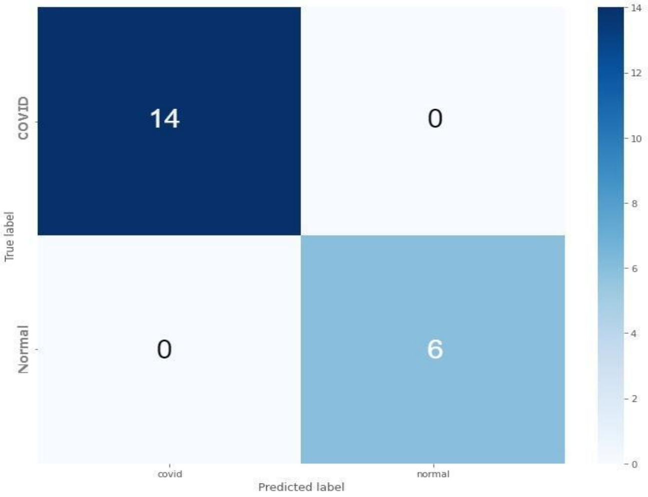


Fig.4. Confusion Matrix.

Figure.4. shows the confusion matrix that was obtained from our model. The confusion matrix defines the performance of the model on a set of data in which the score of certain parameters are known.

**Precision:** This score refers to the number of predictions about patients having COVID-19 was true, the score obtained was 1.00.

$$\text{Precision} = TP/(TP + FP) \qquad (1)$$

**Recall:** It gives a measure of the number of classifications of patients having COVID-19 the model can predict correctly and the score obtained was 1.00.

$$\text{Recall} = Tp/(TP + FN) \qquad (2)$$

**F1 score:** The F1 score being a function of precision and recall determines the number of instances our model accurately classifies without missing a significant number of instances. The F1 score obtained was 1.00.

$$\text{F1 score} = 2 \text{ x } [(\text{Precision x Recall})/(\text{precision} + \text{Recall})] \qquad (3)$$

**Specificity:** It is a percentage of COVID Negative patients that were actually classified as COVID Negative, For our proposed model the score was 100%.

**Sensitivity:**It is the percentage of COVID Positive patients that were actually classified as COVID Positive.For our proposed model the score was100%.

**Table.3**. shows the different scores obtained from our architecture

| Classes | Precision | Recall | F1-score |
|---|---|---|---|
| COVID-Negative | 1.00 | 1.00 | 1.00 |
| COVID-Positive | 1.00 | 1.00 | 1.00 |

## 4. CONCLUSION

In this study we propose to automate the process of detection of COVID-19 using chest CT scans of patients with deep Convolutional Neural Networks. Under the hood, we used a transfer learning technique to leverage the benefits of Efficient Net for training our image classifier to categorize the CT Scans as COVID-19positive or negative. We made use of a transfer learning approach in which we used Efficient Net and customized it by adding layers to accurately classify convict images . We obtained a test and validation accuracy of 100% and equally highscores in other parameters. The data was obtained from a public dataset that was curated from different sources. It was subjected to preprocessing methods like data augmentation, conversion of pixels to hounsfield units, zero centering and normalization to improve feature extraction of our architecture. Our experiment is intended to be a starter work for automatic diagnosis of COVID-19 to assist the medical professional amidst this pandemic to serve the people in a more efficient way. It requires further clinical validations to be used as a fully fledged detection tool.

## ACKNOWLEDGEMENTS

We are sincerely thankful to Adrian Rosebrock, author of PyImageSearch for motivating us to go further with this study and also the authors of the paper COVID-CT-Dataset: A CT Image Dataset about COVID-19 for their efforts to collect CT-scans.

## REFERENCES

[1] Mart́ın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, CraigCitro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensor-flow: Large-scale machine learning on heterogeneous distributed systems.arXivpreprint arXiv:1603.04467, 2016.

[2] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, QianTao, Ziyong Sun, and Liming Xia. Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases.Radiology,page 200642, 2020

[3] Nancy J Brown and Douglas E Vaughan. Angiotensin-converting enzyme in hibitors. Circulation, 97(14):1411–1420, 1998.

[4] Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta,Piotr Doll´ar, and C Lawrence Zitnick. Microsoft coco captions: Data collection and evaluation server.arXiv preprint arXiv:1504.00325, 2015

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet:A large-scale hierarchical image database. In2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.

[6] Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang,and Wenbin Ji. Sensitivity of chest ct for covid-19: comparison to rt-pcr.Radiology,page 200432, 2020.

[7] Ophir Gozes, Maayan Frid-Adar, Hayit Greenspan, Patrick D Browning, HuangqiZhang, Wenbin Ji, Adam Bernheim, and Eliot Siegel. Rapid ai development cy-cle for the coronavirus (covid-19) pandemic: Initial results for automated detec-tion & patient monitoring using deep learning ct image analysis.arXiv preprintarXiv:2003.05037, 2020.

[8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton.Deep learning.nature,521(7553):436–444, 2015.

[9] Yann LeCun, L´eon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition.Proceedings of the IEEE, 86(11):2278–2324, 1998.

[10] Ming-Yen Ng, Elaine YP Lee, Jin Yang, Fangfang Yang, Xia Li, Hongxia Wang,Macy Mei-sze Lui, Christine Shing-Yen Lo, Barry Leung, Pek-Lan Khong, et al.Imaging profile of the covid-19 infection: radiologic findings and literature review.Radiology: Cardiothoracic Imaging, 2(1):e200034, 2020.

[11] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning.IEEE Transac-tions on knowledge and data engineering, 22(10):1345–1359, 2009.

[12] Uwe Schneider, Eros Pedroni, and Antony Lomax. The calibration of ct hounsfield units for radiotherapy treatment planning.Physics in Medicine & Biology,41(1):111, 1996.

[13] Mingxing Tan and Quoc V Le. Efficient Net: Rethinking model scaling for convo-lutional neural networks.arXiv preprint arXiv:1905.11946, 2019.

[14] Sasha Targ, Diogo Almeida, and Kevin Lyman. Resnet in resnet: Generalized Residual architectures.arXiv preprint arXiv:1603.08029, 2016.

[15] Wenling Wang, Yanli Xu, Ruqin Gao, Roujian Lu, Kai Han, Guizhen Wu, andWenjie Tan. Detection of sars-cov-2 in different types of clinical specimens.Jama,2020.

[16] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: a ct scan dataset about covid-19.arXiv preprint arXiv:2003.13865, 2020

## AUTHORS

**Smaranjit Ghose** is currently pursuing Bachelor of Technology in Computer Science and Engineering at SRM Institute of Science and Technology, Kattankulathur Campus. His research interests are medical imaging, explainable ai, machine translation and image enhancements.

**Suhrid Datta** is currently pursuing Bachelor of Technology in Computer Science and Engineering at SRM Institute of Science and Technology, Kattankulathur Campus.His main research interest is in machine learning, computer vision ,reinforcement learning and medical imaging.

# VIRTFUN: FUNCTION OFFLOAD METHODOLOGY TO VIRTUALIZED ENVIRONMENT

Carlos A Petry and Rodolfo J. de Azevedo

Institute of Computing, University of Campinas, Campinas, Brazil

### *ABSTRACT*

*The use of virtual machines (VM) has become popular with substantial growth for both personal and commercial use, especially supported by the progress of hardware and software virtualization technologies. There are several reasons for this adoption like: cost, customization, scalability and flexibility. Distinct domains of application, such as scientific, financial and industrial, spanning from embedded to cloud systems, taken advantage of this kind of machines to meet processing computational demands. However, there are setbacks: hardware handling, resources use, performance and management. This growth demands an effective support by the underlying virtualization infrastructure, which directly affects the hosts' capacity in datacenters and cloud environment that support them. It is evident that the host native processing performs better than VMs, especially when using accelerator devices, where the common solution is to assign each device to a specific VM, instead of sharing it among multiples VMs. Beyond performance issues inside the host, we need to consider the VM performance when using accelerator devices. In this context, it is necessary to provide efficient mechanisms to manage and run VMs which can take advantages of high-performance devices, like FPGAs or even from software resources on the host. To assist this challenge, this paper proposes a methodology to improve communication performance of applications running on the VMs, VirtFun. To do so, we developed a framework able to offload pieces of application's code (vFunction) to host by means of secure data sharing between the application and device. The results achieved in our experiments demonstrated significant acceleration capacity for the guest application vFunction. The speedup reached 340% compared to conventional network execution, reaching maximum slowdown of 2.8% in the worst case and near to 0% in the best case considering the native execution.*

### *KEYWORDS*

*Virtualization, performance, virtual machine, shared-memory.*

## 1. INTRODUCTION

Cloud computing and virtualization provide increase use of software and hardware computational resources like storage, network, and accelerator devices. Virtual machines (VM) extend support to several domains such as commercial, financial, and scientific spanning from embedded to cloud systems. An impacting factor to execute tasks of client applications falls on the performance of the virtual machine sustained by the host resources. Developers usually employ approaches such as increase of computational resources and distribute processing to meet design requirements and performance issues. However, once VMs rely on the host resources, the greater the amount of existing VMs, the greater will be the pressure on the host machine that support the VMs. Therefore, this significantly increases the processing demand on the host's computing resources. Systems such as processors and memory increase systematically their computational

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 59-74, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101106

performance and energy efficiency, but resources like network, storage and I/O devices does not have the same progress. Improving virtualization and hardware infrastructures leads to better support for multi-client environment, processing acceleration, resource management, flexibility and security. In addition, optimizing the use of those resources becomes essential to achieve these improvements. In this sense, accelerator resources, especially hardware accelerators, have increased their adoption as an alternative to speed up tasks' execution. A large spectrum of applications can improve performance and energy efficiency by offloading part of their computation to accelerator devices like GPUs and FPGAs. FPGA has significantly increased its use as tasks accelerations. Furthermore, due to the high demand for both scientific and commercial areas, cloud computing has been employing FPGAs as an effective alternative to increase client application performance, especially for tasks that benefits from pipeline.

An important feature for virtualization is the possibility of sharing data between virtual machines and the host system, especially for guest OS running on the same physical machine where the accelerators will be placed. Many approaches allow communication between guest and host OSs, but usually exchange small amounts of data or rely on mechanisms such as OS sockets and network TCP/IP. These methods require copying data between guest and host OSs, causing communication overhead due to the need for serialization and deserialization of data, beside of multiple buffers instances along the data path.

Therefore, this article presents a methodology to allow offload pieces of application's code (vFunction) to accelerator resources inside the host machine by means of the memory sharing binding the application and the resource. Our approach enables to share and manage data exchange efficiently from guest applications located on the same physical machine (host). The proposed approach takes into account security and isolation, providing performance similar to native application execution, besides allowing to develop applications using the standard development tools.

We do not communicate application with the accelerator resources by means of network mechanisms. Instead, the data exchange takes place though regions mapped in memory, where the guest applications will perform I/O operations writing and reading from the mapped memory. Developers can offload pieces of application to vFunctions, seeking to increase performance. We implemented the proposed methodology as a framework that uses several mechanisms like hypercalls, service, mapping and device drivers to enable a communication channel with such functions. The framework has a management and control agent named virtual function daemon (vFunD) which provides a huge administrative and control services. The device driver deals with vFunction both in software and in hardware, acting as an interface between service and the resource (accelerator). The exchange of data between application and vFunction takes place through reading and writing operations in the shared memory region.

The rest of the paper is organized as follow. Section 2 provides a state of art revision of approaches to accelerators resources and high-level software interface efforts. Section 3 presents the proposed methodology and framework implementation explanation. Section 4 provides the experiments carried out to test and validate the effectiveness of the proposed methodology validates by the framework implementation. Section 5 presents the conclusions of the article, future works, and application possibilities of the proposed methodology

## 2. RELATED WORK

### 2.1. Studies Related to Accelerators

Putnam et al. [1][2] presented a mechanism to improve datacenters, building reconfigurable FPGA to accelerate large-scale Bing services to face pressure on computing resources. The experiment was evaluated based on throughput and latency compared to the original implementation, exhibited performance gains, increasing the throughput by 95% and reducing latency by 29%, and also an increase of 10% in energy consumption, but the 95% performance gain compensated this difference.

Chen et al. [3] described an effort to include FPGAs in the Cloud to increase performance by accelerating multiple application domains. They mentioned issues to integrate these devices into the cloud: resources abstractions, sharing accelerators, interfacing applications and accelerators and security concerns. To overcome such issues, the authors address four requirements. Abstraction, FPGAs seen as resources, not low-level programmable devices. Sharing, to share FPGAs between multiples clients it is necessary to have a robust isolation mechanism. Compatibility, there are several design workflows for FPGAs which require a common interface to support variety of devices. Security, it is suitable that FPGAs execute independently and in isolation.

Byma et al. [4] proposed a FPGA accelerators integration into cloud OpenStack [5]. They read across multiples physical FPGAs accessed similarly to a virtual machine. They classified FPGAs into three categories: (i) infrastructure, where clients are unaware of them; (ii) appliances, where accelerators run in a box, e.g. Memcached [6][7]; (iii) computing resources, which allow clients to allocate hardware easily. They chose to use computing resources by better serving the cloud diversity.

Fahmy et al. [8] presented an approach to integrate FPGAs into datacenters and cloud to achieve efficiency in performance and power compared to CPUs and GPUs [9][10], targeting streaming applications. The authors faced challenges to integrate FPGAs: (i) dynamic reconfigurable support, (ii) effective communications between multiples accelerators, (iii) optimize FPGA resources over the time, and (iv) easy integration between accelerators and applications. While other works focus on static and monolithic accelerators, they proposed FPGAs as general deployed resources.

Asiatici et al. [11] proposed a methodology and a runtime system to make it easier to implement applications in FPGA, using a high-level API, a hardware and software execution model, and a shared memory model. The framework dynamically manages multiples accelerators allowing designing and mapping multiple accelerators in FPGA, enabling three development levels: DSL, HLS, and RTL. Applications host code run on the host CPU, and offloaded accelerator codes to the FPGA.

Several works proposed to integrate FPGA-accelerated to the cloud infrastructure [3] [4] [12] [14] using OpenStack framework to allow clients to create VMs are able to access FPGAs in a virtualized way. However, their approaches are difficult to integrate in the existing systems. The FPGA virtualization mechanism has to modify the host operating system or insert high overhead. These works do not support communication between virtualized hardware regions. Clients requiring a higher amount of logic might have issues placing their designs because the amount of resources into each virtual FPGA is limited.

Mbongue et al [16] proposed a paravirtualized virtio-based framework to communicate VMs and FPGAs, leveraging partial reconfiguration to run client's hardware VF. They were able to allow designers to request dynamically additional FPGA resources, adding a security layer and allowing runtime reassignment of FPGA resources.

A very common approach to manage FPGAs is by provisioning the FPGAs through OpenStack framework and letting the client connect and program the FPGA through traditional IP or MAC address of the FPGA [4][13][14], allowing the clients to use remote procedure calls or socket connections to communicate with the FPGA.

Some alternatives to Virtio-vsock are virtio-serial and virtual networking proposed by Hajnoczi [17]. The former is a virtual serial device that establishes connections between hosts and guests, but that leads to a limited number of channels. The latter is an approach for guest and host communication using a virtual network, but they can be extremely complex to develop.

## 2.2. High-Level Interfaces for Accelerators

The design of hardware modules is a highly specialized and time-consuming task. Various efforts, both in academia and industry, have been employed to make hardware development more accessible to non-hardware specialist developers. There are different abstraction levels to design FPGA modules, start from low-level using HDLs like Verilog and VHDL until high-level languages, as C++ and OpenCL [20].

Rise the abstraction level is the main approach to deal with hardware complexity. This methodology provides user-friendly interfaces facilitating the task of handling hardware inherent complexity. To improve development time, it is mandatory to optimize time-consuming FPGA design tasks. One of the efforts to address this challenge is being driven by Intel, the open source framework OPAE [21][22]. OPAE provides a high-level interface between hardware accelerators and client applications through the use of a software stack [23]. OPAE was developed keeping in mind to facilitate the CPU-to-FPGA interaction, especially targeting the Intel HARP [24][25] project that integrates on the same die the Xeon-CPU and FPGA Stratix. The framework provides an API among client applications and FPGA accelerators modules. It also offers an accelerator simulation environment (ASE) [26], a hardware and software co-simulation environment for any Intel FPGA, employed to test and prototype user designs [27]. In addition, the OPAE SDK [28] provides a service-oriented approach to use on user applications [23]. OPAE includes a set of drivers, APIs, and tools to take care of hardware specific actions like discovering, accessing, and reconfiguring the accelerator modules [21].

Although OPAE makes it easier to design FPGA modules, there are some issues to overcome. The OPAE's design flow does not support sharing accelerators among multiple applications [29]. Applying acceleration approaches such as [30][31] can lead to contention in software stack. Although OPAE offers higher abstraction, the designer still needs to have substantial knowledge of the underlying hardware details. Moreover, there is a lack of intrinsic support to deploy multiple acceleration modules in the same design, leading to manually instantiate each accelerator in a set of modules [27].

## 3. Virt-Fun: Accelerator Offload Methodology

This section presents details of the proposed methodology and the framework implemented to validate this methodology. We explain how a client application, running on a virtual machine on the same physical host (co-located) make use of vFunction.

### 3.1. Layers Assessment

Figure 1 shows the infrastructure considering four abstraction layers, from the client application on top to the hardware on bottom. Solid lines represent the flow related to data, whereas dashed lines indicate the flow related to control issues.

At topmost, the App Layer represents the client applications running on virtual machines, where each of them can execute several applications. Furthermore, there may be multiple instances of VMs running multiple applications each one at the same time. In this layer, client develop and submit their applications with the possibility to offload pieces of the code to the vFunctions on the co-located host. As usual, developers can write code in a variety of languages supported by the platform.

In the next layer, VMOS comprises a number of running VMs (VM1, VM2) and supporting client applications (depicted as: 1A, 1B, 2A, 2B) from a number of clients (e.g.: Cli1, Cli2). Each VM executes applications of only one client, and each client can launch a number of applications as they need or the platform supports. Each VM has a memory region to share data with user-defined vFunctions, however each client application maintains their data in a distinct region.

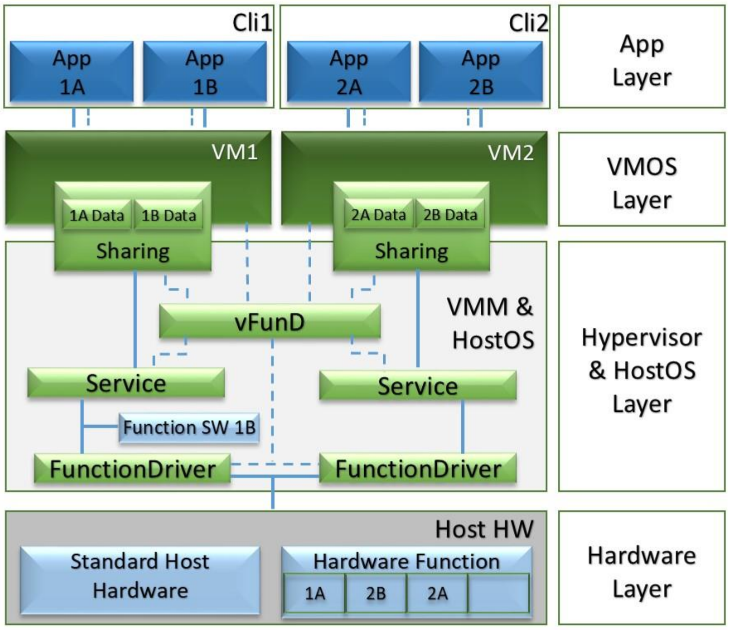


Figure 1. Overall layered framework view.

In the third layer, the VMM & HostOS integrates the main components of the framework. If a client decides to offload the vFunction to software, a version of that function will stay in place of the hardware version. The management software, vFunD, performs the administrative and security tasks by registering, launching, controlling and releasing all vFunctions. The manager will also provide security and isolation between all shared regions. The service component runs similarly to a daemon and is responsible for handling communication requests to/from the vFunction driver. In addition, it takes care of status conditions and command actions built-in in the communication protocol. FunctionDriver is a high-level driver responsible for supporting the low-level communication with hardware, or software version, communicating with the OS device-driver where the vFunction was instantiated. In case of hardware vFunction, the service uses the FunctionDriver to interact with the low-level device-driver supplied by the hardware manufacturer, included in the board support package (BSP) library.

The lower layer, Hardware, represents the physical host, e.g., CPU, memory and storage, etc, together with the underlying devices, labeled as Hardware Function. In the context of this work, we target especially the FPGAs devices.

## 3.2. Communication Flow

In this subsection, we present the execution flow of the proposed methodology applied to the framework, starting from the guest application and following until reach the vFunction on the host. Figure 2 shows a timed diagram consisting of three columns separated by two vertical lines. The leftmost column indicates the actions taken by the client in the context of the guest application, where the right column is responsible for actions belonging to the framework on the host. The central column shows the dynamics of events generated on both sides of the processing.

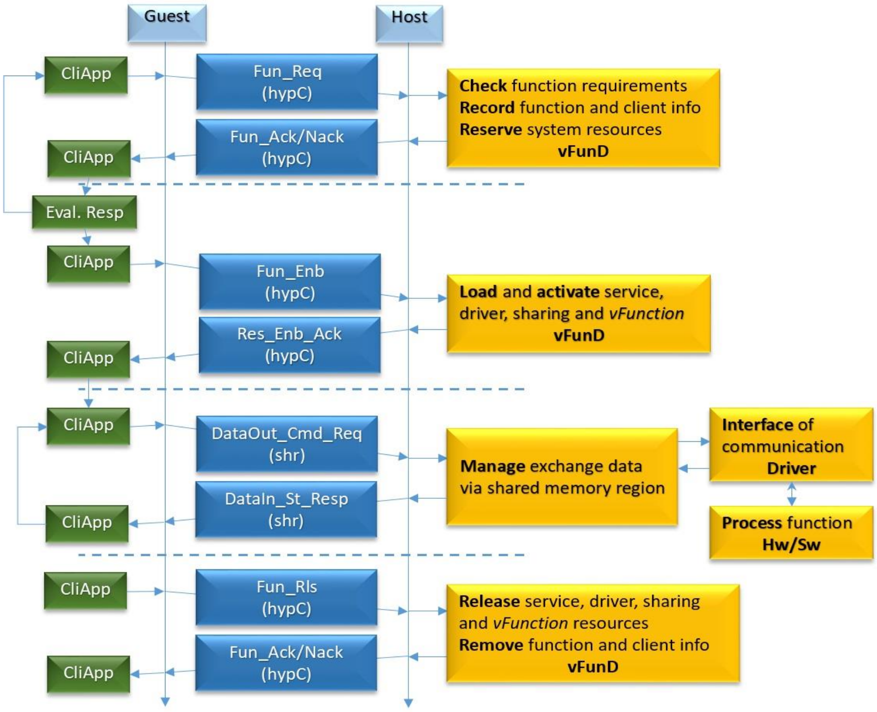

Guest
Host
CliApp
Fun_Req (hypC)
Check function requirements
Record function and client info
Reserve system resources
vFunD
Fun_Ack/Nack (hypC)
CliApp
Eval. Resp
CliApp
Fun_Enb (hypC)
Load and activate service, driver, sharing and vFunction
vFunD
Res_Enb_Ack (hypC)
CliApp
CliApp
DataOut_Cmd_Req (shr)
Manage exchange data via shared memory region
Interface of communication Driver
DataIn_St_Resp (shr)
Process function Hw/Sw
CliApp
CliApp
Fun_Rls (hypC)
Release service, driver, sharing and vFunction resources
Remove function and client info
vFunD
Fun_Ack/Nack (hypC)
CliApp

Figure 2. Time diagram showing the execution flow of the framework.

The execution follows the client-server model, thus the client takes all actions and the server returns the responses. Briefly, the sequence of events happens in four phases as follows.

### 3.2.1. Negotiation Phase

The client application requires that part of its computation to be executed in a vFunction on host. To do so, it invokes a standard systemcall, which in turn invokes a new hypercall implemented into host kernel to respond to this request, this the unique invasive method employed by the framework. The hypercall forwards the demand to the device resource manager (vFunD), that checks all requirements related to the request, such as resource availability, processing capacity, credential, etc. If successful, it reserves the suitable resources and records the request details, like client credential (clientID), application and function identity (AppID, functionID), and vFunction metadata (functionLoc). At the end of this phase, vFunD returns the request status including ack or nack conditions along with status reason.

### 3.2.2. Activation Phase

In this phase, the application analyses the vFunD return, and if function request becomes available, the application can proceed to next phase. Otherwise, the client application can request the same function in software, instead of in the hardware. However, if the reason for failure has been unrecognized resources, possibly this would prevent the execution to continue. On an ack return, the application issues another systemcall, this time requesting the vFunction activation. Again, vFunD processes the systemcall and load the service, driver, vFunction resources, and activates the shared memory region. Moving forward, vFunD updates the request status and launches the vFunction making it available. Finally, it returns the request status including ack or nack conditions along with the status response. On the guest side, once the vFunction is available the application advances to the next phase, where the vFunction computation will be exploited.

### 3.2.3. Computation Phase

The computation phase allows the application to exchange data with the vFunction in the host. Communication transfers data through the shared memory region in accordance to the commands defined in the aforementioned protocol and check the state conditions emitted by the service. The command set sends application's data to the vFunction to processes them in the host, while the command get takes the result of the computation. The command rst allows the application clear data inside vFunction if the project foresees such functionality. There are three essential status conditions for service: wait, when service detects that the vFunction is ready to process new requests; busy, indicates that the device is busy and performing other requested computations; done, stating that the device has finished a requested computation.

### 3.2.4. Release Phase

In this phase, the application notifies the vFunD to release all resources and information related to the ordered vFunction, also through a systemcall. vFunD processes the order calling a hypercall to stop the service, unload driver and vFunction resources and, deallocate the shared memory region. In addition, vFunD remove administrative entries such as credential, resources and any other information related to the vFunction previously instantiated. Finally, it returns the request status including ack or nack conditions along with the status reason. On the application side, the release confirmation enables the application to remove the vFunction.

## 3.3. Implementation and Data Flow

In virtualization, the guest OS runs over the same physical machine (co-located) where the host and hypervisor are also running. However, the guest OS may need to communicate with a resource that is located on the same physical machine or on a remote machine. Resources locally available can provide some desirable advantages and features. Locally provided resources make it possible to implement a file system shared between gest and host OSs. This approach is under development, but in the final stage, through of the VirtioFS project [18][19], which should be part of the Linux kernel mainline soon. Based on this approach, we can implement the data sharing between the guest application and the vFunction on the host through a shared resource by means of a file system mapped to a shared memory region. Through a memory mapped I/O region, the application on guest and the service on host can exchange data using simple variable assignments. This mechanism makes it possible to provide a local data flow mechanism without face the overhead of traditional communication, as the TCP/IP.

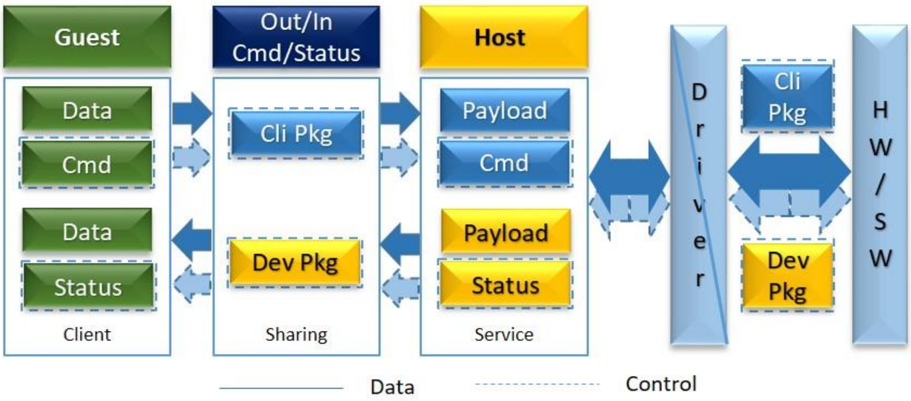


Figure 3. Data flow overview of the framework.

Within the application depicted to the left of Figure 3 (Client, green part), the developer accesses the shared resource by opening files via the traditional open() function and maps them into memory, assigning the returned address to a data structure by the mmap() function. We chose to separate the mapping in two parts: (a) dataout, used to send data for processing together with the client command (set, get, rst); (b) datain, used as a data receiver containing the result of computation together the state condition. The Figure 4 shows a code snippet of the above-mentioned dynamic.

```
(a) opening and mapping data sharing:
fd_out = open(file_out, O_RDWR | O_CREAT | O_SYNC);
dataout_srv = (argout_t*) mmap(NULL, BUFFER_SIZE,
   PROT_READ | PROT_WRITE, MAP_SHARED, fd_out, 0);
fd_in = open(file_in, O_RDONLY | O_SYNC);
datain_srv = (argin_t*) mmap(NULL, BUFFER_SIZE,
   PROT_READ, MAP_SHARED, fd_in, 0);

(b) assigning values to data share:
dataout_srv->cli_data.field1 = value1;
dataout_srv->cli_cmd = e_set;

(c) reading data sharing values:
while (datain_srv->srv_st == e_req_wait);
```

Figure 4. Dataflow code snippet.

Still in the center of Figure 3, the service (yellow host column) executes the dynamic similar to that performed by the client, regarding access to shared data. However, shared data is carried in reverse, and client commands are identified and manipulated to grant the management and communication with the vFunction resource.

```
<filesystem type='mount' accessmode='passthrough'>
  <driver type='virtiofs' queue='1024'/>
  <binary path='/path_to_virtiofs_daemon/virtiofsd'>
    <cache mode='always'/>
    <lock posix='on' flock='on'/>
  </binary>
  <source dir='/path_on_host/vm-share/app-share'/>
  <target dir='app-share-ID'/>
</filesystem>
```

Figure 5. Guest OS directives to enable the data sharing structure.

The central part of Figure 3 (dark blue column) corresponds to the shared data mechanism based on VirtioFS configured according to the rules outlined in Figure 5. To work properly, it is necessary to map a directory on the host OS inside the guest OS file system. There are different ways to configure this mapping, for this work we chose to use the XML writing directives describing the guest VM for the hypervisor in this file. A priori, one can attach the mount point into any directory path inside the guest. However, it is suggested to map the shared resource within the user's home or in the /var path. Permissions and ownership of files inside of hierarchy will be defined by the guest operating system.

The guest mount point will be restricted to the host directory and the directories below, in other words the guest sees the shared resource as the root mount point. It is not possible to escape from this hierarchy due to two security mechanisms: sandbox that implements kernel namespace for processes, and seccomp that restricts system calls invoked by the processes. vFunction provides the computation resources for guest applications through the driver and the specialized hardware or software on the host, as shown at the rightmost Figure 3 (light blue part).

Another choice made for this work was to split the driver functionality into two parts and, denoted in the driver box in Figure 3, by the diagonal line over component. This led to the device driver refactoring in two parts: (i) device critical functions, such as highly-device dependent features (low-level driver); (ii) remaining non-critical (highest level) functionalities (high-level driver). Devices manufacturers, here especially FPGAs, make their low-level drivers available through the platform's BSP. Design and synthesis tools, such as Quartus and Vivado, automatically produce high-level headers interfaces to integrate the service back-end as low-level interfaces. The high-level driver consists of three primary functionality: (i) device_driver_init, which registers and enables the driver making it accessible; (ii) device_driver_exit, which disables and removes the driver instance freeing the resources on the host OS; (iii) driver_function, which implements the communication interface with the low-level manufacturer driver. If client chooses to offload functions in software, the driver will communicate directly with that vFunction version carried out as a daemon.

```
/var
 ├─ vFun
     ├─ admin       ├─ service          ├─ vmcli
     │   └── vfm    │  └─ cli1          │  └─ cli1
                    │        └─ app1A   │       └─ app1A
                    │        └─ app1B   │       └─ app1B
                    │  └─ cli2          │  └─ cli2
                    │        └─ app2A   │       └─ app2A
                    │        └─ app2B   │       └─ app2B
```

Figure 6. Guest OS directives to enable the data sharing structure.

Figure 6 shows the file and directory host OS infrastructure provided to support the vFucntion's functionalities. The root hierarchy (/var) was split into three branches: administrative components (admin), service provision (service), and shared resources (vmcli). Administrative components branch will contain the vFunD management software indicated in the Hypervisor and HostOS layer of Figure 1. It controls and implements all requests related to applications vFunction, which offload pieces of its code to specialized devices on the host.

The service provision branch holds the interface software acting as service for each vFunction request. Inside this branch, each client will have its own directory named by the clientID that will hold one subdirectory for each client's requested function, named according the vFunctionID. Within each subdirectory there will be a service daemon executable to interface the application and the vFunction. Each service daemon controls the communication between application and vFunction resource, attached to the high-level driver of that resource. Finally, shared resource branch will have one directory for each client named by the clientID, similar to the service provision branch, but that will become the root mount point for each client inside the guest VM. Within that mount point, there will be a subdirectory for each function requested by the client, named by the vFunction ID. This directory will be linked to the shared resource and will become the shared memory region. The daemon service executable will consider this local to enable the structures dataout and datain to store and access data.

## 4. Experimental Results

### 4.1. Environment

We implemented a prototype to validate the proposed methodology to offload VM virtualized functions to the host. We used the following hardware and software setup. HW: Intel Core i7 processor, 16GB of RAM and a SSD storage drive of 480GB. SW: GNU/Linux Ubuntu server 20.04 with kernel 5.4.0 and KVM support enabled, hypervisor QEMU 4.2.50 and Libvirt library 6.3.0. We used both software from sources in order to make available the latest functionalities, especially the VirtioFS support, not yet available in the mainline packages.

### 4.2. Methodology

Currently, we validated the framework by offloading two distinct vFunctions to software version: Factorial and Fibonacci. Although the primary goal is to virtualize functions in hardware, the software implementation is sufficient to validate the methodology, since such implementations are irrelevant from the framework point of view. We compared the behavior of three domains of execution. Host-Host: vFunctions are invoked and executed both on the host (native), this is the reference domain. Guest-Host: vFunctions are invoked inside the VM and executed on the host (virtualized), the target domain. Guest-Net: vFunctions are invoked by the VM application through the network and executed on the host (network). We conducted the experiment execution considering four parameters:

- **Domain**: representing the three aforementioned communication approaches, Host-Host, Guest-Host and Guest-Net.
- **Instance**: which comprises a round of calculation characterized by two arguments: repetition, the number of calculation replay, and iteration, the amount of calculation invocations.
- **Time**: the time spent by each instance execution, obtained by the arithmetic mean of five iterations.
- **Value**: number sent to the vFunction fixed as 51 for Factorial and 93 for Fibonacci, chosen for generating the largest result fitted into a 64-bit variable.

## 4.3. Evaluation

Tables 1 and 2 present the execution times of the Factorial and Fibonacci vFunction instances, measured in seconds. The first column indicates the names of the function and the three domains of execution, while the additional columns show the values for five instances executed in the experiment, referenced as: repetitions x iterations. We start processing instances on Host-Host domain, recording the execution times for the reference domain (native), shown in the second row of both tables. Here, the execution time increases as the number of iterations grows, which it is expected since each invocation of the function leads to an additional overhead. For example, comparing the execution time for the Factorial vFunction on Host-Host domain, the time increases by 58.7%, considering the smallest and largest results (invocations 1:2k). We expected this situation since the more invocations, the greater is the overhead to handle them. We also measured instances executions for Guest-Host and Guest-Net domains, recording results for Factorial and Fibonacci vFunctions in that tables, shown by the next two rows.

Table 1. Results of the execution times of the instances for the factorial vFunction considering the three domains.

| Factorial | 500M x 1 | 50M x 10 | 5M x 100 | 500K x 1k | 250K x 2k |
|---|---|---|---|---|---|
| Host-Host | **16,40** | **16,45** | **16,79** | **20,87** | **26,02** |
| Guest-Host | **16,41** | **16,57** | **16,81** | **21,44** | **26,54** |
| Guest-Net | **15,98** | **16,52** | **20,84** | **65,40** | **114,81** |

Evaluating instances execution with 1 and 10 iterations (500Mx1 and 50Mx10 columns), we observed that the execution time remains virtually the same for all domains, around 16s for Factorial and 19s for Fibonacci. This indicates that the system resources are able to handle properly the invocations for this workload. On the other hand, for instances executions starting from 100 iteration this scenario changes, reaching close to 619%, considering the lower and higher results for the Guest-Net domain (invocations 1:2k). However, the same does not happen if we compare the Host-Host domain with the Guest-Host, where the increase reached 61.7%, considering the same invocations case (1:2k).

Table 2. Results of the execution times of the instances for the Fibonacci vFunction considering the three domains.

| Fibonacci | 500M x 1 | 50M x 10 | 5M x 100 | 500K x 1k | 250K x 2k |
|---|---|---|---|---|---|
| Host-Host | **19,68** | **19,78** | **20,03** | **23,92** | **29,62** |
| Guest-Host | **19,69** | **19,80** | **20,08** | **24,87** | **30,13** |
| Guest-Net | **19,18** | **19,79** | **26,17** | **70,99** | **118,16** |

Figures 7 and 8 present a graphical evolution of all instance's executions for every domain. One can see that the leftmost two bars (500Mx1 and 50Mx10 instances) remains almost the same indicating that executions in the three domains, represented by colorized bars, have the same execution time, therefore are equivalent in performance. However, for the executions of the rightmost three sets of bars (5Mx100, 500Kx1K and 250Kx2k instances) the Guest-Net domain stands out for its variation, reaching about 340% in the case of Factorial. This does not happen with other two domains (Host-Host and Guest-Host) that grows in equal proportions. This demonstrates that the Guest-Net approach is not scalable, which is evidenced by the non-linear growth of the gray bars of both figures.

Comparing the execution time growth between the Guest-Host and Host-Host domains, one can see that it behaves in a proportional rate, as can be seen by the shapes sequence of blue and red bars shown in Figures 7 and 8. We confirm this condition in the 2nd and 3rd lines in Tables 1 and 2, where the values, in both lines, increase virtually in the same proportion. This demonstrates the effectiveness of the proposed methodology, since the growth of time to run instances on the Guest-Host domain (virtualized) behaves in a proportional way to the Host-Host reference domain (native).

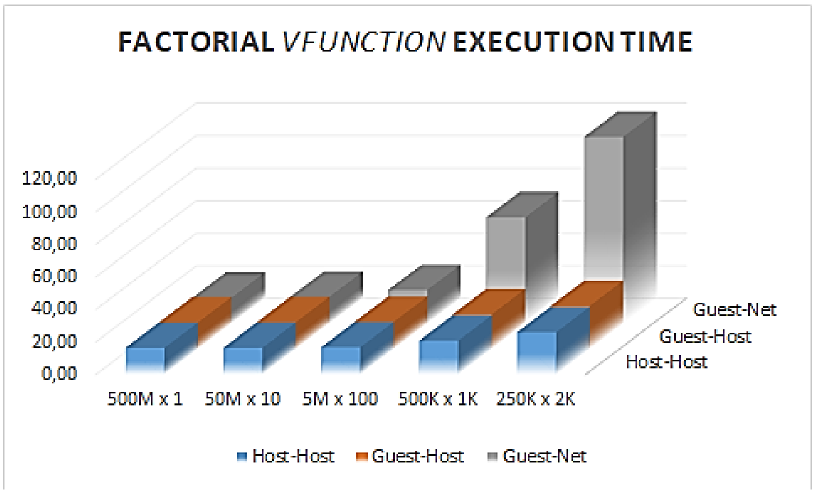


Figure 7. Execution time of instances for all domains of factorial vFunction, in seconds.

We can demonstrate that the vFunctions execute with the equal performance if we compare the Guest-Host domain and Host-Host (native) domain. Furthermore, the methodology effectiveness can be observed through Figure 9, which shows the executions of the instances compared as percentages for both vFunctions applications. The executions percentages in the Host-Host domain as well as the Guest-Host tend to zero, depicted by the labels GH-HH of Figure 9.

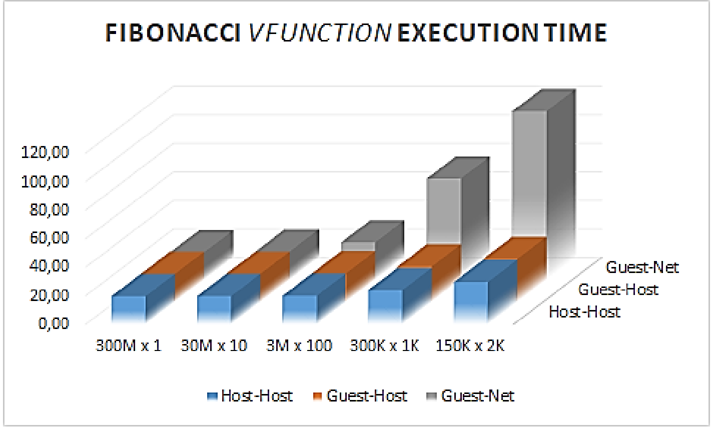
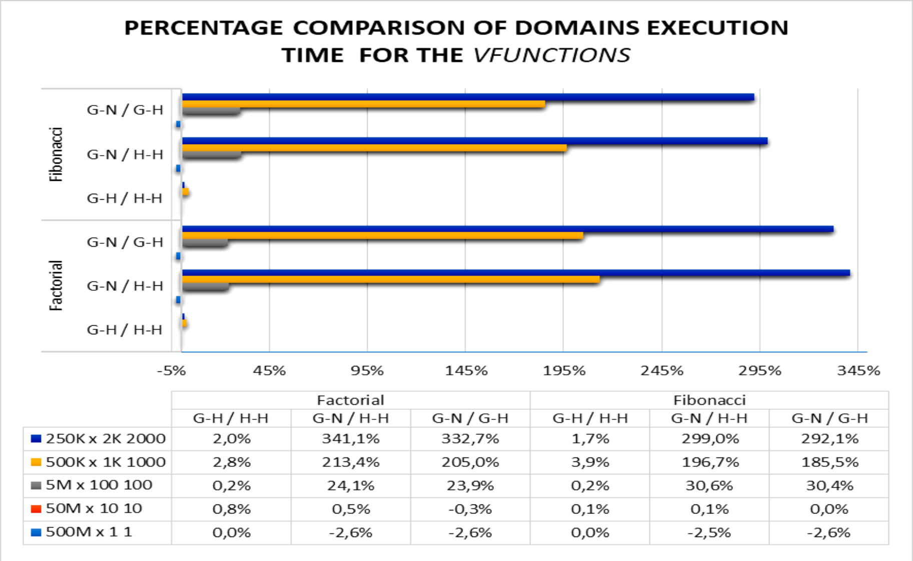


Figure 8. Execution time of instances for all domains of Fibonacci vFunction, in seconds.

On the other side, labels containing the GN (Guest-Net) label present high (non-linear) execution time growth, indicating that for applications with high demands the increase time becomes a bottleneck. Since the performance of domains Guest-Host and Host-Host is proportional, the virtualized domain represents a substantial speedup mechanism compared to the traditional domain (network) and an optimum choice to offload vFunctions on the co-located host and administrate the interaction between them.

Note: H, G and N letters in Figure 9 represent Host, Guest and Network, respectively. In addition, G-H/H-H, and other equivalent expressions, represent the percentage calculated between the instance's values of the Guest-Host domain and the Host-Host.



| | Factorial | | | Fibonacci | | |
|---|---|---|---|---|---|---|
| | G-H / H-H | G-N / H-H | G-N / G-H | G-H / H-H | G-N / H-H | G-N / G-H |
| ■ 250K x 2K 2000 | 2,0% | 341,1% | 332,7% | 1,7% | 299,0% | 292,1% |
| ■ 500K x 1K 1000 | 2,8% | 213,4% | 205,0% | 3,9% | 196,7% | 185,5% |
| ■ 5M x 100 100 | 0,2% | 24,1% | 23,9% | 0,2% | 30,6% | 30,4% |
| ■ 50M x 10 10 | 0,8% | 0,5% | -0,3% | 0,1% | 0,1% | 0,0% |
| ■ 500M x 1 1 | 0,0% | -2,6% | -2,6% | 0,0% | -2,5% | -2,6% |

Figure 9. Execution times percentage comparison for each vFunctions for all domains.

## 5. Conclusion

In this article, we present a methodology to allow pieces of an application's code (vFunction) to be offloaded to acceleration hardware resources. Alternatively, this code can also be delegated to a software version on the same host, that was our choice to validate the proposed methodology. We developed a framework to validate the methodology and conducted experiments to validate the effectiveness of the proposal. The experiment made use of two vFunction and five variations of executions, performed over three domains. We provided a noninvasive solution (except for a new hypercall) in which the client can use the standard development tools to implement applications. The experiments demonstrated that, for applications with about 100 iterations or more, the speedup of execution time reach substantial gains, over 340% for 2k iterations, comparing the Guest-Host domain and the Guest-Net. The experiments also indicated that the acceleration increases in a linear way when we compare the native and virtualized domains. Therefore, the proposed methodology proved to be an excellent option to offload vFunctions and administrate the guest-host interaction getting good performance results.

Based on the promising results, future works consider integrating the framework with hardware devices, providing the underlying backend especially for FPGAs, since these devices have recent improvements and new useful features. Additionally, we intend to integrate the framework together with the backend in the cloud, providing a complete multi-tenant computing infrastructure to enable the vFunction accelerator to cloud VMs.

## Acknowledgements

This work is supported by the Sao Paulo Research Foundation (FAPESP) (2013/08293-7), CAPES (2013/08293-7), and CNPq (fellowships 438445/2018-0, 309794/2017-0).

## References

[1] A. Putnam, A. M. Caulfield, E. S. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmaeilzadeh, J. Fowers, G. P. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J. Y. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong, P. Y. Xiao, and D. Burger. A reconfigurable fabric for accelerating large-scale datacenter services. In 2014 ACM/IEEE 41st International Symposium on Computer Architecture(ISCA), pages 13-24, June 2014.

[2] A. Putnam, A. M. Caulfield, E. S. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmaeilzadeh, J. Fowers, G. P. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J. Y. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong,P. Y. Xiao, and D. Burger. A reconfigurable fabric for accelerating large-scale datacenter services. IEEE Micro, 35(3):10-22, May 2015.

[3] Fei Chen, Yi Shan, Yu Zhang, Yu Wang, Hubertus Franke, Xiaotao Chang, and Kun Wang. Enabling fpgas in the cloud. In Proceedings of the 11th ACM Conference on Computing Frontiers, CF ’14, pages 3:1-3:10, New York, NY, USA, 2014. ACM.

[4] S. Byma, J. G. Steffan, H. Bannazadeh, A. L. Garcia, and P. Chow. Fpgas in the cloud: Booting virtualized hardware accelerators with openstack. In 2014 IEEE 22nd Annual International Symposium on Field-Programmable Custom Computing Machines, pages 109-116, May 2014.

[5] OpenStack Overview. https://www.openstack.org/software/, 2020. Online: accessed 21-Feb-2020.

[6] Brad Fitzpatrick. Distributed caching with memcached. Linux J., 2004(124):5, August 2004.

[7] Brad Fitzpatrick. Memcached, a distributed memory object caching system. https://memcached.org/, 2020. Online: accessed 24-Feb-2020.

[8] S. A. Fahmy, K. Vipin, and S. Shreejith. Virtualized fpga accelerators for efficient cloud computing. In 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom), pages 430-435, Nov 2015.

[9] S. Asano, T. Maruyama, and Y. Yamaguchi. Performance comparison of fpga, gpu and cpu in image processing. In 2009 International Conference on Field Programmable Logic and Applications, pages 126-131, Aug 2009.

[10] S. Kestur, J. D. Davis, and O. Williams. Blas comparison on fpga, cpu and gpu. In 2010 IEEE Computer Society Annual Symposium on VLSI, pages 288-293, July 2010.

[11] M. Asiatici, N. George, K. Vipin, S. A. Fahmy, and P. Ienne. Virtualized execution runtime for fpga accelerators in the cloud. IEEE Access, PP(99):1-1, 2017.

[12] J. Weerasinghe, F. Abel, C. Hagleitner, and A. Herkersdorf, "Enabling fpgas in hyperscale data centers," in Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015 IEEE 12th Intl Conf on. IEEE, 2015, pp. 1078–1086.

[13] J. Weerasinghe et al., "Network-Attached FPGAs for Data Center Applications," in FPT, Dec 2016.

[14] N. Tarafdar, N. Eskandari, T. Lin, and P. Chow, "Designing for fpgas in the cloud," IEEE Design & Test, vol. 35, no. 1, pp. 23–29, 2018.

[15] N. Tarafdar et al., "Enabling Flexible Network FPGA Clusters in a Heterogeneous Cloud Data Center," in FPGA '17. ACM, 2017

[16] J. Mandebi Mbongue, F. Hategekimana, D. Tchuinkou Kwadjo and C. Bobda, "FPGA Virtualization in Cloud-Based Infrastructures Over Virtio," 2018 IEEE 36th International Conference on Computer Design (ICCD), Orlando, FL, USA, 2018, pp. 242-245, doi: 10.1109/ICCD.2018.00044.

[17] S. Hajnoczi, "virtio-vsock Zero-configuration host/guest communication". https://vmsplice.net/~stefan/stefanha-kvm-forum-2015.pdf. Online: accessed 05-Mar-2020.

[18] Red Hat, Inc, "virtiofs: virtio-fs host<->guest shared file system". https://www.kernel.org/doc/html/latest/filesystems/virtiofs.html. Online: accessed 20-Feb-2020.

[19] Virtio-fs WebSite, "Virtio-fs - shared file system for virtual machines". https://virtio-fs.gitlab.io/. Online: accessed 15-Feb-2020.

[20] M. Gémieux, M. Li, Y. Savaria, J. David and G. Zhu, "A Hybrid Architecture With Low Latency Interfaces Enabling Dynamic Cache Management," in IEEE Access, vol. 6, pp. 62826-62839, 2018.

[21] Intel Open Source, "OPAE-Open Programmable Acceleration Engine". https://01.org/OPAE. Online: accessed 20-May-2020.

[22] Intel Open Source, "OPAE repository". https://github.com/OPAE. Online: accessed 20-May-2020.

[23] P. Colangelo, E. Luebbers, R. Huang, M. Margala and K. Nealis, "Application of convolutional neural networks on Intel® Xeon® processor with integrated FPGA," 2017 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, 2017, pp. 1-7.

[24] Intel Look Inside, "IvyTown Xeon + FPGA: The HARP Program". https://cpufpga.files.wordpress.com/2016/04/harp_isca_2016_final.pdf. Online: accessed 20-May-2020.

[25] Y. Choi, J. Cong, Z. Fang, Y. Hao, G. Reinman, P. Wei, "A Quantitative Analysis on Microarchitectures of Modern CPU-FPGA Platforms", 2016 ACM Design Automation Conference (DAC), Austin, TX, 2016, pp. 1-6.

[26] Intel Corpoation, "Intel Accelerator Functional Unit Simulation Environment Quick Start User Guide (ASE)". https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/ug/ug-qs-ase.pdf. Online: accessed 20-May-2020.

[27] L. Bragança, F. Alves, J. Penha , J. Penha, G. Coimbra, R. Ferreira, J. Nacif, " Simplifying HW/SW integration to deploy multiple accelerators for CPU-FPGA heterogeneous platforms", 2016 ACM International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS), Samos, Greece, 2018, pp. 97-104.

[28] OPAE SDK, "OPAE SDK source code repository". https://github.com/OPAE/opae-sdk. Online: accessed 20-May-2020.

[29] S. Rezaei, K. Kim and E. Bozorgzadeh, "Scalable Multi-Queue Data Transfer Scheme for FPGA-Based Multi-Accelerators," 2018 IEEE 36th International Conference on Computer Design (ICCD), Orlando, FL, USA, 2018, pp. 374-380.

[30] M. Jacobsen, D. Richmond, M. Hogains, and R. Kastner, " RIFFA 2.1: A Reusable Integration Framework for FPGA Accelerators," 2012 ACM ACM Transactions on Reconfigurable Technology ans Systems, 8,4, , pp. 1-23.

[31] M. Vesper, D. Koch, K. Vipin and S. A. Fahmy, "JetStream: An open-source high-performance PCI Express 3 streaming library for FPGA-to-Host and FPGA-to-FPGA communication," 2016 26th International Conference on Field Programmable Logic and Applications (FPL), Lausanne, 2016, pp. 1-9.

## AUTHORS

**Carlos A Petry** graduated in Computer Science at the University of Passo Fundo in 1990. He completed his master's degree in Computer Science at the Pontifical Catholic University of Rio Grande do Sul in 2009. He is currently pursuing his doctorate at the State University of Campinas. Professor at the Federal Institute of Rio Grande do Sul in the Computer Science course in the areas of Computer Architecture, Digital Systems, Computer Theory and Compilers, oriented undergraduate students.

**Rodolfo J de Azevedo** graduated in Computer Engineering from the Federal University of Espírito Santo in 1998, PhD in Computer Science from the State University of Campinas in 2002 and post-doctorate from the University of Washington - USA in 2010. Has a scholarship in Research Productivity from CNPq level 2 Associate professor at the State University of Campinas - UNICAMP, guiding students of the master's and doctorate in Computer Science in the areas of Computer Architecture, Dedicated Systems, Hardware Design and Use of Technology in Education.

# EVALUATION OF COMPANY INVESTMENT VALUE BASED ON MACHINE LEARNING

Junfeng Hu, Xiaosa Li, Yuru Xu, Shaowu Wu and Bin Zheng

Faculty of Science, School of Mathematics,
Beijing University of Technology, Beijing, China

## ABSTRACT

*In this paper, company investment value evaluation models are established based on comprehensive company information. After data mining and extracting a set of 436 feature parameters, an optimal subset of features is obtained by dimension reduction through tree-based feature selection, followed by the 5-fold cross-validation using XGBoost and LightGBM models. The results show that the Root-Mean-Square Error (RMSE) reached 3.098 and 3.059, respectively. In order to further improve the stability and generalization capability, Bayesian Ridge Regression has been used to train a stacking model based on the XGBoost and LightGBM models. The corresponding RMSE is up to 3.047. Finally, the importance of different features to the LightGBM model is analysed.*

## KEYWORDS

*Company investment value assessment, XGBoost model, LightGBM model, Model fusion.*

## 1. INTRODUCTION

Company investment value assessment is an outcome of active market in combination with modern enterprise systems, which can guide investors to understand the intrinsic value of a company. It includes both theoretical analysis and practice which help investors identify valuable investment projects, and make correct and reasonable investment decisions.

At present, traditional methods for company investment value evaluation mainly include free cash flow discount method, economic added value model, price-earnings ratio method, and so on. In particular, factor analysis and analytic hierarchy process are examples of empirical evaluation methods for investment value [1, 2, 3]. With the increase of information content and the development of big data analysis techniques, effective methods have also been developed to reduce uncertainty in real-time decision-making [4, 5]. In [6], a linear information model for value evaluation has been proposed. Random forest and support vector machine are used to establish a high-precision enterprise investment evaluation system [7]. Information transparency is shown to have a significant influence on the company's investment value based on unbalanced panel random-effects regression [8]. In [9], both financial data and non-financial data are trained by machine learning to establish a comprehensive evaluation model for enterprise investment value.

In this work, based on a data set containing comprehensive company information, a variety of data mining techniques and machine learning algorithms are used to develop effective company investment value evaluation models.

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 75-81, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101107

## 2. DATA AND METHODS

In this section, data processing, including pre-processing and feature extraction, will be discussed. Then, a detailed description of the models used in this study will be given.

### 2.1. Data Source

Data in this paper are from IEEE ISI World Cup 2019, including industrial and commercial information, annual report, financial information, tax information, equity information, legal information, intellectual property information, business information, land purchase information and other data of 3500 listed company (a total of 37 excel sheets and 1 enterprise rating form) [10]. These data come from official statistical platform, the data is real and credible.

### 2.2. Data Pre-processing

We perform data exploratory analysis to discover the inherent data characteristics, which helps us choose appropriate techniques for data pre-processing and analysis. In particular, the quality of data set after data pre-processing is very important for feature extraction in the next step. Hence, data pre-processing has a great impact on model results. The following is a list of problems involved in the data pre-processing and the corresponding processes taken:

- Label, Sample duplication problem: Deduplication.
- Missing value filling problem: Filling 0 or -99.
- Category variable processing: Label encoding and One-hot encoding.
- A Unit conversion: 1 Dollar = 6.7*1 Yuan, 100 Million Yuan = 100,000,000 Yuan.
- Date conversion: Convert to a timestamp, Separation of year, month, day, convenient extraction of time characteristics.
- Credit rating conversion: Advanced certification enterprise = 4, General certified enterprise = 3, General credit enterprise = 2, The remaining = 1.
- Exception string handling: For example '--', '\xad'. replace with -99.
- Number extraction: Mostly Regular Expression.

### 2.3. Feature Extraction

To facilitate feature extraction, each Excel form is extracted separately. Then, all features are combined into one Excel form. The features extracted in this paper are shown in TABLE I, where the first column is the variable names in the original table, and the second column is the corresponding feature extraction.

In the TABLE 1, count is the number of occurrences of some data, mean is the average of data, nunique is the number of elements in the data set, max is the maximum value of the data, min represents the minimum value of data, std is the standard deviation of data, skew means the skew of the data, median is the median of the data. In total, there are 436 features extracted from the data set.

Table 1. Original variable and its features

| Original Variables | Feature Extraction |
|---|---|
| Company Number | count: The company number in each table |
| Product Type | count, nunique |
| Registered Capital (Ten Thousand Yuan), Number of Employees | Registered Capital (Ten Thousand Yuan), Number of Employees |
| Registered Capital Currency (Regular), Operating State, Industry Categories (Code), Industry of Small Class(Code), Type, Province Code, City Code, Area Code, Whether the Listed, Registration Authority Area Code | LabelEncoder |
| Industry Categories (Code) | OneHotEncoder |
| Cancellation Reason | 1- No missing value, 0- Missing value |
| Date of Establishment | The time interval between the date of establishment and the term of operation |
| Land Use, Administrative Region | nunique |
| Total Land Supply Area, Transaction Price (Ten Thousand Yuan) | Median、mean、max、min、std、skew |
| Total Area | mean |
| Economic Division | count: Each economic division |
| Tax Year A | count |
| Credit Rating | mean |
| Annual Report Year | Count, mean, max, std: Annual reports from 2013 to 2017 |
| Label of Competing Products, Competitive Product Rotation, Detailed Address of Competing Products, Operation Status of Competing Products | count, nunique |
| Information of Subscribed Capital Contribution, Information of Paid-in Capital Contribution | mean: Logarithm |
| Are There Any Websites or Outlets, Whether the Enterprise Has Investment Information or Purchase Equity of Other Companies, Whether There Is a Change of Shareholders' Equity in the Limited Liability Company This Year, Whether to Provide External Guarantee, Whether the Number of Employees Is Open | sum |
| Trademark | count: Each state of the trademark |
| Application Date | mean,max,std,skew,kurt: The first difference of the application date<br>nunique, max, min, mean, median: Year of application, Day of application, |
| Earnings Per Share, Net Assets Per Share (Yuan), Provident Fund Per Share (Yuan), Undistributed Profit Per Share (Yuan), Operating Cash Flow Per Share (Yuan) | mean, std |
| Asset-liability Ratio (%), Current Liabilities/Total Liabilities (%), Liquidity Ratio, Quick Ratio | mean, std |
| Total Revenue (Yuan) | mean |
| Total Asset Turnover (Times), Days of Receivables Turnover (Days), Inventory Turnover Days (Days) | mean |

| Original Variables | Feature Extraction |
|---|---|
| Assets: Monetary Capital (Yuan), Assets: Fixed Assets (Yuan), Assets: Intangible Assets (Yuan), Assets: Total Assets (Yuan), Liabilities: Total Liabilities (Yuan) | mean |
| Label, Provinces, Name of the Certificate | nunique |
| Coupon Rate (%), Total Planned Issuance (100 Million Yuan) | mean |
| Patent Type | count: Each type of patent type |
| State | count: Each type of state |

## 2.4. Feature Selection

When the number of features increases, it may cause "dimension disaster" and hence decrease model learning performance [11]. To avoid such problems including the over-fitting problem, many feature selection methods have been developed, such as filter, wrapper, and embedded method. Here, a gradient boosting decision-tree model is chosen to reduce feature dimensionality. This method combines the advantages of the filter and wrapper methods, and uses the parameters inside the learner to sort the features. This effectively improved the performance of the learner and the computing efficiency [12].

## 2.5. Models

This section introduces the main models used in this paper: XGBoost model, LightGBM model and Stacking model.

First, XGBoost (Extreme Gradient Boosting) is a popular gradient boosted trees algorithm. Traditional GBDT (Gradient Boosting Decision Tree) uses first-order derivative information for optimization [13]. XGBoost uses a second-order Taylor expansion of the loss function, which adds second derivative information in addition to the retained first-order derivative. Hence, it can speed up model convergence on the training set [14].

XGBoost allows for setting sample weights. By adjusting these weights, we can pay more attention to some samples and use many strategies to prevent overfitting, such as introducing regularization term, Shrinkage and Column Subsampling, etc. Support of parallelism is one of the great strengths of XGBoost, which allows nodes of the same hierarchy to run in parallel. XGBoost was also designed to handle sparse data effectively.

Second, LightGBM (Light Gradient Boosting Machine) was released by Microsoft Asia research Institute in 2016. It is an open source, fast and efficient promotion framework based on decision tree algorithm. It is used in sort, classification and regression, and other machine learning tasks, and supports efficient parallel training [15]. LightGBM contains gradient-based one-side Sampling (GOSS) and Exclusive Feature Bundling (EFB). GOSS is used to filter partial data. The role of EFB is to bind and merge mutually exclusive features, thus achieving dimension reduction. The so-called mutual exclusion means that there are some features in the feature space that do not take non-zero values at the same time [16, 17]. LightGBM uses the leaf-wise method with depth limitation to improve the accuracy of the model. By changing the decision rules of the decision tree algorithm, LightGBM provides direct native support for categorical features without transformation, hence greatly improves the training speed.

Third, stacking algorithm is an integrated learning algorithm proposed by Wolpert in 1992 [18]. For Bagging and Boosting method the same learning algorithm is used in a single training study

[19, 20]. Stacking algorithm combines a number of different learning algorithms to improve generalization performance. It can be considered as a special kind of portfolio strategy, similar to the Voting and Blending [21]. Moreover, the stacking algorithm is constructed by cross validation and hence robust.

## 3. Results

This work first applies XGBoost and LightGBM for 5-fold cross-validation, and then use stacking algorithm for model fusion to improve stability and generalization ability. The performance of these models on feature set with or without feature selection is reported. Here, Root-Mean-Square Error (RMSE) is selected to measure model performance. The corresponding formula is given in Equation 1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{\mathrm{mod}el,i})^2}{n}} \qquad (1)$$

where $X_{obs,i}$ is the true value, $X_{\mathrm{mod}el,i}$ is the predicted value, and $n$ is the sample number.

The results of models performance are shown in TABLE 2. It can be seen from TABLE 2 that LightGBM model is superior to XGBoost model when using the same feature set. Model fusion can achieve higher accuracy. Also, the accuracy is improved after feature selection. Hence, more features does not necessarily lead to higher accuracy. A better selected feature set may help to achieve better model performance.

Table 2. Model Results

| **Models** | **RMSE** | |
|---|---|---|
| | ***No feature selection, number of features is 436*** | ***With feature selection, number of features is 66*** |
| XGBoost | 3.109 | 3.098 |
| LightGBM | 3.065 | 3.059 |
| Stacking fusion | 3.056 | 3.047 |

The importance of features for LightGBM is ranked, and the top 10 are shown in Fig. 1. Registered Capital is the most important, followed by the standard deviation of Current Liabilities/Total Liabilities, the mean of Total Asset Turnover, the standard deviation of Earnings Per Share and the mean of Provident Fund Per Share. The importance of features can help investors quickly select features they are interested in and make decisions based on various factors.

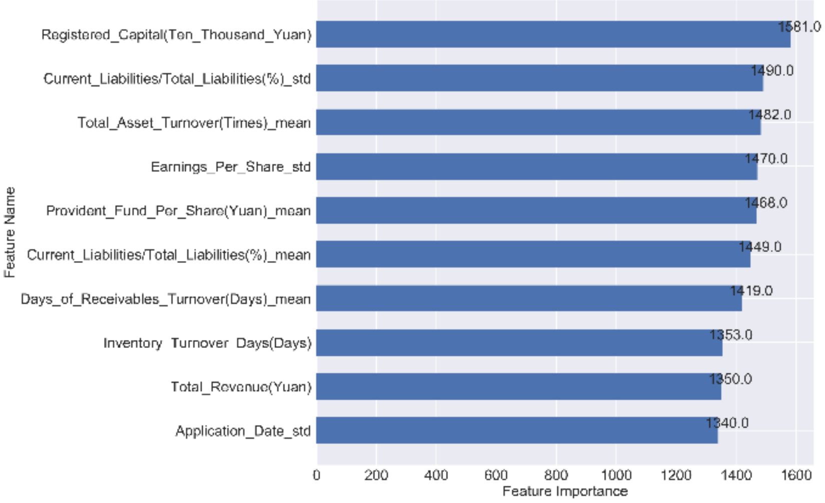


Fig 1. The importance of features for LightGBM

## 4. CONCLUSIONS

In this paper, several company investment value evaluation models are proposed by mining the comprehensive company information, extracting valuable features, and applying machine learning algorithms. The stacking model can achieve high precision. The obtained features importance value from LightGBM is instrumental for company investors. In order to find more valuable features and improve the accuracy of the proposed models, advanced feature engineering techniques as well as deep learning algorithms will be explored in future work.

## ACKNOWLEDGEMENTS

The research in this paper has been completed successfully, thanks to the platform provided by IEEE International Conference on ISI and Shenzhen Artificial Intelligence and Data Science Research Institute, as well as the data support.

## REFERENCES

[1] Z. Wang, H. Yang, & H. Meng. Literature review of enterprise investment value assessment model. MEI, 2018(14), pp.79.

[2] Z. Wang, Y. Zhu, & Y. Zhang, Research review of enterprise value assessment methods, ICSSED, 2019(314), pp.38-41.

[3] L. Beisland, A review of the value relevance literature. OSBJ, 2009, 2(1), pp.7-27.

[4] C. Hang, E. Garnsey, & Y. Ruan. Opportunities for disruption. Technovation, 2015(39), pp.83-93.

[5] A. Urbinati, M. Bogers, V. Chiesa, et al, Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. Technovation, 2019, pp.21-36.

[6] G. Feltham & J. Ohlson, Valuation and clean surplus accounting for operating and financial activities, CAR, 1995, 11(2), pp.689-731.

[7] M. Zhao, Research on the investment value of equipment manufacturing enterprises based on machine learning, Master's Thesis, Jilin. Univ, 2018,

[8] T. Yang, C. Chou, & Y. Yang. Investigation of enterprise value using information transparency: The case of optoelectronic industry, JCAF, 2020(31), pp.83-99.

[9] C. Zhang, X. Zhang, & Z. Yang, Enterprise investment value analysis based on machine learning model of rapidminer, J. Phys, 2020(1584).

[10] IEEE ISI World Cup 2019: http://ww w.linkx.ac.cn/#/competition.

[11] X. Wang & X. Hu, Overview on feature selection in high-dimensional and small-sample size classification. Comput. Appl, 2017, 37(9), pp.2433–2438+2448.

[12] Z. Wu & Y. Dong, Contrastive analysis of features selection on network video traffic classification, Comput. Eng. Appl, 2018, 54(6), pp.7-13.

[13] T. Chen & C. Guestrin, "Xgboost: A scalable tree boosting system," ACM, 2016, pp.785-794.

[14] C. Li, W. Zhang, & J. Lin. Research on star/galaxy classification based on XGBoost algorithm, Acta Astronomica Sinica, 2019, 60(2), pp.73-82.

[15] X. Ma, J. Sha, & X. Niu, An empirical study on the credit rating of P2P projects based on LightGBM algorithm, J. Quant. Tech. Econ, 2018, 35(5), pp.144-160.

[16] K. Mo, N. Wang, H. Li, et al, Network intrusion detection system model based on LightGBM, JISR, 2019, 5(2), pp.152–156.

[17] G. Ke, Q. Meng, T. Finley, et al, LightGBM: A highly efficient gradient boosting decision tree, NIPS, 2017, pp.3147–3155.

[18] D. Wolpert, Stacked generalization, Neural Networks, 1992, 5(2), pp.241–259.

[19] L. Breiman, Bagging predictors, KAP, 1996(24), pp.123-140.

[20] L. Breiman, Arcing Classifiers, Ann. Stat, 1998, 26(3), pp.801–824.

[21] X. Zhou, L. Ding, R. Wan, et al. Research on classifier ensemble algorithms. J. Wuhan. Univ (Natural Science Edition), 2015, 61(6), pp.503–508.

# A WIDE BAND MICROSTRIP MONOPOLE SLOT ANTENNA FOR CHIPLESS RFID APPLICATIONS

Chaker ESSID[1], Hedi SAKLI[2,3] and Nizar SAKLI[3]

[1]Tunisia Polytechnic School, Carthage University, SERCOM
Research Laboratory, Tunisia
[2] National Engineering School of Gabes, MACS Research Laboratory,
University of Gabes, Gabes, 6029, Tunisia
[3]EITA Consulting, 5 Rue du Chant des Oiseaux, 78360 Montesson, France

## *ABSTRACT*

*A new design of wideband microstrip antenna, where slots are placed on structure, is proposed. Also, a newly structure of a monopole antenna based on the noches form is designed using the HFSS. It has been found that the characteristics of new microstrip antenna are comparable to the conventional patch antennas, whereas its gain, directivity, and radiating efficiency are remarkably improved which make it to be useful in RFID chipless applications. The proposed antenna operates from 11.45 to 13.28 GHz, and 14.61 to 19.55 GHz can be used in RFID chipless applications, and it has a bandwidth about 677 MHz. The return loss of the proposed antenna is indeed below -10 dB. Prototype for all antennas are fabricated and measured and a good agreement between the measured and simulated results is achieved.*

## *KEYWORDS*

*Slot antenna, notches, monopole antenna, chipless RFID, multi resonant, wide band*

## 1. INTRODUCTION

Wireless communications are of great importance in the telecommunications sector. And in these applications where the size, weight, cost, performance, ease of installation and aerodynamic profile are constraining, profiled antennas such as microstrip are required. However, the microstrip antennas generally have narrow bandwidths and, in general, are half-wavelength structures operating at the fundamental resonance mode [1], the various research works are striving to solve the problem of the narrow bandwidth, and various structures have been studied to extend the bandwidth [2], for example, by introducing slots in a microstrip patch configuration. Also, compared with ordinary microstrip patch antenna, the printed monopole antenna has become a suitable choice due its wider bandwidth, lower radiation loss and lower dispersion [3]. They are also more likely to join various additional structures by etching slots on the radiating area, with the resulting multiband characteristic. For all that, monopole broadband antenna accompanied with slots is often used kind of structure hooked for multi resonant antenna, not only for its miniaturization of the antenna dimension effectively but also for the ability of generation of multiband characteristic.

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 83-89, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101108

The reason for this is that some microstrip monopole antennas are required in some applications, but not others, i.e., some applications require wide bandwidth, short-range and low-cost tags. Hence, the design and choice of RFID systems are application specific.

Moreover, RFID identification technology has grown considerably in recent decades and it can be applied in many applications such as industrial robotics, wireless gas sensors, wireless humidity sensors [4]. In the RFID literature is classified into several categories low frequency (LF),high frequency (HF), ultrahigh frequency (UHF), superhigh frequency (SHF), and ultra-wide band frequency(UWBF) RFID systems[5][6].Recently, the price of silicon chips has gone down exponentially causing in a significant reduction in cost of the chip based RFID tags. However, the chip based RFID tags are not yet economical enough to completely replace the barcodes for item level tagging [5]. Chipless RFID tags can be a more suitable choice for item level tagging. The chipless RFID tags are also expected to possess higher read range as no RF power from the reader signal is exploited to power up the chip [8]. Multi-resonator structures are employed to encode the data in frequency signature technique which is one of the main data encoding technique of the core's chipless RFID [9].

In this paper, the prospects of improving the bandwidth and the data capacity per unit area further are discussed by placing notches and slots on the radiating element.

## 2. Antenna Design

The wide band monopole slot antenna is based on a squared microstrip patch filled with notches and sub slots structures in figure 1 and 2 respectively. These slots structures behave as miniaturization method, allowing the widening of the bandwidth.

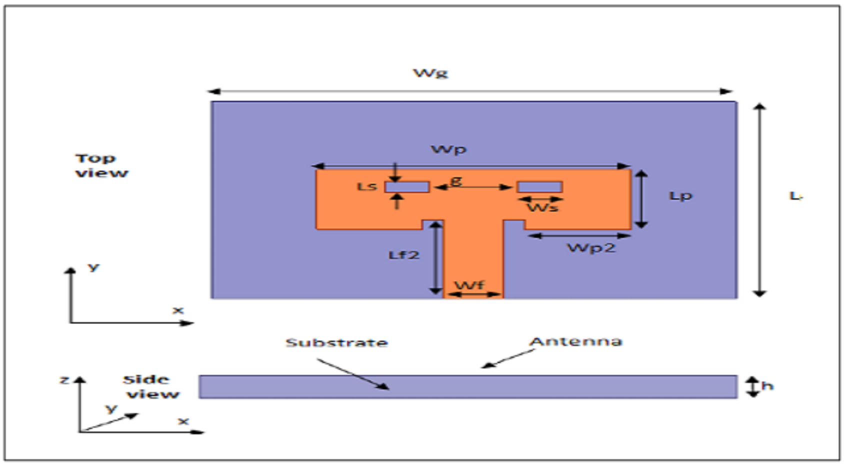


Figure.1. The layout of the proposed monopole wideband antenna: top view

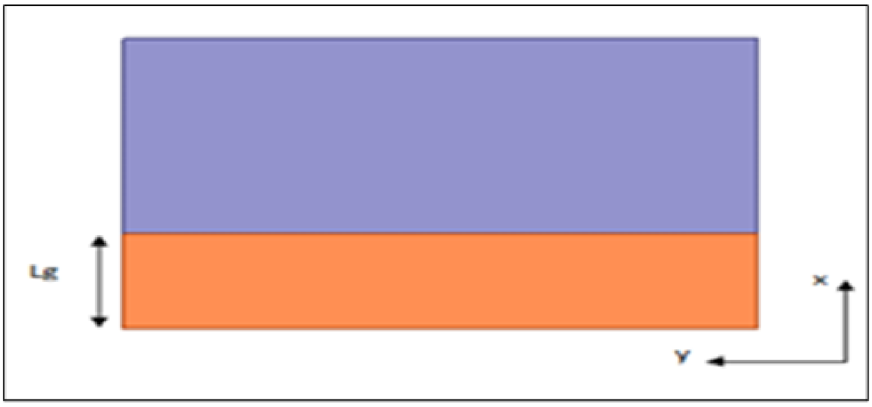


Figure.2. The layout of the proposed monopole wideband antenna: bottom view.

The antenna structure presented in this study can be studied in three different states that have been shown in this part. As shown in figure 3, this antenna is initially provided by rectangular radiation patch with a width of 25mm and a length of 20mm printed on a FR4-epoxy substrate with h=1.6mm and Ɛ=4.4. the patch is fed by a coplanar line with 50Ω input impedance. The dimensions of the line are 7.8*2.86mm$^2$. the port feeding line is places at the middle of the edge of the patch. This first design will work only on 5.8GHz and 11.8GHz frequency band reffering to figure 6.

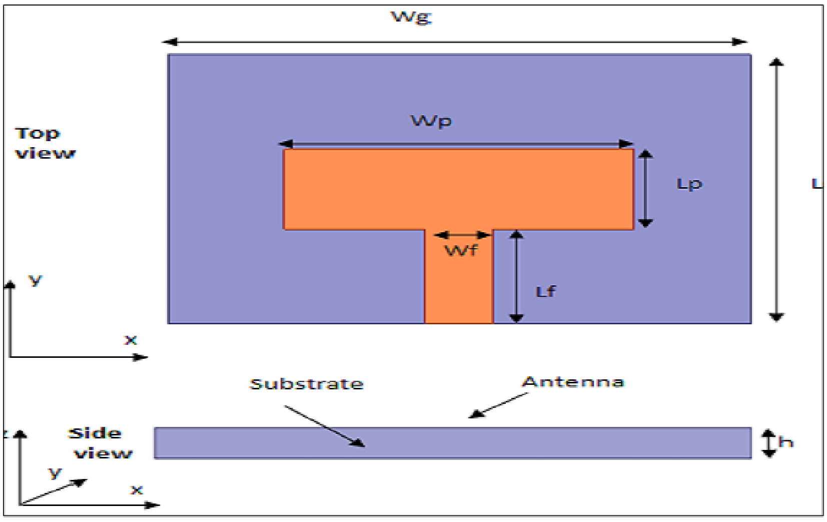


Figure.3. the design of the conventional antenna structure

By adding two rectangular slots to radiation patch, antenna will cover three frequencies of 5.8, 13 and 17 GHz resonant frequency. The design step is shown in fig 4. The dimensions of the rectangular slots are 1.88*1mm$^2$ spaced 3.75mm.

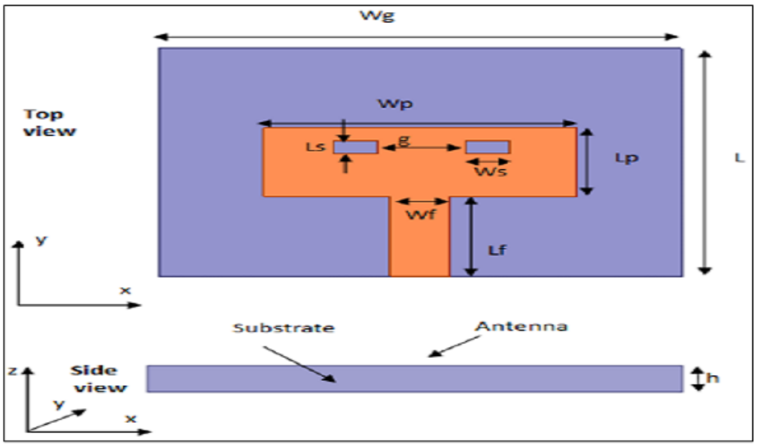


Figure.4. The design of the "antenna + 2 slots" structure

The antenna parameter values are shown in table 1. Antenna three shown in fig 1 and it is the last stage of design process that the final structure of antenna will be obtained by creating two notches on the radiation patch. In this design 5.8, 12 and 16GHz resonant frequency are obtained.

Table1. Antenna parameters.

| | | | |
|---|---|---|---|
| $L_f$ (mm) | 7.8 | G (mm) | 0.5 |
| $L_{f2}$(mm) | 7.8 | $L_s$(mm) | 3.75 |
| $W_f$(mm) | 2.86 | $W_{s2}$(mm) | 1 |
| L(mm) | 20 | $L_g$(mm) | 6.5 |
| $W_g$(mm) | 25 | $W_p$(mm) | 15 |
| $L_p$(mm) | 6 | $W_{p2}$(mm) | 5.07 |

## 3. Experimental Results

Prototypes of these antennas have been manufactured in figure 5. The experimental S parameters of these antennas are also plotted in figure 6. It can be seen that simulations and measurements agree.

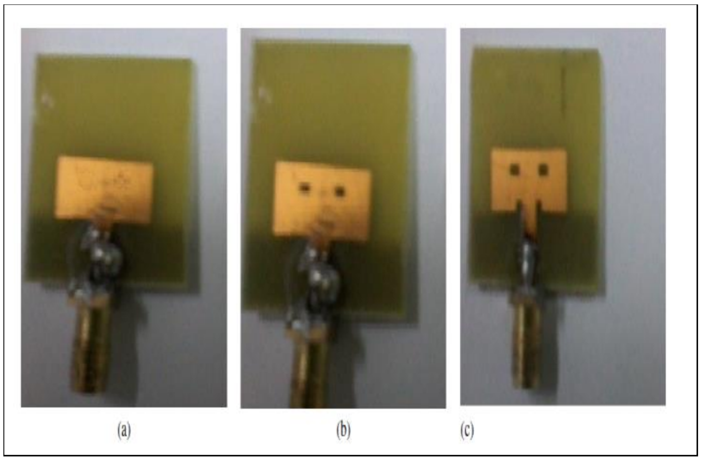


Figure5. Fabricated prototype of the proposed antennas, (a) monopole antenna, (b) "antenna + 2 sub slots", (c) "notch antenna + 2 sub slots"

Figure 6 shows the simulated and measured insertion loss versus frequency of the multi resonating antenna. From this, we can see five distinct resonant nulls due to the slots and notches forms. Each notch and slot can be used to a particular resonance which can be used for data en coding.

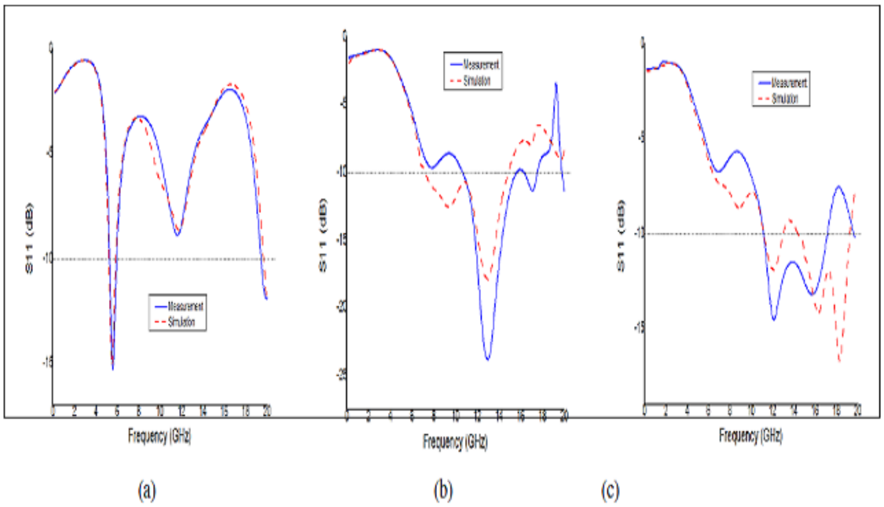


Figure 6. Measured and simulated reflection coefficients of the proposed antennas, (a) monopole antenna, (b) "antenna + 2 sub slots" , (c) "notch antenna + 2 sub slots"

The second prototype covers a frequency band ranging from 7.27 to 14.85 GHz, while the measured antenna covers a frequency band from 10.2 to 14.8 GHz. The last antenna has two frequency bands: the first is ranging from 11.79 to 13.46 GHz, while, the second is situated between 15.15 to 15.49 GHz and has a bandwidth of 677MHz. It is noted that the measured antenna covers a frequency band from 11.79 to 15.2 GHz.

Simulation results prove 5 equally spaces resonant notched are observed in the 4 to 20GH frequency band. Each notch encodes a unique data bit.

Figure.7 and figure.8 show the gain of each antenna in E and H planes.

First, figure.7 (a) and figure.8 (a) shows the radiation patterns of monopole antenna at 7 GHz in the E-plane. The measurements seem accurate more than those of simulation ones.

Second, figure.7 (b) and figure.8 (b) present the radiation pattern of "antenna + 2 sub slots" at 8.4 GHz, the results are the same. In addition, figure. 7(c) and figure.8 (c) present the radiation patterns of "antenna notched + 2 sub slots" at 12 GHz; measurement and simulation results are the same behavior.

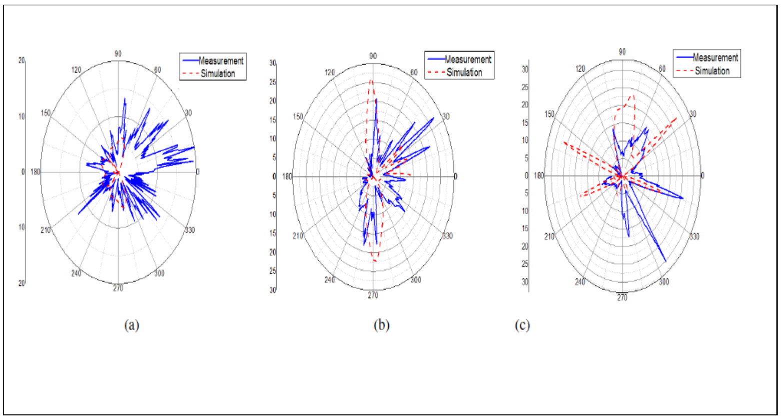


Figure.7. Measured and simulated far field normalized radiation patterns of the proposed antenna in E plane,(a) monopole antenna at 7 GHz, (b) “antenna + 2 sub slots” at 8.4 GHz, (c) “notch antenna + 2 sub slots” at 12 GHz

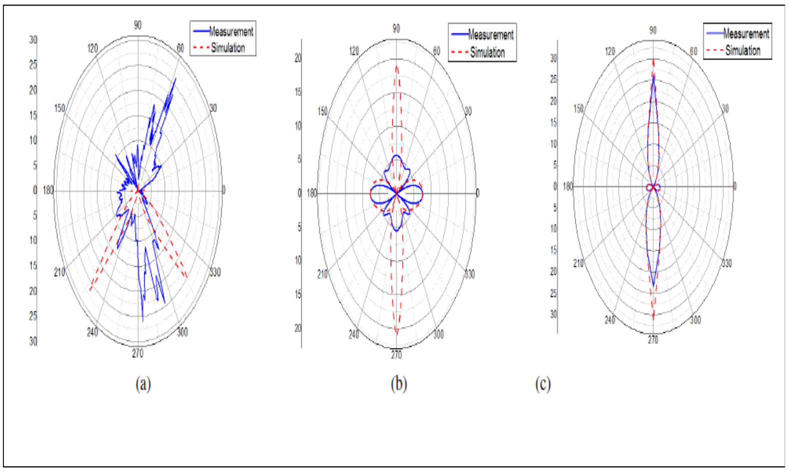


Figure.8. Measured and simulated far field normalized radiation patterns of the proposed antenna in H plane,(a) monopole antenna at 7 GHz, (b) “antenna + 2 sub slots” at 8.4 GHz, (c) “notch antenna + 2 sub slots” at 12 GHz

However, comparing our proposed monopole antenna with conventional patch antenna, it is noted that the conventional patch antenna uses microstrip technology, but this prototype is structured based on the monopole design which is advantageous as regards the microstrip technology. Indeed, this technology presents the advantage of a compactness (reduced volume), cost-effective production and simplicity of components implement.

Thus, taking an example from the literature reported in [7] can be used to display this comparison. In their work, the authors present a patch antenna disposed on FR4 epoxy substrate with a thickness of 1.6 mm and a permittivity Ɛ=4.4. The overall size of the antenna is 50*40 mm$^2$, the width and the length of the patch are 24 mm and 22 mm respectively. This antenna is fed by a microstrip line with a width and a length of 2.75 mm and 15 mm respectively. Thus, this patch antenna covers the frequency 6.5 GHz. Moreover, most conventional patch antenna possesses oversized dimensions particularly when the application referred is low frequency and it seems covering a single frequency band which is so narrow. In our case, the proposed antenna has a relatively small size, a wide frequency band, a best gain and is more directives compared to that of the conventional patch antenna.

## 4. CONCLUSIONS

In this paper, a novel chipless RFID antenna is presented, which can be configured as monopole wide band antennas. Proposed approach offers enhanced data capacity. A 5-bit chipless RFID antenna is also presented. High gain and wide bandwidth were found to be 4.78dB and 677 MHz respectively and will be used to prolong the read range.

## REFERENCES

[1] Tayeb A. Denidni Qinjiang Rao,"Ultra-wideband slot antenna for wireless communication system", *International Journal of RF and Microwave Computer Aided Engineering*, Volume 16, Issue 4.

[2] Kang Ding ,Cheng Gao,Tong bin Yu, De xin Qu, "CPW fed C-shaped slot antenna for broadband circularly polarized radiation", *International Journal of RF and Microwave Computer-Aided Engineering* Volume 25, Issue 9.

[3] Broggi Max Ammann and Zhi Ning Chen., (2003) "A wide-band shorted planar monopole with bevel". ,*IEEE Trans. Antenna.*

[4] Adbulkawi, W. M., &Sheta, A. F. A. (2018, April), "Printable Chipless RFID Tags for IoT Applications", *1st International Conference on Computer Applications & Information Security (ICCAIS) IEEE* (pp. 1-4).

[5] A. Vena, E. Perret, and S. Tedjini, (Dec. 2011), "Chipless RFID tag using hybridcoding technique," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 12,pp. 3356–3364.

[6] M. M. R. Rezaiesarlak. (2015), "Chipless RFID Design Procedure and Detection Techniques"

[7] JA Evans and MJ Ammann. (2006), "Reduced-size reconfigurable tri-band printed antenna with cpw tapered-feed and shorting post", *Microwave and optical technology letters*, 48(9):1850-1853.

[8] Abdelnasser A Eldek, Atef Z Elsherbeni, Charles E Smith, and K-F Lee.(2002), "Wideband rectangular slot antenna for personal wireless communication systems", *Antennas and Propagation Magazine, IEEE*, 44(5):146-155.

[9] Wonseob Kim, Seokjin Hong, Hoon Park, and Jaehoon Choi. (2007), "Planar monopole antenna with wide impedance bandwidth for mobile handset application", *Microwave and optical technology letters*, 49(4):779-781.

# QUALITY OF SERVICE-AWARE SECURITY FRAMEWORK FOR MOBILE AD HOC NETWORKS USING OPTIMIZED LINK STATE ROUTING PROTOCOL

Thulani Phakathi, Francis Lugayizi and Michael Esiefarienrhe

Department of Computer Science,
North-West University, Mafikeng, South Africa

### ABSTRACT

*All networks must provide an acceptable and desirable level of Quality of Service (QoS) to ensure that applications are well supported. This becomes a challenge when it comes to Mobile ad-hoc networks (MANETs). This paper presents a security framework that is QoS-aware in MANETs using a network protocol called Optimized Link State Routing Protocol (OLSR). Security & QoS targets may not necessarily be similar but this framework seeks to bridge the gap for the provision of an optimal functioning MANET. This paper presents the various security challenges, attacks, and goals in MANETs and the existing architectures or mechanisms used to combat security attacks. Additionally, this framework includes a security keying system to ascertain QoS. The keying system is linked to the basic configuration of the protocol OLSR through its Multi-point Relays (MPRs) functionality. The proposed framework is one that optimizes the use of network resources and time.*

### KEYWORDS

*Routing protocols, MANETs, Trust framework, Video streaming, QoS.*

## 1. INTRODUCTION

A MANET is an autonomous type of system which has separately connected sets of self-configuring nodes that may be activated by putting to use various techniques e.g. Bluetooth or WLAN. MANET is autonomous in behaviour because each node is regarded as its host and router at the same time [1]. They rely on direct communication and multi-hops for communication between distant nodes within the network making them scalable and robust. These advantages make them more flexible to accommodate many nodes, decentralize administration and their setup can be placed anywhere and at any time [2]. Contrary to other Wireless systems, MANETs do not have a central authority that monitors the forwarding of traffic. MANET systems consist of an infrastructure-less system of associated nodes connect through wireless links [1]. Nodes move arbitrarily or rather in a random faction [3]. Security in MANETs is crucial from the node level to the network level. According to [4], because of the dynamic nature of MANETs, trust can be used as a measure for nodes that want to provide an acceptable level of trust in that relationship among themselves. Security in a MANET is way too challenging than in traditional network environments infused with a central controller because of the dynamic topological nature and characteristics of MANETs. MANETs are primarily used in the army and security-based applications e.g. covert missions, emergency, and rescue missions [8] [6] [9]. The availability of network resources, integrity, and confidentiality depends on the

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 91-108, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101109

security mechanisms that are put in place. The design of MANETs makes them vulnerable to security attacks. The vulnerabilities come through non-secure boundaries and compromised nodes although many other factors contribute to the weakening nature of MANETs. The best approach to mitigating attacks is prevention and avoidance algorithms, not security mechanisms that remove attacks as these tend to require more resources. Network resources in MANETs must be optimally used to achieve Quality of Service. The needs for the provision of QoS are increasing with applications that involve voice and video and it is most appropriate to support these through the implementation of ad hoc network environments. QoS was first brought to attention in 1994 as a phenomenon that has the overall requirements of a network connection, as well as response time during service times, network detriments such as echo, interrupts, signal to noise ratio and also loudness levels. QoS is generally the network's assurance to ascertain a specific level of execution to a data transmission [10]. To achieve QoS, the concept of routing is important. Routing is regarded as the act of steering information from a source node to a terminal node in the network. [10]. One intermediate node within the network is experienced during the movement of information. The most important aspect is achieving good QoS. It is impossible to say a characteristic like this one can be completely run over. It is possible to achieve a greater QoS to such an extent that its dynamic nature would not be such a limitation thus a robust and efficient security framework is needed. The framework would guard against malicious activity in the network amongst nodes. This work seeks to close that gap by building a framework that not only looks at the security but also the Quality of Service in video streaming applications over MANETs.

This work is arranged as follows: Section 2 discusses the various routing protocols (RPs) in MANETs, Section 3 is on trust in MANETs and Section 4 presents security frameworks studies. Section 5 presents Typical security attacks in MANETs. Section 6 gives a highlight of related works while Section 7 presents the proposed framework in detail.

## 2. OPTIMIZED LINK STATE ROUTING FEATURES

### 2.1. OLSR Protocol

OLSR is simply the optimization of traditional link-state protocol for MANETs. It falls under the category of proactive routing protocols. With the OLSR protocol, every network node chooses a neighbouring node-set, commonly termed as the multipoint relays (MPR), which rebroadcasts the packets that were initially transmitted. To this end, neighbouring nodes that are not found in the MPR set have the instruction to only read and process the packets [14]. According to Saravanan and Vijayakuma [18], OLSR keeps tracks in the pathfinding table to provide a route if necessitated. MPRs are primarily responsible for declaring and forwarding link-state information, forwarding and controlling traffic, providing effective mechanisms for broadcasting control traffic by minimizing the frequency of required transmissions [19]. OLSR utilizes two types of control messages: Hello and Topology Control (TC). Hello messages are used to the information concerning the link status and the host's neighbours [20].

### 2.2. OLSR Architectural Design

OLSR has a cross-layered design just like that of the OSI (Open Systems Interconnection) model in networks. From a designer's perspective, there are two relative choices in the design process of the protocol. The first option is to design the protocol by the rules of the reference architecture and that is the higher layer being able to access services provided by the lower layer with no consideration of how such service is made available. Secondly, the routing protocols can be developed in violation of the original architecture and that is by allowing cross-layer communication between layers. This deliberate violation is called the cross-layer design.

## 3. SECURITY GOALS IN MANETS

The goal of every system is to achieve excellent quality of service and adherence to security targets to protect client data or information [24].

### 3.1. Availability

Availability is one security target for every system. An authorized user will request us of the node and in an operable state. The node is therefore supposed to provide its services by its design. Attacks may seek to disrupt the node's operation and also use up some of the node's resources but the node must be able to survive those attacks and be available when requested

### 3.2. Confidentiality

This security feature ensures the unavailability of certain features to unauthorized entities or users. The information is restricted to only authorized personnel. A message that an as the source will only be decrypted at the destination node. Many cryptographic attacks may try to reveal the message contents. An ideal system will be able to protect the contents of such information from unauthorized users.

### 3.3. Non-repudiation

This feature ensures that the source node will not be able to interfere with an occurred action like deny the authenticity of a message sent. It also facilitates the detection of malicious nodes. Many of the existing algorithms are based on reputation and trust e.g. CONFIDANT.

### 3.4. Authentication

Authentication ensures user validation and avoiding impersonation. The malicious node could impersonate a legitimate node by using the node's MAC address or even an IP address to obtain authentication and also launch its attack at a higher level.

### 3.5. Integrity

Integrity is a security feature that ensures that the original contents of the data are maintained and not altered in any way. An effort to intercept the data being transmitted, either by human beings or malicious nodes is an act against the trustworthiness of the network. Dropping attacks are usually launched by a malicious node but the node is compelled to cooperate in the system.

### 3.6. Anonymity

This feature is for privatizing the true identity of a node to ensure privacy and confidentiality. In most cases, the source of packets is kept private.

## 4. VULNERABILITIES IN MANETS

MANETs are prone to various [13] security vulnerabilities that pose to gain unauthorized access to the user's data. Vulnerabilities of a system may be termed as weaknesses possessed by a system. MANETs are more vulnerable because they rely on wireless technology, unlike traditional wired networks. Attacks in MANETs can be categorically put in two; namely active and passive attacks. The severity of these attacks differ. Wireless networks are most vulnerable to all sorts of attacks internally and externally as compared to [28] traditional networks (wired networks) due to limited physical security, scalability, mobile nodes, dynamic topology, lack of centralized management, and threats emanating from compromised nodes. The vulnerability of the network highlights a weakness in the security architecture or system. MANETs operate in very dynamic environments. Security is very important in MANETs even though the environments' hostility makes it difficult to achieve most security properties as proposed by many authors [24].

### 4.1. Central Controller

The lack of a centralized controller that could act as a monitoring-server is one vulnerability that comes with MANETs. This makes it complex in terms of security provision against attacks as in most instances the network environment is huge and highly dynamic.

### 4.2. Dynamic Topology

As mentioned earlier, the network environment in MANETs is dynamic. This may, in turn, affect the trust relationship among nodes. Malicious nodes that may be compromised within the network are also difficult to spot as the nodes are mobile.

### 4.3. Power Limits

Mobile nodes rely solely on battery power and such may pose so many problems. A node may behave maliciously within the network and could be suspected of being an internal attacker but only to discover that it behaves selfishly because of limited power supply.

### 4.4. Resource Availability

Resource constraints are the primary reason why some services are not utilized in MANETs. For example, secure communication is needed but most often it is difficult to provide it because of the dynamic environment. Ad hoc security mechanisms and architectures are needed to prevent attacks from flooding the network.

## 5. SECURITY ATTACKS IN MANETS

Attacks in MANETs can be categorically put in two; namely active and passive attacks. The severity of these attacks differ. Wireless networks are most vulnerable to all sorts of attacks internally and externally as compared to [28] traditional networks (wired networks) due to limited physical security, scalability, mobile nodes, dynamic topology, lack of centralized management, and threats emanating from compromised nodes. The vulnerability of the network highlights a weakness in the security architecture or system.

### 5.1. Active Attacks

These are known to disrupt the normal operation of a network [21]. The attacker actively alters the network's normal operation. The attacker acts as one of the stations in the network. In this way, it able to exploit any other node and uses it to its advantage. It can feed nodes fake packets or even denial of service (DoS). The active attacker can:

- Fabricate messages
- Replay packets
- Modify packets
- Drop packets
- Node Impersonation
- Insert infected code

### 5.2. Passive Attacks

Passive attacks are characterized by their inability to [19] actively participate in causing harm to the network. The attacker monitors the network to attain information. They do that so that they may get node information, for example, how nodes are communicating and their geographical location within the network. They do not just attack the network. At first, they acquire enough information before launching an attack. Once they acquire information, they easily hijack it and launch an attack. They can decrypt weakly encrypted data, acquire passwords, private and public keys, monitor communication routes, and message flow among entities [29]. It may be hard for the user to identify a passive attack as it does not necessarily alter anything regarding user data or traffic.

## 6. Related Works on Security Frameworks in MANETs

Hurley-Smith et al. [31] proposed a security protocol called Security Using Pre-Existing Routing for Mobile Ad hoc Networks (SUPERMAN). The initial protocol design was to solve issues like the authentication of nodes, secure network access control, and secure network communication through existing routing protocols. SUPERMAN was designed to bring together communication security and routing at the network layer. Their security protocol is unique from others in the sense that it combines routing and communication security at the network layer. This is in contrast to existing approaches that may require additional protocols to protect the network. SUPERMAN was created to give security to all data communicated over a mobile ad-hoc network. It may not apply to other networks. In a nutshell, it provides protection and efficiency. One efficient method it employs is that it protects application data and routing, ensuring that the network provides trustworthy, confidential communication, and reliable to all true nodes [31].
Kaur et al. analyzed [33] security threats MANETs. Their security objective called CBDS was successfully carried out on Blackhole and Grey hole attacks before and their trial was proven successful in the case of Sleep deprivation and denial of service attacks. Their simulation results have showed increased detection for CBDS and an enhanced response [35]. The limitation within the framework was its implementation of two attacks and no proof is given out in terms of its validity towards other active and passive attacks.

Monica et al [35] in their work analyzed, simulated and three different attacks based on many parameters. These attacks were Blackhole, Denial of service (DoS, and wormhole. The comparison was made for their throughput, End to End Delay, and Packet Delivery Ratio (PDR)
Authors in [5] proposed a MANET trust model that uses a combination of direct, indirect, and mutual trust values among network nodes to reflect the behavior of sensor nodes. The aim was to test out the effectiveness of a secured node can be routed within the network. QoS metrics were used to evaluate the trust level. This provided an accurate recommendation for packet forwarding and thus reducing the rate of dropped packets. A performance evaluation was conducted and the trust model achieved reduced packet loss, reduced energy consumption, and an increased network throughput despite having malicious nodes in the network. This meant that that proposed algorithm improved the overall network performance as compared to an existing single-trust based model. The proposed model filtered out malicious nodes.

Sahu et. al [7] proposed a cross-layer security framework that prevents malicious attacks in the network. The proposed framework used throughput, end to end delay, jitter, and packet drop ratio as QoS metrics where-in different scenarios were evaluated. Their framework included a design and implementation of a Neighbour Node Surveillance Real-Time MAC(NNSRT-MAC) protocol at MAC Layer. Key contributions include the design and implementation of a QoS framework that doesn't use a complex algorithm to prevent over-reservation, QoS degradation, flooding attack, state table starvation. The results gave better results in terms of the QoS metrics in both malicious-free and malicious scenarios.

Madhavan et. al [11] proposed an algorithm called GA-ACO (Genetic Algorithm-Any Colony Optimization) to optimize QoS by using a secure agent-based multicast routing scheme to optimize parameters by combining GA and ACO techniques. This hybrid technique outperformed existing protocols like AODV and OLSR.

Authors in [12] proposed an efficient multi-hop and relay-based communication framework. Using Brodatz Texture database, CT scan images, and Brain MRI scan images as input. The algorithm was designed to operate 3 stages. The first stage involved selected regions using the spatial candidate region detection. The second stage involved applying average entropy feature space for the detection of the cluster centre. The final stage involved spatial density-based clustering of images carried out by tracking down dense regions. This method produced better clustering results and PSNR rates. The improvement of QoS was based on Random Repeat Trust Computational Approach using direct and indirect trust. The proposed framework showed more than 30% effectiveness as compared to the existing system.

Tygi et. al [17] proposed a Proposed Local Adjustment AODV in high mobility environments with scarce network resources and ruptured explored routes The algorithm controls the flooding of control packets and its maintenance during transmission with maximum adjustment locally when a node responsible of forwarding packets is out of range in terms of transmission. Metrics used were control overhead, packet delivery ratio, and energy consumption. The proposed algorithm outperforms AODV in terms of the QoS metrics evaluated.

Authors in [25] investigated the routing protocol ZRP by improving an existing algorithm called zone-based routing with parallel collision guided broadcasting protocol The network's topology is controlled using an estimate of the node's energy dropout rate. The energy efficiency is measured enhanced to find an optimal QoS routing path and reduced overhead. The proposed protocol gave improved results in terms of performance as compared to other experimented protocols

The architecture presented in this work is a conceptual model of the proposed framework. The framework's cross-layered design is strongly in-line with OLSR's architectural design in that the higher layer can access services provided by the lower layer with no consideration of how the service is made available. The overall framework utilizes the TCP/IP model to fully articulate each stage. The layers exhibited are the data link layer, application layer, physical layer, network layer, and the transport layer. model. The layers exhibited are the data link layer, application layer, physical layer, network layer, and the transport layer.

## 7. PROPOSED QOS FRAMEWORK & ITS OPERATIONS

This section presents a conceptual (fig.1) and a high-level overview of the proposed security framework and the different components or technologies associated with it. It also presents assumptions that we made in connection with the framework.

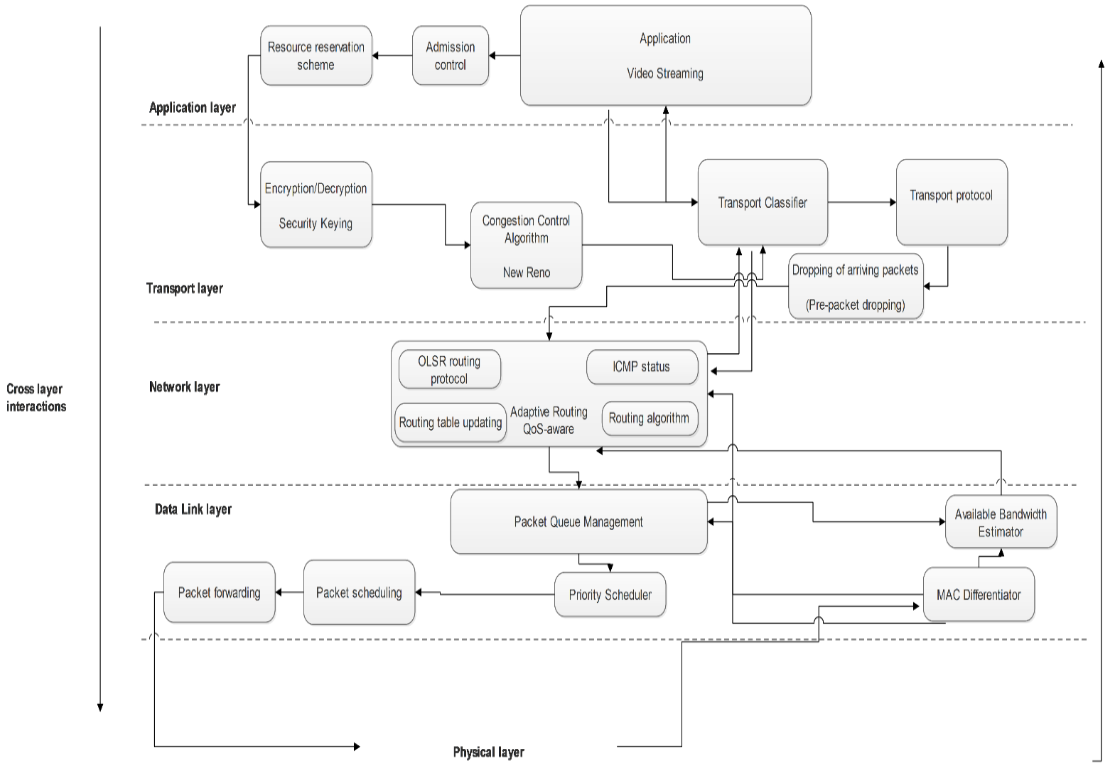


Figure 1. QoS-aware security framework

## 7.1. Application Layer

The application layer is mainly responsible for generating relevant traffic i.e. CBR, video, voice, email, and HTTP. To fully ascertain QoS, a resource reservation scheme is implemented. A resource reservation scheme will ensure that QoS for high priority sessions is guaranteed and that sufficient bandwidth is administered throughout the transmission phase to fulfill the fundamental requirements of QoS. The assigned resource reservation scheme will guarantee QoS performance as it decreases the chances of high priority session collisions while using the bandwidth. Security threats like malicious nodes, worms, and viruses ar. The end goal of the framework is the assurance of QoS within the network. This can be achieved through:

- QoS Medium Access Control (MAC) scheduling
- Admission Control
- QoS-aware routing
- Traffic policing

These four mechanisms are covered throughout the framework at different layers in terms of implementation. The proposed QoS-aware routing protocol solution will be based on:

- OLSR protocol and QoS
- OLSR protocol and MAC protocol

## 7.2. Transport Layer

At the transport layer, there are different activities involving the movement of packets from the application to the network layer. It focuses on end to end communication during data encryption. This can be done using the Transport Control Protocol (TCP) or User Datagram Protocol (UDP) protocol.

### 7.2.1. Congestion Control Algorithm

This is used to control the congestion within the network. In our work, we implement New Reno, an algorithm that improves retransmission during the quick-recovery phase of TCP Reno.

### 7.2.2. Transport protocol

A transport protocol has the prime responsibility of establishing and facilitating the movement of data from one node to the other. In this work, the TCP protocol is used because of the streaming data traffic.

### 7.2.3. Encryption

This is an effective way of achieving data security. A secret key will be used to have access to an encrypted file. This is called Decrypting. This work uses AES (Advanced Encryption Standard) algorithm.

## 7.3. Network Layer

The network layer is more central to the realization of a fully functional system whereby there is bi-directional communication between the transport and network layer moreover the network layer and the data link layer. This includes:

### 7.3.1. Adaptive routing (QoS-aware)

A process of determining the most efficient path in which a data packet can use in a network to reach its destination

### 7.3.2. Routing protocol

It determines how MANET nodes should communicate with each other by round-robin fashion that enables them in selecting optimal routes between any two nodes within the network.

### 7.3.3. ICMP status

This relays messages about the status of our IP address e.g. Destination Unreachable, Time exceeded and Trace Route

### 7.3.4. Packet scheduler and buffer

This contains the actual memory that is used to store packets. Additionally, the scheduler automatically builds a protective front (firewall) against hostile nodes and thus protecting network resources from saturation.

### 7.3.5. Packet classifier

This process strategically categorizes packets into flows. It contains a set of rules categorizing packets according to their header fields

### 7.3.6. Routing algorithm

This is a set of stepwise operations implemented to direct internet traffic more efficiently. It mathematically determines the best path to take during routing in the MANET.

### 7.3.7. Link state table domain

This a table that contains link information about all known MANET nodes exercising routing functionalities.

### 7.3.8. QoS scheme

These are mechanisms utilized for the attainment of an acceptable level of QoS in the network.

### 7.3.9. Admission control

This is a very important component of the proposed framework. It is key in the provisioning of QoS in the MANET because it determines the fair provisioning of network resources and the extent at which they are utilized. and if QoS characteristics are delivered. Admission control can be considered as a validation process where the checking is done before the establishment of a connection to calculate the sufficiency of network resources for a proposed connection.

## 7.4. Data Link Layer

The IEEE 802.11 standard is utilized at the data link layer. The MAC plays a huge role in linking the network and physical layer. The MAC address plays a central role in handling queues during routing and bandwidth estimation during cross-layer interactions between the data link layer and the physical layer. The responsibility of the link layer in this framework is towards the MAC Access control and it performs the following activities;

- Packet scheduling and forwarding
- Priority classification
- Packet Queueing
- Bandwidth Estimation

## 7.5. Framework Operations

The conceptual view of the framework presented above demonstrated some of the key stages in the actual prototype. The prototype in Fig 4.1 follows the same design mechanism like the one in Fig 3.6. The TCP/IP model is at the core of the design of the framework. At the upper level is the application layer which contains video traffic generation, admission control, and a resource reservation scheme which links up with the keying system at the transport layer.

The keying system is responsible for security encryption and decryption to prevent attackers from having access to information passed on from the source to the destination. The utilized encryption method is an improved AES standard as shown in Fig 3.7 of chapter 3. A proper data handover architecture was designed to fully support MANET environments. The alternative route in the security structure was designed to maintain data integrity and prevent passive attacks which may phish for information in the channel and observe activity within the channel. The next step would be the congestion control algorithm which in this instance is New Reno, an algorithm that improves retransmission and helps share. The transport classifier then takes over to the transport protocol which in this instance is Real-Time Messaging Protocol. This protocol is chosen over UDP and TCP based on its excellent delivery in video streaming traffic although traditional networks would opt for TCP as it guarantees packet delivery other than UDP which does not allow retransmission. Cross-layer interactions between layers (application and transport) continue because of the transport classifier which links up with the application.

The network layer is primarily responsible for QoS aware routing and that involves a technique called adaptive routing. Adaptive routing determines the best/optimal path a data packet should follow in the network to reach its preferred destination. This is achieved by using the shortest path algorithm which takes the data packet to destination with minimal congestion and thus efficiently using the network resources. This algorithm allows nodes to calculate routes in given network topology and thus saving time and minimizes overhead size. Adaptive routing improves network performance as routes adjust automatically in response to dynamic network topology. The nodes exchange route information and updates during adaptive routing.

This is necessary to fully adhere to QoS requirements. The routing protocol, OLSR, facilitates routing in the network layer. It is the highest decision making entity that is responsible for directing all packets. The routing table is constantly updated as the node moves randomly within the network. Most of the operations that happen in our framework depend on the network level. The ICMP status is for monitoring the IP connection of the network.

The data link layer provides more of a link address (MAC) to associate with the IP address shown at the ICMP status. The framework has an available bandwidth estimator. These addresses work to ensure that packets are properly scheduled, forwarded, and allocated appropriate resources (bandwidth) and that each node’s unique identifier (MAC address) is linked with an IP address. There are so many cross-layer interactions within this layer as the physical layer is also involved. When packets are forwarded to the physical layer then the process of moving from the physical to application layer begins. The packets will move from the MAC differentiator to the QoS aware mechanism, link up with the transport classifier, and lastly the application, where it all started.

### 7.5.1. Security Keying

There are several threats to the security of a MANET but again resource allocation plays a huge role build-up to a practical security structure. The Key management approach, AES, in this work is implemented because it prioritizes primary data protection and security. In traditional networks, Watchdogs and controllers are used to impose a well-functioning structure but involving such tools would surely drift away from the MANET architecture as MANETs have no central authority but every node acts out as its independent router and regulator. The proposed framework aims at achieving an acceptable level of QoS by securing the network from active and passive attacks. Figure 2 shows a keyed source node sending packets towards its intended destination.

The node broadcasts information to the network and as expected would have alternative route links (marked as Atl. Link in the figure). The information. Alt. Link 4 was able to decode the information from the source and will now be used as a backup in an event that the received information was incomplete or compromised. It then sends the information to the terminal node. The terminal node will then compare the two and compare the signal received from both the original link and the alternative link. This approach not only saves network resources but fully subscribes to the operation of MANETs in terms of the broadcast mechanisms used. There is no need for additional network devices to fully carry security. The primary focus would now be to provide good QoS in terms of end to end delay, packet loss, and throughput. The keying system depends on the functionality of the routing protocol called Multi-Point Relays (MPRs).

Every network workstation chooses a neighbouring node-set known as the MPR. OLSR is considered to be a proactive routing protocol for mobile ad-hoc networks. It uses a link-state algorithm for routing and thus making routes available immediately whenever it needs them. OLSR uses multipoint relays (MPRs) to minimize the overhead from broadcasting control messages. Through MPRs, the protocol can significantly reduce the number of retransmissions that are needed to broadcast a message to all network workstations. OLSR provides shortest path routes through the flooding of a partial link state. OLSR periodically maintains destination routes in the network. Another advantage of OLSR is that it regulates the reactivity of topological variations by decreasing the highest time interval for periodic control message transmission. This means more optimization can be achieved in denser network environments. This result is far better as compared to the traditional link-state algorithm.

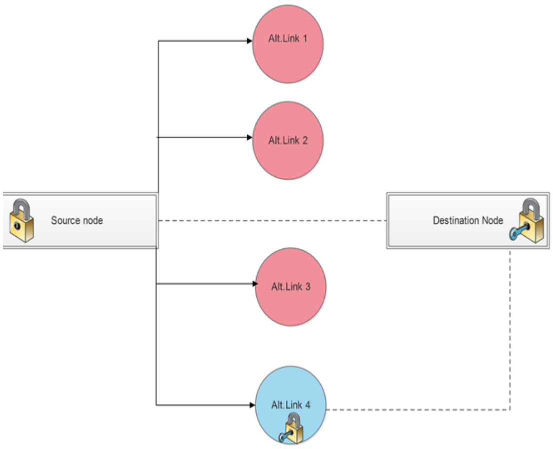


Figure 2. Security keying system

The security keying system used in the framework is the AES Standard for both encryption and decryption. AES [1]is chosen on the basis that it consumes fewer resources (e.g. battery power), fast, and most effective algorithms.

AES is sometimes referred to as the Rijndael block cipher operating on matrix blocks with 8-bit entries of size:

$$4 \times N_b \qquad (1)$$

Whereby $4 \leq N_b \leq 8$ is the block length.$N_b$ represents the number of bits.

Encryption scrambles the message and outputs it as unrecognizable data. Decryption takes the encrypted data which would be in the form of unrecognizable data and adds a source key to output the original message. The AES is proposed because of its less resource constraint, lightweight, and less computationally demanding features when it comes to routing. The AES is made up of a couple of initialization steps. The first step is the key expansion where the key is broken down to multiple subkeys and the other step is the initial round which involves substitution and transposition. The keys can be broken down into the following:

Table 1 AES algorithm processing properties [2]

| Bits | Cycles | Rating |
|---|---|---|
| 128-bits | 10 | Fastest |
| 192-bits | 12 | Slower |
| 256-bits | 14 | Slowest but most secure |

In this work, the 128-bit version of key management is implemented. A series of rounds are performed using the multiple sub-keys made during the expansion phase. The number of times the rounds are made depends on the size of the key we select as highlighted in Table 1. As part of the contribution of this work, we introduced an intermediate trusted node between the source and the destination. This unique node can relay the same message packets issued from the source node to the destination. It provides an alternate route to secure the integrity of the message. This is done to ensure that at the destination, all packets are delivered. In an event whereby Node A, for example, cannot get all the packets to Node D, then Node X acts as surety for complete packet delivery depending on the routing table as repetitive packets are discarded. This would mean that Node X is equipped with encrypted relay capabilities to Node D.

Node X acts as a Multi-Point Relay (MPR) node. MPR’s are trusted nodes within the network to relay routing information to the intended destination. This would practically mean that Node A selects Node X as an MPR then retransmits control packets from Node A. In the network, each transmitting node could have one or more of these MPR nodes. These are nodes that are selected by their 1-hop neighbours to retransmit all the broadcast messages it receives from that particular node provided that message is not a duplicate and that the message has a “Time to Live” field greater than one. Routes are selected by MPRs to avoid data packet transfer problems over uni-directional links. Each node will select its MPR set by using its 1-hop neighbours.

A group of MPR nodes is called the MPR set. The set is chosen such that it covers all symmetric 2-hop nodes and a coverage strictly confined within the radio range. An MPR set of Node X is denoted as MPR (N). The other nodes within the network not selected as MPR process control packets as it is a broadcast environment but do not forward the packets. If Node A has selected Node X and Y as its MPRs then it is safe to say:

$$\text{Node A: } MPR(A) = \{X,Y\} \qquad (2)$$

Where X and Y represent Node X and Node Y, MPR (A) is the set of MPR nodes belonging to Node A.

The MPR nodes are select based on a neighbour basis to the transmitting node. Each transmitting node uses HELLO messages to determine its MPR set. These HELLO messages are periodically broadcasted to one-hop neighbours and not forwarded. Through the neighbour list in the HELLO messages received, nodes can determine 2-hop neighbours and an MPR set. This MPR set is assigned a sequence number and the sequence number is incremented each time a new set is calculated. An MPR set is re-calculated when there is a change in 1-hop or 2-hop neighbourhood detected. MPR nodes are the only ones allowed to generate and propagate Topology Control (TC) messages. The advertisement before sending TC messages is not sent to all links in the network. MPRs minimizes the control traffic overhead of OLSR through retransmission of control messages. The technique significantly reduces the rate of transmissions needed to flood a message to all network nodes. The introduction of MPRs is to minimize the overhead of flooding messages and reducing redundant retransmissions in the network.

The encryption and decryption [2] methods may appear similar but in essence function differently and need to be separated. Rounds are several repeated AES repeated at a set number of times. Encryption has the following steps from round 0 till 9:

- Byte substitution
- Shift rows
- Mix columns
- Add Round key

In pseudo C notation, the above is derived as:

```
Round(State, RoundKey)
{
ByteSub(State);
ShiftRow(State);
MixColumn(State);
AddRoundKey (State, RoundKey);
}
```

For the last round execution (round set 10) for AES Encryption is presented in the following order as shown below:

- Sub byte
- Shift Row
- Add Round Key

In pseudo C notation, we can represent it as:

```
FinalRound(State,RoundKey)
{
ByteSub(State) ;
ShiftRow(State) ;
AddRoundKey(State,RoundKey);
}
```

Decryption has the following steps for the rounds 11 till 15:

- Add Round key
- Mix Columns
- Shift columns
- Byte substitution

For the last round of operation AES decryption has the following steps:

- Inverse Shift Rows
- Inverse Sub bytes
- Add Round Key

When reducing the number of rounds performed, this may reduce power consumption but would harm the security of the protocol by making it less secure. The 10 rounds of key expansion imposed on this work are done to strengthen the security of the protocol. Naturally, seven rounds or more can be considered fairly secure and energy-efficient.

### 7.5.2. Congestion Control Algorithm

In the framework, the presence of a congestion control algorithm is clearly outlined. Traffic load is one of the most important factors to be considered when addressing QoS. This is because the network resources and scalability thereof are tested within the transport and network layers respectively. When there are too many packets within a network, this may cause packet delay and loss. Packet delay and loss may affect the performance of the system and such a situation is called congestion. It is for that reason the cross-layer interactions are important among the transport and network layer as both layer share in the responsibility of safeguarding traffic load. When there is congestion, it would mean that the available network resources are limited/less than the load. It is the responsibility of the network to resolve congestion. For efficient operation, it is better to reduce load as highlighted in the framework (figure 4.1)

$$\text{Efficiency= Pre-packet dropping of arriving packets} \qquad (3)$$

Another possible solution towards congestion control would be to increase resources which is the job of the Admission control component of the framework. Admission Control is done before a connection is established so it is virtually impossible to increase resources in the middle of transmission as such a validation process is performed before transmission.

In this work, a congestion control algorithm called New Reno is used. It is an algorithm derived from the algorithm called Reno, which was proposed because of the inefficiency of Tahoe. Old Tahoe is the combination of the slow start and congestion avoidance algorithm. The later version of Old Tahoe is called Tahoe. Tahoe is an algorithm that works on duplicate ACK whereby retransmission happens without waiting for a timeout. During packet dropping, Reno enters fast recovery multiple times and thus decreasing the congestion by half. In scenarios where multiple packets are being dropped Reno does not, however, increase throughput. New Reno, the modified version, uses TCP to store a sequence of number that belongs to the highest data packet. New Reno implements fast recovery in the case of three duplicate acknowledgments and improves retransmission during the fast recovery phase of TCP Reno. When the partial ACK arrives, the congestion window is severely reduced by the amount of acknowledged data after the retransmitted packet. This acknowledged data is then called new data as shown in equation 3:

$$cwnd = cwnd - \quad + SMSS \qquad (4)$$

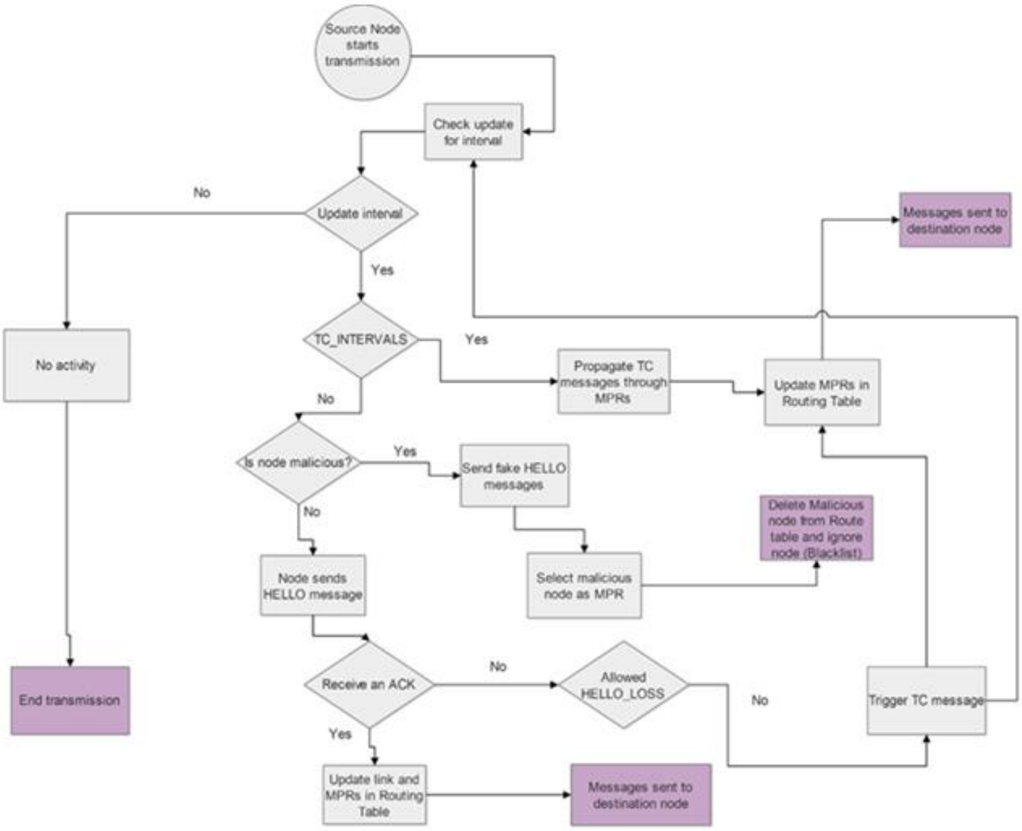
Where, Cwnd being the congestion window, is the new data and SMSS being the Sender Maximum Segment Size.

### 7.5.3. QoS-aware adaptive routing

The routing protocol, OLSR, facilitates all routing in the network. OLSR works using a link-state algorithm that constantly works with the routing table. The routing table is flexible to adaptive routing as the topology is dynamic and it needs constant updating. The packets will use an optimal path as directed by the MPRs to the destination. In this work, the routing flow of the protocol to get rid of other processes that may consume more of the limited network resources hence the use of the AES algorithm was modified. Nodes can exchange updates and route table information. Adaptive routing allows the routing path to change over time as the topology in which the nodes operate in is ever-changing. Another role player component in our framework is packet switching. Packet switching is regarded as a higher-level decision-making entity responsible for driving packets from source to destination.

The routing protocol, OLSR's flow chart in the proposed framework is shown in Figure 3 where we propose a few changes towards some of the traditional operations of the protocol to accommodate QoS and increase efficiency in terms of performance. Figure 3 shows the flow chart of our modified OLSR protocol. PRs still play a critical role in terms of propagating TC messages through to the routing table. An un-authenticated node will be regarded as a malicious node and will be isolated from the network. At first, the node will be sent a fake HELLO message then selected as an MPR then blacklisted at the routing table. OLSR needs constant updating of its route table due to the nature of MANETs and its dynamic topology.

After routing, packets are sent to the data link layer where there are packet queue management, MAC differentiator, and available bandwidth estimator. Under packet queue management, there is packet scheduling, priority scheduling, and packet forwarding.


Source Node starts transmission
Check update for interval
No
Update interval
Yes
Messages sent to destination node
No activity
TC_INTERVALS
Yes
Propagate TC messages through MPRs
Update MPRs in Routing Table
No
Is node malicious?
Yes
Send fake HELLO messages
No
Delete Malicious node from Route table and ignore node (Blacklist)
Node sends HELLO message
Select malicious node as MPR
End transmission
Receive an ACK
No
Allowed HELLO_LOSS
No
Trigger TC message
Yes
Update link and MPRs in Routing Table
Messages sent to destination node

### 7.5.4. Packet Queue Management

When packets arrive at the data link layer they are processed according to the First-In-First-Out rule. The first job to come in is scheduled first. The packet scheduler is responsible for providing the actual memory used for storing packets and providing a firewall used against malicious nodes whose intent is to selfishly use up network resources.

## 8. CONCLUSIONS

In this paper, we have presented and discussed a security framework from the perspective of QoS by using MANET routing protocol, OLSR. The architecture of the QoS-aware security framework was presented and its components were explained with the aim of QoS delivery in video streaming applications. The functions within the framework were explained in accordance to their respective interconnected layers. It is of paramount importance that whatever scheme is used in this architecture, the network resources are spared as best as possible since all nodes are moving randomly and in unfriendly topologies. Most approaches to network security do not consider the aspect of QoS hence our contribution to develop a QoS-centric framework that will not only look into the security aspect but also QoS delivery in the network. Computing the QoS of a MANET is due to the dynamic topology in which the mobile nodes operate in.

A custom flow chart of the routing protocol, OLSR, was also presented as part of our contribution to QoS-aware adaptive routing and improving the existing OLSR routing protocol. As future work, the utmost intention is to evaluate the proposed framework at both the network and application layers. This is a work in progress paper on its final evaluation stages.

## 9. ACKNOWLEDGMENTS

This work would not be possible without the support from the Faculty of Natural and Agricultural Sciences at the North-West University and our partners at TETCOS, India who provided the test-bed for our on-going practical and evaluation of the proposed framework.

### REFERENCES

[1] S. Semplay, R. Sobti, and V. Mangat, "Review: Trust management in MANETs," International Journal of Applied Engineering Research, vol. 7, p. 2012.

[2] D. S. Aarti, "Tyagi,“Study Of Manet: Characteristics, challenges, application and security attacks”," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, pp. 252-257, 2013.

[3] A. Rajaram and D. S. Palaniswami, "A trust based cross layer security protocol for mobile Ad hoc networks," arXiv preprint arXiv:0911.0503, 2009.

[4] P. Goyal, V. Parmar, and R. Rishi, "Manet: vulnerabilities, challenges, attacks, application," IJCEM International Journal of Computational Engineering & Management, vol. 11, pp. 32-37, 2011.

[5] Manoranjini, A. Chandrasekar and S. Jothi, "Improved QoS and avoidance of black hole attacks in MANET using trust detection framework", Automatika, vol. 60, no. 3, pp. 274-284, 2019. Available: 10.1080/00051144.2019.1576965.

[6] G. D. Delgado, V. C. Frías, and M. A. Igartua, "Video-streaming transmission with qos over cross-layered ad hoc networks," in 2006 International Conference on Software in Telecommunications and Computer Networks, 2006, pp. 102-106.

[7] Sahu and S. Sharma, "Secure and Proficient Cross Layer (SPCL) QoS Framework for Mobile Ad-hoc", International Journal of Electrical and Computer Engineering (IJECE), vol. 9, no. 4, p. 2603, 2019. Available: 10.11591/ijece.v9i4.pp2603-2613.

[8] Y. Singh and M. V. Siwach, "Quality of Service in MANET," Int. J. Innov. Eng. Technol, 2012.

[9] N. Sharma, S. Rana, and R. Sharma, "Provisioning of Quality of Service in MANETs performance analysis & comparison (AODV and DSR)," in Computer Engineering and Technology (ICCET), 2010 2nd International Conference on, 2010, pp. V7-243-V7-248.

[10] J. Sen, "A survey on reputation and trust-based systems for wireless communication networks," arXiv preprint arXiv:1012.2529, 2010.

[11] P. Madhavan, "Framework for QOS Optimization in MANET using GA-ACO Techniques," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 529-532, doi: 10.1109/ICACCS.2019.8728310.

[12] Nivedita and N. Nandhagopal, "Improving QoS and efficient multi-hop and relay based communication frame work against attacker in MANET", Journal of Ambient Intelligence and Humanized Computing, 2020. Available: 10.1007/s12652-020-01787-5.

[13] [13] I. Ahmad, H. Jabeen, and F. Riaz, "Improved quality of service protocol for real time traffic in manet," arXiv preprint arXiv:1308.2797, 2013.

[14] S. Xu, P. Guo, B. Xu, and H. Zhou, "QoS evaluation of VANET routing protocols," Journal of Networks, vol. 8, pp. 132-139, 2013.

[15] S. Singh, P. Kumari, and S. Agrawal, "Comparative Analysis of Various Routing Protocols in VANET," in Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on, 2015, pp. 315-319

[16] N. S. M. Usop, A. Abdullah, and A. F. A. Abidin, "Performance evaluation of AODV, DSDV & DSR routing protocol in grid environment," IJCSNS International Journal of Computer Science and Network Security, vol. 9, pp. 261-268, 2009.

[17] Tyagi, S. Som and S. Khatri, "Reliability-based dynamic multicast group formation provisioning local adjustment ensuring quality of service globally in MANETs", International Journal of Parallel, Emergent and Distributed Systems, pp. 1-15, 2019. Available: 10.1080/17445760.2019.1650040.

[18] V.Saravanan and D. Vijayakumar, "Performance Of Reactive And Proactive MANET Routing Protocols With Trajectories," in International Journal of Engineering Research and Technology, 2012.

[19] H. Singh and M. Singh, "Effect of Black Hole Attack on AODV, OLSR and ZRP Protocol in MANETs," International Journal of Advanced Trends in Computer Science and Engineering, vol. 2, 2013.

[20] P. Palta and S. Goyal, "Comparison of OLSR and TORA routing protocols using OPNET Modeler," in International Journal of Engineering Research and technology, 2012.

[21] L. Li and L. Lamont, "A lightweight service discovery mechanism for mobile ad hoc pervasive environment using cross-layer design," in Pervasive Computing and Communications Workshops, 2005. PerCom 2005 Workshops. Third IEEE International Conference on, 2005, pp. 55-59.

[22] J. Toutouh, J. García-Nieto, and E. Alba, "Intelligent OLSR routing protocol optimization for VANETs," IEEE transactions on vehicular technology, vol. 61, pp. 1884-1894, 2012.

[23] E. Winjum, A. M. Hegland, P. Spilling, and O. Kure, "A performance evaluation of security schemes proposed for the OLSR protocol," in Military Communications Conference, 2005. MILCOM 2005. IEEE, 2005, pp. 2307-2313.

[24] D. Spanos, "Intrusion Detection Systems for Mobile Ad Hoc Networks," 2018.

[25] Tamil Selvi and C. Suresh GhanaDhas, "A Novel Algorithm for Enhancement of Energy Efficient Zone Based Routing Protocol for MANET", Mobile Networks and Applications, vol. 24, no. 2, pp. 307-317, 2018. Available: 10.1007/s11036-018-1043-x.

[26] P. Rajakumar, V. T. Prasanna, and A. Pitchaikkannu, "Security attacks and detection schemes in MANET," in Electronics and Communication Systems (ICECS), 2014 International Conference on, 2014, pp. 1-6.

[27] S. Jain and N. Hemrajani, "Detection and mitigation techniques of black hole attack in MANET: An Overview," International Journal of Science and Research (IJSR), India Online ISSN, pp. 2319-7064, 2013.

[28] G. M. Borkar and A. R. Mahajan, "A secure and trust based on-demand multipath routing scheme for self-organized mobile ad-hoc networks," Wireless Networks, vol. 23, pp. 2455-2472, November 01 2017.

[29] M. S. Khan, M. I. Khan, O. Khalid, M. Azim, and N. Javaid, "MATF: a multi-attribute trust framework for MANETs," EURASIP Journal on Wireless Communications and Networking, vol. 2016, p. 197, 2016.

[30] F. A. Khan, M. Imran, H. Abbas, and M. H. Durad, "A detection and prevention system against collaborative attacks in Mobile Ad hoc Networks," Future Generation Computer Systems, vol. 68, pp. 416-427, 2017.

[31] D. Hurley-Smith, J. Wetherall, and A. Adekunle, "SUPERMAN: Security Using Pre-Existing Routing for Mobile Ad hoc Networks," IEEE Transactions on Mobile Computing, vol. 16, pp. 2927-2940, 2017.

[32] S. Sharma and M. Mahajan, "Security Mechanisms for Mitigating Multiple Black hole Attack In Manets," IJISE International Journal of Innovative Science, Engineering & Technology, vol. 2, pp. 582-588, 2015.

[33] N. Kaur and M. Joshi, "Implementing MANET Security using CBDS for Combating Sleep Deprivation & DOS Attack," Open Science Framework. June, vol. 26, 2017. [5] P. Goyal, V. Parmar, and R. Rishi, "Manet: vulnerabilities, challenges, attacks, application," IJCEM International Journal of Computational Engineering & Management, vol. 11, pp. 32-37, 2011.

[34] M. Lindeberg, S. Kristiansen, T. Plagemann, and V. Goebel, "Challenges and techniques for video streaming over mobile ad hoc networks," Multimedia Systems, vol. 17, pp. 51-82, 2011.

[35] L. L. Monica, "Evaluation of Attacks using different Parameters based on their performance," Evaluation, pp. 106-110, 2018.

# THE GENERAL LAW PRINCIPLES FOR PROTECTION THE PERSONAL DATA AND THEIR IMPORTANCE

Jonatas S. de Souza[1, 2], Jair M, Abe[1], Luiz A. de Lima[1, 2] and
Nilson A. de Souza[2, 3]

[1]Graduate Program in Production Engineering - Paulista University, São Paulo, Brazil
[2]National Association of Data Privacy Professionals - ANPPD - Scientific Committee, São Paulo, Brazil
[3]São Paulo State Supplementary Pension Foundation– PREVCOM, São Paulo, Brazil

## *ABSTRACT*

*Rapid technological change and globalization have created new challenges when it comes to the protection and processing of personal data. In 2018, Brazil presented a new law that has the proposal to inform how personal data should be collected and treated, to guarantee the security and integrity of the data holder. The purpose of this paper is to emphasize the principles of the General Law on Personal Data Protection, informing real cases of leakage of personal data and thus obtaining an understanding of the importance of gains that meet the interests of Internet users on the subject and its benefits to the entire Brazilian society.*

## *KEYWORDS*

*Data, Law General, Personal, Protection, Regulation, Legal.*

## 1. INTRODUCTION

The concern about the protection of people's data has grown over the years, but only after the approval of the Brazilian Law that received the name of Marco Civil da Internet, established by Law No. 12,965, 2014 [1]. In Brazil, a new law has recently been sanctioned and it is generating a lot of discussion in several areas. The General Law on Personal Data Protection, Law No. 13.709 [2] of 14th August 2018, gives the Brazilian population rights and guarantees on how organizations will have to adapt to the collection and processing of personal data, whether by physical or digital means.

Discussing data protection in Brazil has become a challenging task. The state of the crisis provoked by COVID-19 (coronavirus) had a severe impact on companies, which began to adopt measures to make their workforce compatible with the demands existing during social isolation, and the adoption of measures to minimize the risk of the disease spreading among their workforce.

The use of Virtual Private Network - VPN and practices such as BYOD (Bring Your Own Device) have become common to incorporate daily life. There was also an exponential growth of

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 109-120, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101110

e-commerce, home office, webinars, virtual meetings, and numerous activities that started to occur entirely through the Internet.

In the same proportion, the risks associated with the improper use of personal data, data leaks, improper access by third parties, theft of data kept by corporate servers, creation of fake profiles, fake news, among other practices frequently reported were multiplied. The objective of the paper is to present important aspects such as the principles and fundamentals of Brazilian law and to present some real cases on data leaks.

The paper is composed of sections, in Section 2 presents the Theoretical Reference that will address the history of data protection in Brazil and the European Regulation, in Section 3 describes the Principles of Brazilian Law demonstrating the similarity with the European Regulation, in Section 4 the Importance of Brazilian Law showing the fundamentals of the Law and emphasizes the importance of consent of the data holder, in Section 5 the Results and Discussions with real cases of data leaks and countries that already have some legislation on protection of personal data, in Section 6 are the Conclusions bringing the final considerations obtained.

## 2. THEORETICAL REFERENCE

### 2.1. Personal Data Protection in Brazil

In Brazil, the legislation is based on the positivist model of law, adopted by Lusitanian, German and Italian schools that privilege the written law, this reflects in the delay of the implementation of the legislative process (figure 1), which begins with the initial idea, passes through the creation of the bill, then through the bicameral approval and then the presidential sanction, to finally come into force with coercive force.

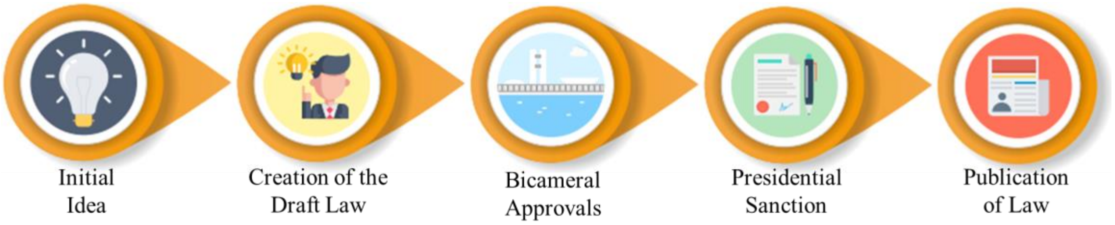


Fig. 1: Law Creation Process.

The first Brazilian initiative on personal data protection was in Article 5 of the 1988 Federal Constitution [3].

Art. 5 All are equal the law, without distinction of any nature, guaranteeing Brazilians and foreigners residing in the country the inviolability of the right to life, freedom, equality, security, and property, under the following terms [3]:

X - The intimacy, privacy, honor, and image of persons are inviolable, and the right to compensation for material or non-material damage resulting from their violation is guaranteed [3].

XII - The secrecy of correspondence and telegraphic communications, data, and telephone communications shall be inviolable, except in the latter case by judicial order, in the cases and the manner established by law for criminal investigation or criminal proceedings [3].

Law No. 9.296, of July 24th, 1996 [4] deals with the interception of telephone communications and regulates clauses XII, Art. 5 of the Federal Constitution.

On September 11th, 1990 [5] Law No. 8.078, known as the Consumer Code (CDC), was enacted, bringing in its Article 43 the guarantee of access to the holder's data, demanding clarity and objectivity of the information and the possibility for the consumer to demand the correction of his registration data [5].

Art. 43 The consumer, without prejudice to the provisions of Art. 86, shall have access to the information existing in registers, files, records, personal data, and consumption filed about him, as well as to their respective sources [5].

Paragraph 1. Consumer registrations and data must be objective, clear, truthful, and in easy-to-understand language, and may not contain negative information for a period longer than five years [5].

Paragraph 2. The opening of the registration, file, record, personal, and consumption data shall be communicated in writing to the consumer when not requested by him [5].

Paragraph 3. The consumer, whenever he finds any inaccuracy in his data and registrations, may demand their immediate correction, and the archivist shall, within five working days, communicate the change to the eventual recipients of the incorrect information [5].

Paragraph 4. Databases and registers relating to consumers, credit protection services, and the similar are considered public entities [5].

Paragraph 5. Once the statute of limitations on the collection of consumer debts has been consummated, the respective Credit Protection Systems shall not provide any information that may prevent or hinder new access to credit with suppliers [5].

Paragraph 6. All information referred to in the caption of this article must be made available in accessible formats, including for persons with a disability, at the request of the consumer [5].
Even bringing some progress on personal data protection, the CDC was still limited in its scope on the subject, which means that the protection would exist in the relationship between supplier and consumer within the scope of the legal concepts established in Articles 2 and Article 3 [5] of the CDC.

On April 23rd, 2014, Law No. 12,965, now known as Marco Civil da Internet [1], was approved, establishing principles, guarantees, rights, and duties for the use of the Internet in Brazil, and has the guarantee of privacy and protection of personal data, and will only make such data available through a court order. In Art. 7, clauses I, II and III, and clauses VII, VIII, IX, and X, deal with the rights of the holders of personal data [1].

Art. 7 Access to the Internet is essential to the exercise of citizenship, and the user has assured the following rights [1]:

I - Inviolability of intimacy and privacy, their protection and compensation for material or moral damage resulting from their violation [1].

II - Inviolability and secrecy of the flow of your communications over the Internet, except by court order, in the form of the law [1].

III - Inviolability and secrecy of your stored private communications, except by court order [1].

VII – Do not provide third parties with your data, including connection records, and access to internet applications, except by free, express and informed consent or in the cases provided by law [1].

VIII - Clear and complete information about the collection, use, storage, treatment and protection of your data, which may only be used for purposes that: a) justify their collection; b) are not prohibited by law, and c) are specified in service contracts or terms of use of internet applications [1].

IX - Express consent on the collection, use, storage, and processing of personal data, which shall occur in a manner detached from the other contractual clauses [1].

X - Definitive exclusion of personal data that you have provided to a certain internet application, at your request, at the end of the agreement between the parties, except for the cases of mandatory storage of records provided for in this Law [1].

The Civil Framework of the Internet also includes aspects of the responsibility for the protection of personal data by access providers and in operations carried out through the Internet, providing for some sanctions, described in Articles 10, 11, and 12 [1].

On August 4th, 2018, Law No. 13,709, called the General Law on Personal Data Protection [6], was approved, providing for the processing of personal data, whether digital or not, to protect the fundamental rights of freedom and privacy and the development personal personality of the individual in society.

## 2.2. General Law on Personal Data Protection

The General Law on Personal Data Protection - LGPD, Law No. 13.709 of 14th August 2018, which would come into force in August 2020, has been postponed by Provisional Measure No. 959/2020 [6] extending the vacatio legis [7] and postponed to 3 May 2021 [8]. The LGPD purpose is to provide guidelines on how personal data will be collected and processed, and to ensure the security and integrity of the data holder, whether digital or not. On 10th July 2018, Project Law 53/2018 - PLC [9] was approved by the plenary of the Federal Senate and was sanctioned on 14th August 2018 by the 37th President of Brazil [2]. Article 1 of the LGPD states that it is prepared to protect the processing of personal data to protect the rights of freedom, privacy, and personality development of the individual. Moreover, it applies to any individual or legal entity that carries out-processing operations such as collection, production, reception, classification, processing, among other activities by physical or digital means in Brazilian territory, or abroad if it is using personal data of individuals living in Brazil.

## 2.3. General Data Protection Regulation

The General Data Protection Regulation 2016/679 [10] – GDPR, of the European Parliament and of the Council of European Union – EU, of 27th April 2016, is a Regulation that is on the protection of individuals about the processing of personal data and the free movement of such data and that repeals Directive 95/46/EC [11], EU companies had two years to comply with the regulation by the date of 28th May 2018. The Regulation applies to all activities involving the processing of personal data using full or partial consent, as well as to the processing of personal data by non-automated means.

# 3. THE PRINCIPLES OF LGPD

For a better understanding of LGPD [2], it is necessary to know the legal bases (principles) that should be observed for any type of data processing activities, the Law is composed of ten principles that are listed in Art 6. A GDPR [10] [12] is also guided by principles [13], which are set out in Article 5, which form the basis for the EU Regulation, and these principles should be linked to data processing.

## 3.1. Principle of Purpose

The LGPD [2], the purpose for which the data will be done must be very specific, explicit, and informed to the holder of the personal data that will be processed [2]. In GDPR, the Purpose Limitation principle [13], the data must be collected for specific, legitimate, and explicit purposes, and may not be processed for other unspecified purposes [10].

## 3.2. Principle of Adequacy

The LGPD [2], is the formality with the holder of the personal data to process personal data [2]. In GDPR, the Storage Limitation principle [13], data may be stored in a database until the end of the data processing and must be informed to the data owner, and after the end of the processing, the data must be deleted from the database. It is also linked to the principle of Bidding [13] that the company that will process the data must comply with the Regulation and with the data holder. [10].

## 3.3. Principle of Necessity

In the LGPD [2], the amount of data for data processing is only relevant, proportional, and not excessive [2]. In the European Regulation, the Data Minimization principle [13], data should be collected following its purpose and only data that are necessary for the processing [10].

## 3.4. Principle of Free Access

The LGPD [2], guarantees that the data holder will have free access to the data in its entirety at any time, and this principle is linked to the GDPR Transparency principle. In the European Regulation there is a right which is described in Article 17 [10], which is called Right to Erasure [10] or Right of Forgetfulness, which gives the "right to be forgotten" to the data holder of the database concerning the purpose of the processing, after the data holder has requested to delete the data, the officer shall delete the data relating to the data holder's request [10].

## 3.5. The Principle of Data Quality

The LGPD [2], guarantees the data owner clarity, accuracy, and relevance and updates the data according to the needs of the data treatment [2]. The Accuracy principle of GDPR [13] that data should always be updated and correct thus maintaining the quality of the data that will be processed and incorrect data will be rectified or deleted [10].

## 3.6. Principle of Transparency

The LGPD [2], ensures that the data owner will have access to all necessary information clearly, accurately, and easy access to data processing [2]. The Transparency Principle of GDPR is divided into three words, Lawfulness, Fairness, and Transparency [10] [13]. The Lawfulness or

Bidding is concerned, data controllers should comply with the Regulation, on Fairness or Loyalty, it is stated that processing should take place fairly with the consent of the data owner, and on Transparency, the data controller will allow him to have access to all information of the data processing [10].

### 3.7. Principle of Security

The LGPD [2], will use techniques for the protection of personal data from unauthorized access or accidental or illicit situations of alteration, destruction, loss, dissemination, and communication [2]. The principle that about security in GDPR is the Integrity principle and Confidentiality [13], the data must be stored securely, guaranteeing the data integrity, and adopting methods of protection against unauthorized processing, loss, accidental damage, destruction, or unauthorized access [10].

### 3.8. Principle of Prevention

It will use methods to prevent data processing damaging [2] [10].

### 3.9. Principle of Non-Discrimination

Data may not be processed for discrimination, illicit or abusive purposes [2].

### 3.10. Principle of Accountability and Reporting

In Brazilian Law [2], it is up to the treatment agent to prove the purpose and which effective methods have been adopted, and he must be able to prove compliance with and enforcement of personal data protection rules, including the effectiveness of these methods [2]. In the European Regulation, the Accountability principle [13], which is the full responsibility of the data processing agent, guarantees the length of the purpose of the processing and has evidence of the necessity of the processing [10].

## 4. THE IMPORTANCE OF LGPD

The LGPD [2], sanctioned in Brazil, was inspired by GDPR of the EU [10] and contains many similarities in its respective principles. In its Art. 2 they show the foundations (figure 2), that served as a basis for the development of the Law [2]:

Art. 2 The discipline of personal data protection based on the according to fundamentals [2]:

I - Respect for privacy [2].
II - Informative Self-determination [2].
III - Freedom of Expression, Information, Communication, and Opinion [2].
IV - The Inviolability of Intimacy, Honour, and Image [2].
V - Economic and Technological Development and Innovation [2].
VI - Free Enterprise, Free Competition, and Consumer Protection [2].
VII - Human Rights, Free Development of Personality, Dignity, and the exercise of citizenship by natural persons [2].

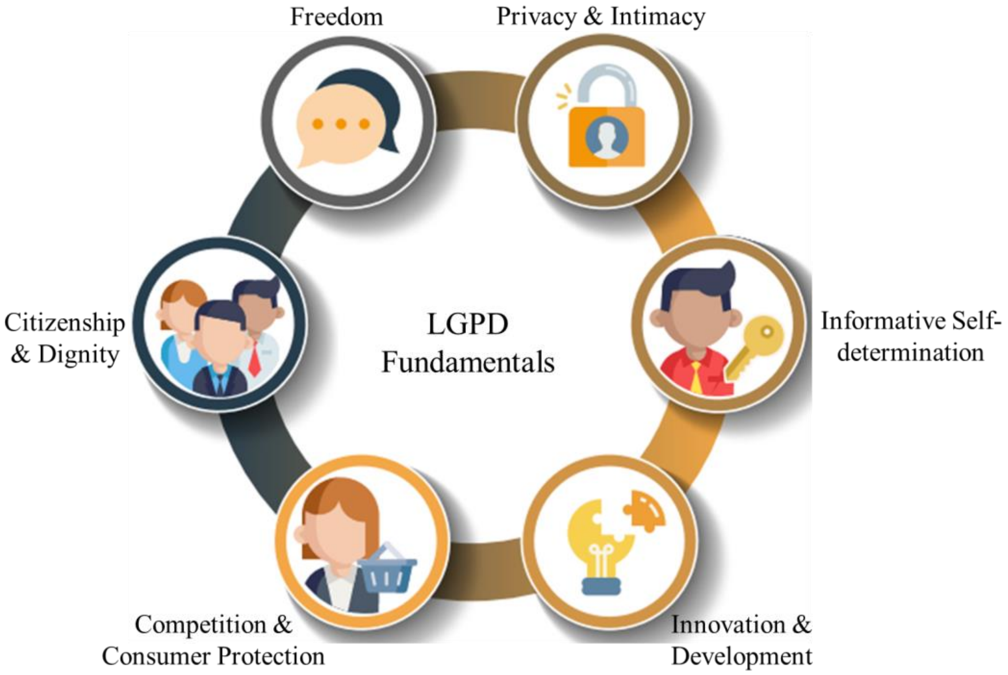


Fig. 2: LGPD Fundamentals [14].

One of the most important points that makes data processing possible is to have the consent of the data holder, according to Article 7, clauses 1 [2], and in Article 8 [2] it is reinforced that the authorization must be in writing or by other means of manifestation of the owner will, and it is stated from Paragraph 2 [2] that in case the authorization is in writing it must be highlighted in the contractual clauses. In the case of processing sensitive personal data, Article 11, clauses 1 [2], states that with the consent of the holder, and Article 14, Paragraph 1 [2], which states must have the consent of parents or legal guardians concerning the processing of personal data of children and adolescents.

In GDPR [10] it is also explicit that for any activities that require a data processing must have the consent of the data holder, in Article 6, Paragraph 1, clauses A [10], it says that data processing will be lawful upon the consent of the data holder for specified purposes previously informed to the data holder, in Article 7 [10] which sets out the conditions applicable to consent, it says that the data processing agent must prove that the data holder has agreed to the specified purposes. As regards the processing of data on children, Article 8 of the European Regulation [10] requires the person legally responsible for the child under the age 16 to consent to the processing. The State may be responsible for giving consent if the child is under the age of 13 and has no family members to answer for him or her.

Without a reference law for the use of personal data, the possibility of abuse in the collection and use of personal data is increased, as well as the encouragement of several other non-specialized bodies to issue their opinions regarding the use of data, which causes great confusion. This is the case, for example, of inspections and inspections by the Public Prosecutor's Office and consumer protection agencies, the issuance of opinions by regulatory agencies, or even judicial decisions based on various sparse legal provisions [1] [5] that seek to define parameters for the processing of personal data.

In the graph (figure 3), shows the level of interest in Internet users searches on the terms LGPD and GDPR over a twelve-month period, where the term LGPD represents the blue line and the term GDPR represents the red line, on the horizontal axis represents the time and on the vertical axis represents the level of search made on the terms, these levels are represented by the numbers 0 (very low), 25 (low), 50 (average), 75 (high) and 100 (very high). This simple analysis shows that the red line had several peaks in some periods, this because the GDPR since 2016 [10] is approved and had an adequacy period of two years and the level of interest is between average and very high, the blue line has had small peaks, this because the LGPD is a new subject in Brazil and this makes the level of interest is between very low and average. In general, users looking for LGPD and GDPR terms are professionals in the juridical environment or Information Technology.

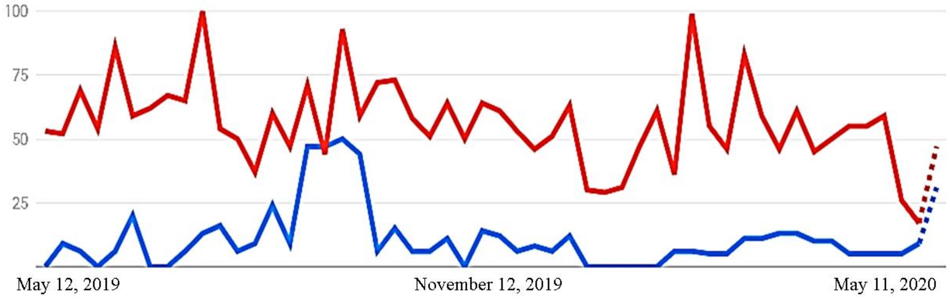


Fig. 3: Level of interest in the terms LGPD and GPDR [25].

## 5. Discussions

It should be noted that both the Brazilian Law and the European Regulation, they provide a guide for data processing and what procedures companies should take to comply with the law if these principles are not followed these companies will be at serious legal risk. An example of non-compliance with the law was the Cambridge Analytica scandal [15], which misused data from 87 million Facebook users (figure 4), manipulated the data without the consent of the data holders, to help win Donald Trump's US presidential campaign, and for the British to vote to leave the European Union, both in 2016 [15], Facebook was asked about data security.

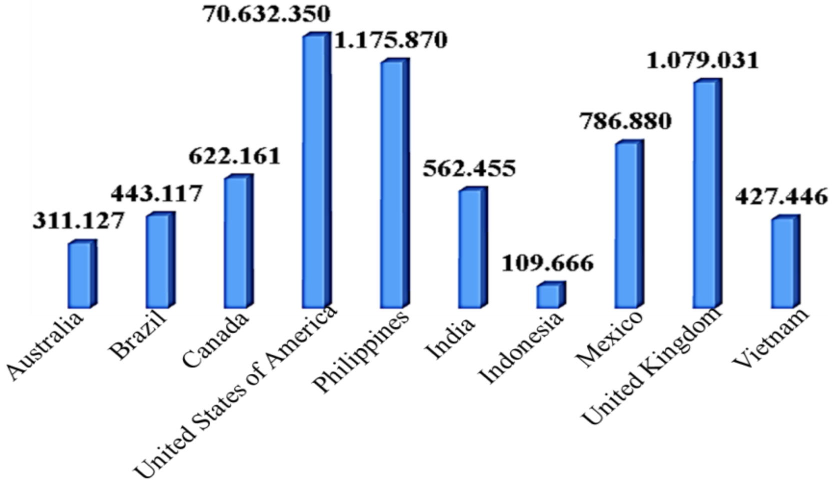


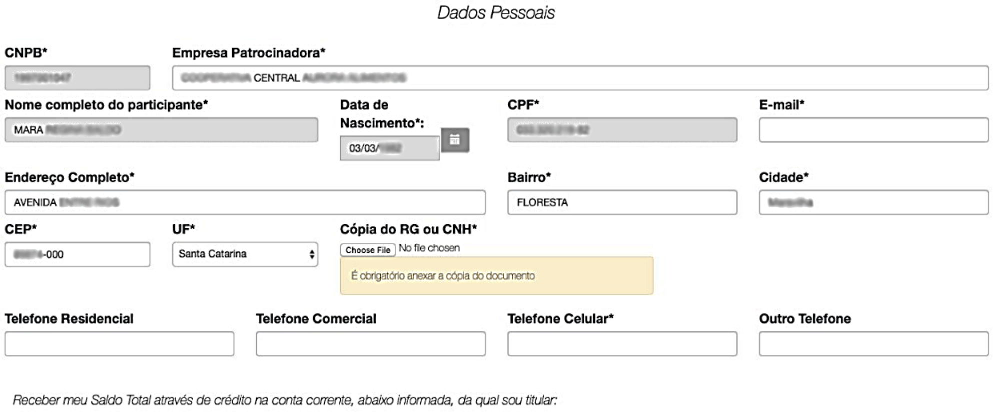
Fig. 4: Number of Facebook users who may have had their data used improperly with Cambridge Analytica [16].

In Brazil, there have been several cases of data loss, such as the case of the Netshoes website, according to the Coordinator of the Commission for Personal Data Protection, Prosecutor Frederico Meinberg, "this is one of the largest security incidents recorded in Brazil" [17], which because of the data leak could put the integrity of 1,999,704 users at risk if the leaked data fell into the wrong hands.

The impact that data leaks go far beyond the financial losses, the exposure of each citizens' information can be irreversible damage that becomes impossible to measure the size of the loss. Without an information security policy, it can cause serious problems such as the invasion of vital systems to steal tax returns, data, making illegal financial transfers, interrupting the strategic operations of a company, or the government.

Another case about data leakage was written by Liliane Nakagawa and published on the website Olhar Digital [18], which displays the news about the banking institution, specifically the Bank of Brazil Provident Fund [17]. According to Nakagawa, data leak that reaches 153 thousand clients - official number of registered in the BB Previdencia platform, according to Bank of Brazil. The source, who identified the security gap, stated that through the private pension system, aimed at companies and public agencies, it is possible to have access to all personal data of participants and, from breaking, editing and registering beneficiaries, all in the name of the registered person himself [18].

After this news, several headlines were reporting the incident, the Exame magazine published on its website, "BB Previdencia website leak exposes data of 153 thousand clients" [19], the newspaper, O Estado de S. Paulo, published on its website, "Security sheet on BB Previdencia website exposes client data" [20] (figure 5).



*Dados Pessoais*

CNPB*: [redacted]
Empresa Patrocinadora*: [redacted] CENTRAL [redacted]
Nome completo do participante*: MARA [redacted]
Data de Nascimento*: 03/03/[redacted]
CPF*: [redacted]
E-mail*:
Endereço Completo*: AVENIDA [redacted]
Bairro*: FLORESTA
Cidade*: [redacted]
CEP*: [redacted]-000
UF*: Santa Catarina
Cópia do RG ou CNH*: Choose File No file chosen
É obrigatório anexar a cópia do documento
Telefone Residencial:
Telefone Comercial:
Telefone Celular*:
Outro Telefone:

*Receber meu Saldo Total através de crédito na conta corrente, abaixo informada, da qual sou titular:*

Fig. 5: Personal Data Empty from the BBTurPrev Plan Withdrawal Page [18].

For these leaks not to occur, companies must have a Data Protection Officer – DPO [2] [10], where the primary function is to ensure that the organization processes the personal data of their employees, their customers, their suppliers or any other individuals securely and reliably according to the data protection rules of law [10].

An LGPD will give the right to protection of the personal data of the respective holders and will give guidelines to the companies on how the treatment should be done. Brazil will be adapting to GPDR and will move the job market for data protection specialists. However, Brazil [21] already has a law for the creation of a supervisory body to verify whether companies comply with the LGPD, but directors have not yet been appointed to the National Data Protection Authority - ANPD and the National Council for Personal Data Protection and Privacy [8]. The European body responsible for supervising undertakings on whether they comply with the European Regulation is the European Data Protection Supervisor - EDPS, an independent supervisory authority established according to EU Regulation 2018/1725, and its task is to ensure that the fundamental rights and freedoms of individuals - in particular their privacy - are respected when EU institutions and bodies process personal data. In the world, there are already some countries [22] outside the EU that have a regulation regarding data protection. On the European Commission's website, it informs countries that are at an appropriate level to the Regulation, the European Commission has recognized Andorra, Argentina [23], Canada (trade organizations), Faeroe Islands, Guernsey, Israel, Isle of Man, Jersey, New Zealand, Switzerland, Uruguay [23] and the United States of America (limited to the Privacy Shield framework) as providing adequate protection [24].

## 6. CONCLUSIONS

Through the Internet Civil Framework, which establishes rights and duties, guarantees and principles for the use of the Internet in Brazil, it does not guarantee data protection and privacy in a well-structured, complete and comprehensive manner, nor is a general regulation on the protection of personal data, and its provisions on data protection not protective in nature.

Some of the challenges identified for implementing the Law in Brazil are legal adjustments and appropriate training, a complete action plan for companies to comply with LGPD, specialized implementation of personal data governance processes, information security technologies, educating Brazilian society about this Law by showing the rights and duties of citizens.

Therefore, there will still be a lot of debates and discussions about LGPD and whether it will adhere to GDPR, and how Brazil will behave with the law when it becomes effective.

## ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) -Finance Code 001.

## REFERENCES

[1] BRASIL “Lei Nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil, Marco Civil da Internet.” Diário Oficial da União, 24 de abril de 2014, Edição 77, Seção 1, p 1-3.

[2] BRASIL “Lei Nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet).” Diário Oficial da União, 15 de agosto de 2018, Edição 157, Seção 1, p 59-64.

[3] BRASIL. Constituição, 5 de outubro de 1988. CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL DE 1988. Diário Oficial da União, 5 outubro de 1988, Edição 191-A, Seção 1, p 1-32.

[4] BRASIL “Lei Nº 9.296, de 24 de julho de 1996. Regulamenta o inciso XII, parte final, do art. 5° da Constituição Federal. - LEI DA ESCUTA.” Diário Oficial da União, 25 de julho de 1996, Edição 143, Seção 1, p 13757

[5] BRASIL. Lei nº 8.078, de 11 de setembro de 1990. Dispõe sobre a proteção do consumidor e dá outras providências. Diário Oficial da União, 12 de setembro de 1990, Edição Suplemento, Seção 1, p 1

[6] BRASIL. “Medida Provisória Nº 959, De 29 De Abril De 2020. Regras para o auxílio emergencial e adiamento da vigência da LGPD.” Diário Oficial da União, 29 de abril de 2020, Edição 8-A, Seção 1 - Extra, p 1.

[7] “Vacatio legis — Senado Notícias”, Senado Federal. [Online]. Available in: <https://www12.senado.leg.br/noticias/glossario-legislativo/vacatio-legis> [Accessed May 11, 2020].

[8] BRASIL. Lei nº 13.853, de 8 de julho de 2019. Altera a Lei nº 13.709, de 14 de agosto de 2018. Para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados; e dá outras providências. Diário Oficial da União, 8 de julho de 2019, Edição 130, Seção 1, p 1-3.

[9] BRASIL Projeto de Lei da Câmara N° 53. 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, 23 de abril de 2014. [Online]. Available in: <https://www25.senado.leg.br/web/atividade/materias/-/materia/133486> [Accessed May 11, 2020]

[10] REGULAMENTO. In: EUR - Lex. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). [Online]. Available in: <https://eur-lex.europa.eu/legal-content/pt/ ALL/?uri=CELEX:32016R0679>. [Accessed May 11, 2020].

[11] JORNAL Oficial das Comunidades Europeias. Directiva 95/46/CE dO Parlamento Europeu e do Conselho de 24 de outubro de 1995 relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desse dado. Luxemburgo, October 24, 1995. [Online]. Available in: <https://eur-lex.europa. eu/legal-content/PT/TXT/PDF/?uri=CELEX:31995L0046&from=PT> [Accessed May 11, 2020]

[12] Gabel, Detlev, & Tim Hickman. GDPR - Handbook: Unlocking the EU General Data Protection Regulation. September 13, 2017. [Online]. Available in: <https://www.whitecase.com/publications/article/gdpr-handbook-unlocking-eu-general-data-protection-regulation?s=Handbook:%20Unlocking> [Accessed May 11, 2020]

[13] The Principles. s.d. [Online]. Available in: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/> [Accessed May 11, 2020]

[14] Setor de Tecnologias Educacionais – SETED, Senac, 2019

[15] Cambridge Analytica teve acesso à 87 milhões de contas. April 4, 2018. [Online]. Available in: <http://www.meioemensagem.com.br/home/ultimas-noticias/2018/04/04/cambridge-analytica-teve-acesso-a-87-milhoes-de-contas.html> [Accessed May 11, 2020]

[16] Schroepfer, Mike. An Update on Our Plans to Restrict Data Access on Facebook. April 4, 2018. [Online]. Available in: <https://newsroom.fb.com/news/2018/04/restricting-data-access/> [Accessed May 11, 2020].

[17] Moreira, Braitner. Netshoes deverá procurar 2 milhões de clientes afetados por vazamento, diz MP. 2018. [Online]. Available in: <https://g1.globo.com/df/distrito-federal/noticia/mp-pede-que-netshoes-tome-providencia-apos-vazamento-de-2-milhoes-de-contas.ghtml> [Accessed May 11, 2020].

[18] Nakagawa, Liliane. "[EXCLUSIVO] Previdência Privada do Banco do Brasil vaza dados de 153 mil clientes", Olhar Digital - O futuro passa primeiro aqui. May 6, 2020. [Online]. Available in: <https://olhardigital.com.br/noticia/-exclusivo-previdencia-privada-do-banco-do-brasil-vaza-dados-de-153-mil-clientes/100395> [Accessed May 11, 2020].

[19] Flach, Natália. "Vazamento de site da BB Previdência expõe dados de 153 mil clientes", Exame. May 06, 2020. [Online]. Available in: <https://exame.abril.com.br/negocios/vazamento-de-site-da-bb-previdencia-expoe-dados-de-153-mil-clientes/> [Accessed May 11, 2020].

[20] Kerber, Diego. "Falha de segurança em site da BB Previdência expõe dados de clientes - Economia - Estadão". May 7, 2020. [Online]. Available in: <https://economia.estadao.com.br/noticias/geral,falha-de-seguranca-em-site-da-bb-previdencia-expoe-dados-de-clientes,70003295592> [Accessed May 11, 2020].

[21] Motta, Matheus. Quais os princípios do GDPR e seu impacto no Brasil? September 4, 2018. [Online]. Available in: <https://medium.com/@adtail/quais-os-princ%C3%ADpios-do-gdpr-e-seu-impacto-no-brasil-5e92cc76a57c> [Accessed May 11, 2020]

[22] Proteção de Dados Pessoais Pelo Mundo. April 17, 2015. [Online]. Available in: <http://pensando.mj.gov.br/dadospessoais/2015/04/protecao-de-dados-pessoais-pelo-mundo/> [Accessed May 11, 2020].

[23] Paixão, Pedro. Proteção de dados na América Latina. July 10, 2018. [Online]. Available in: <https://cio.com.br/protecao-de-dados-na-america-latina/> [Accessed May 11, 2020].

[24] "Adequacy decisions", European Commission - European Commission. [Online]. Available in: <https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en> [Accessed May 11, 2020].

[25] "GOOGLE TRENDS - c Trends Tools" May 11, 2006. [Online]. Available in: <https://trends.google.com.br/trends> [Accessed May 11, 2020]

## AUTHORS

**Jonatas Santos de Souza** has a degree in Information Systems from the Facudade Impacta de Tecnologia – FIT (2016). Has experience in Computer Science, with emphasis on Information Systems, working mainly on the following topics: Artificial Intelligence, Paraconsistent Analysis Network, Paraconsistent Logic, Industry 4.0, and Artificial Neurons. He holds a Post-Graduation in Management and Governance of Information Technology by Anhanguera (2019), and a Master's Degree in Production Engineering by Universidade Paulista - UNIP, being a CAPES scholarship holder. Member of the Scientific Committee of the National Association of Data Protection Professionals- ANPPD.

**Jair Minoro Abe** received B.A. and MSc in Pure Mathematics - University of Sao Paulo, Brazil. I also received the Doctor Degree and Livre-Docente title from the same University. He is currently the coordinator of the Logic Area of Institute of Advanced Studies - University of Sao Paulo Brazil and Full Professor at Paulista University - Brazil. His research interest topics include Paraconsistent Annotated Logics and AI, ANN in Biomedicine, and Automation, among others. He is a Senior Member of IEEE. Leader of the Research Group "Paraconsistent Logic and Artificial Intelligence" cataloged at CNPq Fellowship,

**Luiz Antonio de Lima** Doctor of Science student in Production Engineering Universidade Paulista, Master's degree in Production Engineering in the area of Artificial Intelligence Applied to Software Paraconsistent Measurement Software, Post-Undergraduate Degree in EAD, University Professor, General Coordinator of IT Course and Campus Assistant: 2008-2009. Speaker and Event Organizer: SENAED; NETLOG; WICS; WINFORMA. IT Consultant and/or roles: IT Director, Commercial Director, Project Manager, with clients: WCI-MahlerTI, WCI-Ericsson, WCI-Frema, WCI-Brasil Brokers, WCI-Gt1, The WCI-Consoft-IPESP, WCI-Consoft-SPPREV, WCI-Consoft-PMSP, WCI-Eversystems-Rede Globo / BankBoston, and WCI- Consoft-BankBoston: using best practices in the market: ITIL, COBIT, SIX SIGMA - Black Belt, CMMi, PMBOK, SCRUM, APF, APT, UCP.

**Nilson Amado de Souza** has a degree in Pedagogy from the University of North Paraná (2010). Has experience in Science and Technology, with emphasis on Information Technology POST-GRADUATE - Teaching for Higher Education FALC - Faculty Aldeia de Carapicuíba POST-GRADUATE - Specialization in Business Intelligence Faculty Impacta de Tecnologia - FIT, ITIL Expert certification; ISO 20000 certification; Privacy Data certification; DPO Privacy and Data Protection Foundation certification; Privacy and Data Protection Practitioner Privacy and Data Protection Essentials ISO-27001

# IoT Model for Smart Universities: Architecture, Challenges, and Applications

Amr Adel

Department of Technology & Innovation, Whitecliffe College
of Technology and Innovation, Auckland, New Zealand

*ABSTRACT*

*The utilization and implementation of IoT based technology in the learning environment is effective as all the activities among everyone is related to technology only. Also the adoption of the use of the technology is preferred by many. Hence, proposing a model that can regulate the various process of institution is essential as well as advantageous for the instructors as well as to the learners. Hence, the various aspect that comes up in adopting such a model is briefly described along with the flow that the model will exercise. The benefits, challenges and its application in the field of education is represented well. The researchers can further expand the study and refine as well as carry out future research in this domain.*

*KEYWORDS*

*Internet of Things, Higher Education, E-Learning, Radio-frequency identification, Learning Environment & Learner.*

## 1. Introduction

The advancement of technology in the twenty first century has accelerated the development of technologies as well as the innovations in the same field. Technology have elated several industries starting from construction, healthcare, food as well as the field of education. The effective application of the technologies is helping the students as well as the instructor to ease the process of learning. The concept of e-learning and smart class are the new picture of the modern world. The devices that runs on the concept of IoT are the prime provider of advantages in the learning process. The sensors and RFID tags are the basis of these technologies.

Therefore, the discussion is about the adoption of IoT based model in higher education. This will be started with the evidences of the earlier implementation of the same. The architecture of the proposed model is discussed leading to the flow of the operations of the proposed model, then the analytics that coursed in the learning process, followed by the benefits that the model can provide to all the stakeholders considered in the model along with the challenges they may face while adopting the same. Lastly, the application of the model in the education system is discussed in brief.

## 2. Related Work

According to Sarnok & Wannapiroon, (2018), leaning is connected to the development of the personality of the humans. Hence the ways in which the learning process is carried out is also

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 121-130, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101111

essential to regulate with the changing world and economy. The author has discussed about the opinion of the psychologist that the process of learning is never ending hence, evolving the process with the developed technologies are a must [29]. The journey of learning has the aim to provide “transformative learning” with the developing skills that are required to be a successful person in that very field. The learning in the twenty first century is expected to change and the new set of skills must be introduced. The new process that are quoted by the author are learning that is problem-based, effective interaction, connectivism, learning that is project-based, content delivery through the use of latest technology and constructivism [11]. In today’s world the management of education has different perspective of the classroom that the earlier one. This has been observed in the higher education mostly. The student of different age group and generations are coming in one classroom for learning [25]. It has been observed that the learning process enhance in doing do and sharing of knowledge has been effective for all the participant’s candidates in the room. The internet of everything is referred to the devices that are connected via the internet. The RFID are considered to be the new era helpers of humans. The new and advanced technologies do not only are present in the equipment of home appliances like TV, remote or cars, but also in the field of education. The instructors of education are expected to get familiar with the technical and functional based teaching instruments [30]. Therefore, it will enable the instructors to stimulate the learning process for the new generation students. The use of the devices needed the proper environment in which the students can perform the activities on the digital devices. The perception of connective learning can also be fulfilled through the digital devices irrespective of the place of learning. The authors have provided a framework for connectivism learning with the aid of IOT. The development of smart classroom is formed by the internet of things and the system consist of three parts those are cameras those are IoT based, those will capture the activities taking place inside the classroom to keep track of the environment, safety, etc. [26]. The next on is the system of smart check, it will have the devices like sensors and tags to monitor the activities of the instructors and the students and provide analysis for betterment. The last one is the office system that is IoT based to develop the devices used by the learners in performing better with the new age technology with controlled budget and more usability out of them [19].

As per Miraz et al., 2015, the usability of Internet of things have accelerated in several fields and the field of literature is not away from this too. The future is expected to have the utilization of the technologies at greater levels and new versions of Internet of everything [5]. The benefits that are stated by the use of this technology is that the co0nnecting with every possible object in the world with the possibility of shared knowledge and intelligence [19]. As the acceleration of urbanization flourishes the implementation of the IoT based devices are becoming critical. The e-learning is also developed by the implementation of IoT based devices, I provide easy access of learning materials for the students. It allows monitoring of the students for keeping track of the progress in the education field and giving effective feedback.

The benefits of cloud computing are concrete with for exploring the IoT based leaning. The presence of clod computing is helping in overcoming the difficulties of new age technologies in terms of storing and data accessing purposes. It eases the ability to store information related to education filed infinitely [3]. The services of cloud are well utilized by the students or learners so as to collaborate with other learners irrespective of the location and time zone [22]. The information accessed by the IoT devices are stored in the cloud and hence makes the accessibility easy. One of the feature of cloud is to provide interconnectivity of the devices and information. This allows the learners to collaborate and learn conveniently and effectively. The requirement of physical classroom is also solved by the present of cloud computing, as the teachers can easy conduct online based formal class with the operations of cloud services.

## 3. ARCHITECTURE OF PROPOSED MODEL

The educational model that is developed for the purpose of adopting the IoT based model in the higher education in New Zealand is consist of several participating department of the university, those are collected to other several divisions. The concept of IoT is primarily based on the use of sensors and tags. These devices collect all the activities and events occurring in the university, All the participating entities that are present in the proposed model such as; students, teachers, management department, library, parents, laboratory, gymnasium, classroom, staff, canteen, restroom, auditorium and the Main gates of the university are all equipped with sensors and tags. The basic work of the sensors is to capture the activities around, those activities may be in any form such as physical, electrical or chemical and then convert them to electrical signals. Those signals will be stored in the cloud of the university as a set of information of one entity of the model. Therefore, in our model the entities will be surrounded by the sensors either in the room they are working or in the form of wearable devices. Then all the information captured are transferred through three possible stages and then the timer regulates the timing of the information pass and finally store in the cloud where the analysis of the data and information are performed and revert back whenever needed by the authority of the university. As the data generated by the devices are obtained in huge quantity the implementation of the concept of Big data is also valuable to manage the inventory activities and make the education process more effective, attractive and interesting experience for the students. The model (Fig.1) will work on single server and the network bandwidth for the transmission of data will be wide enough to regulate all the activities smoothly.

### 3.1. The Proposed Model

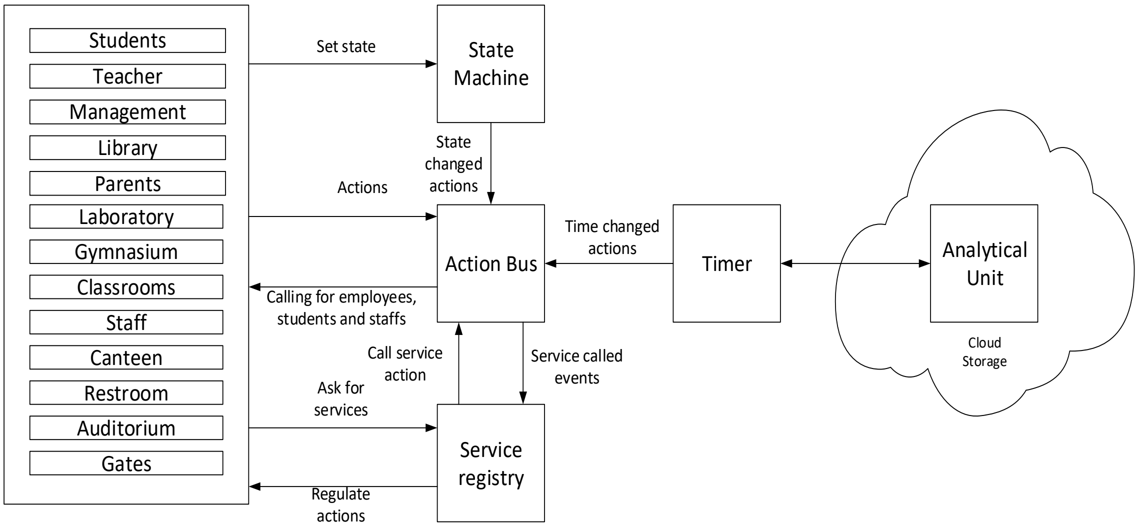


Figure 1. Educational Model IoT-based

The sensors associated with the participating entities in the proposed model have the basic functionality of gathering data and transfer them with encrypted protection to the other department as represented in the model. Whenever a sensor captures a set of information the state machine changes the state of that sensor to show that it has performed its work. Then the action bus helps in managing the data with the help of the APIs. Then the service registry provides the requested services of the different information collected through the API query. The data that is collected in the huge quantity are first rectifies for its credibility and then sent for the further operations. The data are translated into understandable form and analysed by the analytics of Machine Learning [1]. The flow of data into the proposed model is doe in the possible manner; firstly, the data in captured through the sensors present in the surrounding as well as through the wearable present to the staffs, students, and the instructors. Then the generated data is collected to analyse the credibility of the sources it got captured so that the huge amount of data can be minimized, and further analysis process of the data will take lesser time too. The next phase is the setting the status of the operation, when one operation is completed that is associated with the set machine and the status of the operations set to be completed or incomplete[14]. Furthermore, the operation that are assigned as the state changed are further goes to the Action bus, the stakeholders can ask for the actions that is obtained from the earlier generated and analysed data. Similarly, the action bus also calls for the stakeholder to capture the data directly. Next the service registry is present to regulate the actions of the stakeholders of the system on the basis of the data generate and captured. The actions are called, and service are reverted back to the action bus. The timer sets the fixed time of the data to be entered into the cloud storage and retrieval time [15]. Hence maintaining the proper translation of the data from the stakeholders and back to the system. Then the last step of the system is to store the data in the cloud store with the presence of infinite space and the retrieval of the data and information in need by the stakeholders. Hence the overall system is quite complicated yet simple in the flow of the information to extract the important data and use them for the betterment of the institute in the long run.

### 3.2. Analytics of the Learning Process

The analytics of learning are the basis of gathering, analysis and distribute information and data of the learners in their own environment. This analytical process occurs to optimize the opportunities it provides further. The proposed system is powered by the Machine learning concept for the smart work optimization. As the ability of the human brain is to keep the short-term memory and long term memory the system also keeps all the data generated in the form of usable data and memory in the system [32]. The learning analytics are present to provide shape to the data obtained. The analytical approach that is commonly used to find the patterns in the data collected by the sensors or the tags are by the structural approach of machine learning [7]. Also, the data mining, factor analysis are other helpful ways in which the analytics are cleared at effective manner. The ability to predict from the obtained data is required to provide manifestation to the proposed model [10]. The already existing data in the system will be used to compare with the new data and predict for the future with the aids of tolls and technologies of the machine learning theories.

All the components of a particular classroom will be connected to a single node and thus the sensors and other devices to collect data can run smoothly on them. From the wearable devices and the standalone physical devices installed in the classrooms will capture the activities of the students and process them in the model as discussed above [20]. Then the probable output will be obtained to have suggestion, improvements about the environment of the student, their learning abilities, and several other aspects that the instructors can work on and make changes as required.

### 3.3. Benefits of the Model

The benefits that the proposed model will bring into the higher education of New Zealand are by the emerging technologies that facilitates the smooth operations of the participating entities. The technological innovation in the twenty first century has bought revolution in the field of education. The ability to provide a cohesive and collaborative approach by the staffs and the student t under one roof is beneficial for provide the new age of knowledge [4]. The benefits the model is expected to reflect for the higher education are as follows:

### 3.3.1.Improved Results of Classroom

The IoT based classroom and overall system of the higher education will allow to acquire more knowledge of all the student around and their capabilities and abilities in terms of performance in the classroom. Those statistics can be analysed well to gain the positivity in the activities and try to explore them and the downsides to be improved in the future.

### 3.3.2. Enhance the Skills of Critical Thinking

The use of the devised for conducting the learning process will be equipped with the sensors and the tags hence the smart features will enable the learners as well as the instructors to enhance their skills from the previous track records [27].

### 3.3.3. Introduction of New Techniques of Solving Problems

The present of new technologies and innovative ways to learn like through the projectors, understand scenarios in the three dimensions, this will enhance the analytical skills of the learners.

### 3.3.4. Increased Interaction of the Learners

The devices will encourage in collaborative learning as well as enable the learners to interact more with the devices and the co-learners. The interaction with the instructors will not be only limited to the classrooms, even the laptops and the mobile phones will be able to access the lecture sessions.

### 3.3.5. Online Library

The books from the libraries will be available in the cloud storage of the education centre. Then the traditional ways of issuing of books will be avoided this was time taking. The availability of the online books will be made available to the students [9]. The students will be given a period of subscription to access the book from the cloud through their mobile devices. The sensors will provide notification of the return date of the book that means subscription will be closed for the particular book. In this way the shift from the paperback to the online based will formulate minimizing the use of paper as a whole.

### 3.3.6. Personalizing lectures

As the sensors will be able to gather data about each student in the classrooms, then providing personalized lecture sessions will be possible. This will enable the student to have better understanding of their week point in learning and help them in scoring better.

### 3.3.7. Authorised Access into System of Institute

when the system will be fully equipped with the technologically advanced devices such as the sensors and other IoT device. All the operations occurring across the institute will be tracked and monitored for the unauthorised access [24]. The proper screening will enable the management to perform regulatory process on the basis of the obtained data [12]. The entry and exist gates can be monitored better. The information system devices will have more security as the continuous monitoring will be followed automatically.

### 3.3.8. Better health of the stakeholders

The devices containing the sensors and RFID tags available to the stakeholders in the form of wearable or physical devices attached at the different departments of the institute will collect all the habits of the members and keep track of the food habits and hygiene [13]. This will allow the management to give update about the health of the members and keep them up-to-date with their health condition. The aim is to reduce the sick possibilities in the institute and maintain healthy participant inside the campus. This will keep the records of the institute at the higher places.

## 3.4. Challenges & Limitations in adopting the model

The difficulties that may occur at the time of adopting the IoT based education by the stakeholders are as follows:

### 3.4.1. Installation Cost

The overall model of the proposed system consists of high cost technologies and hence the arranging all the components together will include a huge amount of money [6]. Hence the installing the overall model together can cause the investment of a huge amount of money of the institute. Also the competition is high for adopting technology based system in every field and especially in the field of education.

### 3.4.2. Lack of knowledge

The authorities present in the institute such as the staffs, instructors and students may not have all the proper knowledge about the devices and the technologies used in the system [31]. Hence the adoption of all the operations of the system may take a longer period of time, this may reduce the progress of the institute. All the people in the institute may not have the complete knowledge of the operations across the model.

### 3.4.3. Security Concerns

As the model is based on the emerging technologies; hence the operation is taking place via the internet only. The internet is filled with several hackers and attackers that causes vulnerability to the devices available on the internet [17]. In the same sense this model will also be vulnerable to the attacks. Hence, the chance of data breach, malicious attacks to the systems are always possible to occur.

### 3.4.4. Management of Huge Data

The system has several number of sensors and RFID tags, the function of these devices is to collect all the possible data in its surrounding. Thus, a huge amount of data is generated at a greater level. The management as well as analysis of the data are a huge concern for the management team of the system [2]. The possibility of data loss or incorrect data occurs due to this problem. This may lead to the incorrect analysis to the data and results will also alter according to those.

### 3.4.5. Privacy Concerns

The privacy policies of the new technologies are way strict in actual practice. This may become a wall in accepting the model by everyone operating in the system [8]. This model involves the presence of information of the students, instructors and the staffs hence any misplace of the information may lead to the privacy disturbance of the concerned authorities of the details. Hence, many may not agree in adopting the model.

## Failure of Software or Hardware

The model is totally based on the operations of the hardware of the institute and the advanced software for the operation of all the activities. The old versions of software may become hindrance in the proper functionality of the model [18]. The failure of the hardware such as non-functionality of the monitor, keyboards, CPU will stop the operations of the model. The disturbance in the institute's network will not let the operation perform as decided. The unavailability of the server which provides power to the system will stop the working of the model.

## 3.5. Application of the Model

The unified architecture of the model enables in providing several application of the physical life of the every individual. This technology is used mainly for enhancing the abilities of the tools for education, provide better experience and obtain most of the usability out of them. Hence, the following applications are:

### 3.5.1 Effective Security System

As the functions of the technology is totally based on the tools that are equipped with all the advanced technologies and the system is totally based on the online platform, the ability to provide secure operation of every activity is possible [21]. The allowance of any tasks with effective encryption and decryption from the sender to the receiver is available. Hence, any transaction containing sensitive data is possible through the system. The system is best for storing the personal details of the students, teachers and the employees of the institute and maintain the credibility of the data as well as the source of their obtaining.

### 3.5.2 Voice Recognition

The sensors and the tags of the IoT devices are have the aim of obtaining the data that is occurring in their surroundings. Along with that the voice of the nearby objects are also getting captured at the same time. Hence, the system allows to convert the captured voice into usable data [28]. The ability of the sensors is utilized and explored well with the presence of the ability of the system to capture the voices and provide the exact output. The advancement of the technology has provided the ability to easily recognize the voices captured from the devices.

### 3.5.3 Data Translation

The techniques of the tools to capture data and analyse them are associated with the advancement of the system proposed. The ability to transform data into understandable state is the speciality of the techniques involved with the proposed system [23]. The translation of the data or the image or the voice captured are possible to obtain is the required for. Hence the proposed system is applicable to the various departments of the institute.

### 3.5.4 Healthcare System

As seen that the system captures the activities around the sensors and then analyse them to provide meaningful output. The propped system is also applicable to use the system to keep track of all the staffs, students, teacher's details in the aspect of their health. Then the behavioural pattern can be generated out from the details taken of each person [16]. The pattern can be used for predicting the possible health issues and with the proper analysis at the time the solution of the complicate health problems can be solved easily.

## 4. Future Directions

The research presents an initial design of Educational Model IoT-Based to provide the necessary support to both students and staff. Based on the challenges stated in section 3.4, the future focus is to improve the performance of the model in terms of software maintained, management of large volumes of data and lack of technical training. An extended work is expected to be made in that regard to facilitate adopting the proposed model into higher education and move to fully online learning & management in educational institutions. Case study is to be discussed in the extended version to find the attractive impacts and results on IoT simulation of the proposed model in higher education.

## 5. Conclusion

It can be concluded that the use of IoT for the purpose of combining it in the model for an e-learning process or providing a smart class for the students of higher education is proposed well. The main element of the model is the presence of sensors and wearable that collects the data and information from it surrounding and processing it to provide predictive actions. The instructors are highly benefits by this concept as they become more familiar with the students and they are able to provide the exact and direct requirement of the student. The advantages of the model are clearly discussed above and it is expected the challenges can be overcoming in the long run of the processing of the system in the educational system.

## References

[1] Ahn, J., Campos, F., Hays, M., & DiGiacomo, D. (2019). Designing in Context: Reaching beyond Usability in Learning Analytics Dashboard Design. *Journal of Learning Analytics*, *6*(2), 70-85.

[2] Amendola, S., Lodato, R., Manzari, S., Occhiuzzi, C., &Marrocco, G. (2014). RFID technology for IoT-based personal healthcare in smart spaces. *IEEE Internet of things journal*, *1*(2), 144-152.

[3] Asghar, M. H., Negi, A., &Mohammadzadeh, N. (2015, May). Principle application and vision in Internet of Things (IoT). In *International Conference on Computing, Communication & Automation* (pp. 427-431). IEEE.

[4] Boninger, F., Molnar, A., &Saldaña, C. M. (2019). Personalized learning and the digital privatization of curriculum and teaching. *National Educational Policy Center). Whitepaper accessed at https://nepc. colorado. edu/publication/personalized-learning*.

[5] Amasha, M. A., Areed, M. F., Alkhalaf, S., Abougalala, R. A., Elatawy, S. M., &Khairy, D. (2020, February). The future of using Internet of Things (loTs) and Context-Aware Technology in E-learning. *In Proceedings of the 2020 9th International Conference on Educational and Information Technology* (pp. 114-123).

[6] Brunner, M., Keller, U., Wenger, M., Fischbach, A., &Lüdtke, O. (2018). Between-school variation in students' achievement, motivation, affect, and learning strategies: Results from 81 countries for planning group-randomized trials in education. *Journal of Research on Educational Effectiveness, 11*(3), 452-478.

[7] Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on industrial informatics, 10*(4), 2233-2243.

[8] Duroc, Y., & Vera, G. A. (2014). Towards autonomous wireless sensors: RFID and energy harvesting solutions. In *Internet of Things* (pp. 233-255). Springer, Cham.

[9] Gunawardena, A. (2017). Brief Survey of Analytics in K12 and Higher Education. *International Journal on Innovations in Online Education, 1*(1).

[10] Halliday, J., & Anderson, M. (2016). Developing a framework for the visualisation of learning analytics in UK higher education. In *Developing effective educational experiences through learning analytics* (pp. 119-142). IGI Global.

[11] Bayani, M. (2020, March). The Influence of IoT simulation in the Learning process: A Case study. *In Proceedings of the 2020 8th International Conference on Information and Education Technology* (pp. 104-109).

[12] Husamuddin, M., & Qayyum, M. (2017, March). Internet of Things: A study on security and privacy threats. In *2017 2nd International Conference on Anti-Cyber Crimes (ICACC)* (pp. 93-97). IEEE.

[13] Jang, Y., Kim, J., & Lee, W. (2018). Development and application of internet of things educational tool based on peer to peer network. *Peer-to-Peer Networking and Applications, 11*(6), 1217-1229.

[14] Joksimović, S., Kovanović, V., & Dawson, S. (2019). The journey of learning analytics. *HERDSA Review of Higher Education, 6*, 27-63.

[15] Khan, B. H., Corbeil, J. R., &Corbeil, M. E. (Eds.). (2018). *Responsible analytics and data mining in education: Global perspectives on quality, support, and decision making*. Routledge.

[16] Khan, M. S., Islam, M. S., & Deng, H. (2014). Design of a reconfigurable RFID sensing tag as a generic sensing platform toward the future Internet of Things. *IEEE Internet of things journal*, *1*(4), 300-310.

[17] Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, *58*(4), 431-440.

[18] Li, S., Da Xu, L., & Zhao, S. (2015). The internet of things: a survey. *Information Systems Frontiers, 17*(2), 243-259.

[19] Liu, M., Zou, W., Li, C., Shi, Y., Pan, Z., & Pan, X. (2019). Using learning analytics to examine relationships between learners' usage data with their profiles and perceptions: A case study of a MOOC designed for working professionals. In *Utilizing learning analytics to support study success* (pp. 275-294). Springer, Cham.

[20] Loeb, S., & Byun, E. (2019). Testing, accountability, and school improvement. *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 94-109.

[21] Madakam, S., Lake, V., Lake, V., & Lake, V. (2015). Internet of Things (IoT): A literature review. *Journal of Computer and Communications*, *3*(05), 164.

[22] Maksimović, M. (2018). IOT concept application in educational sector using collaboration. *Facta Universitatis, Series: Teaching, Learning and Teacher Education*, *1*(2), 137-150.

[23] Matharu, G. S., Upadhyay, P., & Chaudhary, L. (2014, December). The internet of things: Challenges & security issues. In *2014 International Conference on Emerging Technologies (ICET)* (pp. 54-59). IEEE.

[24] Meenakumari, J., &Kudari, J. M. (2015). Learning Analytics and its challenges in Education Sector a Survey. *Int. J. Comput. Appl*, 0975-8887.

[25] Miraz, M. H., Ali, M., Excell, P. S., & Picking, R. (2015, September). A review on Internet of Things (IoT), Internet of everything (IoT) and Internet of nano things (IoNT). In *2015 Internet Technologies and Applications (ITA)* (pp. 219-224). IEEE.

[26] Nouby, A., &Alkhazali, T. (2017). The effect of designing a blended learning environment on achievement and deep learning of graduate students at the Arabian Gulf University. *Open Journal of Social Sciences*, *5*(10), 248-260.

[27] Pak, K., & Desimone, L. M. (2019). Developing principals' data-driven decision-making capacity: Lessons from one urban district. *Phi Delta Kappan*, *100*(7), 37-42.

[28] Roselli, L., Mariotti, C., Mezzanotte, P., Alimenti, F., Orecchini, G., Virili, M., & Carvalho, N. B. (2015, January). Review of the present technologies concurrently contributing to the implementation of the Internet of Things (IoT) paradigm: RFID, Green Electronics, WPT and Energy Harvesting. In *2015 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet)* (pp. 1-3). IEEE.

[29] Sarnok, K., &Wannapiroon, P. (2018). Connectivism learning activity in ubiquitous learning environment by using IoT for digital native. *Veridian E-Journal, Silpakorn University (Humanities, Social Sciences and arts)*, *11*(4), 405-418.

[30] Tyler-Wood, T. L., Cockerham, D., & Johnson, K. R. (2018). Implementing new technologies in a middle school curriculum: a rural perspective. *Smart Learning Environments*, *5*(1), 22.

[31] Wang, K. H., Chen, C. M., Fang, W., & Wu, T. Y. (2018). On the security of a new ultra-lightweight authentication protocol in IoT environment for RFID tags. *The Journal of Supercomputing*, *74*(1), 65-70.

[32] Wise, A. F. (2019). Learning analytics: Using data-informed decision-making to improve teaching and learning. In *Contemporary technologies in education* (pp. 119-143). Palgrave Macmillan, Cham.

# A PROCESS FOR COMPLETE AUTONOMOUS SOFTWARE DISPLAY VALIDATION AND TESTING (USING A CAR-CLUSTER)

Malobika Roy Choudhury

Innovation and Technology, SAP Labs India Pvt Lmt.,
Bengaluru, Karnataka, India

## *ABSTRACT*

*Every product industry goes through the process of product validation before its release. Validation could be effortless or laborious depending upon the process. Here in this paper, a process is defined that can make the task-independent of constant monitoring. This method will not only make the work of test engineers easier it will also help the company meet stringent release deadlines with ease. The method explores how to complete visual validation of the display screen using deep learning and image processing. In the example, a method is discussed wrt a car-cluster display screen. The method breaks down the components of the screen then validates each component against its design and outputs a result predicting whether the displayed content is correct or incorrect. The models like You-Only-Live-Once, Machine Learning, Convolution Neural Networks-Conv2D, and image processing techniques like Hough circle/Hough lines are used to predict the accuracy of each display component. These sets of algorithms compile to provide consistent results throughout and are being currently used to generate results for the validation process.*

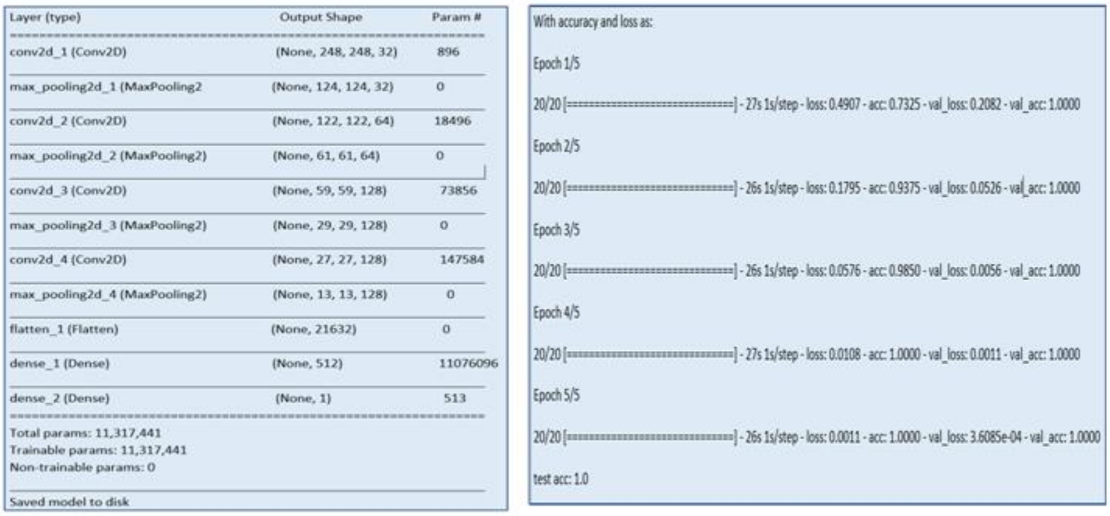
## *KEYWORDS*

*Convolution Neural Networks, You-Only-Live-Once, display-validation.*

## 1. INTRODUCTION

Before a product is released to the market it undergoes a lot of cycles. From development to release it goes through a massive amount of testing and validation. Most of the validation and testing is usually accomplished by test engineers who try and make sure that the final product is market-ready and follows Lean Software Principles. But to achieve this task they must spend hours of their time into visually validating the product. This paper aims to provide a method that can help reduce this time, help reduce last-minute defects (helps reduce cost, saves reputation), and make testing truly Automated. Several methods have been attempted towards this area but seldom discussed in the public forum as it advances to be propriety. In the behavior Driven testing field methods of hard coding are used to test a feature. Several tools are available in the market which are preferred like Test.ai, Testim.io, and other APIs that help in Automation testing. Although most companies have confidential data and tend to not use the available tools for validation, they instead have their tools. Another method deployed is where the validation is outsourced, and the client company provides expensive solutions. This paper wants fully automated visual testing accessible by all.

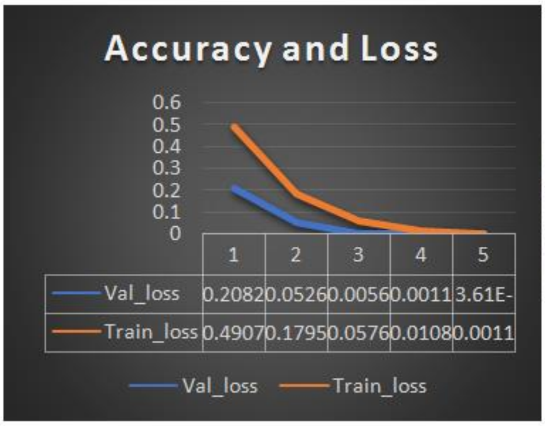
Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 131-139, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101112



```
Layer (type)                          Output Shape              Param #
=======================================================================
conv2d_1 (Conv2D)                     (None, 248, 248, 32)      896
_______________________________________________________________________
max_pooling2d_1 (MaxPooling2          (None, 124, 124, 32)      0
_______________________________________________________________________
conv2d_2 (Conv2D)                     (None, 122, 122, 64)      18496
_______________________________________________________________________
max_pooling2d_2 (MaxPooling2)         (None, 61, 61, 64)        0
_______________________________________________________________________
conv2d_3 (Conv2D)                     (None, 59, 59, 128)       73856
_______________________________________________________________________
max_pooling2d_3 (MaxPooling2)         (None, 29, 29, 128)       0
_______________________________________________________________________
conv2d_4 (Conv2D)                     (None, 27, 27, 128)       147584
_______________________________________________________________________
max_pooling2d_4 (MaxPooling2)         (None, 13, 13, 128)       0
_______________________________________________________________________
flatten_1 (Flatten)                   (None, 21632)             0
_______________________________________________________________________
dense_1 (Dense)                       (None, 512)               11076096
_______________________________________________________________________
dense_2 (Dense)                       (None, 1)                 513
=======================================================================
Total params: 11,317,441
Trainable params: 11,317,441
Non-trainable params: 0
_______________________________________________________________________
Saved model to disk
```

```
With accuracy and loss as:

Epoch 1/5

20/20 [==============================] - 27s 1s/step - loss: 0.4907 - acc: 0.7325 - val_loss: 0.2082 - val_acc: 1.0000

Epoch 2/5

20/20 [==============================] - 26s 1s/step - loss: 0.1795 - acc: 0.9375 - val_loss: 0.0526 - val_acc: 1.0000

Epoch 3/5

20/20 [==============================] - 26s 1s/step - loss: 0.0576 - acc: 0.9850 - val_loss: 0.0056 - val_acc: 1.0000

Epoch 4/5

20/20 [==============================] - 27s 1s/step - loss: 0.0108 - acc: 1.0000 - val_loss: 0.0011 - val_acc: 1.0000

Epoch 5/5

20/20 [==============================] - 26s 1s/step - loss: 0.0011 - acc: 1.0000 - val_loss: 3.6085e-04 - val_acc: 1.0000

test acc: 1.0
```

Figures 1.1 ,1.2 System CNN-Conv2D Model parameters and training statistics



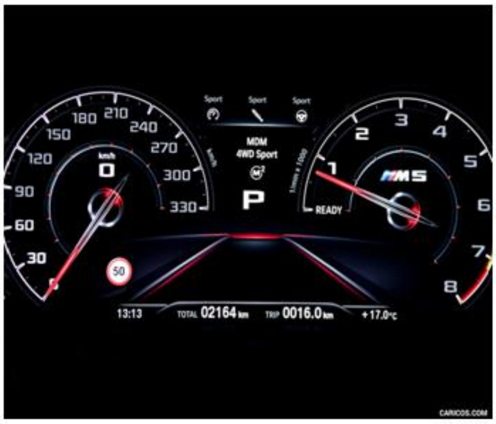


Figure 1.3 Data plotted using results from Figure 1.1 and 1.2.
Figure 1.4 representing an image of car cluster, Such images were used as starting point for YOLO(Image obtained from Reference [11])

During cluster visual validation the simulation of signals for various ECU's is accomplished using some Can-based tool. This tool stimulates RX/TX messages from other subsystems. Not only must messages be validated but also their effect must be studied. This process might involve a few indicators to turn on, speed to change in the speedometer, or some warning being displayed. Each of these is divided into components for regression testing. YOLO is used to divide these components into four major regions of interest. Depending upon the component either positional value approximation or further processing is done. These regions of interest are further passed through convolution [2] models to check for event accuracy. This helps satisfy all the principles of automated testing-helps reduce time, improves efficiency, saves money, and meets deadlines with ease.

The first section discusses the work being done in the respective fields, second describes the algorithm and how it works. The third section describes the dataset creation and accuracy of the model being used. The fourth section discusses the results obtained and explores the challenges and future work.

## 2. Previous Work

The image grabber [2] is mentioned as a tool for validation testing for the Instrument panel cluster. The paper discusses in detail the method of capturing screenshots. It also focuses on the process of saving the screenshots using a frontend. This paper focuses on the validation of screens using video streaming and saving frames. YOLOv is applied over frames of video footage for continuous testing. Another work [3] have used frame by frame comparison, using MSE (Mean Squared Error) and SSIM

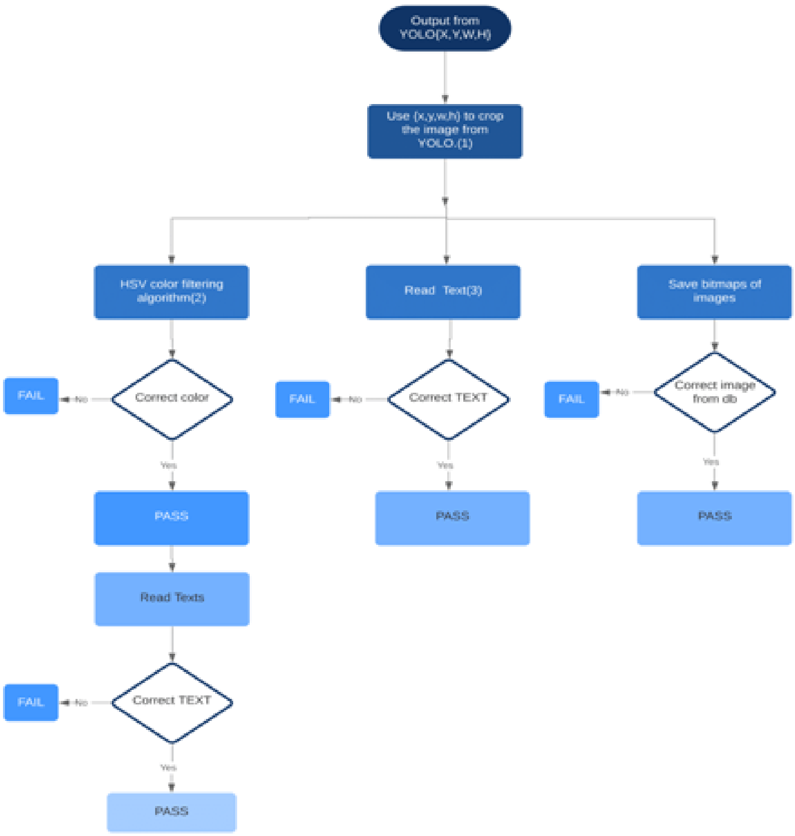


Figure 2.1 Representing the flowchart or system flow

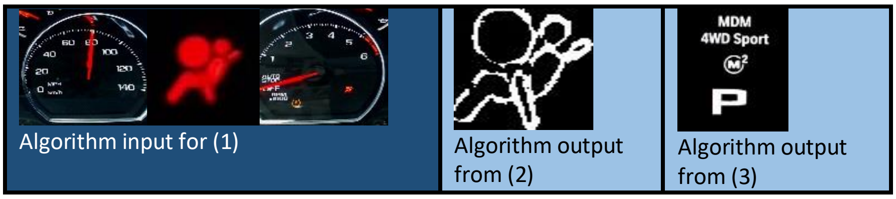


Figure 2.2 Results from algorithm before decision stage

the parameters to extract the similarities between two images. The reference image is compared to the live-stream image, if it fails and exceeds the required MSE, the system will output the test case as failed. Yet, the most seeming drawback with this method is if the brightness or the contrast of room changes the MSE can vary drastically as it is based upon pixel values. Visteon has come close with [4] SIFT in segregating and identifying its text-based regions from images. Before applying YOLO several methods experimented like SIFT or M-SURF, but it resulted in using extreme amounts of computation power. That is why it was discarded. These algorithms can provide good outputs with a high-powered GPU. Most research papers talk about automating the manual signals to automated signals. The major player

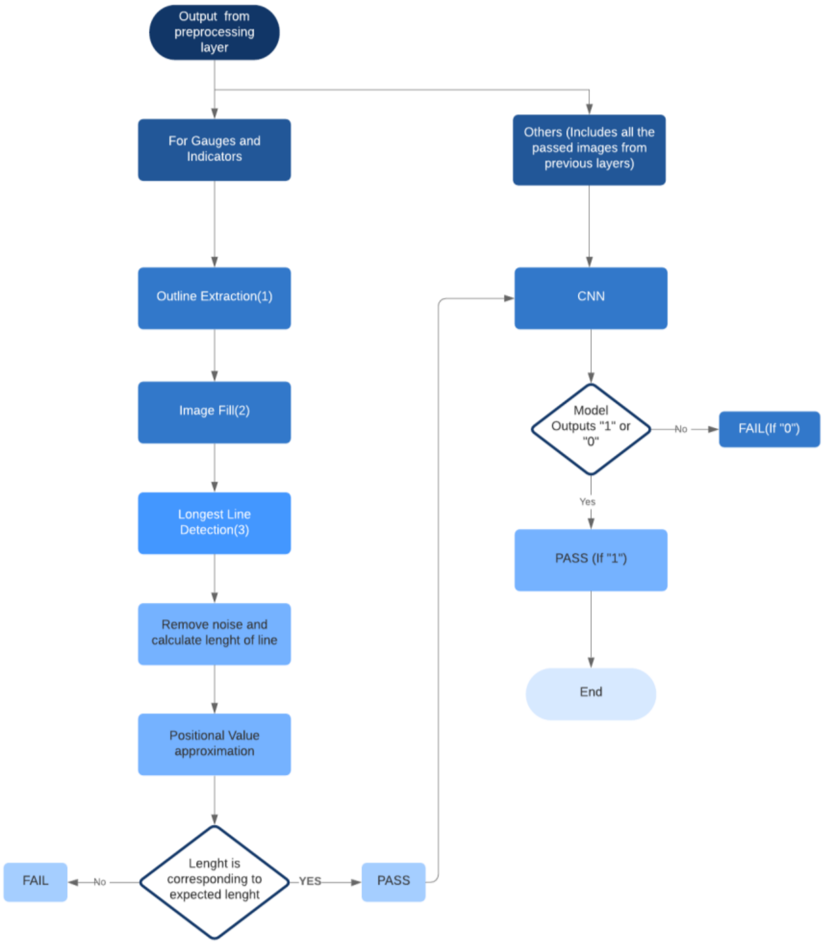

Output from preprocessing layer
For Gauges and Indicators
Others (Includes all the passed images from previous layers)
Outline Extraction(1)
CNN
Image Fill(2)
Model Outputs "1" or "0"
No
FAIL(If "0")
Yes
Longest Line Detection(3)
PASS (If "1")
Remove noise and calculate lenght of line
End
Positional Value approximation
FAIL
No
Lenght is corresponding to expected lenght
YES
PASS

Figure 3.1 Flowchart for Image Processing layer

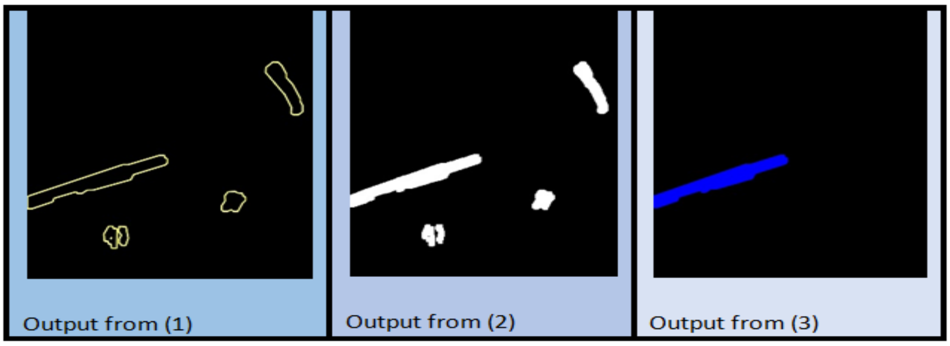


Figure 3.2 Outputs from the image processing layer

in the field is selenium [5], using selenium one can automate the test cases and test for each component. It provides extensive and rich user Interfaces and comfortable experience for testers. But most Fortune 500 companies already have automated test cases, only problem persistence is visual validation and while most companies try to patent their method and have unique ways, challenges remain with how a few algorithms are guaranteed to give results generically.

# 3. A PROPOSED SOLUTION

## 3.1. System Model

The Automation testing can be divided into three major divisions which can be used as three threads interacting with each other based upon inputs and outputs. 1. The Image processing component: This component involves cropping the image, finding contours, extracting bounding boxes, the longest line, or circles from an image. In the method explained OpenCV is used to contour out the regions of interest. These regions of interest are then further used for processing. YOLO is used to extract a box of the desired size from Figure 1.4. Masking techniques are used to get the contour of the needle of the speedometer [Figure 3.1]. As one can see from Figure 2.2 that the needle is red so after masking out red color using HSV [6] filter a mask is applied to extract the image [ Figure 3.2]. Similarly, in a generic car-cluster, there are circular regions to extract, open cv libraries are used to extract Hough circles from image along with Hough lines. This component varies drastically from cluster to cluster. 2. The CNN-Conv2D component. The cluster contains some constant indicator images, telltales, and warnings. These can be verified by using a convolution neural network model either specific to the image or a classification model to indicate the category. CNN-Conv2D comes at the bottom of the architecture because the final prediction of the image or final validation is completed by CNN-Conv2D. Results are interpreted as one or zero. 3. The YOLO [7] component is used to separate the regions in the cluster, this, in turn, helps provide inputs for CNN-Conv2D models.

## 3.2. Algorithm

### 3.2.1. CNN-Conv2D Algorithm

As Referenced in [1] we define a convolution process by assuming a 4-D kernel tensor K with element $K_{i,j,k,l}$ giving the connection strength between a unit in channel i of the output and a unit in channel j of the input, with an offset of k rows and l columns between the output unit and the input unit. Assume the input consists of observed data V with element $V_{i,j,k}$ giving the value of the input unit within channel i at row j and column k. Assume the output consists of Z with the same format as V. If Z is produced by convolving K across V without flipping K , then

$$Z_{i,j,k} = \sum_{l,m,n} V_{l,j+m-1,k+n-1} K_{i,l,m,n} \qquad \ldots(1)$$

$$Z_{i,j,k} = c(K,V,s)_{i,j,k} = \sum_{l,m,n} V_{l,(j-1)\times s+m,(k-1)\times s+n} K_{i,l,m,n} \qquad \ldots(2)$$

Equation (1) and equation (2) help explain the Convolution process as referred to in [1]. This paper does not explore the mathematical backend of the CNN-Conv2D [8] algorithm, it considers the implementation of CNN-Conv2D in detail. Once images from the cluster are obtained either through camera or any other device which requires various CNN-Conv2D models in the background to explain the validity of each image. For this .h5 is used and .json file is used with weights stored from the trained model.

An alternate method explored is to classify between images and map them according to the triggered event. Though it required extreme amounts of computation power. Hence, switching back to event-triggered validation, the algorithm checked for incorrect images on a one-by-one basis. It proved successful in saving time and generating results with lesser use of resources.

Once an event is triggered a screenshot is captured, of the cluster. According to the respective categories images go through various layers of the algorithm and generate results. On the other hand, training models for each of the event was laborious until image processing was used. By using image processing a dataset of about 200*1000 images was generated. Each model used 4 convolution layers and 2 fully connected layers. As seen in Figures 1.1 and 1.2, the model is trained in Anaconda IDE Spyder where the console is displaying the Convolution layers, accuracy, and loss values. Figure 1.3 is the graphical representation of accuracy and loss values through progressive epochs. Figure 1.4 is the input image for the YOLO layer which again uses the CNN algorithm.

CNN-Conv2D has been trained on the images of the correct text-based alert or graphics-based alert which included 100 images for each alert. These images were collected from the live-stream under various light conditions. This was done to prevent the failure of the test cases. The models are trained to be robust and adaptive to any light conditions. Be it day or night the validation process could be carried out at any time.

### 3.2.2. Image processing component

Image processing is used extensively in computer vision to aid the machine learning algorithms. Figure 2.2 is the input for the image processing layer, but only the images having a needle are used in this layer. They not only help in pre-processing but can help extensively in making the dataset more diverse. Various algorithms discussed are extracting the longest line and the largest circle by using Hough Transform [9].

$$r = xcos\theta + ysin\theta \qquad ...(3)$$

Using Hesse’s transform [9], the longest line is obtained in an image where computations are done for pixel values. When the threshold is applied over the values obtained from transform, segregation of different circles and lines helped obtain the required circle.

Similarly, other algorithms can be applying to extract the image. For example: - if the speed of the vehicle is to be calculated from the scale in the speedometer, after extracting the color-filled longest line one can find the speed by calculating $\theta$ and $r$ . Therefore, obtaining the result from the equation below:

$$l = r * \theta \qquad ... (4)$$

Another example -Optical Character Recognition [10] can be deployed for checking the correctness of each alphabet or number for each event.

### 3.2.3. YOLO component

The YOLO [7] is a CNN-Conv2D network that learns to identify objects. Primarily used for object detection, it’s capabilities can be used to obtain Regions of Interest (ROI) from the frame. Each ROI acts as an object for the algorithm. A grid of 9*9 for a 960*800 image is used. YOLO was trained to identify the speedometer, gauges, middle region for warnings, and upper region for indicators. This model can be modifying to generate other output. The vector values for each grid object must be updated to obtain variable results.

## 4. SIGNAL-FLOW

Input frames from the video stream are fed as individual images (Figure 4.1) to the YOLO network. YOLO outputs are a set of ROI's (Figure 4.2) which are fed to the image processing layer but only in conditions where further processing is required like:

- To find the angle of gauge needle (Figure 4.2 Right ROI)
- To find the angle of the speedometer needle (Figure 4.2 Left ROI)

From these angles, the RPM and the speed are calculated, as the radius and length are constant.

- In the case of warnings /text-based alerts validation (Figure 4.2 Center ROI), OCR is used.
- For indicators HSV-based-color-filtering as referred to in [6], open CV models are used to validate the color, but for shapes, CNN-Conv2D is used.

All the components of validation are covered once the above categories are completed.

### 4.1. Training of YOLO network

It is accomplished by feeding the network with 500 images of each of the ROI's. A dataset this large can be cumbersome to build. Hence OpenCV libraries were used to increase contrast, decrease brightness, and include noise to obtain a diverse dataset. During testing of the model unseen images were fed to the network to distinguish the regions. It was only trained initially to output the regions, later updated to output the width, height, x, y-axis of the bounding box for the ROI.

Dimensions obtained were then utilized to crop the regions and fed to the subsequent layers. The complete process is depicted using a flowchart in Figure 2.1. After the images have been fed to CNN Layer the result ("0" for PASS and "1" for FAIL) will conclude if it is "PASS TESTCASE" or "FAILED TESTCASE" which in turn will conclude the process.

## 5. RESULTS

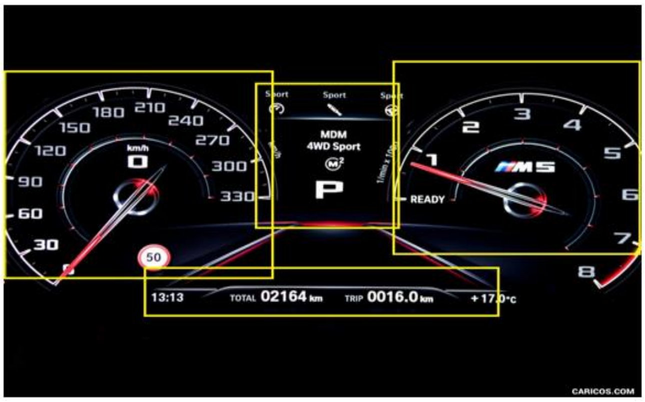


Figure 4.1 Bounding boxes from YOLO, Obtained from Reference [9]

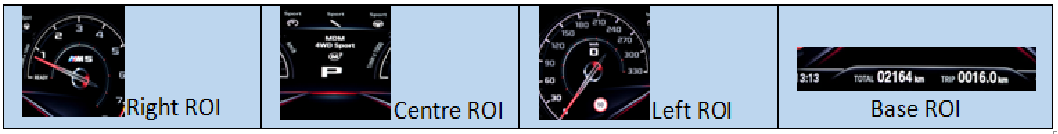


Figure 4.2 ROI's from YOLO/ Output from YOLO

While taking into account the accuracy obtained from the model whereas it was taking 16-17 minutes through manual testing. So, the whole validation process was completed within 5 hours. The entire cost was reduced by hundreds of dollars and efficiency was increased. The tool thus developed was able to generate output at 98% (average from all the models) accuracy overall. The dataset has been tested on the frames of the video stream for car-cluster with different speedometer and gauge values. The 2% where the tool failed was in areas like Optical Character recognition. OCR was not able to identify a 4WD and gave output as AWD. Another challenge was colors; hence a fixed brightness and contrast was set from the camera to detect without errors. The report was generated automatically from the outputs and it completed the validation process. Throughout the process Logitech webcam was used as a camera and a Windows/Linux PC with core Xeon without any external GPU has been used. The solution checked all the parameters of automated testing reduced time, efforts, saved cost, and improved efficiency (Qualitatively and Quantitatively). The various process has claimed to obtain good results, but the solution is validation situation oriented. The methods which can work for one cannot be suitable for the other as the conditions differ drastically with processing power, lighting conditions, and use of various interfaces. Sometimes complete autonomous automation is not the goal, although in cases it is required the paper extensively helps in building solutions with remarkable accuracies.

## 6. STIMULATION

The method is currently deployed in the test environment to test its reproducibility. During its thousand's run, it has not produced any false positives which are crucial to production. The 2% corresponds to the false negatives which can be rechecked if needed by the test engineer. The low brightness and high contrast are being fixed for the test to aid OCR.

Other competing solutions are usually not in-house, the cost, development, as well as maintenance, is expensive. This process gives more power to the user and reduces dependency on third-party software. Moreover, with open-source libraries available it becomes easy to deploy it over a variety of systems under validation.

## 7. CONCLUSIONS

The methods discussed if executed sequentially will help in converting semi-Autonomous testing to fully autonomous testing. Semi-autonomous testing involves automating all signals and messages while the decision is taken by a human. This task is not required in fully autonomous testing where the decision is taken by the program. There can be ways to optimize the flow of algorithms and improve efficiency to 100%, which may require training an in-house OCR. Moreover, process improvement can be achieved by optimizing algorithms and image processing over the video frame.

## ACKNOWLEDGMENTS

I would like to acknowledge and thank the python community in GitHub and other forums. Also, the open-source community of anaconda IDE without whom the paper would not be successful.

## REFERENCES

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016,chapter-9,p. 332.

[2] Nimara, Sergiu & Popa, Dorina & Bogdan, Razvan. (2017). Automotive instrument cluster screen content validation. 1-4. 10.1109/TELFOR.2017.8249425.

[3] Raj, Mohan & Kumar, Sathiesh. (2017). Vision based feature diagnosis for automobile instrument cluster using machine learning. 1-5. 10.1109/ICSCN.2017.8085671.

[4] Deepan Raj M, Prabu A. (2019) [PDF],Available:https://www.visteon.com/wp-content/uploads/2019/01/multilingual-string-verification-for-automotive-instrument-cluster-using-artificial-intelligence.pdf [Accessed:4-May-2020]

[5] Selenium Automation Testing[Online],Available: https://www.guru99.com/introduction-to-selenium.html[Accessed:4-May-2020]

[6] Harrison,Color Filtering OpenCV Python Tutorial[Online],Available:https://pythonprogramming.net/color-filter-python-opencv-tutorial/ [Accessed:5-May-2020]

[7] G. Nishad, You Only Look Once(YOLO): Implementing YOLO in less than 30 lines of Python Code [Online]Available:https://towardsdatascience.com/you-only-look-once-yolo-implementing-yolo-in-less-than-30-lines-of-python-code-97fb9835bfd2[Accessed: 1-May-2020]

[8] E.Allibhai, Building a Convolutional Neural Network (CNN-Conv2D) in Keras[Online].Available:https://towardsdatascience.com/building-a-convolutional-neural-network-CNN-in-keras-329fbbadc5f5[Accessed: 1-May-2020]

[9] OpenCV Documentation, Hesse's Transform[Online],Available: https://docs.opencv.org/2.4/doc/tutorials/imgproc/imgtrans/hough_circle/hough_circle.html[Accessed :4-May-2020]

[10] Optical Character Recognition(OCR) Python documentation, [Online],Available:https://pypi.org/project/pytesseract/[Accessed: 5-May-2020]

[11] Caricos. (2019). 2019 BMW M5 COMPETITION [Online image]. Retrieved October 2, 2019, from https://www.caricos.com/cars/b/bmw/2019_bmw_m5_competition/images/140.html

## AUTHOR

I am currently working as a backend engineer at SAP Labs India, mostly I work on application development, but deep learning is one of my interests. This is one of the papers regarding my previous work. I hope you enjoy it.

# A Study on the Minimum Requirements for the On-Line, Efficient and Robust Validation of Neutron Detector Operation and Monitoring of Neutron Noise Signals using Harmony Theory Networks[1]

Tatiana Tambouratzis[a], Laurent Pantera[b] and Petr Stulik[c]

[a]Department of Industrial Management & Technology,
University of Piraeus, Piraeus 185 34, Greece
[b]Laboratoire des Programmes Expérimentaux et des Essais en Sûreté,
CEA/DES/IRESNE/DER/SPESI/LP2E/, Cadarache,
F-13108 Saint-Paul-Lez-Durance, France
[c]Diagnostics and Radiation Safety Department, ÚJV Řeža.s.,
Hlavní 130, Řež, 250 68 Husinec,Czech Republic

## *Abstract*

*On-line monitoring (OLM) of nuclear reactors (NRs) incorporates – among other priorities – the concurrent verification of (i) valid operation of the NR neutron detectors (NDs) and (ii) soundness of the captured neutron noise (NN) signals (NSs) per se. In this piece of research, efficient, timely, directly reconfigurable and non-invasive OLM is implemented for providing swift – yet precise – decisions upon the (i) identities of malfunctioning NDs and(ii) locations of NR instability/unexpected operation. The use of Harmony Theory Networks (HTNs)is put forward to this end, with the results demonstrating the ability of these constraint-satisfaction artificial neural networks (ANNs) to identify(a) the smallest possible set of NDs which, configured into (b) the minimum number of 3-tuples of NDs operating on(c) the shortest NS time-window possible, instigate maximally efficient and accurate OLM. A proof-of-concept demonstration on the set of eight ex-core NDs and corresponding NSs of a simulated Pressurized Water nuclear Reactor (PWR) exhibits(i) significantly higher efficiency, at(ii) no detriment to localization accuracy, when employing only (iii) half of the original NDs and corresponding NSs, which are configured in (iv) a total of only two (out of the 56 combinatorially possible)3-tuples of NDs. Follow-up research shall investigate the scalability of the proposed methodology on the more extensive and homogeneous (i.e. "harder" in terms of ND/NS cardinality as well as of ranking/selection) dataset of the 36 in-core NSs of the same simulated NR.*

## *Keywords*

*Nuclear Reactor (NR),On-Line Monitoring (OLM), Neutron Noise (NN), Neutron Noise Signal (NS), Neutron Detector (ND), Computational Intelligence (CI),Artificial Neural Network (ANN),Harmony Theory Network (HTN), 3-tuple of NDs/NSs*

[1] *This piece of research is dedicated to the COVID-19 victims worldwide.*

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 141-154, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101113

## *ACKNOWLEDGEMENTS*

*The research leading to these results (i) has received funding from the H2020 CORTEX Euratom Research and Training programme 2014-2018 under grant agreement No 754316, (ii) is supported by the Research Office of the University of Piraeus.*

## *LIST OF ACRONYMS*

| | |
|---|---|
| artificial neural network | ANN |
| computational intelligence | CI |
| cross-correlation | CC |
| ex-core/out-of-core | ex- |
| harmony theory network | HTN |
| in-core | in- |
| neutron detector | ND |
| neutron noise | NN |
| neutron (noise) signal | NS |
| nuclear reactor | NR |
| on-line monitoring | OLM |
| pressurized-water NR | PWR |
| principal component analysis | PCA |
| simulated annealing | SA |
| temperature | T |

# 1. INTRODUCTION

## 1.1. On-Line Monitoring of Nuclear Reactors

Ever since the early installation and operation of nuclear reactors (NRs), on-line monitoring (OLM) has received special attention from nuclear engineering scientists, researchers as well as NR operators [1]. To date, OML covers the entire spectrum of safety

- from “operational”, concerning the sustainable, controllable and maximally efficient chain-reaction with thermal neutrons from – and on – fissile material [2],
- to “radiation protection”, concerning safety of the NR personnel at the local level, as well as ecological wellbeing of the flora, fauna and environment at the global level [3].

OLM of NRs encompasses the prompt processing and analysis of neutron noise (NN) signals(NSs) – as these are captured by NN detectors (NDs) –for ensuring the timely, non-invasive, consistent, reliable and (ideally) directly reconfigurable identification (as well as resolution) of various NR problems, including

(a) failing equipment (e.g. instrumentation, sensors, transmitters, NDs) and
(b) deviating-from-normal and/or inconsistent operation (e.g. aberrant coolant flow and/or temperature measurements).

Essential information on instrument calibration and verification, as well as on instrumentation/equipment/plant condition monitoring, can be found in [4-9]. Additionally, two relevant – complementary to one another, yet each comprehensive in its own focus and domain of interest – reviews of the literature on human/operational and computational intelligence (CI)-based OLM appear in [10] and [11], respectively.

### 1.2. Organization of Presentation

The remainder of this piece of research is organized as follows:

- Section 2 introduces the state-of-the-art relating to NR/OLM. A pertinent selection of innovative OLM approaches is presented, which (i)epitomizes the extensive range of preferred methodologies for tackling the multitude of OLM issues that may arise during NR operation and (ii) underscores the interplay between ND-/NS-derived information and custom-made OLM decisions/derivations/solutions. Subsequently, the motivation for the proposed methodology, as well as the advancement offered by the implemented encoding of the problem variables and constraints, is introduced in this Section.

- The NR set-up and characteristics of/dependencies between the out-of-coreNDs/NSs (ex-NDs/ex-NSs) of the dataset of [12] (which has been used for demonstrating the proposed OML approach) are detailed in Section 3.

- Section 4 provides a comprehensive introduction to the HarmonyTheory Network (HTN) [13] and its custom-made implementation for the concurrent selection of the minimal number of a) ex-NDs per se and b) 3-tuples [14] of the ex-NDs of a), which - combined with c) the shortest possible (of length 256) sliding time-window of the corresponding NSs – implement consistent, global, prompt and precise, non-invasive and directly reconfigurable OLM. Problem decomposition and gradual upscaling is implemented for endowing NR operation with computational (space- and time-) efficiency, at no compromise to the optimality of the returned ex-ND configuration/solution.The advantages of the proposed approach are further supported by a critical presentation of the obtained results.

- Section 5concludes the presentation by summarizing the main characteristics and innovation of the proposed approach, reporting on the importance of the findings, drawing key-conclusions and stating future extensions to the presented research.

## 2. NUCLEAR REACTORS & ON-LINE MONITORING

### 2.1 State-of-the-Art

The extensive range of NR operation-related prerequisites and restrictions/controls/ constraints has resulted in (I) a multitude of nuclear protection guidelines, initiatives and standards,(II) primary information and reviews of the state-of-the-art on multi-sensor coordination (e.g. [15-16]) as well as on modelling, estimation and control (e.g. [17]). Over the last decade, both research and development have delved into the implementational characteristics and properties that are necessary for rendering OLM consistently correct, maximally efficient, robust to missing – yet sensitive to erroneous– information, as well as capable of swift reconfiguration whenever deemed necessary.

A selective – yet representative – collection of pieces of research which employ, validate and co-ordinate sets of collaborating NDs (via the corresponding NSs) is provided next, with each implementation accompanied by a brief exposition of problem statement, execution and novelty:

- Principal component analysis (PCA) has been employed in [18] for mathematically modeling the relationships that hold between topologically related sets of self-powered NDs, culminating into an operational “detection & isolation scheme” for four types of simulated faults (bias, drifting, precision degradation and complete failure).

- A hybrid scheme, combining (i) Kalman filtering for estimating prompt neutron flux variations and(ii) the generalized likelihood ratio for detecting and diagnosing ND faults, has been tested successfully in [19], demonstrating robust on-line correction of the step change on simulated neutron flux data concurrently implementing moving control-rods and fluctuating power demands.

- Three (the observation, dependency and state) sub-models of a sensor model have been implemented in [20] for effectively co-ordinating, as well as integrating, competitive and/or disparate pieces of NS-derived information which include uncertainty in the ND observations.

- The combination of recurrent PCA and k-means clustering of NDs has been put forward in [21] for the consistent detection and severity evaluation of failing NDs.

- A CI/fuzzy-logic-based decentralized multi-sensor detection system with reduced energy demands [22]has been found successful in attaining a superior level of detection accuracy.

## 2.2 Problem Statement & Aims - Motivation for the Implemented Research – Problem Representation - Proposed Advancement in the State-of-the-Art

The validation of (a) correct operation of the in- and ex-core NDs (in-NDs and ex-NDs, respectively) [1] and (b) soundness of the captured NSs (in-NSs and ex-NSs, respectively) per se, constitutes a prerequisite of successful OLM which is based on the agreement of the measured NSs with expected reference values[2], salient characteristics of signal evolution and inter-signal comparisons. Further to rendering OLM completely automated as well as autonomous, an additional major motivation is the advancement of the state-of-the-art by also maximizing the time- and space-efficiency of OML, an endeavour that is instigated in this piece of research via the selection and subsequent utilization of

(i) the minimal set of NDs and
(ii) the minimum number of 3-tuples[3] of "collaborating" NDs [14] (derived from the minimal set of(i)) which is required for consistently implementing OLM over (I) the entire NR, (II) the full spectrum of possible NR operating modes (e.g. footnote 2), (III) the extensive variety of coolant flow-regimes(e.g. bubbly, churn), and
(iii) the shortest sliding time-window that is capable of on-line (real-time) –yet consistent – capture of the time-evolution of the NSs which pertain to the selected 3-tuples of NDs of (ii).

The concomitant satisfaction of (i)-(iii) exposes the minimal set(s)[4] of NDs (and corresponding NSs) which is/are necessary – as well as sufficient – for accomplishing accurate, efficient as well as swift, non-invasive and directly reconfigurable NS-derived information processing and custom-made problem resolution over the entire NR and the full spectrum of possible NR operating/coolant flow-regime conditions (as described in point (ii) of this Section), with the optimum 3-tuple configurations of NDs/NSs being based on the current information acquired

[2] the ranges and trajectories of the signals are fully determined by NR construction (the NR transfer function) for each mode of operation (start-up, shut-down, stand-by, transient vs. steady-state etc.) as well as coolant flow-regime conditions

[3] it has been shown in [12] that 3-tuples of appropriately selected in-NDs are necessary as well as sufficient for the concurrent detection of erroneous in-NSs and/or malfunctioning in-NDs of the 3-tuple

[4] the NDs/NSs of these sets may well be distinct over different modes of NR operation as well as location of instability, in order to better capture the phenomena under development (also see footnote 2)

directly, as well as exclusively, from the NSs in the form of CCs, on-line observed NS deviations, inter-signal comparisons etc.

# 3. Data Characteristics and Problem Encoding

## 3.1. Dataset Description

The data used for demonstrating the proposed approach constitutes the complete set of eight ex-NSs which have been collected by the corresponding ex-NDs of the pressurized-water NR (PWR) described in [12]. The special interest in ex-NSs – rather than in the 36 in-NSs of the same dataset which has been used for demonstrating the 3-tuple configuration of [14]–is based on the following criteria:

(a) on the one hand, the relatively small number of ex-NDs allows the implementation/evaluation of a proof-of-concept study concerning the application of the proposed approach to the entire set of ex-NDs of [12], which can then be transferred to other, more extensive, sets of in- as well as ex-NDs;

(b) on the other hand, the lack of the high frequency component[5] between/across ex-NDs renders OLM significantly more challenging in terms of timeliness and validity of response, especially when compared to in-NDs during rapidly evolving phenomena [3];

(c) as a result of (b),the cross-correlation (CC) coefficients between ex-NSs (shown in Table 1) are significantly lower and more varied than those between the in-NSs of the same dataset, thus placing further demands as far as (i) the concurrent satisfaction of the combination of pertinent sources of NS information and (ii) the processing of considerably more requirements, are concerned;

(d) it is important to determine whether, how and with what gain in terms of computational complexity, the 3-tuple configuration can be applied – successfully as well as confidently – to the full set of ex-NDs and ex-NSs, especially given the comparatively low CC-values between ex-NSs.

**Table 1**. The CC coefficients between the set of eight ex-NS of [12], revealing three clusters of “sufficiently” correlated – i.e. collaborating – NDs and demonstrating two overlapping parts between pairs of neighbouring clusters.

| ex-NS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.0000 | 0.9974 | 0.9345 | 0.9353 | 0.5966 | 0.5926 | 0.5756 | 0.5654 |
| 2 | 0.9974 | 1.0000 | 0.9530 | 0.9553 | 0.6435 | 0.6418 | 0.6233 | 0.6157 |
| 3 | 0.9345 | 0.9530 | 1.0000 | 0.9989 | 0.8179 | 0.8132 | 0.7994 | 0.7886 |
| 4 | 0.9353 | 0.9553 | 0.9989 | 1.0000 | 0.8174 | 0.8158 | 0.7992 | 0.7916 |
| 5 | 0.5966 | 0.6435 | 0.8179 | 0.8174 | 1.0000 | 0.9973 | 0.9993 | 0.9956 |
| 6 | 0.5926 | 0.6418 | 0.8132 | 0.8158 | 0.9973 | 1.0000 | 0.9969 | 0.9988 |
| 7 | 0.5756 | 0.6233 | 0.7994 | 0.7992 | 0.9993 | 0.9969 | 1.0000 | 0.9970 |
| 8 | 0.5654 | 0.6157 | 0.7886 | 0.7916 | 0.9956 | 0.9988 | 0.9970 | 1.0000 |

[5] “… the high frequency component of the neutron noise allows [facilitates] the detection of phenomena in the near vicinity of an in-core detector” [3]

### 3.2. Dataset Analysis – Clustering-Derived Assumptions and Problem-Dependent Constraints

As can be observed in Table 1, the application of a “threshold” of 0.8 to the CC values between all the pairs of the eight ex-NSs reveals three pair-wise overlapping clusters of “sufficiently” correlated ex-NSs, with NS_cluster1, NS_cluster2 and NS_cluster3 comprising ex-NS1 to ex-NS4, ex-NS3 to ex-NS6 and ex- NS5 to ex-NS8 (coloured red, blue and green), respectively, as well as two overlapping areas (coloured darker shades of red and green, respectively).

In theory and, since all the pairs of ex-NSs belonging to the same cluster demonstrate similar behaviour, any such pair could be used for the consistent prediction of every other ex-NS of the same cluster. However, the criterion of “sufficiently high” CC values is not adequate per se, as explained in the following remark and accompanying example which demonstrates the importance of implementing ex-NS/ex-ND selection beyond values and thresholds alone, thereby further substantiating the rationale behind/necessity for the 3-tuple methodology:

- As can be derived from Table 1, the selection of ex-NS2 is “preferable”[6] to that of ex-NS1 as the CC values of ex-NS2 with the remaining ex-NSs (ex-NS3 through to ex-NS8) are higher than those of ex-NS1, with the same observation holding for ex-NS4 over ex-NS3, ex-NS5 over ex-NS6 and ex-NS7 over ex-NS8.
- Even so, such an (purely CC-value based) selection of ex-ND2, ex-ND4, ex-ND5 and ex-ND7[7] does not allow the implementation of valid ex-NS/ex-ND monitoring, as not all the ex-NDs of each of these 3-tuples belong to the same cluster, resulting innone of the {ex-NS2, ex NS4, ex-NS5} or {ex-NS4, ex-NS5, ex-NS7} 3-tuples satisfying the pairwise “CC>0.8” requirement (Table 1).

### 3.3. Problem Representation/Encoding

According to the aforementioned criteria and requirements of sufficient CCs between ex-NSs for the purposes of OLM, only these 3-tuples of ex-NSs which demonstrate values exceeding 0.8 for all (three) of their pairwise-derived CCs are considered operational and – thus – only these are encoded in the present implementation.

Optimization is accomplished in a divide-and-conquer fashion, expressed as the joint utilization of the combination of the minimum possible:

- number of 3-tuples of “collaborating” ex-NDs in the manner demonstrated in [14],
- number of sufficiently (and, ideally, as highly as possible) correlated ex-NSs per se, which – configured into 3-tuples of NDs/NSs – are capable of monitoring the entire area of interest and
- time-step of 1 which, combined with a statistically adequate time-window (of 256 for the present dataset), guarantees the on-line, efficient and robust concurrent identification of ex-ND- (as well as of ex-NN/ex-NS-) related NR faults.

Problem decomposition and gradual upscaling is implemented for endowing NR operation with computational (both space- and time-) efficiency, at no compromise to the optimality of the returned ex-ND configuration/solution.

[6] in the sense that it is more highly correlated with the proximal ex-NSs and, thus, more appropriate for performing ex-NS/ex-ND validity checks at the vicinity of the corresponding 3-tuples

[7] as the ex-NDs corresponding to the ex-NSs with the highest correlations with the other NSs

## 4. Harmony Theory Network Construction/Problem Encoding for the Implementation of On-Line Monitoring

### 4.1. Harmony Theory Networks

The HTN [13] constitutes a semantically constructed[8] two-layer artificial neural network (ANN) architecture which is adept at optimising under constraints (e.g. [23-24]). Instead of training, the problem is mapped directly to the nodes and connections of the HTN during construction in such a manner that the (semantic, problem-specific) collective compatibility of the activation values between connected nodes of the two layers quantitatively expresses the degree of "harmony"/fitness[9] of the HTN state, as this is calculated based on the satisfaction of the constraints dictated by the problem/data per se and encoded in the HTN connections between the two HTN layers. The main characteristics of HTN construction and operation for OLM are briefly defined here:

- The lower HTN layer comprises the "representational feature" nodes (RFs), which encode the problem-related information at the desired/appropriate level of description/ encoding. In the present case, each RF stands for a 3-tuple of NDs which collects information (a) from the corresponding 3-tuple of NSs, as well as (b) from every other "sufficiently" (>0.8) correlated[10] NS of the same cluster(s) as the given NSs of the 3-tuple, in order to determine the expected/anticipated values and trajectories of as many NSs as possible and, consequently, to be able to further provide reliable decisions upon normal operation of the corresponding NDs of the 3-tuple per se. Each RF can acquire one of two states, namely +1 (active) if the encoded 3-tuple of NDs reflects both valid ND operation and agreement in the evolution of the corresponding 3-tuple of NSs, or -1 (inactive) if either (or both) of valid ND operation and agreement in the evolution of the 3-tuple of encoded NSs cannot be established.

- The upper HTN layer comprises the knowledge atoms (KAs), with each KA encoding a NS of the present dataset and acquiring one of two states, +1 (active) if the NS represented by the given KA can be consistently monitored by at least one 3-tuple of NDs encoded ina connected as well as currently active RF of the lower layer, and 0 if monitoring of the given NS cannot be established. Following HTN settling/convergence, the set of active KAs represents the maximal set of mutually compatible NSs that can be (i) successfully monitored by at least one 3-tuple of NDs and (ii) validated as far as anticipated/correct NR operation (encompassing both ND functionality and NS monitoring) is concerned.

- The HTN connections are – by construction – bidirectional, symmetric and strictly limited between RFs/NDs and KAs/NSs; additionally, for the problem-at-hand they are exclusively positive, with the weights of all the connections emanating from the same KA being normalized as well as equal to each other (as is customary in HTNs), thereby encoding the reinforcing relationships between connected RF/KA pairs via:

**a)** Direct monitoring of any given NS (encoded in a KA) by each RF which contains (in its 3-tuple of encoded NDs) the ND corresponding to the encoded NS. This NS/ND relationship

[8] in the sense that the nodes, the connections between nodes as well as the connection weights are assigned in a meaningful (expressive of the problem givens and constraints between givens), mathematical (predicate-logic-&normalization-based) manner

[9] i.e. compatibility between the activation values of connected HTN nodes of the two layers (as the HTN architecture does not allow within-layer connections)

[10] in which case, "sufficient" (>0.8) CC with at least two of the NSs of the corresponding 3-tuple of NSs is required

is implemented in the HTN via positive, reinforcing connections between the KAs and the relevant RFs (shown as fine black lines in Fig.1).

**b)** Indirect monitoring of any given NS (encoded in a KA) by each RF which does not contain (in its 3-tuple of encoded NDs) the ND corresponding to the encoded NS, yet contains at least two NDs that belong to the same cluster as the corresponding ND. This NS/ND relationship expresses the capability of the NDs of the 3-tuple to "indirectly" – yet securely – monitor and validate the given NS. Shown in Fig. 1 as bold red lines, these connections represent the transitive, "propagating" relationships between every KA and each connected RF that contains a 3-tuple of NDs that monitors at least two NSs from the same cluster as the given NS.

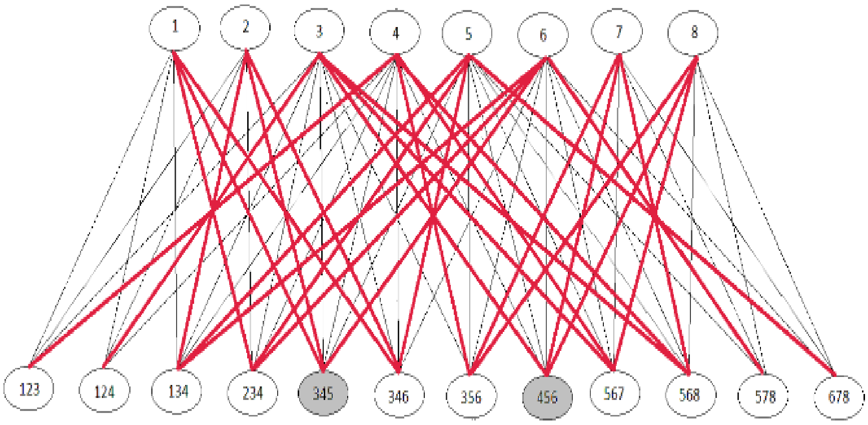


**Fig. 1.** HTN encoding of the (a) 12 sufficiently correlated (>0.8) ND 3-tuples in the RFs of the lower layer, (b) eight captured NSs in the KAs of the upper layer, (c) compatibility between the 3-tuples of sufficiently correlated NSs which can be used for deciding upon expected values and shapes of the NSs

The fine black lines (representing direct monitoring) between the nodes of the two layers of the HTN show three connections for the "exterior" NSs (1, 2, 7 and 8, corresponding to KAs 1, 2, 7 and 8, respectively) and six for the "interior" NSs (NSs 3, 4, 5 and 6, corresponding to KAs 3 though to 6, respectively). On the other hand, the bold red lines (representing the indirect connections) amount to three and four for the "exterior" and "interior" NSs, respectively.

- The <u>HTN fitness function</u> quantitatively expresses the "quality" of each HTN state (set of activation values over the entire sets of RFs and KAs), which - for the problem-at-hand - is implemented at two levels of optimization:

**a)** The KA-based level quantitatively expresses the "quality" of each HTN state in terms of the number of active KAs, per se, thus expressing the degree/level of compatibility between the active nodes of the two layers, and conveying the collective ability of the set of active RFs (3-tuples of NDs) to monitor as well as to verify as many as possible (and, ideally, all) of the NSs;

**b)** The RF-based level: in case of "ties" in (a), i.e. if more HTN states than one exist with the same maximum number of active KAs, the HTN state with the smallest number of active RFs is selected. Furthermore, if the same total number of 3-tuples of NDs is employed by more than one "best" solutions, the total number of the NDs per se (which are encoded in the active RFs of these solutions) are compared and the HTN state representing the smallest possible number of (i) 3-tuples of NDs as well as (ii) NDs per se, is selected; in the highly unlikely case where more than one such solutions co-occur, thresholding based on the lowest CC value between NSs belonging to the tied solutions is implemented for selecting the optimal configuration of NDs and 3-tuples of NDs.

The HTN operation characteristics for the present problem include (i) the elementary advancement (by 1) of the sliding time-window along the data and, consequently, on the time-window employed for the calculation of the CC matrix, thus guaranteeing timely and robust - yet still sensitive - responses to changes within the NR (including changes in the NSs and/or the NDs per se); (ii) the application of simulated annealing (SA) [25] to the HTN, employed at each instance of the sliding time-window as follows. The HTN is initialized with the assignment of randomly assigned ±1 values to the RFs at an inaugural "high" value of the temperature parameter T of the HTN. The propagation of the assigned RF values to the connected KAs and the calculation of the total activations of the KAs is followed by thresholding for extracting the active KAs; this threshold is gradually raised (as T is lowered) during the SA process, such that one of the best – of maximal harmony – HTN state(s) is converged upon[11]. SA is appropriate for this task (and problem representation), as the occasional convergence of the HTN upon a sub-optimum problem-state as the current solution for a given instance of the sliding time-window can be tolerated, as it is smoothed over/corrected by the previous and next HTN decisions. It should be mentioned, nonetheless, that a careful investigation and coordination of the HTN parameters and a conservative scheme of decrementing the T parameter during HTN settling is capable of minimizing premature, as well as imperfect, convergence to a non-optimum solution by initially considering the entire problem space and gradually focusing upon the more – and, eventually most – "promising" sections of the problem space, thus – as a rule – converging upon a state of maximal harmony.

### 4.2. Harmony Theory Network Implementation via Step-Wise Optimization in Terms of Accuracy and Efficiency - Advancements, Advantages and Limitations

The implementation of a serial (incremental) procedure of 3-tuples of NDs is not applicable to the present problem, as each sequentially selected next 3-tuple of NDs – even if "best" per se – far from guarantees the attainment of an optimal final solution. As, however, the implementation of combinatorial optimization does not necessarily scale up well (in terms of time complexity as well as of successful convergence) to problems of practical interest, an alternative problem-decomposition methodology involving a sequence of HTNs (at most as many as the monitored NSs) has been implemented and successfully tested for the problem-at-hand. The implemented HTNs may run either sequentially or in parallel (depending on the aim and/or time/space-computational complexity requirements/potential of the problem encoding), with each (the ith, i=1, 2, …, 8) HTN (I) being set so as to allow for exactly i active RFs (i.e. i=1, 2, …, 8 3-tuples of NDs) for the maximization of H and (II) the SA settling procedure of each HTN, in turn, being automatically adapted according to this added constraint, thus configuring the activation values of the nodes of both layers such that the RFs that are converged upon by the HTN reveal the 3-tuples of NDs which maximize H for the specific value of i. It is also important that such an implementation provides important information on the landscape of the problem space and, thus, on the interconnections and inter-constraints that apply between NSs (as well as between NDs) at different stages of OLM in terms of the numbers and identities of successfully monitored NSs by specific 3-tuples of NDs.

[11] in a probabilistic sense, where the probability of accepting an "inferior" HTN state during operation diminishes along with the drop in T

**Table 2.** A demonstration of HTN fitness of the optimal HTN-derived selection of NDs/NSs (highlighted 345-456 3-tuples) for successful OLM. All the alternative pairs of 3-tuples using the same ex-NDs/NSs (3, 4, 5 and 6) have been also tabulated for demonstrating the variation in HTN fitness that is reached by each such pair of 3-tuples, hence confirming HTN operation and settling to the HTN state of maximum H as well as highlighting the differences in H observed over the different combinations of pairs of 3-tuples.

| pair of CCs ND/NS between 3-tuples of NSs | 345-346 | 345-356 | 346-356 | 345-456 | 346-456 | 356-456 |
|---|---|---|---|---|---|---|
| 34 | 0.9989 (×2) | 0.9989 | 0.9989 | 0.9989 | 0.9989 | - |
| 35 | 0.8179 | 0.8179 (×2) | 0.8179 | 0.8179 | - | 0.8179 |
| 36 | 0.8132 | 0.8132 | 0.8132 (×2) | - | 0.8132 | 0.8132 |
| 45 | 0.8174 | 0.8174 | - | 0.8174 (×2) | 0.8174 | 0.8174 |
| 46 | 0.8158 | - | | 0.8158 | 0.8158 (×2) | 0.8158 |
| 56 | - | 0.9973 | 0.9973 | 0.9973 | 0.9973 | 0.9973 (×2) |
| HTN fitness ϵ[0 1] | **0.7174** | **0.7332** | **0.5336** | **0.7997** | **0.6001** | **0.6160** |
| occurrence (%) | **6.4** | **11.2** | **-** | **81.2** | **-** | **1.2** |
| 250 trials | **16** | **28** | **0** | **203** | **0** | **3** |

The settling procedure of the sequential HTN reported in this piece of research is described next. For the first iteration, the HTN SA-based "settling" procedure is initialized by imposing the activation value of +1 to a single, randomly selected, RF (i.e. a single 3-tuple of NDs of the lower level); the selected RF propagates its activation value to the connected KAs, the activation value of each KA is calculated and thresholded according to the binarization principle of HTN activation for the nodes of both HTN layers, and the entire set of activations is propagated back to the nodes of the lower layer, followed by roulette-wheel [26] selection of a single RF, which is assigned the activation value of 1, and with all the other RFs being assigned activation values of -1 (inactive RFs). The process of limiting the number of active RFs to 1 is repeated according to the HTN settling procedure described in Section 4.1., whereby convergence upon (one of) the 3-tuple(s) of NDs that monitor(s) the most NSs is achieved. If this number equals the total number of NSs, the procedure terminates and the 3-tuple of NDs is returned. Otherwise, this procedure is repeated from scratch, each time

- incrementing (by 1) the number of active 3-tuples of NDs (RFs) that are to be used concurrently for performing the second (or next) HTN iteration and
- retrieving the pair, triplet etc. (for the second, third iteration, etc., respectively) of 3-tuples of NDs of the lower HTN level which maximizes the number of active KAs (i.e. monitored NSs),

up until all the NSs are monitored by the set of selected 3-tuples of NDs, in other words, all the KAs of the HTN are assigned activation values of +1) for the current number – and identity – of 3-tuples corresponding to the active RFs. The transparent construction of the HTN allows monitoring of the settling process as well as direct identification of (a) the 3-tuples of NDs (active RFs) and (b) NSs (active KAs) that optimise the computational (time- as well as space-) efficiency of OLM, where selection is based on the activations of the set of RFs prior to thresholding.

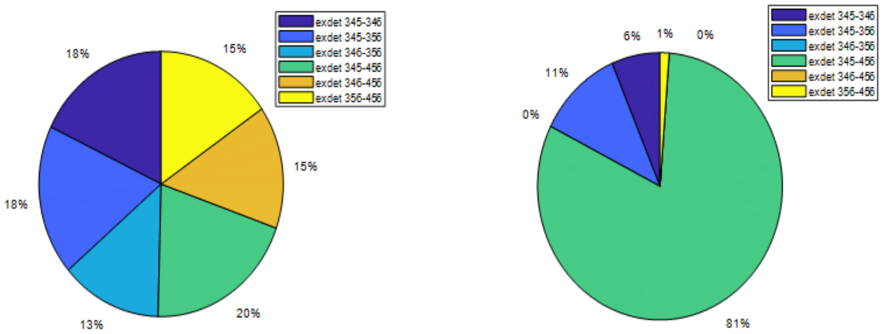


**Fig. 2.** Relative (%) (a) HTN fitness and (b) frequency of HTN settling of the pairs of 3-tuples of NDs/NSs comprising NDs/NSs 3, 4, 5 and 6 over 100 trials (shown in Table 2).

For the present dataset, the implementation and settling procedure of the HTN terminates after two operation/settling iterations: the 3-tuple selected at the first HTN/iteration is identified as{ex-NS3,ex-NS4, ex-NS5}, which maximises the CC values between the selected NSs for time-step 1 and sliding time-window of 256. The second iteration/HTN converges upon the second 3-tuple of NSs, resulting in {ex-NS3, ex-NS4, ex-NS5} and {ex-NS4, ex-NS5, ex-NS6}). This configuration minimizes the number of involved NSs/NDs to 4, while maximizing both the coverage over all eight NSs and the CCvalues within the two 3-tuples.

Table 2 presents a comparison of all the pairs of 3-tuples which contain the same NDs/NSs (3, 4, 5and 6) as the optimal – as well as prevalent (81.2%) – HTN solutions, i.e. are “neighbours” to the optimal k, i.e. in this case the 3-tuple of the first HTN is also included in the optimal configuration of 3-tuples of the second HTN solution. As can be seen, the second-best choice {ex-NS3, ex-NS4, ex-NS5} and {ex-NS3, ex-NS5, ex-NS6} has a slightly lower H (by 6.65%), yet convergence upon this configuration of NDs/NSsis more than seven times less likely to occur than it is for the best configuration, thus further demonstrating the ability of the proposed HTN methodology to magnify– and, thus, distinguish between –small differences in HTN fitness. It is indicative that the identity of the selected NDs/NSs is the same in both configurations, with the superiority of the former pair of 3-tuples amounting to 0.0042x2(normalized difference between the two sets of RF activation values).

In a complementary fashion, Fig. 2 illustrates the % (relative) (a) HTN fitness and (b) frequency of HTN settling, of all the possible configurations of pairs of 3-tuples resulting from NDs/NSs 3, 4, 5 and6, revealing the magnification of H for small differences in CCs of pairs of involved NSs, with the optimal configuration being selected slightly more than 8 out of every 10 trials, and the remaining configurations appearing only occasionally, again in relation to their difference (in terms of H) with the optimal solution.

## 5. Conclusions – Future Directions/Extensions

The feasibility of creating an appropriately selected and fully operational minimal subset of sufficiently correlated ex-NDs employing the minimal time-window of 1 has been demonstrated on the set of eight ex-NDs and ex-NSs of [2]. Maximally efficient (with minimum response time) as well as accurate OLM of NRs is promoted, which is non-invasive and directly reconfigurable based on the characteristics and data-derived (e.g. statistical) properties of the NSs per se, encompassing (I) prompt detection of situations involving faulty NDs and/or NS anomalies, outliers or patterns indicating unexpected/abnormal operation etc. and (II) the accurate characterisation of NR irregularities in/deviations from expected operation. Moreover, the proposed methodology achieves the consistent (III) minimisation of the time- and space-complexity of OLM.

Further to a proof-of-concept, the results obtained from this investigation provide a directly implementable lower bound on the accuracy and consistency of operation of the proposed approach, especially in cases where (i) more NDs per se and/or (ii) more highly correlated NSs are available and/or (iii) longer time windows are implemented for the identification of phenomena of interest which evolve more gradually in time. It is also important that it remains possible – at any time – to (I) enrich the proposed OLM methodology with more NDs/NSs that are “sufficiently” correlated with the NS(s) of interest for expanding upon the initial findings derived from the reduced set of NDs, and/or to (II) exclude NDs/NSs which are redundant or have been detected as erroneous/faulty.

Future research shall focus upon fine-tuning the proposed implementation for such phenomena as transients, transitions between flow regimes and other occurrences during NR operation, as well as on whether the same (or a similar) configuration and implementation can be applied to the significantly larger, yet more highly correlated, set of 36 in-core NDs/NSs of the same dataset of [10], where the significantly higher CCs between in-NSs are compensated by the significantly larger number of (36)NDs/NSs to be simultaneously considered and optimised.

Demonstrating the feasibility of such an endeavour also paves the way for the hardware (H/W) implementation of maximally accurate as well as swift, automated decision-making upon the location(s) of NR instability and malfunctioning ND(s).

The entire computing/programming of the signal processing and HTN simulations has been implemented in the Matlab environment [27].

## References

[1] Eiler J., Glockler O., (2008). On-line monitoring for improving performance of nuclear power plants, Part 1, Instrument channel monitoring. Vienna, International Atomic Energy Agency, IAEANuclear Energy Series, ISSN 1995–7807; no. NP-T-1.1

[2] https://en.wikipedia.org/wiki/Fissile_material (retrieved April 21st, 2020)

[3] https://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/safety-of-nuclear-power-reactors.aspx

[4] https://www.energy.gov/sites/prod/files/2013/10/f4/QSR-RadiationProtection.pdf

[5] Hashemian, H.M. (2011). On-line monitoring applications in nuclear power plants. Progress inNuclear Energy 53, 167-181

[6] Behringer K., & Crow R. (1980). Practical application of neutron noise analysis at boiling waterreactors, Wurenlingen, Switzerland. Swiss Federal Institute for Reactor Research Ch-5303 EIR-Bericht Nr 385

[7] Al Rashdan A., Smith J., St. Germain S., Ritter C., Agarwal V., Boring R., Ulrich T., & Hansen J.(2018). Light water reactor sustainability program: development of a technology roadmap for

onlinemonitoring of nuclear power plants INL/EXT-18-52206. U.S. Department of Energy, Office ofNuclear Energy

[8] https://www.energy.gov/sites/prod/files/2019/08/f65/ne-2019-advanced-sensors-instrumentation-summaries.pdf (retrieved April 21st, 2020)

[9] Pázsit I., &Demazière C. (2010). Noise Techniques in Nuclear Systems, Handbook of NuclearEngineering, Cacuci D.G. (ed.), Springer ISBN: 978-0-387-98130-7, 1629-1737

[10] Ma J., & Jiang J. (2011). Applications of fault detection and diagnosis methods in nuclear powerplants: a review, Progress in Nuclear Energy 53, 255-266

[11] Tambouratzis T., Giannatzis G., Kyriazis A., &Siotropos P. (2020). Applying the computationalintelligence paradigm to nuclear power plant operation: a review (1990-2015), International Journalof Energy Optimization and Engineering 9, 27-109

[12] Chionis D., Dokhane H., Ferroukhi H., Girardin G., &Pautz A. (2018). A PWR neutron noisephenomenology: part I – simulation of stochastic phenomena with SIMULATE 3K, Proceedings ofthe "PHYSOR 2018", IL, U.S.A.: American Nuclear Society, 1001-1012

[13] Smolensky P., (1986). Information processing in dynamical systems: Foundations of harmonytheory, Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1:Foundation, ed. McClelland J.L., Rumelhart, D.E., MITPress/Bradford Books

[14] Tambouratzis T., Chionis D., &Dokhane A., (2018). General regression neural networks for the concurrent, timely and reliable identification of detector malfunctions and/or nuclear reactor deviations from steady-state operation, in Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bengaluru, India, November 18th-21st, 2018, 524-531

[15] Xiong N., &Svensson P., (2002). Multi-sensor management for information fusion: issues andapproaches, Information Fusion 3, 163–186

[16] Li W.G., Wang Z., Wei G.L., Ma L.F., Hu J., & Ding D., (2015). A survey on multisensor fusionand consensus filtering for sensor networks, Discrete Dynamics in Nature and Society, 1–15

[17] Maybeck S. (1982). Stochastic Models, Estimating, and Control. River Edge, NJ: Academic Press

[18] Peng X.J., Li Q., &Wang K., (2015). Fault detection and isolation for self powered neutron detectorsbased on principal component analysis, Annals of Nuclear Energy 85, 213-219

[19] Sagar Y.V., Mishra A.K., Tiwari A.P., &Degweker S.B., (2015). Online fault detection anddiagnosis of in-core neutron detectors using generalized likelihood ratio method, IEEE Transactionson Nuclear Science, 62, 3311-3323

[20] Durrant-Whyte, H.F., (2016). Sensor models and multisensor integration, The International Journalof Robotics Research, 7, 97–113

[21] Battacharyya A., Yogi V., Singla S., BhushanM., Kelkar M.G., Tiwari A.P., Pramanik M., &BelurN., (2017). Adaptible, on-linemodels to detect and estimate gross error SPNDs, 2017 Indian ControlConference, January 4th-6th, 2017, 149-154

[22] Luo J.H., & He X.T., (2018). A soft–hard combination decision fusion scheme for a clustereddistributed detection system with multiple sensors, E Sensors, 18, 4370 (doi:10.3390/s18124370)

[23] Tambouratzis T. (1998). A consensus-function artificial neural network for map colouring, IEEETransactions on Systems, Man, and Cybernetics, vol. 28, 721-728

[24] Tambouratzis T. (1999). A novel artificial neural network for sorting, IEEE Transactions onSystems, Man, and Cybernetics, vol. 29, 271-275

[25] Kirkpatrick S., Gelatt Jr C.D., &Vecchi M.P. (1983). Optimization by simulated annealing, Science,vol. 220, 671-680

[26] Bäck T. (1996). Evolutionary Algorithms in Theory and Practice (1996), p. 120, Oxford Univ. Press

[27] MATLAB R2019.b, The MathWorks Inc. Natick, Massachusetts, U.S.A.

## AUTHORS

**Tatiana Tambouratzis**

BSc in Mathematics (University of Athens, Greece), MSc in Intelligent Systems and PhD in Artificial Neural Networks and Artificial Vision (Brunel University, U.K.). Researcher Grade IV, III, II (Institute of Informatics and Institute of Nuclear Technology - Radiation Protection, NCSR Demokritos, Attiki, Greece); Assistant and Associate Professor (Department of Industrial Management & Technology, University of Piraeus, Piraeus, Greece); Visiting Researcher (Department of Nuclear Engineering, Chalmers University of Technology, Göteborg, Sweden).Scholarships from the Greek Scholarship Foundation, SERC award (U.K.), National Research Institute (Greece) and the Royal Society (U.K.), Fulbright Foundation (U.S.A.), Teaching Mobility Scholarship Program (U.S.A.). Fellow of the Institute of Mathematics and its Applications (IMA), membership no. 23668; member of the Institute of Electrical and Electronics Engineers (IEEE), membership no. 93994512. Member of the editorial advisory board of Progress in Nuclear Energy (area: Fission Technology).

**Laurent Pantera**

PhD in Heuristics and diagnostics for complex systems (University of Technology of Compiègne, UTC, France). Experimentalist at the CEA (French Alternative Energies and Atomic Energy Commission), he focused its activity on nuclear data processing and uncertainty calculation. He was in charge of the online gamma spectrometry measurements and experimental uncertainties assessment in the framework of the PHEBUS PF (Fission Products) international experimental programme to improve the understanding of the phenomena occurring during a severe accidents in a light water reactor. Afterwards, he joined the experimental CABRI International Programme (CIP) whose aim is to study the behavior of fuel rods at high burnup under Reactivity Initiated Accident (RIA).

**Petr Stulik**

Group Leader
Department of Diagnostics and Measurements
ÚJV Řež, a. s.
Hlavní 130, Řež
250 68, Husinec
Czech Republic

Disciplines: mechanics, experimental physics, nuclear physics
Skills & expertise: mechanics, experimental physics, nuclear physics; pattern recognition, data analysis, data mining an knowledge discovery, statistical data analysis, data mining and knowledge discovery, statistical data analysis, information extraction Big Data, data processing, information filtering.

# PENALIZED BOOTSTRAPPING FOR REINFORCEMENT LEARNING IN ROBOT CONTROL

Christopher Gebauer and Maren Bennewitz

Humanoid Robots Lab, University of Bonn, Bonn, Germany
{cgebauer,maren}@cs.uni-bonn.de

## *ABSTRACT*

*The recent progress in reinforcement learning algorithms enabled more complex tasks and, at the same time, enforced the need for a careful balance between exploration and exploitation. Enhanced exploration supersedes the requirement to hardly constrain the agent, e.g., with complex reward functions. This seems highly promising as it reduces the work for learning new tasks, while improving the agents performance. In this paper, we address deep exploration in reinforcement learning. Our approach is based on Thompson sampling and keeps multiple hypotheses of the posterior knowledge. We maintain the distribution over the hypotheses by a potential field based penalty function. The resulting policy is more performant in terms of collected reward. Furthermore, is our method faster in application and training than the current state of the art. We evaluate our approach in low-level robot control tasks to back up our claims of a more performant policy and faster training procedure.*

## *KEYWORDS*

*Deep Reinforcement Learning, Deep Exploration, Thompson Sampling, Bootstrapping*

## 1. INTRODUCTION

The outstanding performance of deep reinforcement learning firstly shown by Mnih *et al.* [1] has opened a wide range of possibilities. Especially in the field of robot control this is promising, as it heavily reduces the requirements to implicitly state the desired behavior in more complicated situations. For example, popular non-learning methods are based on the dynamic window approach for navigation [2] or on differential dynamic programming for low-level control [3]. These approaches solve the task to interact with the environment sufficiently, but need to be carefully fine-tuned and are limited to the designed representation of the environment. The lack of understanding and integration of correlations between events in the interaction with the environment lead to a very stable reactive behavior but instability when more foresightedness is required. This is especially given for sequences of actions that rather dependent on high-level decisions as in human-robot interaction, when social acceptable behavior is addressed.

A reinforcement learning agent with enhanced exploration skills is very promising, especially in tasks including more elaborated behavior. While classical approaches are able to avoid collision with humans [4], learning is rather able to solve this task in a socially compliant manner [5]. Even though the results are promising and lead to good navigation policies, the process of learning is inert, due to the absence of deep exploration. With lack of deep exploration, the agent favors to greedily search in the known action space for the optimal solution instead of deeply explore the unknown capabilities first. As gradient-based methods have been established in recent years, the greedy behavior is originated in the local convergence given by the basic assumptions of their derivation. In navigation tasks, the resulting policy of poorly explored local optima rarely results in the expected behavior. Even though Chen *et al.* [5] showed the promising opportunities of

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 155-165, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101114

reinforcement learning in socially-aware navigation tasks, more research is required for robust end-to-end machine learning solutions [6].

While neural networks of great size inherit the potential to map almost any non-linear function, the major lack is to efficiently force the agent to explore its capabilities as well as the environment itself. Common methods address this problem by a detailed reward design including preprocessing [7] or guided learning steps with increasing difficulty [8]. Another approach includes network extensions to learn auxiliary tasks [9] or an internal reward based on reconstruction error of the current state [10].

In this paper, we introduce a novel boostrapped version of twin delayed deep deterministic policy gradient (TD3) [11] based on Thompson sampling [12] to increase deep exploration and increase the performance of the resulting neural network in terms of maximizing the expected return. Thompson sampling addresses the balance between exploitation and exploration by randomly sampling the policy parameters from a posterior distribution and acting according to it for one episode. The uncertainty inherited in the posterior distribution naturally induces exploration due to the resulting uncertainty in the optimal action. Furthermore, we penalize the similarity of the hypotheses to maintain the posterior distribution. In comparison to Zheng *et al.* [13], our novel bootstrapping method reduces the required computational resources while still improving the performance of the resulting agent. All the claims are backed up with an experimental evaluation.

## 2. Related Work

To address the problem of deepening the exploration, multiple approaches have been developed in the last few years. The most widely used one is curriculum learning [8], which improves the final policy by increasing the difficulty of the task over time. For example, Kulhanek *et al.* [14] applied this method by increasing the complexity of the environment at given milestones during training of a vision-based navigation policy. Nevertheless is this approach not addressing the learning algorithm itself but modifies the information stream of the training samples and therefore is always usable as an extension.

Another concept is based on outputs from additional heads using the same encoded state as the policy. The encoded state represents, e.g., the output of a convolutional neural network and is shared among all consequential networks. A head is the neural network that uses the encoded state to generate any desired output, e.g., action commands. The purpose of additional heads is to influence the shared network structure without direct usage of the head's output. Jaderberg *et al.* [15] introduced this idea as auxiliary tasks and optimized the additional heads via self-supervised learning using available quantities from the environment. Mirowski *et al.* [9] improved this method by introducing further auxiliary tasks, especially to predict the depth based on an RGB image. Both approaches require the agent to receive such quantities from the environment as well as are strongly bounded to a specific task.

Another method is based on internal usage of auxiliary outputs from the neural network. For example, Pathak *et al.* [10], building upon Stadie *et al.* [16], focused on reconstructing the relevant information of the next state by learning the dynamics of the environment. Exploration is induced by adding a bonus reward on states that have been reconstructed worse, representing its novelty. The approach is known as curiosity learning and was applied very effectively to robot navigation by Zhelo *et al.* [17]. Variational information maximizing exploration [18] uses curiosity learning and extends it by directly maximizing the expected information gain due to the corresponding action. All these methods have especially the problem to be unable to differentiate between stochastic dynamics and uncertainty, as it is not directly approximating latter.

Blundell *et al.* [19] used Bayesian neural networks, building upon Hinton *et al.* [20] and Graves *et* *al.* [21]*, to introduce uncertainty into the current weights. This naturally induces exploration due* to the resulting distribution over possible actions with respect to the current state and the uncertainty towards the inherited weights. Nevertheless is this method currently limited to applications where a network is trained to fit a known target value. Henderson *et al.* [22] applied this concept to different actor-critic algorithms by training the critic under weight uncertainty. In general, the critic evaluates the actions, which are mapped directly from states using the actor. This uncertainty, even though not applied to the actor itself, heavily improved training stability, as the actor depends on the critics performance.

Another class of approaches is based on Thompson sampling, which addresses exploration by considering multiple hypotheses of the next optimal action conditioned by a given state and the posterior knowledge. Osband *et al.* [23] applied this technique to deep reinforcement learning, which is known as bootstrapped deep Q-learning. The key concept is a shared network and multiple instances that return the action-value function $Q$, representing the heads. Thompson sampling is applied by drawing a specific head before each episode and acting according to it. The multi-head structure ensures an estimation of the posterior knowledge inherited by the neural networks to introduce uncertainty. Furthermore is each head trained on its own subset of the complete dataset $\mathcal{D}$. Zheng *et al.* [13] extended this to the actor-critic method by replicating multiple actor-critic pairs, known as double bootstrapped deep deterministic policy gradient (DBDDPG). Both approaches achieve a clear increase in performance compared to non-bootstrapped agents. Our approach is based on the actor-critic methods as well, but extends DBDDPG by adjusting the critic structure and penalizing the similarity of each bootstrapped actor. Furthermore, we reduce the computational cost in context of training and application, while still improving the resulting performance regarding the expected return.

## 3. Our Approach

In this section we first describe all preliminaries regarding the reinforcement learning setting and describe our approach in detail afterwards.

### 3.1. Preliminaries

We model the problem as a Markov Decision Process (MDP), where an agent interacts with the environment. Based on the current state $s_t \in \mathcal{S}$ the agent applies action $a_t \in \mathcal{A}$, according to the policy $\pi(\theta) : \mathcal{S} \rightarrow \mathcal{A}$ defined by its parameters $\theta$. At the end of timestep $t$, the agent receives the reward $r_t \in \mathcal{R}$ and the next state $s_{t+1}$ according to the state-transition probability distribution $P : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$. The discounted return is defined by $R_t = \sum_{i=0}^{T} \gamma^i r_{t+i}$, where $t + T$ represents the final time step and $\gamma \in [0, 1]$ is the discount factor.

In general the objective of reinforcement learning is to maximize the expected return $J(\pi) = \mathbb{E}[R_0|\pi]$, where the gradient $\nabla_\theta J(\pi)$ is used to update the policy. For deterministic policy gradient (DPG) [24] the gradient is not defined to directly depend on the policy, but rather points in the direction of the action-value function's gradient $\nabla Q$ at the current sample point

$$\nabla_\theta J(\pi) \quad = \quad \mathbb{E}[\nabla_a Q(s, a)|_{a=\pi(s_t)} \nabla_\theta \pi(s)]. \tag{1}$$

The action-value function $Q$ is defined by $Q(s_t, a_t) = \mathbb{E}[R_t|s_t, a_t]$ with its parameters $\theta_Q$ and updated using temporal difference:

$$\begin{aligned} L(Q) &= \mathbb{E}\left[(Q(s, a) - y_t)^2\right] \\ \text{with} \quad y_t &= r_t + \gamma Q(s_{t+1}, \pi(s_{t+1})) \end{aligned} \tag{2}$$

The target $y_t$ and therefore the loss function $L$ depend on the policy and the action-value function itself. For deep deterministic policy gradient (DDPG) [25], the neural network based extension of DPG, this introduces an instability towards the update as the numerous number of network parameters are adjusted concurrently. To overcome this, Mnih *et al.* [1] introduced a target network that is updated less frequently by partially copying the network parameters. The target network is denoted by a quotation sign, as, e.g., $Q'$, and applied to all trained models.

Another problem is the overestimation bias of the action-value function excessively increasing over training. As the policy update depends on the stability of the action-value function, the overestimation leads to a general instability of the training. Twin delayed deep deterministic policy gradient (TD3) [11] countermeasures this effect by adding another critic and taking the minimal action-value estimate as target. Both of the action-value functions are trained separately with identical, but modified target action-value:

$$y_t = r_t + \gamma \min_{i=1,2} Q'_i(s_{t+1}, \pi'(s_{t+1}) + \epsilon_t) \tag{3}$$

This improves stability by almost eliminating the overestimation bias during training. To further improve generalization, as deterministic policies usually tend to naturally overfit, the smoothing target noise term $\epsilon_t$ is added to the action estimation of the target policy in the target function [11]. It is drawn from a clipped Gaussian distribution. TD3 is the underlying algorithm used for our novel boostrapping approach based on Thompson sampling.

## 3.2. Bootstrapped Actor

The benefit of Thompson sampling is originated in the optimization of multiple correlated hypotheses based on the given dataset $\mathcal{D}$. Each hypothesis corresponds to its own set of function parameters $\theta_i \in \theta$ and trains on a subset $\mathcal{D}_i$ to ensure the maintenance of the distribution over the posterior knowledge $\hat{P}(\theta|\mathcal{D})$. The hat denotes the approximation of the true posterior distribution, as we do not know the true distribution, but rather draw multiple hypotheses assuming to be distributed according to $P$. Before each interaction with the environment, a hypothesis is chosen and acted upon on. This introduces an uncertainty and, therefore, a more elaborated exploration compared to greedy policies.

Our core contribution is a boostrapped version of TD3 based on Thompson sampling, in the following referred to as Multi-TD3. We instantiate a number $N \in \mathbb{N}$ of actors with a random set of parameters $\theta_i$, where $i \in N$, drawn from the prior distribution $P(\theta)$. These actors represent our hypotheses and are forming the distribution $\hat{P}(\theta|\mathcal{D})$, which then is dependent on the prior distribution $P(\theta)$. To maintain the distribution $\hat{P}(\theta|\mathcal{D})$ we update each actor separately based on its subset $\mathcal{D}_i$ defined by the mask $m_t$ drawn from a Bernoulli distribution [23]. The Bernoulli distribution is defined by the masking probability $p$, a new hyperparameter. The mask $m_t$ has the size of $N$ and indicates for each actor whether the specific sample is part of its subset or not. In our case, this results in a dataset $\mathcal{D}$ defined by the collection of tuples $\{s_t, a_t, s_{t+1}, r_t, m_t\}$. While each actor only depends on its subset $\mathcal{D}_i$, we combine the loss of all actors to update the shared network and therefore consider the whole dataset $\mathcal{D}$. The shared network is hindrance for clear separation of the different hypotheses, however, it is usually intractable to train complete copies of the same network in parallel.

We apply Thompson sampling by drawing randomly from the pool of actors before each episode, representing a hypothesis based on the posterior knowledge. In contrast to pure Thompson sampling, we choose during the entire episode the action according to the drawn policy $\pi_i$, neglecting its optimality, instead of redrawing before each interaction as suggested by Osband *et al.* [23]. While the randomness due to neglection of the argmax operator in the deterministic action inference is desired for data collection during training, the policy shall be greedy during evaluation. Therefore, we search greedily in each timestep for the optimal parameter

$$\theta^* \quad = \quad \underset{\theta_i}{\mathrm{a}rgmax}\, Q_1(s_t, \pi(s_t, \theta_i)). \tag{4}$$

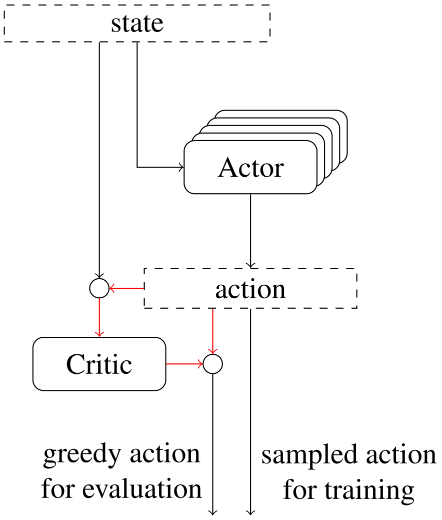


Figure 1: General information flow in Multi-TD3, illustrating the action generation in evaluation and during training.

Type your text

**Algorithm 1:** Multi-TD3

```
Input: Number of heads $N$, replay buffer $\mathcal{D}$, masking
        probability $p$, exploration noise $\epsilon$, smoothing
        target noise $\epsilon_t$
Initialize parameter $\theta_{Q_{j=1,2}}$, $\theta_{i\in N}$ from $P(\theta)$
for each iteration do
    if new episode then
        Pick actor $n \sim$ uniform$\{N\}$
    Apply $a_t \sim \pi_n(s_t) + \epsilon$
    Receive $r_t$ and $s_{t+1}$
    Sample mask $m_t \sim$ Bernoulli$\{p\}$
    Store $\{s_t, a_t, s_{t+1}, r_t, m_t\}$ in $\mathcal{D}$

    Sample minibatch $\mathcal{B}$ from $\mathcal{D}$
    for $i \in N$ do $\hat{a}_i \sim \pi'_i(\mathcal{B}) + \epsilon_t$
    $y \leftarrow \max_{i\in N}\{\ \min_{j=1,2}\ Q'_j(\mathcal{B}, \hat{a}_i)\}$
    Update each critic according to Eq. (2)
    Update each actor according to Eq. (6) using
      Eq. (1)
    Update the target networks
```

$Q_1$ refers to one of the two critics used in TD3, where either are usable. Eq. (4) satisfies the assumption on greedy maximization using argmax regarding the action in DPG. In Fig. 1 we illustrate the comparison of the information flow for action generation in training and evaluation. The additional loop, marked in red, for the greedy search of optimality brings in an overhead due to the linear cost in $N$. The greatest speedup in comparison to DBDDPG [13] is achieved by uniformly sampling an action during data collection for training instead of searching greedily for it. Additionally we are speeding up the greedy network inference by not increasing the number of critics to $N$, which results in a reduction of the runtime complexity from quadratic with $N$ to linear.

As mentioned, the critic structure is not modified and stays identical to TD3. However, we modify the target function in Eq. (3) to still minimize over both critics and additionally to maximize over all predicted next actions according to each actor

$$y_t = r_t + \gamma \max_{i\in N}\left[\min_{j=1,2} Q'_i(s_{t+1}, \pi'_i(s_{t+1}) + \epsilon_t)\right]. \tag{5}$$

The resulting complete optimization problem is given by

$$\begin{aligned}\min_{\theta_i,\theta_{Q_j}} -\sum_{i=1}^{N} \mathbb{E}\left[Q_1(s, \pi(s, \theta_i))\right] + \sum_{j=1,2} \mathbb{E}\left[(Q_j(s, a, \theta_{Q_j}) - y_t)^2\right] \\ +\beta \sum_{i=1}^{N} \sum_{j=1,i\neq j}^{N} \frac{1}{||\pi_i - \pi_j||_2^2},\end{aligned} \tag{6}$$

and summarized in Alg. 1 as well as visualized in Fig. 1. The last term is a penalty to further support the maintenance of the distribution $\hat{P}(\theta|\mathcal{D})$. It is inspired by the entropy cost used in, e.g., Schulman *et al.* [26]. It forces the distribution of the possible actions, given the current state, towards a uniform distribution. Since we apply deterministic policies, entropy is not applicable, as no distributions over actions is available. However, we reformulated this to minimize the repellent force between the deterministic policies induced by their potential field, known from identically charged point particles in the field of electrostatic [27]. This equals minimizing the inverted Euclidean norm and results in a uniform spread of the policies across the action space. Therefore, maximizing the potential of the actions according to the deterministic policies is similar to maximizing the entropy of the policy's distribution. To prevent a division by zero, a lower bound is set for the Euclidean norm. The parameter $\beta$ scales the cost and is a newly introduced hyperparameter.

## 4. Experimental Evaluation

The main focus of this work is to increase deep exploration in the application of robot control tasks. In Sec. 4.1. we introduce the neural network structure and its hyperparameters we used in our experiments. The complete network structure, as well as the optimization algorithm is implemented using Tensorflow [28]. In Sec. 4.2. we compare our novel bootstrapping approach to the state-of-the-art bootstrapping method for deterministic actor-critic methods [13], as well as common unbootstrapped deterministic policies [25], [11]. We evaluated the performance in different low-level control tasks simulated in PyBullet [29] and wrapped with OpenAI gym [30]. In Sec. 4.3. we evaluate the reduced runtime complexity and especially the reduced training time in comparison to DBDDPG [13] as well as the overhead towards the unbootstrapped TD3 [11].

### 4.1. Deep Network Architecture

For DDPG and TD3 two different sets of hyperparameters turned out to be superior in our experiments. The one applied to DDPG and DBDDPG is similar to Lillicrap *et al.* [25]. We apply two dense layers with $\{$**512**,256$\}$, while the bold number represents the shared layer for DBDDPG. The structure for the actor and critic is identical. The hyperparameters are given by the discount factor $\gamma = 0.95$, actor learning rate $l_{r,\pi} = 10^{-4}$, critic learning rate $l_{r,Q} = 3 \times 10^{-4}$, the exploration noise $\epsilon = 0.2$, the target update factor $\tau = 0.01$ and the size of the minibatch $\mathcal{B}$ with 128. We apply kernel and bias regularization via L2-regularization with $l_2 = 10^{-4}$. The replay buffer has a size of $10^6$ for all approaches. For DBDDPG the degree of bootstrapping is given by $N = 5$, as suggested in the original paper [13], and the masking probability by $p = 0.5$.

For TD3 and our approach we choose slightly different hyperparameters, which have been tuned during hyperparameter validation. In general, the modified hyperparameters increase the performance with the cost of more unstable training. Instability refers to higher variance over the course of training or even the absence of any progress. DDPG and DBDDPG did not train with the modified hyperparameters. TD3 did train with the basic hyperparameter setting, but performed much better with our hyperparameters. The hyperparameters have been tuned for the unbootstrapped algorithms and remain unchanged when boostrapping is applied.

The network still consists of two dense layer, but with $\{$**256**,256$\}$ nodes. Again the bold number represents the shared layer for our approach. The modified hyperparameters are the discount factor $\gamma = 0.99$, critic learning rate $l_{r,Q} = 8 \times 10^{-4}$, the exploration noise $\epsilon = 0.1$, the target update factor $\tau = 0.005$ and the size of the minibatch $\mathcal{B}$ with 256. The smoothing target noise is given by $\epsilon_t = 0.2$ and $\epsilon_{t,clip} = 0.5$. We train our agent with a bootstrapping degree of $N = 10$ and a masking probability of $p = 0.3$. The potential penalization factor $\beta$ varies across the environments and needs to carefully tuned. However, mostly a value of $\beta = 10^{-6}$ is a good start for hyperparameter search.

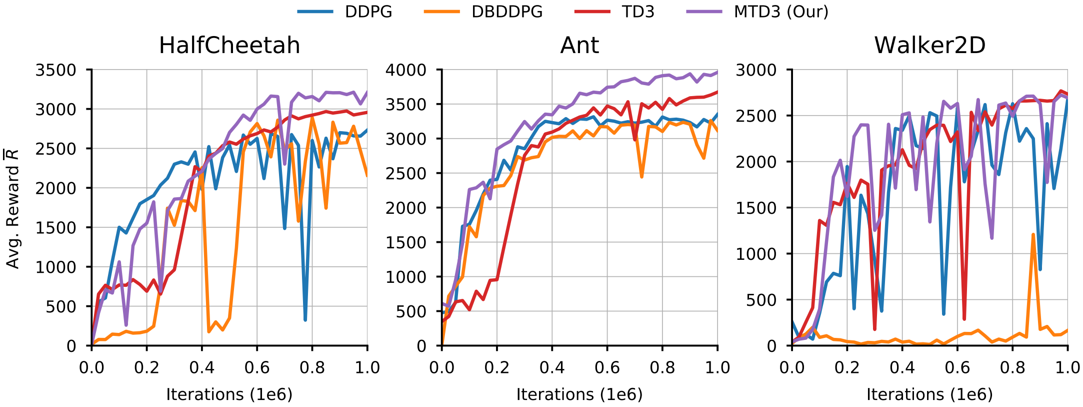


Figure 2: Averaged return over the course of training for a variety of low-level control tasks from the OpenAI gym [30] simulated with PyBullet [29]. As can be seen, enhanced exploration due to our approach clearly improves the performance of the agent.

## 4.2. Bootstrapped Performance

In this section, we evaluate the improved exploration by comparing the evolution of the received reward during the course of training. Each environment represents a low-level control task of a simplified robot. The received observation consists of the joint angles and joint velocities. The actions correspond to the applied torques, one for each joint. The reward is calculated based on the forward traveling speed and a fix alive bonus to consider the length of the episode. Additionally, a penalty for higher effort is applied, which consists of the torque magnitude and the impact forces on the ground. The episode ends and the environment is reset, when the policy leads the robot into an absorbing state or a maximum length of 1000 timesteps is reached. An absorbing state is defined as a state that the robot is unable to leave within the given action space. These robot environments first appeared in Schulmann *et al.* [31]. The underlying engine to simulate the kinematics of the robot and the contacts with the groundplane is PyBullet [29], while the environment is wrapped and accessed by the agent using Gym [30].

The baselines are represented by DDPG and TD3, while TD3 clearly outperforms DDPG. This is originated in the superior target function and stabilized critic optimization. Due to DBDDPG being based on DDPG, not only the absolute difference to our approach regarding the received reward is of interest, but also the difference in relative improvement towards unbootstrapped methods. Where absolute refers to the direct comparison and relative to the comparison of the improvement towards the unbootstrapped base. To reduce statistical drawbacks due to randomness in initialization regarding the comparability of all algorithms, each experiment is conducted four times for each algorithm under identical conditions and the best trials are compared in Fig. 2. The data is generated by evaluating the agent during the course of training and averaging the final return over multiple episodes.

Our approach achieves a clear increase in performance regarding the expected return. We are outperforming all of the other approach due to a more elaborated exploration. However, when no beneficial effect due to bootstrapping is noticeable, the overhead produced by our approach does not decrease the final performance and achieves the same results as TD3. This especially can be seen in the Walker2D environment, when compared to the effect of DBDDPG. In direct comparison DBDDPG, or DDPG in general, suffers from greater instability and therefore high variance over the course of training. Furthermore does DBDDPG not manage to increase the performance in comparison to DDPG, when the hyperparameters from DDPG are applied as shown in our experiments. Therefore, our approach not only outperforms DBDDPG in absolute measure

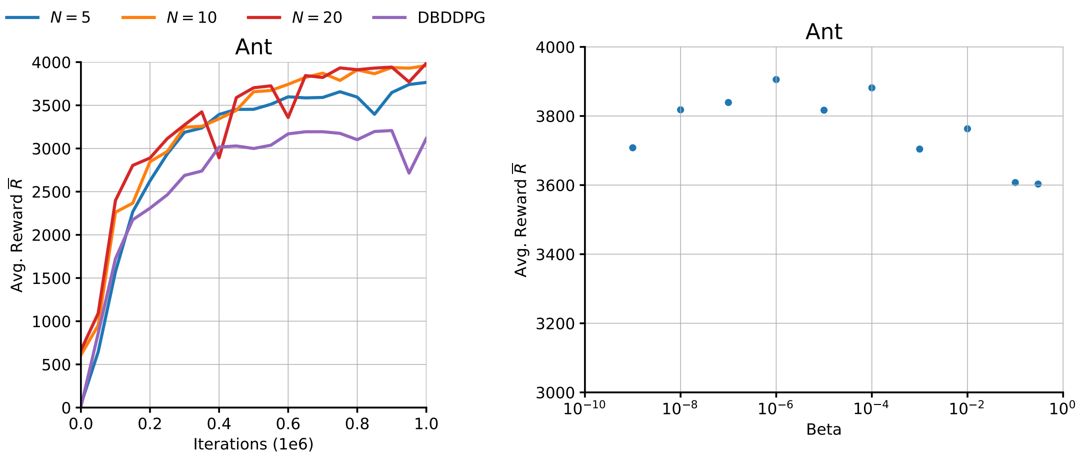


Figure 3: Illustration of the effect of the degree of bootstrapping (left) and the influence of the potential based penalty (right) for the environment Ant. As can be seen, with increasing $N$ the beneficial effect due to enhanced exploration increases. The potential based penalty coefficient improves the expected return until it constraints the training to heavily.

but also in relative measure compared to the unbootstrapped methods. We are still improving the performance of the agent compared to TD3 even though the hyperparameters are optimized explicitly for TD3.

The exploration is especially supported by the potential-based penalty as it ensures a greater maintenance of the posterior distribution. This effect is visible in Fig. 3 on the right for $N = 5$. Each data point represents the mean reward over the last five evaluations given a certain beta. While only the penalty coefficient beta is varying, the average reward starts to increase and decrease after an optimal value of beta. While first being supportive towards the posterior distribution the penalty prevents optimization if too great. Another essential parameter is the degree of bootstrapping $N$, which increases the number of hypotheses and therefore increases the possible diversity. As Fig. 3 on the left shows, increasing the degree of bootstrapping further enhances the beneficial effect on the resulting performance. However, this effect saturates as can be seen by comparing the curves for $N = 10$ and $N = 20$. All values of $N$ are superior over DBDDPG with $N = 5$.

A major limitation is the source of improvement itself. While exploration is heavily enhanced by uncertainty, at the same time there is no guarantee that each trial under identical condition will benefit from it. The potential-based penalty supports this effect, as it prevents a greater similarity between each hypotheses and emphasizes generality. Furthermore, is it unclear with the current state of the art, how a conjunction of the trained hypothesis, each represented by a neural network, could be possible. Since the local optima usually lay far apart, no strategy for weight combination is commonly known. Even if the optima lay close by each other, a naive combination of the weights will most likely result in a major decrease in performance. Therefore, we evaluate each actor greedily, as explained above, using the critic and act upon the expected to be most beneficial hypothesis, while the conjunction is put to future work.

## 4.3. Training- and Runtime Analysis

Another major contribution of our bootstrapping approach is the much faster training time in comparison to the current state of the art due to random policy sampling instead of greedy data collection. To evaluate this, we conducted a shortened training including 10,000 iterations and average the time per iteration. No evaluation takes place, therefore the averaged time includes one interaction with the environment and one optimization step of the neural networks. For the averaged inference time during evaluation the network generates 10,000 actions based on random

Table 1: Runtime (average and std. dev.), normalized by our approach for $N = 5$

| **Application** | **OUR, N = 5** | **OUR, N = 10** | **TD3** | **DBDDPG** |
|---|---|---|---|---|
| Training | $\mathbf{1.0} \pm 0.129$ | $2.31 \pm 0.139$ | $0.609 \pm 0.105$ | $2.817 \pm 0.682$ |
| Evaluation | $\mathbf{1.0} \pm 0.063$ | $1.982 \pm 0.128$ | $0.144 \pm 0.014$ | $3.9603 \pm 0.189$ |

states, while the interaction with the environment is neglected. The averaged time only includes the inference, not the sampling of the random states. All data is summarized in Tab. 1 and normalized by our approach with a degree of bootstrapping $N = 5$, represented by bold digits, to be easily comparable.

As expected, is TD3 in comparison the fastest in training, as least networks have to be optimized. Nevertheless is ours only slightly slower due to the efficient data collection, while the overhead is given by the increased number of actor heads that need to be optimized. In evaluation the overhead becomes more obvious, as we are searching greedily for the optimal action and quantify the performance of each actor using the critic. Our approach is still applicable in real time, since one inference for $N = 10$ takes on average less than 9 ms.

In comparison to DBDDPG, our approach is much faster in training and application. The quadratic cost in evaluation is already for $N = 5$ noticeable and especially the greedy collection of the data heavily slows down the training process in DBDDPG. Already for this degree of bootstrapping our approach is more performant regarding the expected return, while being faster in training and evaluation. Even for an increasing degree of bootstrapping our approach is faster, as is shown for $N = 10$.

## 5. Conclusion

In this paper, we presented a novel bootstrapping approach based on Thompson sampling applied to twin delayed deep deterministic policy gradient (TD3). Our approach trains and infers much faster than the current state of the art, which is DBDDPG [13], and still achieves a seriously improved performance. This especially results from applying the critic structure from TD3 instead of bootstrapping the entire actor-critic structure. Another speedup is achieved by sampling the actions during training based on Thompson sampling and not by greedily searching for the optimal action. To benefit from Thompson sampling, it is important to maintain the posterior knowledge inherited in our actor structure. We address this by adding a potential field based penalty, which induces high cost when the hypotheses agree on the same optimality. A well maintained distribution, given the posterior knowledge, naturally induces deep exploration when acted according to it during data collection.

We evaluated our approach in a variety of low-level control tasks, which strongly back up our claims. The experiments show that we outperform DBDDPG in all of the trained environments with a major decrease in computational cost regarding training and evaluation. Furthermore, we add a clear improvement in comparison to TD3 with the drawback of minor computational cost increase. This benefit is mainly caused by the improved exploration during data collection induced by sampling the current policy based on Thompson sampling, especially in combination with our potential field penalty constraint.

## Acknowledgements

This work has partly been supported by the German Research Foundation under Germany's Excellence Strategy, EXC-2070 - 390732324 (PhenoRob).

## REFERENCES

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, “Human-level control through deep reinforcement learning,” *Nature*, 2015.

[2] D. Fox, W. Burgard, and S. Thrun, “The dynamic window approach to collision avoidance,” *IEEE Robotics and Automation Magazine (RAM)*, 1997.

[3] Y. Tassa, N. Mansard, and E. Todorov, “Control-limited differential dynamic programming,” in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2014.

[4] G. Ferrer, A. G. Zulueta, F. Cotarelo, and A. Sanfeliu, “Robot social-aware navigation framework to accompany people walking side-by-side,” *Autonomous Robots*, 2017.

[5] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, “Crowd-Robot Interaction: Crowd-Aware Robot Navigation With Attention-Based Deep Reinforcement Learning,” *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2018.

[6] V. Dhiman, S. Banerjee, B. Griffin, J. M. Siskind, and J. J. Corso, “A Critical Investigation of Deep Reinforcement Learning for Navigation,” *CoRR*, 2018.

[7] L. Tai, G. Paolo, and M. Liu, “Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2017.

[8] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum Learning,” in *Proc. of the Intl. Conf. on Machine Learning (ICML)*, 2009.

[9] P. Mirowski, R. Pascanu, F. Viola, H. Soyer, A. Ballard, A. Banino, M. Denil, R. Goroshin, L. Sifre, K. Kavukcuoglu, D. Kumaran, and R. Hadsell, “Learning to Navigate in Complex Environments,” *CoRR*, 2016.

[10] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, “Curiosity-driven Exploration by Self-supervised Prediction,” in *Proc. of the Intl. Conf. on Machine Learning (ICML)*, 2017.

[11] S. Fujimoto, H. van Hoof, and D. Meger, “Addressing Function Approximation Error in Actor-Critic Methods,” in *Proc. of the Intl. Conf. on Machine Learning (ICML)*, 2018.

[12] W. R. Thompson, “On the likelihood that one unknown probability exceeds another in view of the evidence of two samples,” *Biometrika*, 1933.

[13] Z. Zheng, C. Yuan, Z. Lin, Y. Cheng, and H. Wu, “Self-Adaptive Double Bootstrapped DDPG,” in *Proc. of the Intl. Jt. Conf. on Artificial Intelligence (IJCAI)*, 2018.

[14] J. Kulhánek, E. Derner, T. de Bruin, and R. Babuška, “Vision-based Navigation Using Deep Reinforcement Learning,” in *Proc. of the Europ. Conf. on Mobile Robotics (ECMR)*, 2019.

[15] M. Jaderberg, V. Mnih, W. M. Czarnecki, T. Schaul, J. Z. Leibo, D. Silver, and K. Kavukcuoglu, “Reinforcement Learning with Unsupervised Auxiliary Tasks,” *CoRR*, 2016.

[16] B. C. Stadie, S. Levine, and P. Abbeel, “Incentivizing Exploration In Reinforcement Learning With Deep Predictive Models,” *arXiv preprint*, 2015.

[17] O. Zhelo, J. Zhang, L. Tai, M. Liu, and W. Burgard, “Curiosity-driven Exploration for Mapless Navigation with Deep Reinforcement Learning,” *CoRR*, 2018.

[18] R. Houthooft, X. D. Chen, Y. M. Duan, J. Schulman, F. D. Turck, and P. Abbeel, “Variational Information Maximizing Exploration,” in *Proc. of the Conf. on Neural Information*

*Processing Systems (NIPS)*, 2016.

[19] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, “Weight uncertainty in neural networks,” in *Proc. of the Intl. Conf. on Machine Learning (ICML)*, 2015.

[20] G. E. Hinton and D. van Camp, “Keeping the neural networks simple by minimizing the description length of the weights,” in *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, 1993.

[21] A. Graves, “Practical variational inference for neural networks,” in *Advances in Neural Information Processing Systems 24*, 2011.

[22] P. Henderson, T. Doan, R. Islam, and D. Meger, “Bayesian policy gradients via alpha divergence dropout inference,” *CoRR*, 2017.

[23] I. Osband, C. Blundell, A. Pritzel, and B. V. Roy, “Deep Exploration via Bootstrapped DQN,” *CoRR*, 2016.

[24] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, “Deterministic Policy Gradient Algorithms,” in *Proc. of the Intl. Conf. on Machine Learning (ICML)*, 2014.

[25] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning.” in *Intl. Conf. on Learning Representations (ICLR)*, 2016.

[26] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms.” *CoRR*, 2017.

[27] E. Shech and E. Hatleback, “The material intricacies of coulomb’s 1785 electric torsion balance experiment,” July 2014.

[28] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[29] E. Coumans and Y. Bai, “PyBullet, a Python module for physics simulation for games, robotics and machine learning,” http://pybullet.org, 2016–2019.

[30] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, “OpenAI Gym,” 2016.

[31] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, “High-dimensional continuous control using generalized advantage estimation,” in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2016.

# ANALYSIS OF THE DISPLACEMENT OF TERRESTRIAL MOBILE ROBOTS IN CORRIDORS USING PARACONSISTENT ANNOTATED EVIDENTIAL LOGIC Eτ

Flavio Amadeu Bernardini[1], Marcia Terra da Silva[1], Jair Minoro Abe[1],
Luiz Antonio de Lima[1] and Kanstantsin Miatluk[2]

[1] Graduate Program in Production Engineering Paulista University,
Sao Paulo, Brazil
[2]Bialystok University of Technology, Bialystok, Poland

## *ABSTRACT*

*This article proposes an algorithm for a servo motor that controls the movement of an autonomous terrestrial mobile robot using Paraconsistent Logic. The design process of mechatronic systems guided the robot construction phases. The project intends to monitor the robot through its sensors that send positioning signals to the microcontroller. The signals are adjusted by an embedded technology interface maintained in the concepts of Paraconsistent Annotated Logic acting directly on the servo steering motor. The electric signals sent to the servo motor were analyzed, and it indicates that the algorithm paraconsistent can contribute to the increase of precision of movements of servo motors.*

## *KEYWORDS*

*Paraconsistent annotated logic, Servo motor, Autonomous terrestrial mobile robot, Robotics.*

## 1. INTRODUCTION

The article focuses mainly on the proposal to develop an algorithm with Paraconsistent Logic for the control of a directional servo motor of an autonomous mobile robot. Given the demand for high investments in new technologies in industrial areas, particularly in the increasing use of robots in the automotive sector [1], this proposal can collaborate with the reduction of the need for maintenance in the factory's internal automated transport.

Regarding mobile robots, they are defined as non-fixed automatic devices capable of moving and interacting with the environment. They are classified according to the environment in which they move, which can be: terrestrial, aerial, aquatic, or underwater [2]. Concerning their displacement environments, they can be in industries [1], where mobile robots can help in the supply of assembly line components; domestic, working in cleaning activities; and at the post office, where they transport items from storage points to distribution points or vice versa.

The design of these robots applies Artificial Intelligence (AI) in decision making. Considering the need for allowing the robot to face uncertainty, alternative systems using non-classical logics, as the so-called Fuzzy systems, are frequently applied. According to [3], paraconsistent logic is one of these non-classical logics, and has applications in software development, neural

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 167-175, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101115

computing, automation and robotics. In applications that require AI technology in decision-making, the Paraconsistent Artificial Neural Network is already used, including robots [4]. Some published works highlight the pioneering of Paraconsistent Annotated Evidential Logic Eτ applied in a series of robots. The first robot, named Emmy, was built in 1999. After that, in 2004 a second prototype enhanced the first one and a third one was designed in 2009 improving the navigation system. The differential of the present work, when compared to the previous ones, consists in the use of servo motor for the steering control of the prototype of the robot [5].

## 2. BASICS CONCEPTS

The following is a brief approach in the main concepts that based this work.

### 2.1. Servo Motor with Microcontroller

Servo motors are precise, high torque electromechanical devices with a rotating motion proportional to an electrical signal. These movements are monitored by a rotary resistive sensor that has the function of returning the information of the servo's real position to an electronic control circuit [6]. Servo motors can be applied in industrial robotics or precision mechanical machines, such as used in machining centers. The actuator system consists of a direct current motor and gearbox that reduce speed and increase the torque applied to the servo control rod. As it is well known, the main determining factors of the technical characteristics for the correct application of the servo motor are the speed of rotation, the degree of freedom, the torque, the material that makes up the gears, as well as the consumption of electrical energy. Microcontrollers are control devices that can be programmed to meet the requirements of robotics projects using the Integrated Development Environment software.

### 2.2. Design and Control of Mechatronic Systems

The design process of Mechatronic Systems (MS) and other engineering objects, in general, contains several phases [7][8]. In the second phase Conceptual Design (CD) the main design activities aim to generate and evaluate the system's conceptual model, and the main design concept. The Detailed Design (DD) is the third phase and comprehends the mechatronic subsystems' concrete model creation, numeric calculations and, synthesis, and analysis. Lastly, comes the production phase. One of the main requirements for the conceptual model of the MS is that the model should allow the easy transfer from the conceptual description at the CD phase to the concrete models of MS structural and functional design during the DD phase (the synthesis and analysis). MS conceptual model should also take into account MS several levels and present them in a regular formal basis, i.e. lower level – MS structure, current level – MS aggregated dynamic representation as a unit in its environment, higher level – environment construction and technology, MS coordinator and its coordination processes, i.e. design and control.

Traditional mathematics, AI, and other nowadays models [7][8] do not meet all the above requirements. They do not allow describing robotic and mechatronic systems on all their levels in one common formal basis. So, Hierarchical Systems (HS) technology and created MS model [7-10] are coordinated with known mathematical and AI models, thus meeting all the above requirements. Models of MS structure, MS as a unit in its environment, and MS environment model are presented in the common HS formal basis. The models are connected by HS coordinator, which performs the design and control tasks on its selection, learning, and self-organization strata.

Moreover, conceptual model of MS presents the connected descriptions of MS subsystems of various nature, i.e. mechanical, electrical, and computer. HS technology was implemented in this paper for the case of the Terrestrial Mobile Robot (TMR) conceptual design and control. More attention was paid to the servomotor (electro-mechanical mechatronic subsystem) design and control.

### 2.3. Paraconsistent Annotated Evidential Logic Eτ

Historically, since Aristotelian thought, logic has contributed to correct thinking and in the world observations are not limited to false and true states, and often seeks to relate reasoning with knowledge. Over time, logic has been divided into classical and non-classical, and within the latter, paraconsistent logic has occupied a prominent place, as it deals with the principles of contradiction, in addition to the basic principles of Aristotle's classical logic [11]. Based on the concepts of Paraconsistent Logic, The Paraconsistent Annotated Evidential Logic Eτ works with propositions of type p (μ, λ), where p is a proposition and μ, λ € (0, 1) (closed range). Intuitively, μ indicates a degree of favorable evidence and λ indicates a contrary degree of evidence of proposition p. Based on the values of the degree of favorable evidence, the degree of unfavorable evidence, the properties of the Paraconsistent Annotated Evidential Logic Eτ are applied to calculate the degree of certainty and degree of uncertainty. Then, these values will be used as a reference for decision making in various applications, such as robotics for example.

## 3. METHODOLOGY

Initially, to ensure a satisfactory sequence in the preparation and execution of this work, the chosen methodology was divided into four stages. The first was a literature search on servomotors, microcontrollers, and Paraconsistent Annotated Logic. Next, conceptual design models of TMR servomotor and its control system were created using HS technology. After these studies, the C language program was prepared for the microcontroller to generate the specific signal to control the servo motor. Tests were performed with the oscilloscope to verify the quality of the signal generated by the microcontroller, as well as to observe the movement of the servo motor. In addition, a logic C programming based on Paraconsistent Annotated Evidential Logic Eτ was applied to verify its efficiency in servomotor decision making. Thus, the utility of logic in controlling the direction of the autonomous robot was verified. In conclusion, the paraconsistent annotated logic in the C programming of the microcontroller was applied to verify the effectiveness of the logic in the decision making of the servomotor, and the results showed good efficiency in controlling the direction of the autonomous robot.

## 4. THE AUTONOMOUS TERRESTRIAL MOBILE ROBOT DESIGN

The robot will be equipped with six ultrasonic sensors connected to a microcontroller that, through a specific electrical signal, will control the robot's directional servo motor.

### 4.1. Mechatronic Design and Control of TMR Servomotor

In this paper, HS technology, and developed MS conceptual model are used for TMR design and control, including all its mechatronic subsystems. In the design process, at the CD phase, the servomotor subsystem was presented by the dynamic system $(\rho,\varphi)$ [7, 8], which was transformed to state-space representation at the DD phase, see Figure 1. The final results are presented in the form of equations (1) and (2).

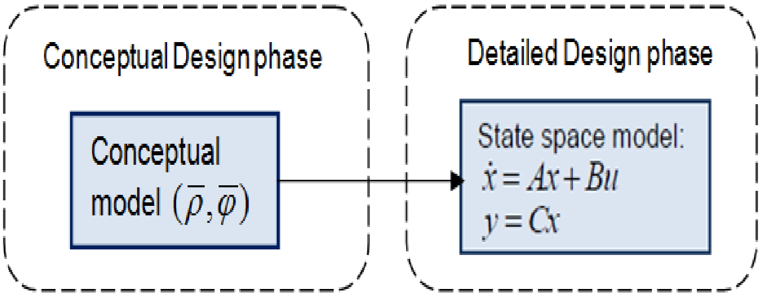


Figure 1. Conceptual model transformation at the design phases.

$$\dot{x}=\begin{bmatrix}\dot{i}_a\\ \dot{\omega}_m\end{bmatrix}=\begin{bmatrix}\frac{-L_a}{R_a} & \frac{K}{R_a}\\ \frac{-K}{B} & \frac{-K}{B}\end{bmatrix}\begin{bmatrix}i_a\\ \omega_m\end{bmatrix}+\begin{bmatrix}\frac{-1}{R_a}\\ 0\end{bmatrix}E$$

$$y=\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}i_a\\ \omega_m\end{bmatrix}+0 \tag{1}$$

The first state equation of (1) corresponds to $\bar{\varphi}$ function, and the output equation corresponds to the reaction $\bar{\rho}$ of $(\bar{\rho},\bar{\varphi})$ representation at CD phase. State space equations (1) can be transformed at DD phase to the following transfer function (2) if necessary:

$$G(s)=\frac{\omega_m(s)}{e_a(s)}=\frac{K_i}{s^2JL_a+sJ_mR_a+K_iK_b} \tag{2}$$

In equations (1) and (2), $i_a$ is armature current, $L_a$ is armature inductance, $R_a$ is armature resistance, $V_a(t)$ is input voltage, $E_b$ is back emf, $K_b$ is voltage constant, $T_L$ is load torque, $T_m$ is motor torque, $\theta_m$, $\omega_m$ are motor angular change and velocity respectively, $K_i$ is a moment constant, $J$ is the motor moment of inertia, $B$ is a friction constant, $K$ is a constant.

The conceptual $(\rho,\varphi)$ model of the servomotor control system was transformed to Paraconsistent Logic model and implemented at DD phase by the developed program unit written in C language. This Paraconsistent Logic program unit was used to control the Dynamixel AX-12A servomotor selected in the design (synthesis) process of TMR. The control results are presented below and show the effectiveness of Paraconsistent Logic model application and the method proposed.

## 4.2. Servo Motor for Robot Control

Therefore, programming the microcontroller in C language can be idealized in the Integral Development Environment after consulting the microcontroller and servo motor manufacturer's manual. The general characteristics of the electrical signal sent by the microcontroller to a servo motor, as well as the respective positions assumed by it, can be seen in Figure 2. In the figure, the first image shows a high signal that lasts 1 ms, followed by a low one. The total period of the signs is the sum of one high and the low that follows. In this case, the configuration of the microcontroller is set to emit signs with a total period of 20 ms, comprising a high-level signal varying its pulse width from 1 ms to 2 ms, and a low-level signal with the corresponding amplitude. To each sign received, the servo motor responds with an angle of movement, varying from 0 to 180 degrees.

In practice, the microcontroller program alternates the high and low levels of the microcontroller output pin to form the signal that will be applied to the servomotor. The high and low-level intervals depend on the load values of the microcontroller time recorder, and the duration of these high-level intervals is controlled using the Positive Duty variable.

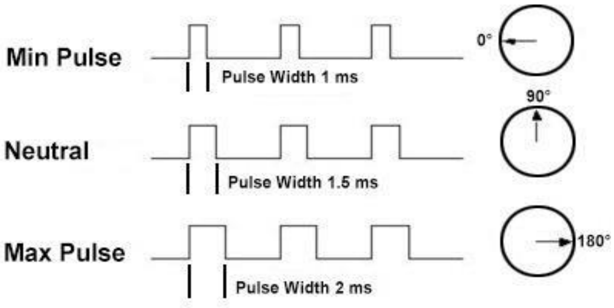


Figure 2. Electric signal control servo motor

## 4.3. Application of Paraconsistent Annotated Logic in Servo Motor to Robots

The initial proposal of the Paranalizer algorithm [12] should be used in the servo motor controller [6] (Dynamixel AX-12A with a set of elements) for the best performance of the angles. The Paranalizer makes subtle adjustments of movement possible, which helps the maintenance of the servo motor working in conformity with the manufacturer's technical guide.

Paranalizer

```
int paraAnalisador (float mi, float lambda) {
Normalization of the degrees of evidence for the value range between 0 and 1 mi = mi / 100;
Favorable degree of evidence - range of values between 0 and 1 lambda = lambda / 100;
Contrary degree of evidence - range of values between 0 and 1 float Gce = mi - lambda;
Gce - Degree of certainty - Gce = mi - lambda - range of values comprised by - 1 to + 1
float Gin = ((mi + lambda) - 1);
Gin - Degree of uncertainty - Gin = mi + lambda - 1 - range of values comprised by - 1 to + 1
int state = 0;
Extreme and non-extreme logical states - float module_Gce;
Value in the module of the degree of certainty
float module_Gin;
Value in the module of the degree of uncertainty
if (Gce < 0)
module_Gce = Gce * (-1);
else
module_Gce = Gce;
if (Gin < 0)
module_Gin = Gin * (-1);
else
module_Gin = Gin;
Determination of extreme states
```

```
Proposition: Free Front
if(Gce >= vcve)
{state = 1};
True - won't hit
else if(Gce <= vcfa)
{ state = 2};
False - will hit - stop, reverse and turn right and then left
else if(Gin >= vcic)
{state = 3};
```

## 5. Results

The practical tests were satisfactory on microcontroller signals generation to control servomotor, and Figure 3 shows the images captured from the Tektronics oscilloscope model TDS-1002 C-EDU that was used in the tests. The vertical cursors indicate Δt of 2.040ms for the 180º angle and a Δt of 1.020ms for a 90º angle of the servomotor, very close to the values required by the manufacturers.

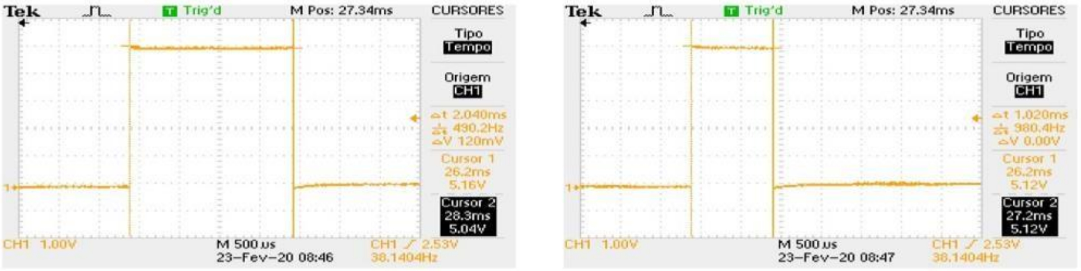


Figure 3. Waveforms generated by the microcontroller for servomotor control.

With the paraconsistent logic applied to the servo motor [8] (Dynamixel AX-12A with a set of elements), it is possible to keep the servo motor within the quality specifications proposed by the manufacturer and help in the displacement (direction) of the land mobile robots in runners. The next program shows the possibility in implementing para- consistent logic in servo control. It can be seen in the examples below that the states of the information can vary depending on the signal of each sensor: True, False, Paracompleteness – information is not sufficient to make a decision, Inconsistent – information is contradictory. This implementation will be complete at the appropriate time of the research. Paraconsistent Logic will be applied when sensors are close to indicate a movement that requires an angle lower then 90º or higher then 180º, as well as long duration of limit angle, which could provoke a possible failure or reduction of the servo motor's life. Examples:

```
1-Inconsistent - turn slightly to the right, obstacle to the left wide open
else if(Gin <= vcpa)
{state = 4};

2-Paracompleteness - turn slightly left, right obstacle wide open
else if( (Gce >= 0) && (Gce < vcve) && (Gin >= 0) && (Gin < vcic) && (Gce >= Gin))
{state = 5};
```

Tending

1-Almost true tending to inconsistent - turning too much to the right, obstacle to the left next -

turning more than state 3

```
else if((Gce >= 0) && (Gce < vcve) && (Gin >= 0) && (Gin < vcic) && (Gce < Gin))
{state = 6};
```

2-Inconsistent tending to true - turn slightly right, obstacle left open - turn less than state 5

```
else if((Gce >= 0) && (Gce < vcve) && (Gin > vcpa) && (Gin <= 0) && (Gce >= modulo_Gin))
{state = 7};
```

3-Almost true tending to Paracompleteness - turning a lot to the left, obstacle to the right next - turning more than the state 8

```
else if((Gce >= 0) && (Gce < vcve) && (Gin > vcpa) && (Gin <= 0) && (Gce < modulo_Gin))
{state = 8};
```

4-Paracompleteness tending to the true - turn left slightly, obstacle open right - turn more than state 4

```
else if((Gce > vcfa) && (Gce <= 0) && (Gin > vcpa) && (Gin <= 0) && (modulo_Gce >=
modulo_Gin))
{state = 9};
```

5-Almost false tending to paraconsistent - stop turning too much to the left - almost hitting, an obstacle to the right too close

```
else if((Gce > vcfa ) && (Gce <= 0) && (Gin > vcpa) && (Gce < Gin) && (Gin <= 0))
{state = 10};
```

6-Paracompleteness tending to false - stop and turn a little to the left, obstacle to the right open very close

```
else if((Gce > vcfa) && (Gce <= 0) && (Gin >= 0) && ( Gin < vcic) && (Gce >= Gin))
{state = 11};
```

7-Almost false tending to inconsistent - stop and turn too much to the right, an obstacle to the left too close

```
else if((Gce <= 0) && (Gce < vcfa) && (Gin >= 0) && (Gin < vcic) && (Gce < Gin))
{state = 12};
```

8-Inconsistent tending to false - stop and turn slightly to the right, an obstacle to the left open too close

```
{return state};
```

## 6. CONCLUSION

The work showed the applicability of the developed algorithms based on Paraconsistent Annotated Evidential Logic Eτ algorithms in the DD phase in the robot's servomotor and, thus, contributes to the servomotor's efficiency and assist in driving decision making. Additionally, the application of Paraconsistent Logic allows to maintain the servomotor working within the manufacturer specifications, which contributes for a longer life cycle. As future work, the algorithm must be improved to ensure the use of the servomotor within its technical specifications and keep the perspective of the device's life. This article has provided possibilities that will be explored in the next phase with new experiments extensively.

## ACKNOWLEDGEMENTS

"This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) -Finance Code 001"

## REFERENCES

[1] Ruggero, S. M., dos Santos, N. A., Sacomano, J. B., & da Silva, M. T. (2019). Investments in the Automotive Sector and Industry 4.0. Brazilian Case. In IFIP International Conference on Advances in Production Management Systems (pp. 650-657). Springer, Cham.

[2] Hexmoor, Henry. (2013). "Essential Principles for Autonomous Robotics." Synthesis Lectures on Artificial Intelligence and Machine Learning 7(2): 1–155. https://doi.org/ 10.2200/S00506ED1V01Y201305AIM021.

[3] Abe, J. M., Akama, S., & Nakamatsu, K. (2015). Introduction to annotated logics: foundations for paracomplete and paraconsistent reasoning Vol. 88. Springer.

[4] Torres, C. R., Abe, J. M., Lambert-Torres, G., da Silva Filho, J. I., & Martins, H. G. (2009). Autonomous mobile robot emmy iii. In New Advances in Intelligent Decision Technologies (pp. 317-327). Springer, Berlin, Heidelberg.

[5] Carvalho, Fábio R.; Abe, Jair M. (2018). A Paraconsistent Decision-Making Method, Smart Innovation, Systems and Technologies Vol. 87, Springer International Publishing. https://doi.org/10.1007/978-3-319-74110-9, Library of Congress Control Number: 2018933003.

[6] Bayindir, Ramazan, Ersan Kabalci, Orhan Kaplan, e Yunus Emre Oz. (2012). "Microcontroller based electrical machines training set." In 2012 15th International Power Electronics and Motion Control Conference (EPE/PEMC), Novi Sad, Serbia: IEEE, S3e.12-1-DS3e.12-4. https://doi.org/ 10.1109/EPEPEMC.2012.6397366.

[7] Miatliuk, K. (2015). Conceptual model in the formal basis of hierarchical systems for mechatronic design. Cybernetics and Systems 46 (8), 666-680.

[8] Miatliuk, K. (2017). Conceptual Design of Mechatronic Systems. WPB, Bialystok, available online: https://pb.edu.pl/oficyna-wydawnicza/wp-content/uploads/sites/4/2018/02/Miatluk_publikacja.pdf

[9] Miatliuk, K., Kim ,Y.H., Kim, K., Siemieniako, F. (2010). Use of hierarchical system technology in mechatronic design. Mechatronics 20 (2), 335-339.

[10] Miatluk, K., Nawrocka, A., Holewa, K., Moulianitis, V. (2020) Conceptual design of BCI for mobile robot control. Applied Sciences 10 (7), 2557.

[11] Abe, J. M. (2010). Paraconsistent logics and applications. In 4th International Workshop on Soft Computing Applications. p. 11-18, IEEE.

[12] Abe, J. M. (2015). Paraconsistent intelligent-based systems: New trends in the applications of paraconsistency. (Eds.), Vol. 94. Springer.

## AUTHORS

**Flavio Amadeu Bernardini** holds a degree in Mathematics - Integrated College of Science as Human, Health and Education of Guarulhos (2007). Specialization course in Industrial Automation - National Industry Service (SENAI) College of Tecnologia Mechatronics (2015). He is currently an instructor of professional practices in the area of Electronics - SENAI - Regional Department of Sao Paulo. He has experience in electronic maintenance. Studying with a Master's program in Production Engineering - Paulista University – UNIP.

**Márcia Terra da Silva** is Full Professor at the Post Graduation Program of Production Engineering of Universidade Paulista (PPGEP-UNIP). In the last 20 years, she has been developing research in the area of Service Operations Management, with focus on professional services as healthcare and educational services, and has published several articles in Brazilian and international journals. Currently, she researches High Education organization and management, with leading interest in workers' qualification to the demands of the Industry 4.0.

**Jair Minoro Abe** received B.A. and MSc in Pure Mathematics **-** University of Sao Paulo, Brazil. Also received the Doctor Degree and Free-Teacher title from the same University. He is currently coordinator of Logic Area of Institute of Advanced Studies - University of Sao Paulo Brazil and Full Professor at Paulista University - Brazil. His research interest topics include Paraconsistent Annotated Logics and AI, ANN in Biomedicine and Automation, among others. He is Senior Member of IEEE.

**Luiz Antonio de Lima** is Doctor of Science student in Production Engineering Paulista University, Master degree in Production Engineering in the area of Artificial Intelligence Applied to Software Paraconsistent Measurement Software, Post-Undergraduate Degree in EAD, University Professor, General Coordinator of IT Course and Campus Assistant: (2008-2009). University Professor: 02 Postgraduate Course and 12 Higher Courses in 43 disciplines; Speaker and Event Organizer: SENAED; NETLOG; Wics; WINFORMA.

**Kanstantsin Miatliuk** is a Professor of Bialystok University of Technology, Poland. He received his M.Sc. degree in Robotics from the Belarusian State Technical University, Minsk, Belarus (1988), his Ph.D. degree in Automation and Robotics from AGH University of Science and Technology, Krakow, Poland (2006) and his D.Sc. degree in Machines Constructing in Warsaw, Poland (2019). He is IEEE member. K.Miatliuk was a visiting professor in Kyung Hee University, Korea (2008, 2010) University of Southern Denmark, Denmark (2012) and University of Las Palmas Gran Canaria, Spain (2016). K.Miatliuk participated in numerous EU R&D projects. His research interests include mechatronics and robotics, systems science and computer science.

# NON-NEGATIVE MATRIX FACTORIZATION OF STORY WATCHING TIME OF TOURISTS FOR BEST SIGHTSEEING SPOT AND PREFERENCE

Motoki Seguchi[1], Fumiko Harada[2] and Hiromitsu Shimakawa[1]

[1]College of Information Science and Engineering,
Ritsumeikan University, Kusatsu, Shiga, Japan
[2]Connect Dot Ltd., Kyoto, Japan

## *ABSTRACT*

*In this research, we propose a method of recommending the best sightseeing spot through watching stories of sightseeing spots. It predicts the rating for each sightseeing spot of a target tourist based on Non-negative Matrix Factorization on the story watching times and ratings of tourists. We also propose to estimate the degree of the target tourist's preference for a sightseeing spot. Tourists visit a sightseeing spot for a certain purpose of tourism. The preferences of tourists appear prominently in their purposes of tourism. In addition, the degree of the tourists' preferences for sightseeing spots differs depending on the sightseeing spot. If we can estimate the degree of preference of a tourist, it will be possible to recommend a sightseeing spot that can achieve his purpose of tourism.*

## *KEYWORDS*

*Sightseeing, Recommendation, Interest Estimation, Story Watching, Preference.*

## 1. INTRODUCTION

In recent years, the development of the global tourism industry has been remarkable [1]. One of the backgrounds for the development of the tourism industry is the enhancement of tourist information. Tourist information is transmitted not only by organizations such as companies and mass media but also by various people through social networking service (SNS). The amount of tourist information is increasing every day. With the spread of smartphones, anyone can easily obtain tourist information. However, it is difficult to find the best sightseeing spot for oneself from the huge amount of tourist information.

To solve such the problem, various research on recommendation methods of the best sightseeing spots have been developed [2], [3], [4], [5], [10], [13], [14]. However, these researches aim to improve the accuracy of recommendation, and they did not discuss the effort tourists spend to select a sightseeing spot. There are also many tourist guide apps that help tourists select sightseeing spots. However, they also provide a wide variety of information, which complicates tourists' select of sightseeing spots. Tourists want a method that allows them to select a sightseeing spot that suits their wishes with little effort.

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 177-192, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101116

Tourists visit sightseeing spots for a certain purpose. The purpose of tourism of a tourist is always explicit, but variable. Therefore, it is necessary to recommend sightseeing spots that suit the purpose of tourism at that time. The preferences of tourists appear prominently in the purposes of tourism. In addition, the degree of the tourists' preferences for sightseeing spots differs depending on the sightseeing spot. If we can estimate the degree of preference of a tourist, we will be able to recommend sightseeing spots that can achieve the tourist's purpose of tourism.

It is possible to use storytelling marketing for recommending sightseeing spots [5]. The storytelling marketing can arouse tourists' empathy by using stories from a third-party perspective on sightseeing spots. Storytelling marketing is useful because it allows tourists to receive detailed information about attractions of sightseeing spots [7]. In addition, reviews of various sightseeing spots can be used as stories [11]. Therefore, many posts about sightseeing spots written on SNS can be used as stories.

In this research, we propose a method of recommending the best sightseeing spot by using non-negative matrix factorization on the story watching times of tourists. This method estimates the best sightseeing spot from the similarity of the behavior regarding the watching of the story between the target tourist who will visit a sightseeing spot and the tourists who have visited the sightseeing spot in the past. In this study, the only task to be given to the target tourists to recommend sightseeing spots is to watch the stories. Tourists can use this method as if they were enjoying surfing the internet on SNSs, and the load on the tourists is small. In this study, we propose the methods not only to recommend the best sightseeing spot, but also to estimate the degree of the tourists' preferences for each sightseeing spot. If we can estimate the degree of a tourists' preference for a sightseeing spot, we will be able to recommend the other sightseeing spots which brings him the attractiveness same to that of the sightseeing spot according to his preference.

In this paper, Section 2 describes the current state of tourism. Section 3 describes related work. Section 4 proposes the method of estimating the best sightseeing spot and degree of tourist preference for sightseeing spots. Section 5 describes the experiments that evaluate the proposed method and the results. Section 6 describes the evaluation of experimental results and their consideration.

## 2. CURRENT STATE OF TOURISM

### 2.1. Purpose of Tourism

Tourists visit sightseeing spots for a purpose of tourism, such as "I want to see a beautiful scenery" and "I want to heal fatigue." Van Harssel's research [6] classified tourism into the following 10 types, based on the purposes of tourism that tourists mainly aim at.

1) Nature trip
2) Cultural trip
3) Social trip
4) Activity trip
5) Recreational trip
6) Sports trip
7) Special trip
8) Religious trip
9) Health trip
10) Ethnic trip

For example, the nature trip applies to the case that a tourist wants to see a beautiful scenery. The Health trip applies to the case that he wants to heal fatigue. Naturally, tourists want to achieve their purposes of tourism. Therefore, tourists should select the tourism type according to the purpose of tourism. In addition, any sightseeing spot has a suitable tourism type. Therefore, tourists cannot achieve the purpose of tourism unless they select a sightseeing spot that matches the tourism type corresponding to the purpose of tourism.

On the other hand, people have a preference that is a desire to be satisfied in daily life. One person's preference is "eating" and another person's preference is "seeing." When preferences are not satisfied, people induce extraordinary behavior to satisfy them. It is considered that one of the behaviors is that people are going to travel [10]. Therefore, tourists' preferences are prominent for purposes of tourism. For example, the tourism purpose of "healing fatigue" would strongly include the preference of "healing." If we can grasp the potential preference and its degree of each tourist that he does not want to say, we can consider that the preference is not satisfied in recent daily life. Therefore, if we can recommend the tourist a sightseeing spot where he will be able to achieve satisfaction of the preference as the purpose of tourism, he can be pleased much with the tourism.

## 2.2. Attractions Associated with Sightseeing Spots

Sightseeing spots have various attractions. "Nature," "environment," "facility," and "events" are often cited as attractive factors of sightseeing spots. Mill [9] roughly divided the attractions of sightseeing spots into the following 8 categories.

1) Sun, sea and resort
2) Landscape
3) Animal
4) Hot springs and health resorts
5) Urban attractive conditions
6) Local attractive conditions
7) Sports event
8) Systematically developed attractive conditions

Hudman and Hawkins [8] divided the attractive factors in the 8 kinds of attractions into the following twelve categories.

1) Buildings and their environment
2) Cultural activities
3) Religion
4) Politics
5) Science
6) Nature
7) Climate
8) Scenery
9) Outdoor life
10) Outdoor recreation and sports
11) Entertainment
12) Health and hot springs

There is a correspondence between the preferences of tourists and the attractions associated with sightseeing spots. For example, tourists who have a strong preference to "seeing" strongly like sightseeing spots that have a strong attraction of "scenery."

Sightseeing spots provide various attractions in various degrees. On the other hand, how to feel the attraction is different for each person. The preferences that tourists have for sightseeing spots differ depending on how they feel the attraction. For example, a person who feels that a national park is strongly associated with the attraction of "scenery" has a strong preference of "seeing." A person who feels that the park is strongly associated with the attraction of "entertainment" has a strong preference of "playing." It is possible to estimate the attractions of a sightseeing spot by estimating the tourists' preferences for the sightseeing spot because the attractions of sightseeing spots and the preferences of persons correspond. If we can estimate the attractions associated with the sightseeing spot for a person, we can recommend him another sightseeing spot that he will feel the same attractions as that sightseeing spot.

## 3. RELATED WORK

In recent years, the tourism industry has grown [1]. It can be said that the enrichment of tourist information has contributed significantly among many factors for growth. However, as tourist information continues to increase, it has become difficult for tourists to obtain suitable tourist information. Tourists have to select the most suitable one for themselves from the vast amount of tourist information. Therefore, it is difficult for tourists to find the best sightseeing spot for themselves. Therefore, in recent years, research that recommends the best sightseeing spot to tourists has become popular.

There are methods of recommending sightseeing spots using location information obtained by GPS [13], [14]. These methods recommend the best tourist plan to tourists by learning the movement history of the tourists in the sightseeing spots. However, radio waves from positioning satellites cannot be captured everywhere and lack stability. For example, there is a large error in the location information in undergrounds or in forests where radio waves are hard to connect. In some cases, there are even things that cannot be supplemented. Also, these studies do not consider tourists' preferences.

There is also sightseeing spots recommendation methods that use personal data of tourists such as time spent for sightseeing [2], [3], [4]. Because these methods allow users to input personal data such as budget into a device before sightseeing, personalized tourist information can be provided. However, tourists must decide in advance the time and budget to spend on sightseeing. These methods cannot be used when tourist time and budget are undecided, because personal data of tourists are insufficient. Moreover, some tourists may find it annoying to input their personal data into the device. Furthermore, these methods do not take into account the preferences of tourists.
No matter how wonderful the attraction of a sightseeing spot is, if the attraction of the sightseeing spot does not match the preference of a tourist, the sightseeing is worthless for him. It is necessary to consider a tourist's preference and propose a method of recommending a sightseeing spot that has an attraction that matches the preference.

# 4. A Recommendation Method of Sightseeing Spots Through Watching Stories

## 4.1. Recommendation of Sightseeing Spots and Estimation of Preferences Using Story Watching Time

In this research, it is assumed that the user watches stories of sightseeing spots of interest. Therefore, it is considered that the user's interest in sightseeing can be estimated by the length of the story watching time. In this research, we propose a method of recommending the best sightseeing spot to the target user by analyzing the target user's story watching time with the past user's data. In this research, the user who is going to visit sightseeing spots from now on is called the target user. The user who has visited a sightseeing spot in the past is called the past user. In this research, we consider recommending a sightseeing spot to the target user based on the similarity of watching behavior between past users and the target user. In this research, we consider only that the accuracy of recommendation to the sightseeing spots itself is improved, but also whether the degree of the user's preference for the sightseeing spots can be estimated. If the latter can be achieved, it is possible not only to recommend a target user a suitable sightseeing spot visited by past users, but to recommend another sightseeing spot having the same attraction as that of the suitable sightseeing spot according to the preference of the target user. The overall diagram of the proposed method is shown in Figure 1.

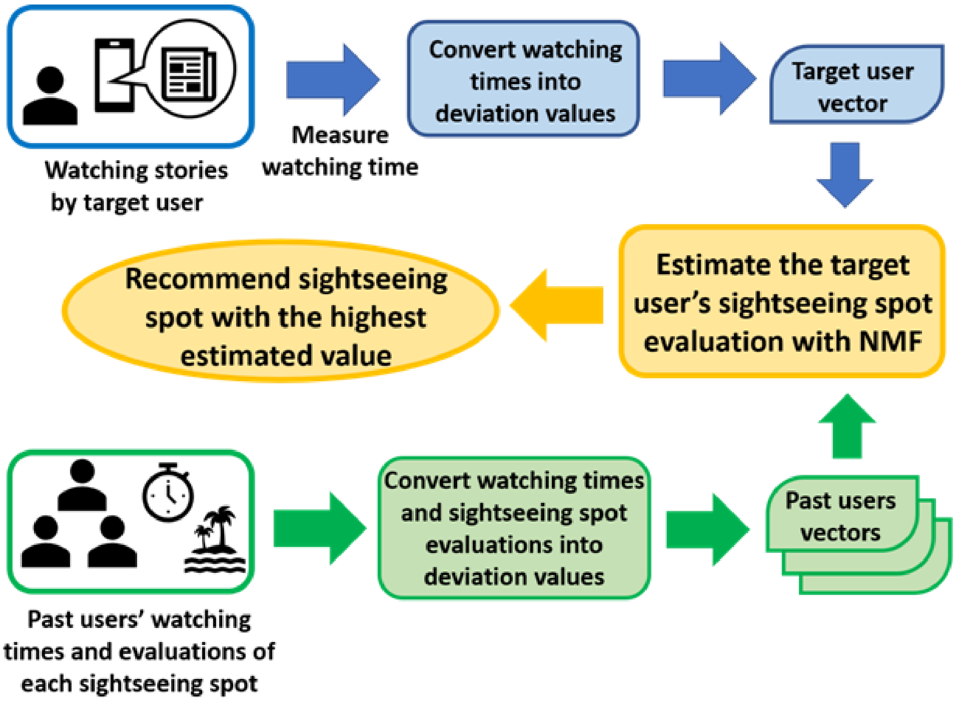

Watching stories by target user
Measure watching time
Convert watching times into deviation values
Target user vector
Estimate the target user's sightseeing spot evaluation with NMF
Recommend sightseeing spot with the highest estimated value
Past users' watching times and evaluations of each sightseeing spot
Convert watching times and sightseeing spot evaluations into deviation values
Past users vectors

Figure 1. Method outline

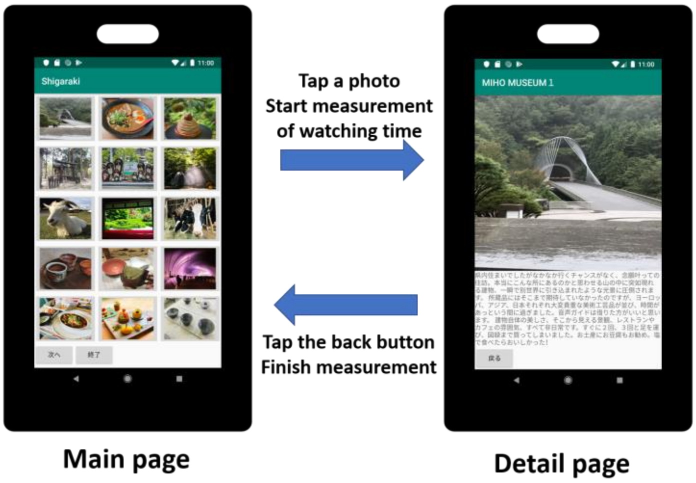
In order to provide the target user with basic data for recommending sightseeing spot, past users' data is firstly generated as follows. Past users freely select the interesting stories from the many prepared stories of the sightseeing spots and watch them for a short time. After that, the past users actually visit the sightseeing spots and evaluate each sightseeing spot. The story watching times and evaluations are recorded as the past users' data. On recommendation to the target user, the target user selects the interesting stories from the prepared stories and also watches them freely for a short time. The watching time of each story watched is recorded. In the proposed method, non-negative matrix factorization (NMF) [12] is used to estimate the evaluations of the target user for each sightseeing spot from the past user's data and the target user's story watching times. The sightseeing spot with the highest estimated evaluation value is recommended as the best sightseeing spot for the target user. By using NMF, we can recommend the best sightseeing spot only by imposing watching the stories on the target user. In this research, we also consider whether NMF can be used to estimate the degree of the user's preference for sightseeing spots.

With the proposed method, the only task for target users to recommend sightseeing spot is to watch interesting stories. Target user do not wear special sensors or answer questionnaires. Therefore, by using this method, it is possible to recommend a best sightseeing spot without imposing a heavy load on the target user.

## 4.2. Story Watching by Target User

In the proposed method, the target user watches the stories in order to enjoy recommendation of the best sightseeing spot. In this research, the stories about each sightseeing spot are collected from the posts about the sightseeing spot in SNSs such as Instagram. Each story consists of text-format experiences in the sightseeing spot and photographs of the sightseeing spot written by various people. Because there are many people who posted stories on SNSs, there are many different stories about a same sightseeing spot. Since there are many different stories, it is possible to grasp sightseeing spots from various viewpoints. Therefore, it is possible to prevent a biased view.

In this research, we provide the target user with many different stories. The target user can freely watch the stories on the smartphone. Figure 2 shows the screenshot in the story watching on a Smartphone.



県内住まいでしたがなかなか行くチャンスがなく、念願叶っての往訪。本当にこんな所にあるのかと思わせる山の中に突如現れる建物。一瞬で別世界に引き込まれたような光景に圧倒されます。所蔵品にはそこまで期待していなかったのですが、ヨーロッパ、アジア、日本それぞれ大変貴重な美術工芸品が並び、時間があっという間に過ぎました。音声ガイドは借りた方がいいと思います。建物自体の美しさ、そこから見える景観、レストランやカフェの雰囲気、すべて非日常です。すぐに2回、3回と足を運び、図録まで買ってしまいました。お土産にお豆腐もお勧め。塩で食べたらおいしかった！

Figure 2. The screenshot at watching stories on a smartphone

A story is made up of a pair of text-format of experience and photograph. First, only multiple photos are displayed on the screen. Each photo is associated with a particular story. This state is the main page (see the left part in Figure 2). The target user can tap a photo of interest on the main page. On tap a specific photo on the main page, the screen moves to the story page associated with the photo. This page is the detail page (see the right part in Figure 2). On the detail page, both the enlarged photo of the tapped photo and the experience are displayed. On tap the back button on the detail page, the screen returns to the main page again.

The time from tapping a photo on the main page to tapping the back button on the corresponding detail page is the story watching time. On the main page, the photos of each story are displayed in small size and the target user cannot watch the experiences in each story, so it is not in a watching state.

The target user does not have to watch all stories. The target user can select and tap favorite ones from many stories. If he/she tapped once a story but it was a story he was not interested in, he can tap the back button immediately and select another story again on the main page. Therefore, the target user watches the stories that he/she is interested in for a long time, and immediately stops or does not watch the stories that he/she is not interested in. The target user can intuitively watch as many stories as they like. Therefore, this system can be used with the same feeling as if one normally enjoys surfing the internet on SNSs, and the load on the user is small.

## 4.3. Acquisition of Past User’s Data

In this research, past users evaluate each sightseeing spot. This method uses the target user’s story watching time and past user’s data. A past user’s data includes the scores of each sightseeing spot in addition to the story watching time. The past user scores two points, which are the evaluation of each sightseeing spot and the degree of each preference for the sightseeing spot. The evaluation of a sightseeing spot is expressed by a real number from 0 to 100. The closer to 0, the lower the evaluation, and the closer to 100, the higher the evaluation.

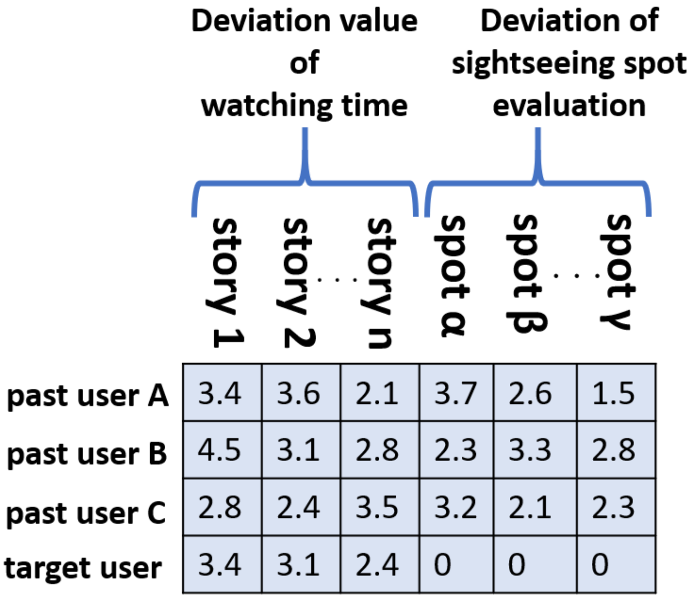
In this research, it is assumed that the target user feels seven preferences of "eating," "making," "playing," "seeing," "healing," "history," and "nature" for any sightseeing spot at individual degrees. The degree of each preference is evaluated on a scale of 5 from 1 to 5, with a degree closer to 1 being lower and a degree closer to 5 being higher. The stories given to the target user and the past users are the same. The past users select favorite stories from a large number of given stories and watch them.

## 4.4. Estimating Sightseeing Spot Evaluation by NMF

The proposed method uses the NMF to estimate the evaluation of each sightseeing spot from the target user’s story watching time and past user’s data.

NMF is an algorithm that decomposes a non-negative matrix X into two non-negative matrices W and H. At this time, the product of the decomposed matrices W and H is an approximation of the matrix X. NMF allows that some elements of X are missing (unknown values). The missing elements are replaced with 0 in advance. The unknown values are estimated when the matrix X is approximated by the product of the matrices W and H. NMF approximates matrix X with the product of matrices W and H. That is, the (i, j)-element of X is represented by the inner product of the i-th row vector of W and the j-th column vector of H. Each row vector of W and each column vector of H are used to compute multiple elements of X. NMF attempts to approximate all elements of X except the missing elements by the inner product of the row vectors of W and the column vectors of H. When the number of missing elements is small enough and the approximation of all non-missing elements is achieved, the inner product of the row vector of W and the column vector of H corresponding to each missing element can be calculated. NMF considers this inner product value to be an estimation of the missing value. That is, in NMF, missing values can be estimated by decomposing the matrix.

In the method, we consider a vector that summarizes the story watching times and the evaluations of sightseeing spots. The vector for the target user and the vector for each past user will be called the target user vector and the past user vector, respectively. The proposed method applies NMF to the matrix that combines the target user vector and the past user vector. An example of this matrix is shown in Figure 3.


Deviation value of watching time
Deviation of sightseeing spot evaluation

|  | story 1 | story 2 | story n | spot α | spot β | spot γ |
|---|---|---|---|---|---|---|
| past user A | 3.4 | 3.6 | 2.1 | 3.7 | 2.6 | 1.5 |
| past user B | 4.5 | 3.1 | 2.8 | 2.3 | 3.3 | 2.8 |
| past user C | 2.8 | 2.4 | 3.5 | 3.2 | 2.1 | 2.3 |
| target user | 3.4 | 3.1 | 2.4 | 0 | 0 | 0 |

Figure 3. Example of matrix for estimating evaluation

The target user has never visited the sightseeing spots and his evaluation of the spots are missing. Therefore, the proposed method estimates these missing evaluations through NMF.

The matrix for NMF is generated from the target and past vectors based on the following idea. The row vectors of the matrix are the past and target user vectors. The past vector of each past user is a (n + m) row vector for n stories and m sightseeing spots. The watching time of i-th story is given as the i-th element of the past user vector. The evaluation of j-th sightseeing spot is given as the (n+j)-th element of the past user vector. The target user vector is generated in the similar manner.

However, NMF does not recommend treating values with different units such as watching time and evaluation as element values in one matrix. Moreover, in NMF, the accuracy of estimation increases when the variations in the values are similar. Evaluation of sightseeing spots is subjective. Among some sightseeing spots, some users evaluate them with a large variation while some other users evaluate them with a small variation. In order to improve the estimation accuracy, it is better to make the size of the evaluation variations uniform. The variation among watching times should be uniformed. Therefore, the variations in the watching times and in the evaluation of sightseeing spots are adjusted based on the deviation values within the corresponding user. In the proposed method, standardization is performed so that the evaluation value takes a value from 0 to 5 by taking the deviation value with the average being 2.5. The watching times are replaced in the similar manner.

In the matrix generated in the above manner, the elements corresponding to the evaluations of sightseeing spots in the target user vector is missing. Through NMF on this matrix, these missing elements can be estimated as a value represented by the deviation value.

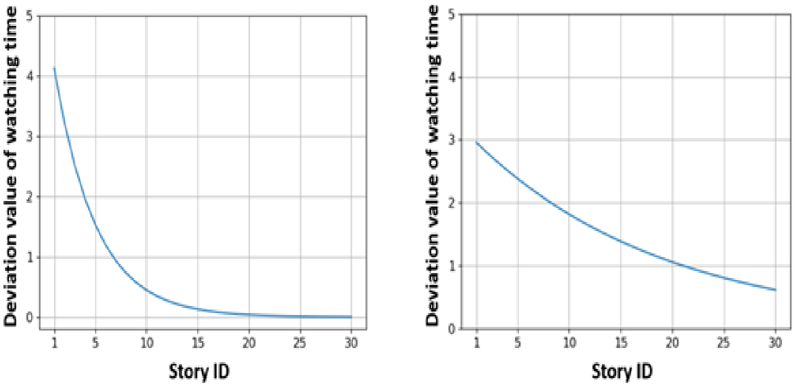
## 4.5. Watching Time for Unwatched Stories

NMF can estimates missing values in the matrix. In the proposed method, the matrix for NMF includes missing elements as the missing data. A past or the target user do not watch all stories, where the corresponding watching times become missing. There are the cases where the users do not watch a story because they are not interested in it, and where they cannot watch it because of spending time for other stories even though they are interested in it. We consider, if they avoided watching a story because they are not interested, to set the corresponding watching times close to 0. We also consider, if they are interested but cannot watch, to set the corresponding watching times as a positive-valued watching time.

If NMF is used, it is possible to estimate these deviation values by setting the story watching time that the target user and the past users have not watched as a missing value. However, the purpose of the proposed method is to estimate the evaluation of the sightseeing spots scored by the target user. Estimating the deviation values of the watching times of stories that has not been watched does not fit for the purpose. Therefore, we calculate the deviation values of the story watching times that the target user and the past users did not watch are found by regressing within each of the user in advance. The deviation value of the watching time of a story that is not watched by a user is calculated from the decay curve expressed by the following equation.

$$y=S(1/a)^{n-1} \qquad (1)$$

y is the deviation value of the watching time and is the objective variable of the regression. n is the explanatory variable of the regression equation, and it is the story ID given based on the length of watching time. In other words, IDs are assigned in the descending order of the length of the watching time. Obviously, the ID of the story that watched in the longest is 1, and the ID of the story that watched the second longest is 2. S is the deviation value of the story watching time that was watched in the longest. a is a regression coefficient and represents the degree of attenuation. The degree of attenuation a here is the degree to which the story watching time is reduced. The degree of attenuation varies from person to person. Examples of the decay curves are shown in Figure 4.



[a]User with largely different watching times [b] User with almost consistent watching times

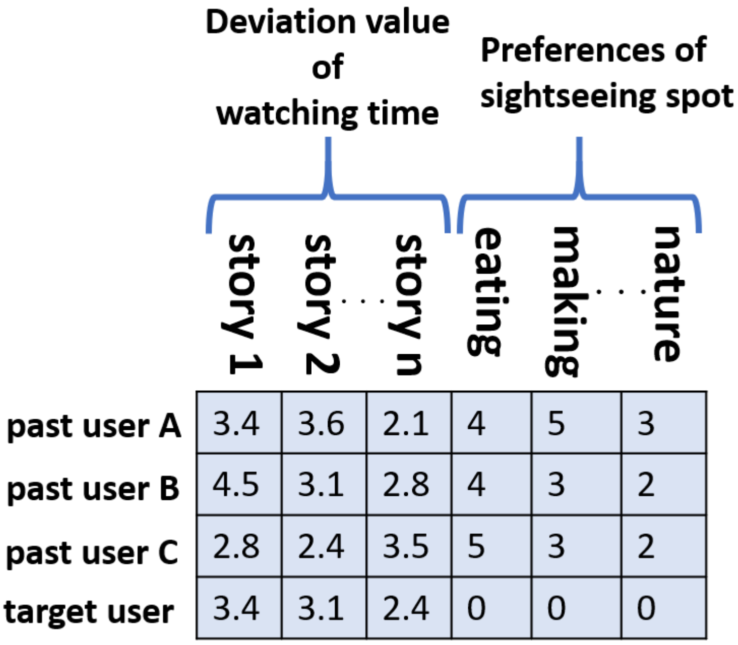
Figure 4. Difference in decay curves depending on user

Figure 4 [a] is an example of a user with significant variations in the story watching times. Since the watching time varies significantly among the stories, the degree of attenuation large. On the other hand, Figure 4 [b] is an example for a user who does not have much variation in watching times for stories. Since the watching time does not vary so much, the degree of attenuation is small. The degree of attenuation of the watching time of each user is found by regressing a for each user. By substituting the total number of stories for n, the deviation value of the virtual watching time of the story that is not watched is obtained. Substituting the total number of stories for n gives the watching time as a very small positive value. If this value is set in the matrix shown in Figure 3 and applied to NMF, this is not considered as a missing value.

## 4.6. Estimating User's Preference for Sightseeing Spots

The proposed method also estimates the degree of each preference that the target user has for each sightseeing spot. As mentioned in Subsection 2.2, there are roughly 10 types of tourism types according to the purpose of tourism. The target user's purpose of tourism is always explicit, but variable. Therefore, it is necessary to recommend sightseeing spots according to the occasion. The preferences that are emphasized differ depending on the purpose of tourism. If we can grasp the degree of each preference that the target user has for sightseeing spots, we can recommend a sightseeing spot that suits the target user's purpose of tourism.

In this research, it is assumed that the target user has seven preferences of "eating," "making," "playing," "seeing," "healing," "history," and "nature" for any sightseeing spot. There are individual differences in the degree of each preference that target user has for sightseeing spots. Therefore, unlike Subsection 4.4, we consider the target user vector that combines the watching times and the degrees of preference for sightseeing spots. Then, the matrix that combines the target user vector and the past user vectors is decomposed by NMF. The degree of each preference that the target user has for the sightseeing spot is estimated. Figure 5 shows an example of the matrix that combines target user vector and past user vectors for estimating preferences.


Deviation value of watching time
Preferences of sightseeing spot
. . .
. . .

| | story 1 | story 2 | story n | eating | making | nature |
|---|---|---|---|---|---|---|
| past user A | 3.4 | 3.6 | 2.1 | 4 | 5 | 3 |
| past user B | 4.5 | 3.1 | 2.8 | 4 | 3 | 2 |
| past user C | 2.8 | 2.4 | 3.5 | 5 | 3 | 2 |
| target user | 3.4 | 3.1 | 2.4 | 0 | 0 | 0 |

Figure 5. Example of matrix for estimating preferences

In this study, the story watching times of the target user and the past users are replaced by the deviation values while the degree of each preference are not replaced by a deviation value. This is because the past users have evaluated the degree of each preference they have for the sightseeing spot they visited in five levels. The evaluation of each sightseeing spot is scored on a 100-point scale while the degree of preference is evaluated on a scale of five. Thus, the variations in preference degree values are uniform. By using NMF, we can estimate the degree of each preference that the target user has for the sightseeing spot simply by watching the story. If we can estimate the degree of each preference that the target user has for the sightseeing spot, we can recommend the sightseeing spot that suits the target user's purpose of tourism.

## 5. EXPERIMENT

### 5.1. Outline of Experiment

We conducted an experiment to verify the usefulness of the proposed method. The following two were verified by the experiment.

- The accuracy of recommending sightseeing spot to target users
- The accuracy of estimating target users' preferences for sightseeing spots

In this experiment, we used three actual sightseeing spots: a museum, a restaurant and a pottery hall in Shigaraki, Shiga Prefecture in Japan. The subjects were nine men and two women in their twenties. The total number is 11. The 11 subjects have never visited the three sightseeing spots. We prepared 30 stories about Shigaraki. We got the stories from Instagram and personal blogs. Each story was chosen to have the same amount of text. This is to prevent a difference in watching time due to a difference in the amount of text.

The experiment was conducted according to the following procedure.

1. Each of the 11 subjects selected interesting stories from the prepared stories within the 3-minute time limit and freely watched the stories on his smartphone. Each story watching time was automatically recorded by the smartphone.
2. The 11 subjects visited the three sightseeing spots. Each of them gave an evaluation to each of the sightseeing spot out of 100. He/she also gave the degree of each of the seven preferences of each of the sightseeing spots in a 5-point scale.
3. We calculated the estimation accuracy of the proposed method through the leave-one-out cross-validation by using one of the subjects as the target user and the remaining 10 subjects as the past users. By applying NMF, we estimated the evaluations of the sightseeing spots scored by the target user. Only the story watching times were used as the target user data. The evaluations for the sightseeing spots were treated as missing values.
4. We calculated the correlation coefficient between the estimated evaluation values of the sightseeing spots and the actual evaluations of the sightseeing spot actually given by each of the target user.
5. We similarly estimated the degrees of the 7 preferences of the target user and calculated the correlation coefficient between the estimated degrees and the actual degrees given by the target user, as Steps 3 and 4.

### 5.2. Sightseeing Spot Recommendation Accuracy

In order to verify the recommendation accuracy of sightseeing spot using the proposed method, we investigated the correlation between the sightseeing spots evaluations actually scored by the subjects and the estimated sightseeing spots evaluations. In this study, the evaluation of each sightseeing spot actually scored by the subject is called the actual evaluation value, and the evaluation of the sightseeing spot estimated using NMF is called the estimated evaluation value. In this experiment, the 11 subjects evaluated the 3 sightseeing spots. There are 11 × 3 = 33 actual evaluation values and estimated evaluation values. We calculated the correlation coefficient and the rank correlation coefficient of these 33 actual evaluation values and estimated evaluation values. The rank correlation coefficient is the correlation coefficient obtained by converting each variable into ranks. In this study, we calculated the correlation coefficient by converting the actual evaluation values and estimated evaluation values of each user into ranks. In addition, a p-value was also obtained by performing a test for no correlation with the significance level 1%. The null hypothesis here is "no correlation." Table 1 shows the correlation coefficient and rank correlation coefficient between the actual evaluation value and the estimated evaluation value. Figure 6 shows the correlation diagram between the actual evaluation value and the estimated evaluation value. The horizontal axis of the figure is the actual evaluation value. The vertical axis is the estimated evaluation value. and the straight line is the regression line.

Table 1. Correlation coefficient and rank correlation coefficient between actual evaluation value and estimated evaluation value and their p-values

| | |
|---|---|
| correlation coefficient | 0.647 |
| p-value of correlation coefficient | $4.64 \times 10^{-5}$ |
| rank correlation coefficient | 0.820 |
| p-value of rank correlation coefficient | $5.07 \times 10^{-9}$ |

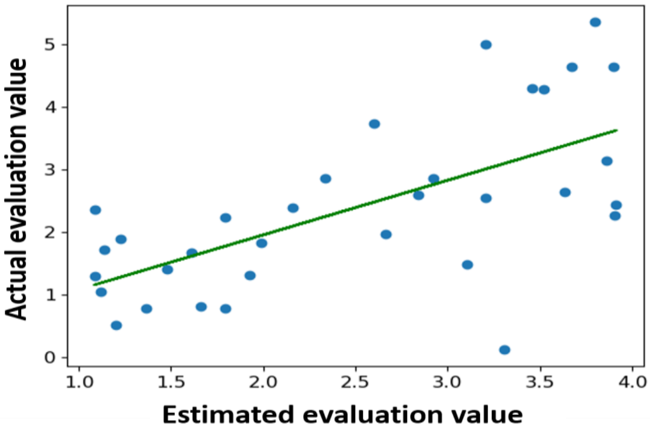


Figure 6. Correlation diagram of actual evaluation value and estimated evaluation value

From Table 1, the correlation coefficient was 0.647 and the rank correlation coefficient was 0.820, both of which show a significant strong positive correlation. The p-value is smaller than the significance level in all cases. In addition, we succeeded in estimating the evaluation ranking of 8 out of 11 subjects. It can be said that the recommendation accuracy is high because a strong positive correlation was found in the correlation between the actual evaluation value and the estimated evaluation value.

## 5.3. Accuracy of Estimating Tourists' Preferences for Sightseeing Spots

We verified the estimation accuracy of each preference that the target user has for each sightseeing spot. We investigated the correlation between the degree of each preference that the target user actually scored to each of the sightseeing spot and the degree of each estimated preference. In this study, the degree of each preference that the target user actually has for the sightseeing spot is called the actual preference value, and the degree of each preference for the sightseeing spot estimated using NMF is called the estimated preference value.

In this experiment, since the 11 subjects gave 7 preference values for each of the 3 sightseeing spots, there are 11 × 7 × 3 = 231 actual preference values and estimated preference values. As described in Subsection 5.2, we calculated the correlation coefficient and rank correlation coefficient among these 231 actual preference values and estimated preference values. We confirmed the relationship between them. In addition, the p-value was also obtained by performing a test for no correlation with the significance level 1%. The null hypothesis is "no correlation." Table 2 shows the correlation coefficient and rank correlation coefficient between the actual preference values and the estimated preference values. Figure 7 shows the correlation diagram between the actual preference values and the estimated preference values. The horizontal axis of the figure is the actual preference value. The vertical axis is the estimated preference value. The straight line is the regression line.

Table 2. The correlation coefficient and rank correlation coefficient between the actual preference values and estimated preference values and their p-values

| | |
|---|---|
| correlation coefficient | 0.485 |
| p-value of correlation coefficient | $5.06 \times 10^{-15}$ |
| rank correlation coefficient | 0.576 |
| p-value of rank correlation coefficient | $8.99 \times 10^{-22}$ |

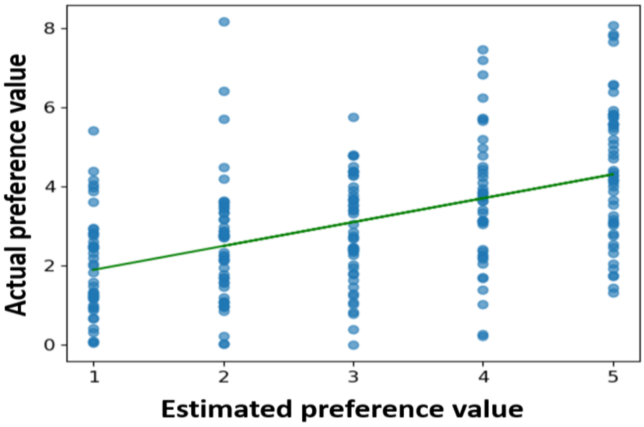


Figure 7. The correlation between actual preference values and estimated preference values

From Table 2, the correlation coefficient was 0.485 and the rank correlation coefficient was 0.576, both of which show a significant weak positive correlation. The p-value is smaller than the significance level in all cases. It was found that there is a generally positive correlation between the actual preference values and the estimated preference values. However, it cannot be said that there is a strong correlation in either result. Therefore, it can be said that the preference can be estimated to some extent, but the accuracy is lower than that of estimating the evaluation of sightseeing spots.

## 6. DISCUSSION

In this research, we proposed a method of recommending the best sightseeing spot to the target user by using the story watching time of the target user and the data of past users. In this study, we did not only recommend the best sightseeing spot but also estimated the degree of the tourists' preference for each sightseeing spot. Here, we consider these estimating accuracies.

### 6.1. Usefulness of Story Watching Time in Recommending Sightseeing Spots

Regarding the recommendation of the best sightseeing spot, a significant strong positive correlation was found between the actual evaluation value and the estimated evaluation value. In addition, the accuracy of recommending sightseeing spots can be said to be high because we succeeded in estimating the evaluation ranking of 8 out of 11 subjects. In the future, the following point can be considered to further improve the recommendation accuracy.

In the experiments in this paper, the time during which the subjects can watch the story was fixed at 3 minutes. Some users may find this 3 minutes long, while others may find it short. A subject who found 3 minutes long may have watched all the stories that he/she was interested in and then have watched the stories that he/she was not interested in until the 3 minutes have passed. A subject who found the 3 minutes short may have spent 3 minutes before watching all the stories of interest. Thus, it cannot be said that the story watching time reflects the user's interest entirely. It was necessary to allow the subjects to finish watching the stories at any time without limiting the watching time of the stories to 3 minutes. By not limiting the watching time, it is considered that the user's interest appears significantly in the watching time of each story.

We used only three sightseeing spots in the experiment: a museum, a restaurant, and a pottery hall. The correlation coefficient and rank correlation coefficient may take higher values due to the small number of sightseeing spots. In the future, it is necessary to increase the number of sightseeing spots and confirm the accuracy.

### 6.2. Estimating Preferences of Tourists for Sightseeing Spots

Regarding the estimation of the preference of the subjects to sightseeing spots, a weak positive correlation was found between the actual preference values and the estimated preference values. However, the estimation accuracy was lower than the recommendation of the best sightseeing spot. This is because the number of the sightseeing spots was three while it was necessary to estimate the seven values in terms of preference. The greater the number of values to be estimated, the harder it is to estimate all of them because the known information available for estimating is limited.

In the proposed method, NMF is used to estimate the user's preference for sightseeing spots. In NMF, the accuracy of estimation increases as the number of data increases. Therefore, it is considered that the accuracy of preference estimation becomes higher by increasing the number of past users' data.

In this research, assuming that the recommended sightseeing spot is Shigaraki, the seven preferences that target users have for each sightseeing spot are "eating," "making," "playing," "seeing," "healing," "history," and "nature." This is based on the website of the Shigaraki Tourism Association [15] and is considered to be specialized in sightseeing spots of Shigaraki. In the future, it is necessary to verify whether these seven preferences are appropriate for the target user for each sightseeing spot in Shigaraki.

Also, all p-values were quite small. This is probably because the number of subjects was 11. The number of subjects should be increased in the future. In addition, the ages of the subjects in this study were all in their 20s, which was quite biased. In the future, subjects of various ages should be recruited.

## 7. CONCLUSION

In this research, we proposed a method of recommending the best sightseeing spot through watching videos of stories of sightseeing spots. This method recommends best sightseeing spot based on the similarity of the behavior regarding the watching of stories of tourist who are going to visit the sightseeing spot and tourists who have visited the sightseeing spot in the past. Moreover, this method does not only recommend a sightseeing spot but also estimates the degree of tourists' preferences for sightseeing spots.

As the experimental result of verifying the usefulness of this method, it was suggested that the story watching time is useful for recommending the best sightseeing spot. It was also suggested that it is possible to estimate the preferences to some extent, although it is inferior to the recommendation of the best sightseeing spot.

In the future, we will increase the number of recommended sightseeing spots to confirm the recommendation accuracy. We will also increase the amount of data of past users to improve the accuracy of preference estimation.

## REFERENCES

[1] International Tourism Highlights, 2019 Edition. World Tourism Organization, Madrid, Spain, 2019.

[2] Ricardo Anacleto, Lino Figueiredo, Ana Almeida, and Paulo Novais Mobile application to provide personalized sightseeing tours. Journal of Network and Computer Applications, Vol. 41, pp. 56–64, 2014.

[3] Igo Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. Where shall we go today? planning touristic tours with tripbuilder. In Proceedings of the 22nd ACM international conference on Information and Knowledge Management, pp.757–762, 2013.

[4] Igo Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. Tripbuilder: A tool for recommending sightseeing tours. In European Conference on Information Retrieval, pp. 771–774. Springer, 2014.

[5] Fabio Clarizia, Saverio Lemma, Marco Lombardi, and Francesco Pascale. An ontological digital storytelling to enrich tourist destinations and attractions with a mobile tailored story. In International Conference on Green, Pervasive, and Cloud Computing, pp. 567–581. Springer, 2017.

[6] J van Harssel. Tourism: an exploration. No. Ed. 2. National Publishers of the Black Hills, Inc., 1986.

[7] Kuo-Lun Hsiao, Hsi-Peng Lu, and Wan-Chin Lan. The influence of the components of storytelling blogs on readers' travel intentions. Internet Research, 2013.

[8] Lloyd E Hudman, Donald E Hawkins. Tourism in contemporary society: an introductory text. Prentice-Hall, Inc., 1989.

[9] Mill, Robert Christie. Tourism: The international business. Prentice Hall Englewood Cliffs, NJ, 1990.

[10] Hiromitsu Shimakawa, Momoko Kato. Recommendation of tour route from tourist motivation improving serendipity occurrence. International Journal of Latest Research in Engineering and Technology, Vol. 3, No. 2, pp. 26–36, 2017.

[11] Rebecca Pera. Empowering the new traveller: storytelling as a cocreative behaviour in tourism. it Current Issues in Tourism, Vol. 20, No. 4, pp. 331–338, 2017.

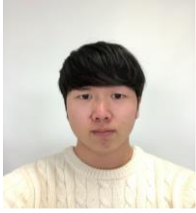
[12] Toby Segaran. Programming collective intelligence: building smart web 2.0 applications. "O'Reilly Media, Inc.", 2007.

[13] Hyoseok Yoon, Yu Zheng, Xing Xie, and Woontack Woo. Smart itinerary recommendation based on user-generated gps trajectories. In International Conference on Ubiquitous Intelligence and Computing, pp. 19–34. Springer, 2010.

[14] Hyoseok Yoon, Yu Zheng, Xing Xie, and Woontack Woo. Social itinerary recommendation from user-generated digital trails. Personal and Ubiquitous Computing, Vol. 16, No. 5, pp.469–484, 2012.

[15] Shigaraki Tourism Association, Hottosuru Shigaraki, https://www.e-shigaraki.org/

## AUTHORS



**Motoki Seguchi** received B.E degrees from Ritsumeikan University in 2020. He is currently pursuing the M.E degree with Ritsumeikan University. He is currently studying data engineering.

**Dr. Fumiko Harada** received B.E., M.E, and Ph.D degrees from Osaka University in 2003, 2004, and 2007, respectively. She joined Ritsumeikan University as an assistant professor in Ritsumeikan University in 2007, and is currently a counselor of Connectdot co.ltd. She engages in the research on real-time systems and data engineering. She is a member of IEEE.

**Prof. Hiromitsu Shimakawa** received Ph.D degree from Kyoto Univ. in 1999. Since 2002, He has worked in Ritsumeikan Univ. as a professor. His research interests include data engineering, usability, and integration of psychology with IT. He is a member of IEEE and ACM.

# DEEP REINFORCEMENT LEARNING FOR NAVIGATION IN CLUTTERED ENVIRONMENTS

Peter Regier Lukas Gesing Maren Bennewitz

Humanoid Robots Lab, University of Bonn, Bonn, Germany
{pregier,maren}@cs.uni-bonn.de

## ABSTRACT

*Collision-free motion is essential for mobile robots. Most approaches to collision-free and efficient navigation with wheeled robots require parameter tuning by experts to obtain good navigation behavior. In this paper, we aim at learning an optimal navigation policy by deep reinforcement learning to overcome this manual parameter tuning. Our approach uses proximal policy optimization to train the policy and achieve collision-free and goal-directed behavior. The output of the learned network are the robot's translational and angular velocities for the next time step. Our method combines path planning on a 2D grid with reinforcement learning and does not need any supervision. Our network is first trained in a simple environment and then transferred to scenarios of increasing complexity. We implemented our approach in C++ and Python for the Robot Operating System (ROS) and thoroughly tested it in several simulated as well as real-world experiments. The experiments illustrate that our trained policy can be applied to solve complex navigation tasks. Furthermore, we compare the performance of our learned controller to the popular dynamic window approach (DWA) of ROS. As the experimental results show, a robot controlled by our learned policy reaches the goal significantly faster compared to using the DWA by closely bypassing obstacles and thus saving time.*

## 1. INTRODUCTION

A prerequisite for nearly all mobile robot applications is collision-free navigation. Typical solutions apply a two-stage approach and use 2D path planning on a cost grid in combination with a low-level motion controller for path tracking and collision avoidance. The low-level controller hereby determines the motion commands for the current time step taking into account the global path and the current robot state as well as the local environment. Typical navigation systems require manual parameter tuning to achieve a good navigation behavior. This tuning requires a significant amount of time and profound knowledge about the navigation software, the robot hardware, as well as the environment conditions, and is a difficult task due to the trade off between time efficiency and safety.

In this paper, we present a self-learned navigation controller realizing collision avoidance and goal-directed behavior. Our approach combines grid-based planning with reinforcement learning (RL) and applies proximal policy optimization (PPO) [1] for the learning task. Our framework hereby uses a global planner to obtain a 2D path from the current robot pose to the global goal. The input of the network consists of the robot's translational and angular velocities, local goals determined from the global path, and a patch of the occupancy grid map containing the obstacles in the robot's vicinity. The outputs are the robot's velocity commands for the next time step. Fig. 1 shows an example situation and a visualization of our approach.

To the best of our knowledge, we present the first solution that integrates global path planning with deep RL to reach a global goal in the environment. Our framework thereby learns the appropriate distance to obstacles and performs regular recomputation of the global path. As a result, no parameter tuning of the navigation controller and inflation of the objects in the map is needed to find the best trade-off between completion time and safety distance to obstacles. When the global

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 193-204, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101117

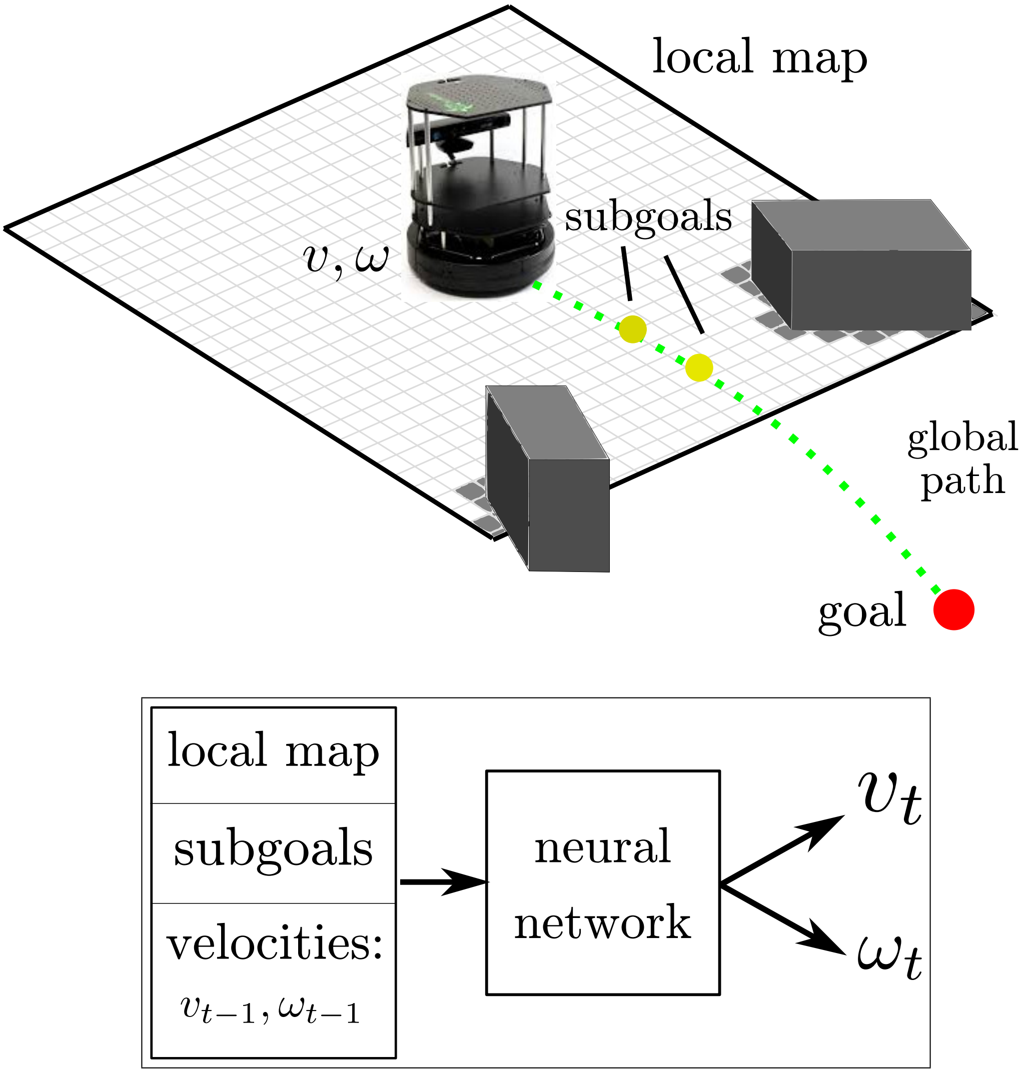


Figure 1: Visualization of our approach, which relies on a global path that is recalculated on a 2D occupancy grid every time step. We use subgoals on the path as part of our observation space and input to a neural network. Additionally, we use a local grid map centered around the robot and the current translational and angular velocities $v_{t-1}$ and $\omega_{t-1}$ as network input. We train the network to learn a navigation policy that outputs the velocity commands $v_t$ and $\omega_t$ for the next time step.

path leads through narrow passages, where collisions are likely to occur, our system learns to drive around those narrow regions.

We integrated our learned navigation policy as a collision avoidance module that can be used with the Robot Operating System (ROS) navigation stack [2]. We thoroughly evaluated our approach in simulation and in a real-world experiment and compared the performance with the dynamic window approach (DWA) [3], which is used in ROS and which is still one of the most popular navigation schemes. As the experimental results show, the robot steered by our learned policy reaches the goal significantly faster than using the DWA. Furthermore, we show that the policy, which is initially trained only on a simple environment, can be transferred to environments of increasing complexity.

## 2. Related Work

In the last few years, several learning approaches for mobile robot navigation have been presented. Sergeant *et al.* [4] proposed to use a state representation based on laser range data and learned translational and angular velocity commands for local obstacle avoidance. The authors trained an autoencoder neural network with human-controlled action commands. Pfeiffer *et al.* [5] presented an end-to-end navigation system learned for simple maps using 2D laser data as input, the velocity commands as output, and a 2D path as teacher. The authors later extended the work by applying subsequent RL training to the learned model [6], thus reducing the training time of RL and avoiding overfitting of the imitation model. Liu *et al.* [7] used a local occupancy map as state representation to learn a navigation policy using a variant of the value iteration networks. Tai *et al.* [8] proposed generative adversarial imitation learning to achieve socially complaint navigation. The authors use depth data to train the network and the social force model to generate a large set of training data. Pokle *et al.* [9] designed a local controller to determine the robot's velocity commands and predicting a local motion plan, while considering the trajectories of surrounding humans. These supervised learning methods all depend on the teacher, e.g., controls provided by humans, a global path planner, or a well-tuned optimization, while the goal of our work is to enable the robot to learn by itself while navigating in the environment.

Gupta *et al.* [10] investigated a mapping and planning navigation network based on visual data that encodes the robot's observations into a birds-eye view of the environment, which makes the method limited to known scenarios. Also the approach presented by Hsu *et al.* [11] was developed for known environments. A CNN processes image data and generates discrete actions to move the robot towards a global goal pose. In contrast to that, we use a binary occupancy grid map as representation, which makes the learned policy applicable to environments not seen in the training data.

Chen *et al.* [12] deployed also PPO for deep RL as we do. The authors rely on height-map observations as state representation for a wheel-legged robot. Due to a high-dimensional robot state, the authors discretize the action space and use a set of navigation behaviors to deal with obstacles of certain, given shapes.

Tai *et al.* [13] presented a method that utilizes the robot's velocities and target positions as state representation for an actor-critic RL approach. The authors developed a local controller relying on sparse laser-range measurements and trained a mapless motion planner. Fan *et al.* [14] proposed to use a set of subsequent laser scans and apply PPO to learn movement commands for navigation through crowds. Those approaches do not consider global path planning as they are designed for local navigation.

Chiang *et al.* [15] applied AutoRL to learn two different navigation behaviors, i.e., path following and driving to a global goal location. The authors do not combine learning with global path planning but use the global goal coordinate as input to the network. In our experiments, the robot got stuck in local minima while using only the global goal as input. Therefore, we use a subgoal on the regularly recomputed global path as additional input.

## 3. Problem Description

We consider a robot moving according to the unicycle model that has to reach a goal location by executing translational and angular velocities. A path planner computes the 2D path to the goal on a global grid map at every time step using the estimated robot pose from a localization system. The RL learning task is to determine the velocity commands for each time step to navigate collision-free and as fast as possible to the goal.

We model the problem as a partially observable Markov decision process (POMDP) defined as the tuple $(\mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$. Here, $s \in \mathcal{S}$ corresponds to the state of the environment including the robot. The state of the environment changes based on the robot's actions $a \in \mathcal{A}$, which are in our case the translational and angular velocity commands $(v, \omega)$, and according to the transition probability $\mathcal{T}(s'|s, a)$. The agent cannot determine the state $s$ but has to rely on observations $\mathcal{O}(o|s', a)$. After every state transition the robot receives a reward $\mathcal{R}(s, a)$.

The actor critic approaches approximate the value function (critic) to be able to update the policy (actor) itself. We use a deep neural network as non-linear function approximator to evaluate the state value function $V^{\pi}$, which determines the expected return for state $s$ when following the policy $\pi$. The goal of RL is to find a stochastic policy $\pi_{\theta}(a_t|o_t)$ that maximizes the expected reward

$$\max \mathbb{E}\left(\sum_{k=0}^{T} \gamma^k \mathcal{R}(s_k, a_k)\right),$$

where $\theta$ is the set of parameters that specify the function approximator, $T$ is the final time step, and $\gamma$ is the discount factor.

The critic network is updated based on the advantage value

$$A_t = A(o_t, a_t) = Q^{\pi_\theta}(o_t, a_t) - V^{\pi_\theta}(o_t). \tag{1}$$

where $Q^{\pi_\theta}(o_t, a_t) = r_t + \gamma V^{\pi_\theta}(o_t')$. Here $r_t$ is the immediate reward at time $t$ and $V^{\pi_\theta}(o_t')$ is the expected return for the observation $o_t'$. The actor network uses the policy gradient (PG) method to update the network weights $\theta$ in order to maximize

$$\max \mathbb{E}(\log \pi_\theta(a_t|o_t) A_t). \tag{2}$$

Proximal policy optimization (PPO) [1] substitutes the $\log \pi_\theta$ term for the policy probability ratio $\Psi = {}^{\pi_\theta}/_{\pi_{\theta_{old}}}$, to achieve stability. To avoid large policy updates that can impede and reset the training process, the probability ratio is constrained to the range of $[1-\epsilon, 1+\epsilon]$ via the *clip* function

$$\eta^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(\Psi \mathrm{A}_t, \mathrm{clip}\left(\Psi, 1-\epsilon, 1+\epsilon\right) \mathrm{A}_t\right)\right]. \tag{3}$$

## 4. Neural Network Approximator for Local Navigation

To learn the navigation strategy that takes into account the global path to the goal and the obstacles in the robot's vicinity, we train a deep neural network approximator that provides the robot's translational and angular velocities. The architecture of this network is described in the following.

### 4.1. Observation Space

The observation space consists of three components as described in the following. The first component is $o_v = (v_{t-1}, \omega_{t-1})$ with $v_{t-1}$ and $\omega_{t-1}$ as the robot's current translational and angular velocities computed at the previous time step. The second component is $o_m$, which corresponds to the $3\,m \times 3\,m$ patch of the 2D occupancy grid map around the robot (see Fig. 2). As resolution of the map we use $0.05\,\mathrm{m}$, thus the grid patch size has a dimension of $60 \times 60$ cells.

Additionally, we use a representation of local 2D subgoals in the observation. The subgoal at the current time step is calculated as the position on the global path that is $1m$ away from the robot and stored in map coordinates. At time step $t$, we transform the global coordinates of the subgoals stored at time steps $t-1$ and $t-5$ into the robot frame to get their relative positions, which serve as third observation component $o_g = (p^x_{-5}, p^y_{-5}, p^x_{-1}, p^y_{-1})$. The representation of the local goal $p^x_{-1}, p^y_{-1}$ indicates the robot's progress that was made towards the goal since the previous time step and is used for the reward calculation. By adding a second subgoal to the observation space $\mathcal{O}$, we noticed an improvement of the navigation policy and speed up of the training. As already noted by Kulhánek *et al.* [16], using information of previous observations helps the system to infer the real state $s \in \mathcal{S}$ of the environment. To summarize, an observation is defined as $o = (o_g, o_v, o_m)$.

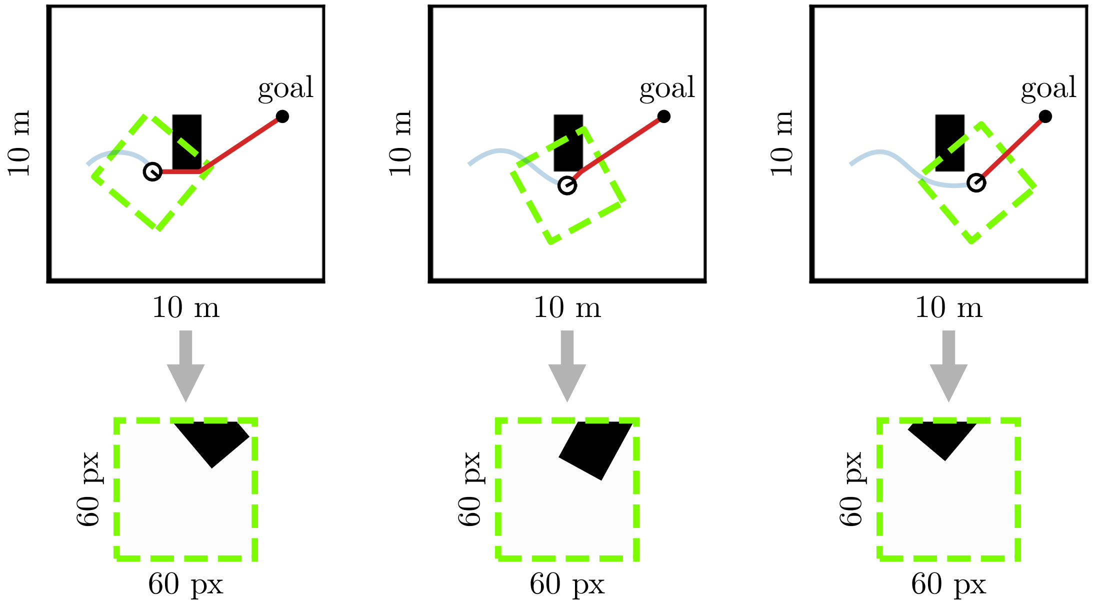


Figure 2: Binary image representation used as input to the network. A $3\,m \times 3\,m$ patch (dashed green) around the robot's pose is cropped from the global occupancy grid map. The robot is at the center of the resulting egocentric image and the viewing direction is to the right side. The global path (red) is computed with the A* search in a binary global map. Interpolated values in the cropped image resulting from the rotation are set to occupied as well as regions outside the boundaries of the global map.

### 4.2. Reward

Our reward function considers task completion, the duration, and the progress towards the goal

$$\mathcal{R}(s,a) = \mathcal{R}_{\mathit{fin}}(s,a) + \mathcal{R}_{\mathit{fix}} + \mathcal{R}_{\mathit{dist}}(s,a). \tag{4}$$

A navigation task ends if the robot arrives at the goal, a collision occurs, or a maximum number of time steps is reached. Accordingly, the reward $\mathcal{R}_{\mathit{fin}}(s,a)$ is defined as follows:

$$\mathcal{R}_{\mathit{fin}}(s,a) = \begin{cases} b & \text{if the goal was reached} \\ -c & \text{if a collision occurred} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$\mathcal{R}_{\mathit{fin}}(s,a)$ is a large positive value if the distance to the final goal is less than $0.3\,m$, a large negative value if the distance between the robot and the nearest obstacle is less than $0.3\,m$, meaning a collision is occurred, and zero otherwise.

$\mathcal{R}_{\mathit{fix}}$ is a fixed negative reward, that penalizes each action to force the robot to finish an episode as fast as possible.

To speed up the training, we use a third reward component

$$\mathcal{R}_{\mathit{dist}}(s,a) = \alpha \cdot \mathcal{D}_{(s,a)}, \tag{6}$$

where $\mathcal{D}(s,a)$ is the function computing the distance between the robot and subgoal $(p^x_{-1}, p^y_{-1})$ and $\alpha$ is a scaling factor.

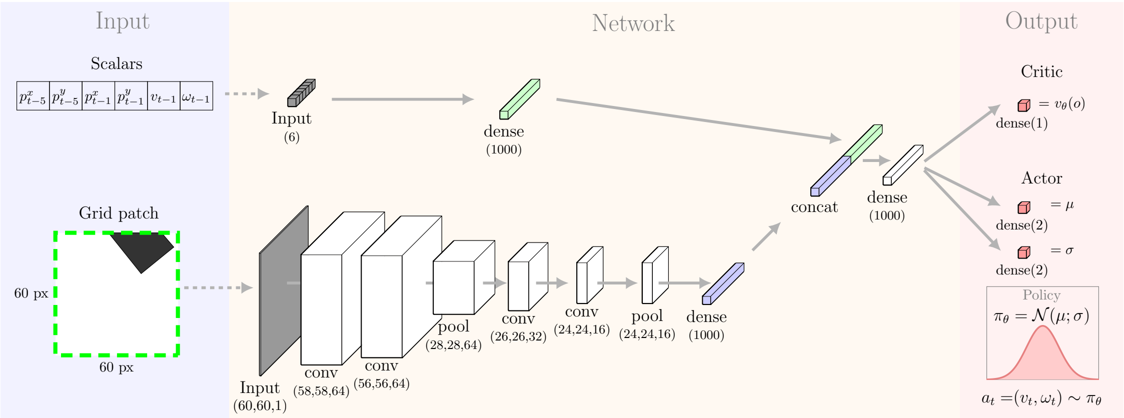


Figure 3: Network structure of the actor-critic scheme. The input consists of scalar values and the grid patch. The scalar values are fed into a single, fully connected dense layer. The binary image of the grid patch (see also Fig. 2) is handled by multiple CNN layers to distinguish obstacle configurations. Then, both branches are concatenated and assembled in a further dense layer. Finally, the critic value $v_\theta(o)$ corresponding to the value function estimator is computed by a last layer. The policy distribution $\pi$ is calculated by the mean and standard deviation of two normal distributions from which $v_t$ and $\omega_t$ are sampled.

## 4.3. Neural Network Structure

Our observation space as described in Sec. 4.1. is divided based on the representation of the data. Typically, the obstacle grid around the robot is represented as a binary image, while the rest of the observation space provides information about the different components of the robot state. Thus, we propose a network architecture that consist of two branches that split the observation space into scalar values and the binary grid patch (left part of Fig. 3). The scalar branch of the network is a single, fully connected neural network layer (green layer in the upper branch in Fig. 3) and encodes the subgoals and robot velocities into a high dimensional feature space to process them in the following layers.

The grid patch is processed by separate CNN layers (lower branch in Fig. 3), that are well suited for processing 2D data structure, e.g., images. The layers can identify 2D relationships between pixel values and encode obstacles in the robot's vicinity. Max-pooling layers after the first two CNNs reduce the shape and compress the information. This layered design is inspired by the network composition of the well-known VGGnetworks for image recognition [17]. The 3D output of the last max-pooling layer is flattened and reduced to a one-dimensional output with another dense layer (shown in blue). Then, we concatenate the outputs of both branches (blue and green) and process them together in an extra fully connected layer. Finally, we normalize the output, which is a standard technique [18].

The actor and critic estimators share the same connected layers. We found out that the parameter sharing between the actor and critic improves the learning speed because there are fewer parameters to learn. For the value function estimator $v_\theta(o_t)$, the shared network output is inserted into a last dense layer to get a single real number which represents the critic value. The final output of the actor network is the policy $\pi(a|o)$ modeled by the two Gaussian distributions $\mathcal{N}(\mu_{trans}; \sigma_{trans})$ and $\mathcal{N}(\mu_{ang}; \sigma_{ang})$. The two mean values are shrunk with an $tanh$ activation function. This scaling forces the values to stay between the desired velocity limits ($[0 : 0.7]$ m/s and $[-0.7 : 0.7]$ rad/s). The $\sigma$ values are the standard deviations of the normal distributions. We apply a sigmoid activation function scaled with 0.5 to guarantee that the bandwidths of the normal distributions do not massively grow.

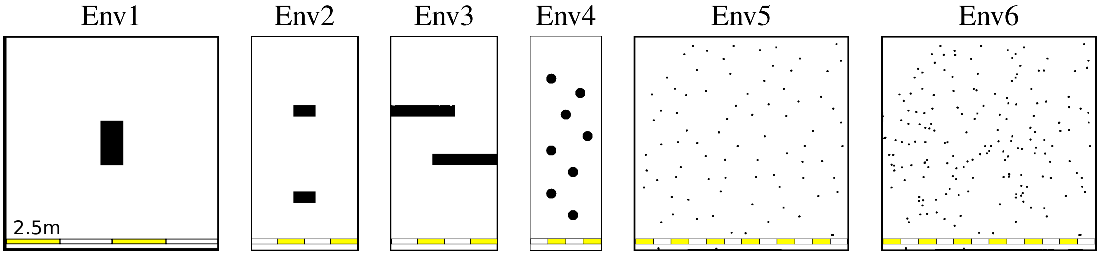


Figure 4: Environments used for training and evaluation. The policy was initially trained only in Env1 and evaluated in all other environments. Afterwards, we used further episodes from Env6 to improve the navigation behavior in highly cluttered scenes as in Env5 and Env6. The evaluation results are depicted in Tab. 1 and Fig. 6.

## 5. Experiments

The implementation of our framework is based on several components. As communication backbone, we use ROS and for the RL approach, we created a simulation environment with *Gazebo* [19]. We implemented the RL in *Python* with the *Tensorflow* library [20]. As mobile platform, we use the Robotino robot by Festo [21].

### 5.1. Training

To train the neural network and learn a policy to follow a global path and reach a goal without collisions, we used a simple environment (see Env1 in Fig. 4). During the training, we sampled the start and goal positions randomly across the free space, where we chose start-goal configurations with a short Euclidean distance at the beginning and later increased the distance for more challenging scenarios. This helps the robot initially to reach preferable states and learn basic navigation in free space, while longer start-goal configurations force the robot to deal with obstacles, as suggested in [22].

We used four simultaneously operating robots to ensure our collected data is independent and identically distributed. Each robot was given different start and goal configurations. Every episode was limited to 1000 time steps, the batch size was 32 and the entire training involved $10^6$ episodes. In Eq. (5), we set the final reward $b$ to 10, $c$ to 50, $r_{fix}$ in Eq. (4) to $-0.1$, and $\alpha$ in Eq. (6) had value of 10, as experimentally determined. The controller run with a frequency of 10 *Hz* during training and testing. The overall training time was about 24 hours using a *Nvidia GeForce GTX 1080.*

### 5.2. Evaluation

After training, we performed experiments in different environments to evaluate the policy learned in Env1 in terms of number of successful runs, which means that the robot reached the goal without collisions, and completion time, both in comparison to the standard ROS [2] navigation stack. The latter uses the DWA [3] to calculate the robot's velocity commands. We configured the DWA with similar restrictions to guarantee similar conditions in terms of acceleration and velocity limits and application of the unicycle robot control. The translational velocity was limited between 0 and 0.7m/s and the angular velocity between $-0.7$ rad/s and $0.7$ rad/s. The acceleration limits for translational and angular steering were set to 1 m/s$^2$ and 1 rad/s$^2$ for both approaches.

Note that the DWA approach needs an inflation radius around obstacles in the 2D grid map. This corresponds to a general safety distance to prevent collisions that could result, e.g., from the discretization of the environment. The inflation parameters usually need to be tuned to achieve a good trade-off between safety and time performance. One advantage of our approach is that it

| Env1 | Env2 | Env3 | Env4 | Env5 | Env6 | Env5* | Env6* |
|---|---|---|---|---|---|---|---|
| 1.0 | 1.0 | 0.99 | 0.99 | 0.75 | 0.22 | 1.0 | 0.84 |

Table 1: Success rate of the trained policy. The evaluation consists of 400 runs for each of the environments shown in Fig. 4. A successful run means that the robot reaches the goal within a certain time limit without any collision. To improve performance in Env5 and Env6, we continued to train the policy on Env6. As shown in the last two columns (Env5* and Env6*), the results of the re-trained policy were seriously better.

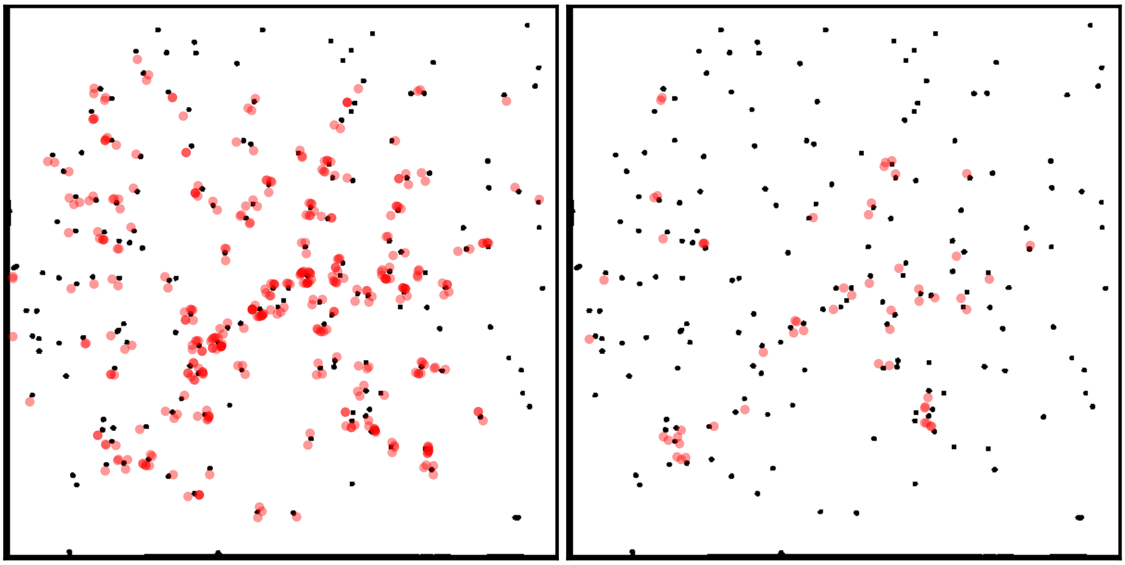


Figure 5: Performance improvement in Env6. (a) Collisions in Env6 with the policy trained on Env1. (b) The policy resulting from the additional training in Env6 shows much fewer collisions.

works on a binary map of the environment without any inflation. Our approach directly learns the appropriate distance to obstacles depending on their local configuration.

Fig. 4 depicts the environments we used in the evaluation. Each map introduces a further level of difficulty. Env2 is similar to the training environment Env1 but the length of the room is doubled and an additional obstacle occurs in the center. The large walls of Env3 can lead the robot into local minima if no global path is used. This map is well suited to test the performance of the learned policy in terms of a reduced completion time while avoiding collisions since fast movements on circular arcs around the obstacle corners are needed to achieve a good navigation behavior. Env4 introduces round obstacle shapes not experienced before. Env5 and Env6 consists of several regions with a high obstacle density. In those maps, it is not always possible to follow the global path computed on a map without obstacle inflation since the path might lead through regions with very close obstacles. Thus, the robot has to learn to bypass the corresponding region by moving away from the global path.

### 5.3. Success Rate

For each environment in Fig. 4, we performed 400 runs with the DWA and with our trained policy. The robot's start and goal configurations were sampled randomly for each run but were the same for the two approaches. The DWA controller was able to reach all goals in all environments without any collisions. The success rates of our trained policy are listed in Tab. 1. Our approach performs equally well in Env1 to Env4. In Env5 and Env6 the performance decreases due to

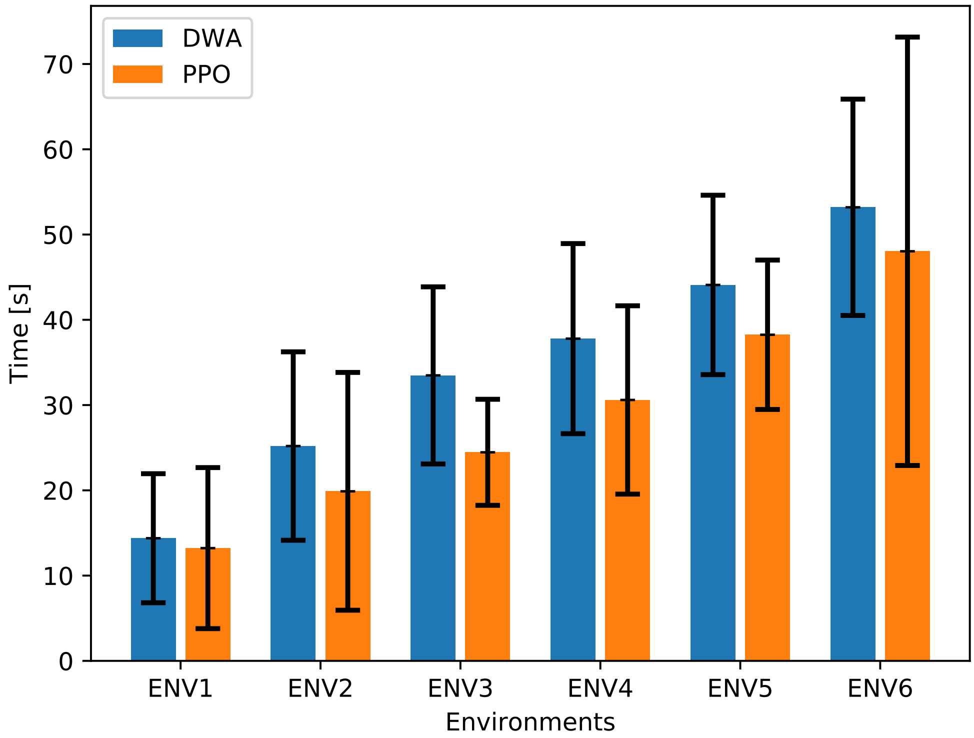


Figure 6: Average completion time for the DWA and our learned policy. The box height shows the average completion time and the whiskers illustrate the standard deviation. As can be seen, our trained policy outperforms the DWA in each environment. The difference is statistically significant in Env3, Env4, and Env5 according to a *paired* t-test at the 0.05 level.

an insufficient generalization resulting from Env1, that leads to increased collision rates. We discovered that situations in which the robot has to depart from the global path did not occur in Env1 and, thus, the robot could not learn a suitable strategy to handle those situations.

To overcome this limitation, we continued the training process with the so far learned policy parameters $\theta$ on Env6. After only 8000 further trained episodes (which corresponds to not even 1% of the initial size of the training set), the performance improved significantly and the results are shown in the last two columns of Tab. 1. In Env5 we could achieve a success rate of 100% with the newly learned policy and in Env6 the robot now reached the goal in 84% of all runs (the results are denoted as Env5* and Env6* in Tab. 1).

The left image of Fig. 5 visualizes for Env6 the positions where the robot collided with obstacles when following the policy learned on Env1. The right image of Fig. 5 shows the collisions after further training on Env6. As can be seen, fewer collisions appear in regions with high obstacle density. The reason is that the robot learned when it is beneficial not to follow the global path into narrow space but rather drive around depending on the obstacle configuration.

### 5.4. Completion Time

Next, we evaluated the completion time of the navigation tasks when using the standard DWA approach and our learned policy. Fig. 6 shows the average completion time for the runs from Sec. 5.3. that were successfully completed by both approaches. Our approach is 16% faster on

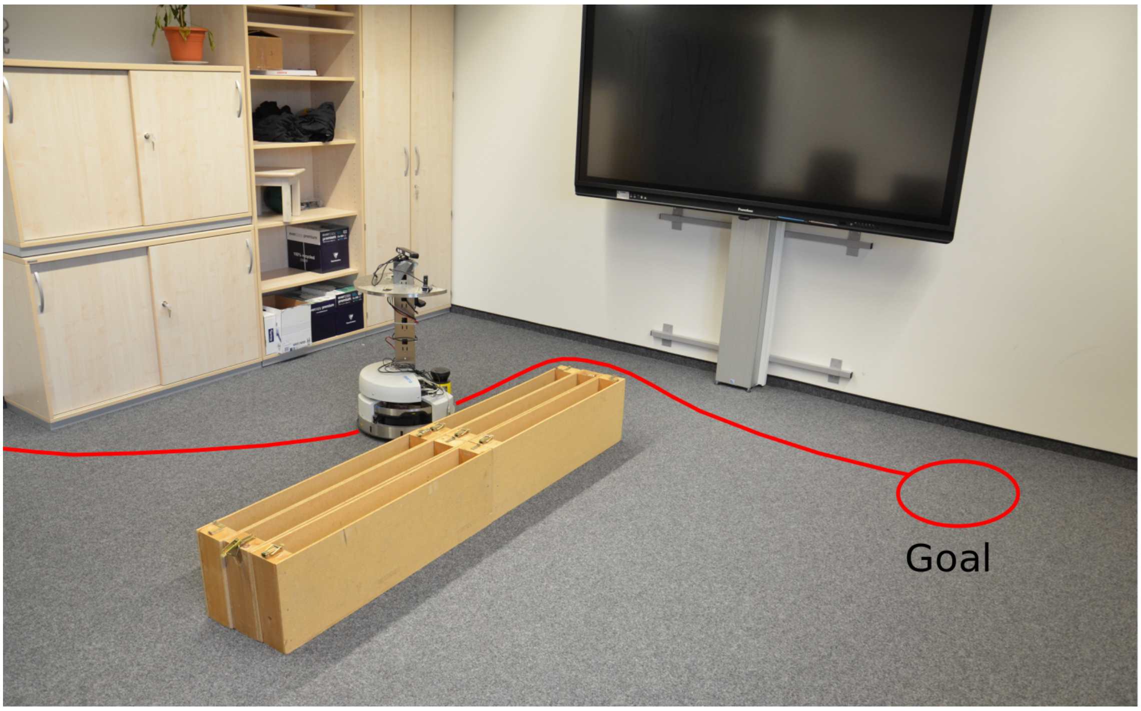


Figure 7: Real-world experiment. The robot entered the office from the left. Based on our learned navigation policy, the robot chooses the best translational and angular velocities to reach the global goal quickly while avoiding the obstacle in the center.

average over all evaluated runs. The difference is statistically significant in Env3, Env4, and Env5 according to a *paired* t-test at the 0.05 level. One reason for the faster performance is that the robot learns the best distance to obstacles, which reduces the trajectory length and leads to time savings, especially in Env3 where our policy performs $26\%$ faster than the DWA.

### 5.5. Real-World Experiment

Finally, we applied our learned policy on a real robot and compared the performance to the DWA. In the experiment, the robot had to enter an office from the corridor and navigate around an obstacle to reach the global goal (see Fig. 7) by following subgoals on the path. An occupancy grid of the environment was mapped before and we applied Monte Carlo localization [23] to obtain the robot pose.

For the evaluation, we performed 10 experiments with similar start and goal configurations for both the standard DWA approach and our trained policy. With both approaches, the robot reached the goal in each run. The DWA approach needed $28.2\,s$ on average to reach the goal location while our approach had a reduced average completion time of $25.5\,s$. The difference was statistically significant according to a *paired* t-test at the 0.05 level.

Our approach saves time by driving closer around obstacles while the standard DWA takes into account a general inflation radius around obstacles.

## 6. Conclusions

In this paper, we proposed a new approach to learn a navigation policy for wheeled robots in an unsupervised manner. We use proximal policy optimization for reinforcement learning to train a network that provides the robot's translational and angular velocity commands for the next time step. Our solution combines global path planning with deep RL to navigate collision-free and reach a global goal in the environment.

Our policy was first trained in a simple environment and subsequently evaluated in environments with increasing complexity. The experimental results demonstrate that our network successfully learned collision-free, goal-directed behavior also in cluttered environments. Furthermore, we compared the performance of our trained policy to the popular dynamic window approach (DWA) with respect to completion time of navigation tasks. On average, the robot controlled by our learned policy completed the tasks 16% faster than the DWA of ROS. In our real-world experiment, we experienced similar results, i.e., the robot performs 10% faster than the DWA using our navigation policy. Our learned strategy safes time by keeping a closer distance to obstacles and choosing appropriate velocities. This is a direct result of the optimization of the motion commands based on the local configuration of the obstacles without any parameter tuning for the navigation controller.

In future work, we plan to incorporate dynamic obstacles, e.g., walking humans, into our framework. An additional sensor would detect the moving obstacles and combine them with the static grid map as input to the CNN.

## ACKNOWLEDGMENTS

We would like to thank Christopher Gebauer for helpful discussions and his support during the experimental evaluation. This work has partly been supported by the German Research Foundation under Germany's Excellence Strategy, EXC-2070 - 390732324 (PhenoRob).

## REFERENCES

[1] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint*, 2017.

[2] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: An open-source robot operating system," in *Proc. of the ICRA Workshop on Open Source Software*, 2009.

[3] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics and Automation Magazine (RAM)*, 1997.

[4] J. Sergeant, N. Sünderhauf, M. Milford, and B. Upcroft, "Multimodal deep autoencoders for control of a mobile robot," in *Proc. of the Australasian Conf. on Robotics and Automation (ACRA)*, 2015.

[5] M. Pfeiffer, M. Schaeuble, J. Nieto, R. Siegwart, and C. Cadena, "From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2017.

[6] M. Pfeiffer, S. Shukla, M. Turchetta, C. Cadena, A. Krause, R. Siegwart, and J. Nieto, "Reinforced imitation: Sample efficient deep reinforcement learning for mapless navigation by leveraging prior demonstrations," *IEEE Robotics and Automation Letters (RA-L)*, 2018.

[7] Y. Liu, A. Xu, and Z. Chen, "Map-based deep imitation learning for obstacle avoidance," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2018.

[8] L. Tai, J. Zhang, M. Liu, and W. Burgard, "Socially compliant navigation through raw depth inputs with generative adversarial imitation learning," *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2018.

[9] A. Pokle, R. Martin-Martin, P. Goebel, V. Chow, H. M. Ewald, J. Yang, Z. Wang,

A. Sadeghian, D. Sadigh, S. Savarese, and M. Vazquez, “Deep local trajectory replanning and control for robot navigation,” in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2019.

[10] S. Gupta, J. Davidson, S. Levine, R. Sukthankar, and J. Malik, “Cognitive mapping and planning for visual navigation,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[11] S.-H. Hsu, S.-H. Chan, P.-T. Wu, K. Xiao, and L.-C. Fu, “Distributed deep reinforcement learning based indoor visual navigation,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2018.

[12] X. Chen, A. Ghadirzadeh, J. Folkesson, M. Björkman, and P. Jensfelt, “Deep reinforcement learning to acquire navigation skills for wheel-legged robots in complex environments,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2018.

[13] L. Tai, G. Paolo, and M. Liu, “Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2017.

[14] T. Fan, X. Cheng, J. Pan, D. Monacha, and R. Yang, “Crowdmove: Autonomous mapless navigation in crowded scenarios,” in *Proc. of the IROS Workshop on From freezing to jostling robots: Current challenges and new paradigms for safe robot navigation in dense crowds*, 2018.

[15] H. L. Chiang, A. Faust, M. Fiser, and A. Francis, “Learning navigation behaviors end-to-end with AutoRL,” *IEEE Robotics and Automation Letters (RA-L)*, 2019.

[16] J. Kulhánek, E. Derner, T. de Bruin, and R. Babuška, “Vision-based navigation using deep reinforcement learning,” in *Proc. of the Europ. Conf. on Mobile Robotics (ECMR)*, 2019.

[17] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. of Int. Conf. on Learning Representations (ICLR)*, 2015.

[18] J. Lei Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” *arXiv preprint*, 2016.

[19] N. Koenig and A. Howard, “Design and use paradigms for gazebo, an open-source multi-robot simulator,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2004.

[20] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[21] Festo. [Online]. Available: https://www.festo-didactic.com/int-en/learning-systems/education-and-research-robots-robotino/

[22] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” in *Proc. of the Int. Conf. on Machine Learning (ICML)*, 2009.

[23] D. Fox, W. Burgard, F. Dellaert, and S. Thrun, “Monte Carlo localization: Efficient position estimation for mobile robots,” *Proc. of the Conference on Advancements of Artificial Intelligence (AAAI)*, 1999.

# MACHINE LEARNING FOR MULTIPLE STAGE HEART DISEASE PREDICTION

Khalid Amen[1], Mohamed Zohdy[1] and Mohammed Mahmoud[2]

[1]Department of Electrical and Computer Engineering,
Oakland University, Rochester, MI, USA
[2]Department of Computer Science and Engineering,
Oakland University, Rochester, MI, USA

### *ABSTRACT*

*According to the Centers for Disease Control and Prevention (CDC), heart disease is the number one cause of death for men, women, and people of most racial and ethnic groups in the United States. More than one person dies every minute and nearly half a million die each year from it, costing billions of dollars annually. Previous machine learning approaches have been used to predict whether patients have heart disease. The purpose of this work is to predict the five stages of heart disease starting from no disease, stage 1, stage 2, stage 3, and advance condition or severe heart disease. We investigate different potential supervised models that are trained by machine learning algorithms and find out which of these models has better accuracy. In this paper, we describe and investigate five machine learning algorithms (SVM, LR, RF, GTB, ERF) with hyper parameters that maximize classifier performance to show which one is the best to predict the stage at which a person is determined to have heart disease. We found that the LR algorithm performs better compared to the other four algorithms. The experiment results show that LR performs the best with an accuracy of 82%, followed by SVM with an accuracy of 80% when all five classifiers are compared and evaluated for performance based on accuracy, precision, recall, and F measure. This predication can facilitate every step of patient care, reducing the margin of error and contributing to precision medicine. Lastly, this paper aims to improve heart disease prediction accuracy, precision, recall and F measure using UCI heart disease dataset. For this, multiple machine learning approaches were used to understand the data and predict the chances of heart disease in a medical database.*

### *KEYWORDS*

*machine learning, ml, cnn, dnn, rnn, jupyter, python, cleveland dataset, gradient tree boosting, gtb, random forest, rf, support vector machine, svm, extra random forest, erf, logistic regression, lr.*

## 1. INTRODUCTION

### 1.1. Machine Learning

Machine learning is the process of teaching a computer system how to make accurate predictions when provided data. It uses algorithms and neural network models to assist computer systems in progressively improving their performance. Machine learning algorithms automatically build a mathematical model using sample data – also known as "training data" – to make decisions without being specifically programmed to make those decisions [1] [2] [6].

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 205-223, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101118

Those predictions could be answering whether a piece of fruit in a photo is a banana or an apple, spotting people crossing the road in front of a self-driving car, whether the use of the word book in a sentence relates to a paperback or a hotel reservation, if an email is spam, or recognizing speech accurately enough to generate captions for a YouTube video [2] [4].

Machine learning is used across many spheres around the world. The healthcare industry is no exception. Machine learning can play an essential role in predicting presence/absence of locomotor disorders, heart diseases and more. Such information, if predicted well in advance, can provide important insights to doctors who can then adapt their diagnosis and treatment to a per patient basis [3] [4] [5].

Machine learning when applied to health care is capable of early detection of disease which would aid to provide early medical intervention. In heart disease prediction, machine learning techniques have played a significant role. Analysis of disease has become vital in health care sectors. The massive data collected by healthcare sectors are preprocessed and analyzed to discover the underlying information in the data for effective decision making and to provide proper medical intervention. The success of machine learning in the medical industry is its capability in analyzing the huge amount of data gathered by the health sector and its effectiveness in decision-making. Since the medical field involves too many manual processes, it has become necessary to automate these procedures. Remarkable advancements in electronic medical records have made it possible. Diagnosing diseases is an intricate job in the medical field [1] [3] [4] [7].

In order to conduct this prediction, a Jupyter notebook was constructed in Python using the publicly available Cleveland dataset for heart disease, which has over 300 unique instances with 76 total attributes. From these 76 attributes, only 14 of them are commonly used for research to this date. In addition, the hyperparameters used in this prediction come from the recommendations by Dr. Olson, “data-driven advice for applying machine learning to bioinformatics problems” [17]. The libraries and coding packages used in this analysis are: SciPy, Python, NumPy, IPython, Matplotlib, Pandas, Scikit-Learn, and Scikit-Image [18] [23].

## 1.2. Heart Disease

Heart disease describes a range of conditions that affect the heart. Diseases under the heart disease umbrella include blood vessel diseases such as coronary artery disease, heart rhythm problems, and congenital heart defects, among others [20] [21].

The term "heart disease" is often used interchangeably with the term "cardiovascular disease." Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect the heart's muscle, valves or rhythm, are also considered forms of heart disease [20] [21] [22].

Heart disease causes roughly 735,000 heart attacks each year in the U.S. killing more than 630,000 Americans. According to the American Heart Association, over 7 million have suffered a heart attack in their lifetime [22].

There are several risk factors for heart disease; some are controllable, others are not. Uncontrollable risk factors for heart disease include male, older age, family history of heart disease, being postmenopausal, and race. About half of Americans (47%) have at least one out of three key risk factors for heart disease: high blood pressure, high cholesterol and smoking [21] [26].

Heart disease is the number one killer of both men and women. Heart disease can happen at any age, but the risk increases as people get older. Children of parents with heart disease are more likely to develop heart disease themselves. African-Americans have more severe high blood pressure than Caucasians, and a higher risk of heart disease. Heart disease risk is also higher among Mexican-Americans, American Indians, native Hawaiians and some Asian-Americans. This is partly due to higher rates of obesity and diabetes [20] [21].

Genetic factors likely play some role in high blood pressure, heart disease, and other related conditions. However, it is also likely that people with a family history of heart disease share common environments and other factors that may increase their risk. Most people with a significant family history of heart disease have one or more other risk factors. Just as you cannot control your age, sex and race, you cannot control your family history; so it's even more important to treat and control any other modifiable risk factors you have [22] [4] [21].The risk for heart disease can increase even more when heredity is combined with unhealthy lifestyle choices, such as smoking cigarettes and eating an unhealthy diet [21] [22] [26].

High blood pressure increases the heart's workload, causing the heart muscle to thicken and become stiffer. This stiffening of the heart muscle is not normal and causes the heart to function abnormally. It also increases risk of stroke, heart attack, kidney failure and congestive heart failure [23].

When high blood pressure is present alongside obesity, smoking, high cholesterol levels or diabetes, the risk of heart attack or stroke increases even more. Some risk factors for heart disease cannot be controlled, like family history, for example. But it's still important to lower the chance of developing heart disease by decreasing the risk factors that can be controlled.

## 2. RELATED WORK

Many researchers have completed a lot of work on data analysis and survivability analysis through Machine Learning (ML) and Data Mining (DM) approaches. Several studies reported that these techniques are significant for future predictions such as in the field of medical diagnosis. In these studies, the authors applied multiple approaches to specific problems and achieve high classification accuracies e.g. in the healthcare industry, these techniques are used for disease prediction.

In [1], [38] author applied Decision Tree (DT), LL, NB, SVM, KNN, PCA, ICA classifier respectively to analyze the kidney disease data. Early detection and treatment of the diseases prevents it from getting to the worst stage making it not only difficult to cure but also impossible to provide treatment. Breast cancer affects many women, so researchers work on different classifiers such that DT, SMO, BF Tree and IBK help to analyze the breast cancer data and examine the performance of the related techniques in order to accurately predict breast cancer using DT [39] and Weka software [9]. RBF Network, Rep Tree, and Simple Logistic DM techniques are used to predict and resolve the survivability of breast cancer patient [40]. Simple Logistic is used for dimension reduction and proposed RBF Network and Rep Tree model used for fast diagnosis of the other diseases.

The prediction of heart disease and patient survivability has been a critical research problem for a few decades. Globally, heart diseases are one of the major cause of deaths. About 80% of deaths in low and middle-income countries are due due to heart diseases [41]. Researchers use multiple DM techniques to develop a prediction model for the survival of heart disease patients. K-mean, C4.5 techniques are used in [8]. NB, J48 DT and Bagging algorithm, CART, ID3 (Iterative

Dichotomized 3) and Decision Table, Logistics Classification, Multilayer Perception and SMO; these three algorithms are respectively used in [41], [9], [10] to predict heart disease patients and their survivability. However, with the advancement of these technologies, the measurements are not sufficient for the prediction of diseases.

## 3. BACKGROUND OF CLEVELAND DATASET

Experiments with the Cleveland dataset have concentrated on simply attempting to distinguish the presence of heart disease from absence [15]. The 14 attributes that were used are listed in Table 1.

Table 1: Cleveland dataset attributes

| Name | Type | Description |
|---|---|---|
| Age | Continuous | Age in years |
| Sex | Discrete | 0 = female 1 = male |
| Cp | Discrete | Chest pain:<br>1 = typical angina, 2 = atypical angina, 3 = non-angina pain, 4 = asymptom |
| Trestbps | Continuous | Resting blood pressure (in mm Hg) |
| Chol | Continuous | Serum cholesterol in mg/dl |
| Fbs | Discrete | Fasting blood sugar>120 mg/dl: 1 = true 0 = false |
| Restecg | Discrete | Resting Electrocardiograph |
| Thalach | Continuous | Exercise Max Heart Rate Achieved |
| Exang | Discrete | Exercise Induced Angina: 1=yes 0=no |
| Old peak ST | Continuous | Depression induced by exercise relative to rest |
| Slope | Discrete | The slope of the peak exercise segment: 1=up sloping 2=flat 3=down sloping |
| Ca | Continuous | Number of major vessels colored by fluoroscopy that range between 0 and 3 |
| Tha | Discrete | 3=normal 6=fixed defect 7=reversible defect |
| Class | Discrete | Diagnosis classes: 0=No Presence 1=Least likely to have heart disease 2=>1 3=>2 4=More likely have heart disease |

The five stages of our prediction of Heart Disease presented are mapped as follows:

Table 2: Five stages prediction

| | |
|---|---|
| 0 | No Heart Disease |
| 1 | Stage1 |
| 2 | Stage2 |
| 3 | Stage3 |
| 4 | Heart Disease presented |

## 4. BACKGROUND ON RECOMMENDATION OF MODEL ALGORITHMS

A study conducted by Randal S. Olson provides insightful best practice advice for solving bioinformatics problems with Machine Learning, “Data-driven Advice for Applying Machine Learning to Bioinformatics Problems” [18]. He analyzed 13 state-of-the-art commonly used

Machine Learning algorithms on a set of 165 publicly available classification problems in order to provide data-driven algorithm recommendations to current researchers.

From his findings, he was able to provide a recommendation of five algorithms with hyperparameters that maximize classifier performance across the tested problems, as well as general guidelines for applying machine learning to supervised classification problems. The recommendations are as follows [1]:

Table 3: Five Machine Learning Algorithms

| **Algorithm** | **Parameters** | **Datasets Covered** |
|---|---|---|
| GradientBoostingClassifer | Loss= devianceLearning_rate=0.1,<br>n_estimators=500<br>max_depth=3, max__features = log2 | |
| RandomForestClassifier | n_estimators=500, max__features = 0.25, criterion=entropy | 19 |
| SVC | C=0.01, gamma=0.1, degree=3, coef0=10.0 | 16 |
| ExtraTreesClassifier | n_estimators=1000, max__features = log2, criterion=entropy | 12 |
| LogisticRegression | C=1.5, Penalty=L1, Fit_intercept=true | 8 |

The database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by DL researchers to this date. The "num" field in the figure refers to the presence of heart disease in the patient. It is integer valued from zero (no presence) to four. Experiments with the Cleveland database have concentrated on attempting to distinguish presence (values 1,2,3,4) from absence (value 0) [15] [23].

1. #3 (age)
2. #4 (sex)
3. #9 (cp)
4. #10 (trestbps)
5. #12 (chol)
6. #16 (fbs)
7. #19 (restecg)
8. #32 (thalach)
9. #38 (exang)
10. #40 (oldpeak)
11. #41 (slope)
12. #44 (ca)
13. #51 (thal)
14. #58 (num) (the predicted attribute)

## 5. APPROACH

The Data is split into 80% training (237 people) and 20% testing (60 people) after we dropped six from missing values. Several different models are evaluated through k-fold Cross-Validation with k-fold = 10 using GridSearchCV, which iterates on different algorithm's hyperparameters:

1. Gradient Tree Boosting (GradientBoostingClassifier).
2. Random Forest (RandomForestClassifier).
3. Support Vector Machine (SVC).
4. Extra Random Forest (ExtraTreesClassifier).
5. Logistic Regression (LogisticRegression).

### 5.1. Gradient Tree Boosting

Gradient boosting is a type of machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error. The target outcome for each case in the data depends on how much changing that case's prediction impacts the overall prediction error [24]:

- If a small change in the prediction for a case causes a large drop in error, the next target outcome of the case is a high value. Predictions from the new model that are close to its targets will reduce the error.
- If a small change in the prediction for a case causes no change in error, the next target outcome of the case is zero. Changing this prediction does not decrease the error.

The name gradient boosting arises because target outcomes for each case are set based on the gradient of the error with respect to the prediction. Each new model takes a step in the direction that minimizes prediction error, in the space of possible predictions for each training case [25].

This algorithm builds an ensemble of trees in a serial approach, where a weak model, e.g., a tree with only a few splits, is trained first and consecutively improves its performance by maintaining to generate new trees [34]. Each new tree in the sequence is responsible for repairing the previous prediction error [30]. Based on the *grid* search, we set the learning parameters as follows:

Table 4: GTB Parameters

| **Parameter** | **Meaning** | **Value** |
|---|---|---|
| Learning rate | Impact of each tree on the final outcome | 0.1 |
| N_estimators | Number of sequential tree to modeled | 500 |
| Max_depth | Max depth of a tree | 3 |
| Max_features | Number of features | Log2 |

### 5.2. Random Forest

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model’s prediction.

The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds. In data science speak, the reason that the random forest model works so well is a large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models [36]. The low correlation between models is the key. Just like how investments with low correlations come together to form a portfolio that is greater than the sum of its parts, uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions. The reason for this effect is the trees protect each other from their individual errors. While some trees may be wrong, many other trees will be right, so as a group the trees are able to move in the correct direction.

This classifier takes bagging of decision tree procedure to evoke a large collection of trees to improve performance. Compared to other similar ensembles, Random Forest (RF) that requires less hyperparameter tuning [27]. Original bagging decision tree yields tree-mutuality, which suffers from the effect of high variance. Hence, RF offers a variance reduction by introducing more randomness into the tree-generation procedure. Based on grid search, we set the learning parameters as follows:

Table 5: RF Parameters

| **Parameter** | **Meaning** | **Value** |
|---|---|---|
| N_estimators | Number of trees in the forest | 500 |
| Max_features | Number of features to consider | 1 |
| Criterion | Function to measure the quality of a split | Entropy |

## 5.3. Support Vector Machine (SVM)

Support Vector Machines (SVMs) are powerful yet flexible supervised machine learning algorithms which are used both for classification and regression. But generally, they are used in classification problems. In 1960s, SVMs were first introduced but later they were refined in 1990. SVMs have their unique way of implementation as compared to other machine learning algorithms. Lately, they are extremely popular because of their ability to handle multiple continuous and categorical variables [29].

An SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH).

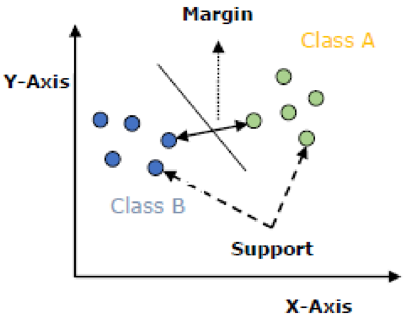


Figure 1: SVM Classes

The import concepts in SVM are:

- Support Vectors – Datapoints that are closest to the hyperplane are called support vectors. Separating line will be defined with the help of those data points.
- Hyperplane – As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.
- Margin – It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors.
- Large margin is considered a good margin and small margin is considered a bad margin.

The main goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH) and it can be done in the following two steps

- First, SVM will generate hyperplanes iteratively that segregate the classes in the best way.
- Then, it will choose the hyperplane that separates the classes correctly.

In practice, SVM algorithm is implemented with kernel that transforms an input data space into the required form. SVM uses a technique called the kernel trick in which kernel takes a low dimensional input space and transforms it into a higher dimensional space. In simple words, kernel converts non-separable problems into separable problems by adding more dimensions to it. It makes SVM more powerful, flexible and accurate. The following are some of the types of kernels used by SVM [28] [29] [30].

- Linear Kernel – It can be used as a dot product between any two observations. The formula of linear kernel is:

  $$K(x, y) = sum(x * y) \quad (1)$$

  From the above formula, the product between two vectors (x) and (y) is the sum of the multiplication of each pair of input values.
- Polynomial Kernel – It is more generalized form of linear kernel and distinguish curved or nonlinear input space. The formula of Polynomial kernel is:

  $$K(x, y) = 1 + sum(x * y)^\wedge d \quad (2)$$

  where d is the degree of polynomial which it be specified manually in the learning algorithm.
- Radial Basis Function (RBF) kernel – it is mostly used in SVM classification. It maps input space in indefinite dimensional space. The formula if RBF is:

  $$K(x, y) = exp\big(gamma * sum(x - y^2)\big) \quad (3)$$

  Gamma ranges from 0 and 1 and it needs to be specified manually in the learning algorithm.

Based on grid search, we set the learning parameters as follows:

Table 6: SVM Parameters

| Parameter | Meaning | Value |
|---|---|---|
| C | Penalty or Regularization Parameter | 0.01 |
| Gamma | Gamma coefficient | 0.1 |
| Kernel | Kernel Type | rbf |
| Degree | Degree of the polynomial function | 3 |
| Coef0 | Independent term in kernel function | 10.0 |

### 5.4. Extra Random Forest (Extra Tree Classifier)

Extra Trees Classifier is an ensemble machine learning algorithm. In other words, it is an ensemble of decision trees and is related to other ensembles of decision trees algorithms such as bootstrap aggregation (bagging) and random forest [31]. The Extra Trees algorithm works by creating a large number of unpruned decision trees from the training dataset. Predictions are made by averaging the prediction of the decision trees in the case of regression or using majority voting in the case of classification [32]. Unlike bagging and random forest that develop each decision tree from a bootstrap sample of the training dataset, the Extra Trees algorithm fits each decision tree on the whole training dataset [31] [32] [37].

Like random forest, the Extra Trees algorithm will randomly sample the features at each split point of a decision tree and select a split point at random.

Based on grid search, we set the learning parameters as follows:

Table 7: ERF Parameters

| Parameter | Meaning | Value |
|---|---|---|
| N_estimators | The number of trees in the forest | 1000 |
| Max_features | The number of features to consider | Log2 |
| Criterion | The function to measure the quality of a split | Entropy |

### 5.5. Logistic Regression

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of the target or dependent variable is dichotomous, which means there will be only two possible classes [33]. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

Mathematically, a logistic regression model predicts P(Y=1) as a function of X. It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, diabetes prediction, cancer detection etc. [17].

#### 5.5.1.Type of Logistics Regressions

Logistic Regression means binary logistic regression having binary target variables, but there can be two more categories of target variables that can be predicted by it. Based on those number of categories, Logistic regression can be divided into following types [38]:

- Binary or Binomial – In such a classification, a dependent variable will have only two possible types either 1 and 0. For example, these variables may represent success or failure, yes or no, win or loss etc.
- Multinomial – In such a classification, dependent variable can have three or more possible unordered types or the types having no quantitative significance. For example, these variables may represent “Type A” or “Type B” or “Type C”.
- Ordinal – In such a classification, dependent variable can have three or more possible ordered types or the types having a quantitative significance. For example, these variables may represent “poor”, “good”, “very good” or “excellent” and each category can have scores such as 0, 1, 2 or 3.

Based on grid search, we set the learning parameters as follows:

Table 8: LR Parameters

| Parameter | Meaning | Value |
|---|---|---|
| C | Inverse of regularization strength | 1.5 |
| Penalty | Specify the norm used in the penalization | L2 |
| Fit_intercept | Specifies if a constant should be added to decision function | True |

## 6. METHODOLOGY

The proposed methodology using fiveclassification techniques; Gradient Tree Boosting, Random Forest, Support Vector Machine (SVM), Extra Random Forest, and Logistic Regression to predict heart disease as the proposed methodology shown in Fig 4. These classifiers are used to improve the prediction. We applied the classifiers in Fig 5to heart disease data that comes from the Cleveland dataset to predictin which of five stages a patient has heart problems. The performance of these classifiers are to evaluate on the bases of accuracy, precision recall, and F measure.

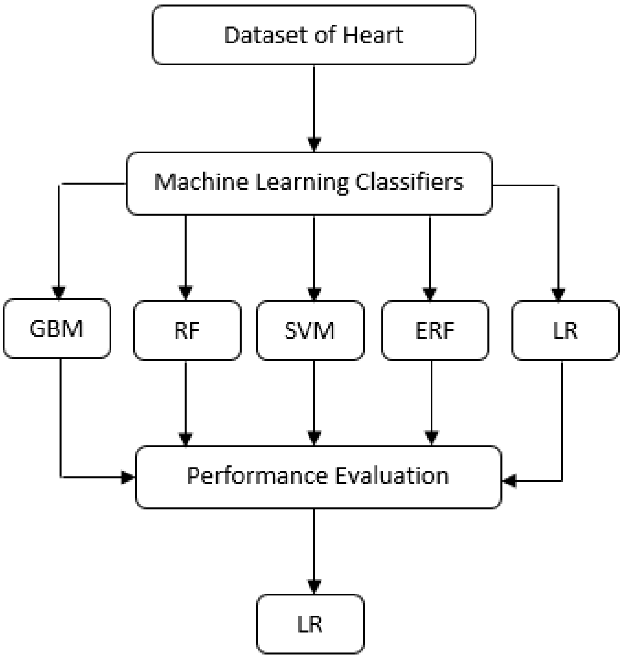


Figure 2: Proposed Methodology

The dataset of heart is taken from UCI repository [23], the classifier taking it as input for disease prediction. These classifiers are implemented in Python language. Python is a powerful interpreter language and a reliable platform for research [19]. The accuracy of prediction increased by comparing the results of these five classifiers using evaluation parameters. The experimental result describes which classifier is best between them.

## A. Evaluation Parameters

Some evaluation parameters in data mining are accuracy, precision, recall, and F measure. Where TPTrue Positive, TN- True Negative, FP- False Positive and FN- False Negative [19].

- Accuracy is defined as the number of accurately classified instances divided by the total number of instances in the dataset as in (4).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

- Precision is the average probability of relevant retrieval as described in (5).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

- The recall is defined as the average probability of complete retrieval as defined in (6).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

- F- Measure is the calculated by using both precision and recall as shown in (7).

$$\text{F Measure} = \frac{2 * (Precision * Recall)}{Precision + Recall} \quad (7)$$

Where all evaluation parameters accuracy, precision, recall, and F measure are calculated from dataset when splitting the dataset into training data and test data.The Pseudo codes for the evaluation parameters are as follow:

```
Def  evaluationParameters(X_train, y_train, X_test, y_test):
        X_train ← fit_transform(X_train)
        Classifier ←sklearn()
        y_pred ← classifier.predict(X_test)
        cm_test ← confusion_matrix(y_pred, y_test)
        y_pred_train ← classifier.predict(X_train)
        cm_train ← confusion_matrix(y_pred_train, y_train)
        training_accuracy = (cm_train[0][0] + cm_train[1][1])/len(y_train)
        test_accuracy = (cm_test[0][0] + cm_test[1][1])/len(y_test)
        training_percision = cm_train[0][0]/(cm_train[0][0] + cm_train[1][0])
        test_percision = cm_test[0][0]/(cm_test[0][0] + cm_test[1][0])
        training_recall = cm_train[0][0]/(cm_train[0][0] + cm_train[0][1])
        test_recall = cm_test[0][0]/(cm_test[0][0] + cm_test[0][1])
                training_f_measure ← (2 * training_percision *
        training_recall)/(training_percision + training_recall))
        test_f_measure ← (2 * test_percision * test_recall)/(test_percision + test_recall))

return (training_accuracy, test_accuracy, training_percision, test_percision, training_recall,
test_recall,            training_f_measure,  training_f_measure)
```

## B. Dataset

To perform the research, heart disease datasetis used. This heart disease dataset contains 14 attributes and 303 instances. This dataset is taken from UCL repository. It's an online repository that contains 412 diverse datasets. UCI provides data for ML to perform analysis in a different direction. The UCI database is known for its extensiveness in data, its completeness, and accuracy [23].

## C. Machine Learning Classifiers:

In this research, five classification methods are implemented in python using the pandas and keras libraries. These models are used to improve prediction. These classifiers are compared to find out which of the five stages best predictsthe chance of heartdisease in patients. In the next section, we briefly describe these classification techniques/ classifiers.

**1)** Logistics Regression (LR): is the predictive analysis to conduct on discrete (binary) values based on a specified set of independent variables. LR describes the data and clarifies the relationship between one (binary) dependent variable and independent variables. It predicts the event occurrence probability by fitting the data into a logit function. Therefore, it is also called logit regression [32].

Input values x are linearly combined using coefficient values b, to calculate an output value p. The output values as predictable lies between 0 and 1. Input data associated with coefficient b (constant value) learned from training data. Where p is the output, b0 is an intercept term and b1 is the coefficient of input value x as shown in (8).

$$\mathrm{P} = \frac{e^{(b0+b1+x)}}{1+ e^{(b0+b1+x)}} \quad (8)$$

**2)** Support Vectors Machine (SVM): is a classification and regression algorithm. In SVM, every data item is plotted in n-dimensional space, a number of dimensions are equivalent to the number of features or attributes. Where n represents the number of attributes. The value of each attribute being the value of certain coordinates. Once plotting all the data items then performed classification by drawing a line or by finding the optimal hyperplane that separates two classes completely. For example, if we have two features of individual like hair and height length. First, we plot these two features in two-dimensional space where every point has two coordinates (these co-ordinates are also known as Support Vectors) [29].

**3)** Random Forests (RF): are ensemble learning technique for regression, regression, classification, and for other tasks. That operate by making a multitude of Decision Tree (DT) at training stint and outputting that is the mean prediction (regression) or mode of classes (classification) of the distinct trees.

Every tree in the forest contributes for a classification. To classify new case based on its attributes. We identify the tree "votes" for that class so the forest indicates the classification of the case that is taking the most votes [37].

Every tree is planted and grown as follows:

- If the number of objects N in training set, the sample of N objects is taken randomly with replacement. This sample act as a training set for growing tree.
- If there is an input variable N,numbern < N is stated that at each node, randomly selection of n variable out of input variable N. So, the best splitting on n is used to split the node. The value of m (node splitter) is constant during gowning the forest.
- Each tree is growing up to the largest magnitude possible so there is no trimming.

**4)** Gradient Tree Boosting: is a machine learning technique for regression and classification problems, which produces a predication model in the form of an ensemble of weak predication models. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function [35].

It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error. How are the targets calculated? The target outcome for each case in the data depends on how much changing that case's prediction impacts the overall prediction error.

**5)** Extra Random Forest (ERF): is very similar to Random Forests (RF) but there are two main differences:

- ERF does not resample observations when building a tree. They do not perform bagging.
- ERF does not use the best split. Like RF, ERF selects a random subset of predictors for each split. Instead of the best split for predictors, ERF makes a small number of randomly chosen splits-points for each of the selected predictors. ERF then selects the best split from this small number of choices.

In ERF, the features and splits are selected at random. Since splits are chosen at random for each feature, it is less computationally expensive than RF [31].

## D. Scaling the Data:

To accomplish the five stages output prediction for a patient to be diagnosed with one of five stages, it is important to scale the data so the machine learning algorithms do not overfit to the wrong features. Using the MinMaxScaler() method on Python, the values are scaled per features based on the minimum and maximum between 0 and 1. This keeps the information from being lost but allows the machine learning algorithms to correctly train with the data. The training data and test data are scaled between 0 and 1 and the output data is scaled between 0 and 1 as well. Then, the scaled output value is mapped as follows:

Table 9: Five Stages

| Output Value | Stage |
|---|---|
| 0 | No disease presented |
| 0 < and <= 0.25 | Stage1 |
| 0.25 < and <= 0.5 | Stage2 |
| 0.5 < and <= 0.75 | Stage3 |
| 0.75 < and <= 1 | Advance disease presented |

## 7. EXPERIMENTAL RESULT

The experiment is conducted for the prediction of heart disease stages by applying various machine learning classifiers. From the experiment results, we identify that Logistic Regression performs better as compared to the other four ML classifiers in the prediction of these diseases. In this experiment, we use multiple stages of heart disease prediction to forecast the stage at which a person is determined to have heart disease. In previous works [42] [43] [44] [45], the study used two outcome predications, either a person has the disease or not; that is represented by (0 ,1) or (true, false).The Pseudocodes for the experiment are as follow:

```
data_frame 🡐 read_CSV_file
X 🡐 data_frame [column: 0 - 12]
y 🡐 data_frame [column: 13]
target 🡐 preprocessing.scale(y)
data 🡐 preprocessing.scale(X)

for k 🡐 0 to data - 1
        if data[k] = 0 then
        data[k] 🡐 'no disease'
        if data[k] > 0 && data[k] <= 0.25 then
        data[k] 🡐 'stage1'
        if data[k] > 0.25 && data[k] <= 0.5 then
        data[k] 🡐 'stage2'
        if data[k] > 0.5 && data[k] <= 0.75 then
        data[k] 🡐 'stage3'
        else
        data[k] 🡐 'disease presented'

X_train, X_test, y_train, y_test 🡐 train_test_split(X, y, test_size=0.2, random_state=0)
svm(X_train, y_train, X_test, y_test)
lr(X_train, y_train, X_test, y_test)
rf(X_train, y_train, X_test, y_test)
gtb(X_train, y_train, X_test, y_test)
erf(X_train, y_train, X_test, y_test)
```

The below Figures show the performance of various evaluation parameters in the prediction of heart disease. The experimental results show the comparison of LR, ERF, GTB, SVM and RF classifiers and evaluate the performance on the bases of accuracy, precision, recall and F measure. In all classifiers, LR performs the best with an accuracy of 82%, followed by SVM with an accuracy of 80%.

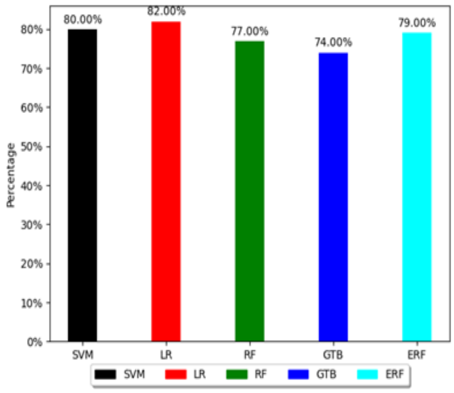


Figure 3: Heart Disease Accuracy

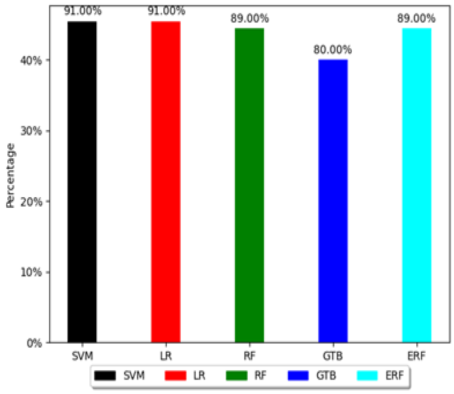


Figure 4: Heart Disease Precision

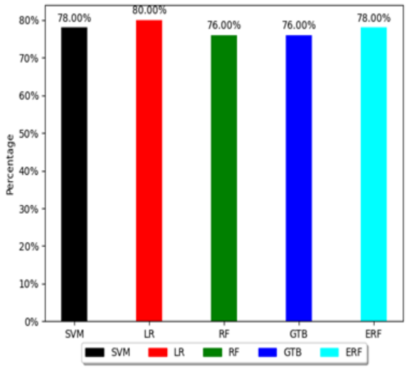


Figure 5: Heart Disease Recall

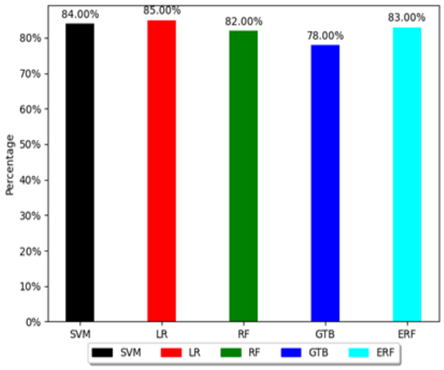


Figure 6: Heart Disease F Measure

Table 10: ML Algorithms Comparison

| Algorithm | Accuracy | Precision | Recall | F Measure |
|---|---|---|---|---|
| SVM | 80% | 91% | 78% | 84% |
| LR | 82% | 91% | 80% | 85% |
| RF | 77% | 89% | 76% | 82% |
| GTB | 74% | 80% | 76% | 78% |
| ERF | 79% | 89% | 78% | 83% |

## 8. CONCLUSIONS

The importance of extracting the valuable information from raw data has very good consequences in many fields of life such as the medical area, business area, and more. In this study, we proposed a multiple stage detection model of heart disease based on five algorithms to compare which one performs better. The proposed method was built by the stacking of five different ensemble learners, such as the Random Forest, Gradient Boosting Machine, Extreme Random Forest, Logic Regression, and Support Vector Machine. The proposed detection model was tested on well-known Cleveland dataset in order to provide a fair benchmark against existing studies. Based on the experimental results, our proposed model was able to outperform heart disease detection methods with respect to accuracy, precision, recall and F measure. The result reflected the highest result obtained showed that Logic Regression has a better result comparing to the other four methods or algorithms. The performance was further enhanced using feature selection techniques. The feature selection techniques helped to improve the accuracy, precision, recall, and F measure of the ensemble algorithms.The experiment results show that LR performs the best with an accuracy of 82%, followed by SVM with an accuracy of 80% when all five classifiers are compared and evaluated for performance based on accuracy, precision, recall, and F measure

## 9. LIMITATIONS

The Cleveland heart dataset from UCI machine learning repository was utilized for training and testing purposes. The ML algorithms SVM, LR, RF, GTB, and ERF were employed for experiments. As far as the dataset are concerned, they need to be amplified as the main limitation in this work is the small size of the dataset. If the dataset has bad data and is not caught, then this would generate bad or inaccurate predictions. The dataset has a limited number of patient records; therefore, the dataset was augmented using appropriate techniques.

## 10.FUTURE WORK

There are many possible improvements that could be explored to improve accuracy, precision, recall, and F measure of this prediction system. Due to time limitations, the following research/ work needs to be performed in the future:

- There is a need of more ML algorithms for comparison.
- Large dataset to be trained
- Build a system or a framework for automation of heart disease predication.
- Real data from health care organizations and agencies needs to be collected and all the available techniques will be compared for the optimum accuracy.
- Use deep learning to structure algorithms in layers to create Artificial Neural Network (ANN) or Convoluted Neural Network (CNN) that can learn and make intelligent decision.

## 11.ACKNOWLEDGEMENTS

This paper and the research behind it would not have been possible without the grace, the bounty, and the blessing of almighty Allah (God) first and foremostand the exceptional support of my professors, Mohamed Zohdy and Mohammed Mahmoud. Their enthusiasm, knowledge and exacting attention to detail have been an inspiration and kept my work on track from my first encounter with machine learning research to the final draft of this paper.

## REFERENCES

[1] R. Garg, "A Comparative Study of Different Classification Algorithms on Kidney Disease Prediction," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 6, no. 2, pp. 741–746, Feb. 2018, doi: 10.22214/ijraset.2018.2132.0

[2] Nick Health, "What is machine learning? Everything you need to know", https://www.zdnet.com/article/what-is-machine-learning-everything-you-need-to-know/, Sep. 2018

[3] G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop," in *2017 International Conference on Inventive Computing and Informatics (ICICI)*, Nov. 2017, pp. 973–979, doi: 10.1109/ICICI.2017.8365283.

[4] P. K. Sahoo, S. K. Mohapatra, and S.-L. Wu, "Analyzing Healthcare Big Data With Prediction for Future Health Condition," *IEEE Access*, vol. 4, pp. 9786–9799, 2016, doi: 10.1109/ACCESS.2016.2647619.

[5] Karan Bhanot, “Predicting presense of Heart Diseases using Machine Learning”, https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c , Feb. 2019

[6] Keith Foote, “A Brief History of Machine Learning”, https://www.dataversity.net/a-brief-history-of-machine-learning/, Mar. 2019.

[7] N. A. R. and D. R. Vincent, “Heart Disease Prediction System Using Ensemble of Machine Learning Algorithms,” *Recent Patents Eng.*, vol. 13, Mar. 2019, doi: 10.2174/1872212113666190328220514.

[8] A.Rairikar, V. Kulkarni, V. Sabale, H. Kale, and A. Lamgunde, “Heart disease prediction using data mining techniques,” in *2017 International Conference on Intelligent Computing and Control (I2C2)*, Jun. 2017, pp. 1–8, doi: 10.1109/I2C2.2017.8321771.

[9] V. Chaurasia, “Early Prediction of Heart Diseases Using Data Mining,” *Caribb. J. Sci. Technol.*, vol. 1, pp. 208–217, 2013, doi: 10.1377/hlthaff.2014.0041.

[10] S. Vijiyarani, S. Sudha, and M. P. Research Scholar, “An Efficient Classification Tree Technique for Heart Disease Prediction,” 2013

[11] A. Chaudhary and P. Garg, “Detecting and Diagnosing a Disease by Patient Monitoring System,” *Int. J. Mech. Eng. Inf. Technol.*, vol. 2, no. 6, pp. 493–499, 2014.

[12] P. R. V. A. M Archana Bakare, “Prediction of Diseases using Big Data Analysis,” *Int. J. Innov. Res. Comput. Commun. Eng. (An ISO Certif. Organ.*, vol. 3297, no. 6, pp. 11449–11455, 2016, doi: 10.15680/IJIRCCE.2016.

[13] A.V. Solanki, “Data Mining Techniques Using WEKA classification for Sickle Cell Disease,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4, pp. 5857–5860, 2014, [Online]. Available: http://www.ijcsit.com/docs/Volume 5/vol5issue04/ijcsit20140504222.pdf.

[14] V. S and D. S, “Data Mining Classification Algorithms for Kidney Disease Prediction,” *Int. J. Cybern. Informatics*, vol. 4, no. 4, pp. 13–25, 2015, doi: 10.5121/ijci.2015.4402.

[15] Heart Disease in Cleveland, https://www.rpubs.com/aepoetry/log_reg_heart

[16] V. Kunwar, K. Chandel, A. S. Sabitha, and A. Bansal, “Chronic Kidney Disease analysis using data mining classification techniques,” in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, Jan. 2016, pp. 300–305, doi: 10.1109/CONFLUENCE.2016.7508132.

[17] G. Caocci, R. Baccoli, R. Littera, S. Orru, C. Carcassi, and G. La, “Comparison Between an Artificial Neural Network and Logistic Regression in Predicting Long Term Kidney Transplantation Outcome,” in *Artificial Neural Networks - Architectures and Applications*, InTech, 2013.

[18] Dheeru Due, UC Irvine Machine Learning Repositoty, https://archive.ics.uci.edu/ml, Sep. 2018

[19] Simon Tavasoli, Machine Learning Algorithms,https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article, Jul. 2020.

[20] WebMD, Risk Factors for Heart Disease, https://www.webmd.com/heart-disease/risk-factors-heart-disease

[21] CDC, Know Your Risk for Heart Disease, https://www.cdc.gov/heartdisease/risk_factors.htm, Dec. 2019

[22] American Heart Association, Understand Your Risks to Prevent a Heart Disease, https://www.heart.org/en/health-topics/heart-attack/understand-your-risks-to-prevent-a-heart-attack#:~:text=An%20inactive%20lifestyle%20is%20a,blood%20pressure%20in%20some%20people , Jun. 2016

[23] Steven Smiley, Diagnositc for Heart Disease with Machine Learning (ML), https://github.com/stevensmiley1989. Feb. 2011.

[24] Jake Hoare, Gradient Boosting Explained - The Coolest Kid on The Machine Learning Block, https://www.displayr.com/gradient-boosting-the-coolest-kid-on-the-machine-learning-block, DisplayR

[25] Wikipedia, Gradient Boosting, https://en.wikipedia.org/wiki/Gradient_boosting, Aug. 2020.

[26] CDC, Heart Disease Facts, https://www.cdc.gov/heartdisease/facts.htm, Jun. 2020.

[27] Bayu Adhi Tama, Sun Im, Seungchul Lee, “Improving an Intellegent Detection System for Coronary Heart Disease Using A Two-Tier Classifer Ensemble”, vol 2020, article ID 9816142, April 2020.

[28] KD Nuggets, What is a Support Vector Machine and Why Would Use it?, https://www.kdnuggets.com/2017/02/yhat-support-vector-machine.html, Feb. 2017.

[29] Savan Patel, Chapter 2: SVM (Support Vector Machine) Theory, https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72, May 2017.

[30] Tutorialspoint, ML – Support Vector Machine, https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_support_vector_machine.htm.

[31] Geek for Geeks, ML – Extra Tree Classifer for Feature Selection, https://www.geeksforgeeks.org/ml-extra-tree-classifier-for-feature-

selection/#:~:text=Extremely%20Randomized%20Trees%20Classifier(Extra,to%20output%20it's%20classification%20result, Jan. 2020

[32] Json Brownlee, How to Develop an Extra Trees Ensemble with Python, https://machinelearningmastery.com/extra-trees-ensemble-with-python/, Apr. 2020

[33] Tutorialspoint, Machine Learning – Logistic Regression, https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_logistic_regression.htm.

[34] Pablo Diez, Smart Wheelchairs and Brain-Computer Interfaces, https://www.sciencedirect.com/topics/engineering/confusion-matrix, 2018

[35] Dan Nelson, Gradient Boosting Classifiers in Python with Scikit-Learn, https://stackabuse.com/gradient-boosting-classifiers-in-python-with-scikit-learn/.

[36] Steven Smiley, Diagnostic for Heart Disease with Machine Learning, https://towardsdatascience.com/diagnostic-for-heart-disease-with-machine-learning-81b064a3c1dd, Jan. 2011.

[37] Frank Ceballos, An Intuitive Explanation of Random Forest and Extra Trees Classifiers, https://towardsdatascience.com/an-intuitive-explanation-of-random-forest-and-extra-trees-classifiers-8507ac21d54b, Jul. 2019.

[38] Sunil Ray, Commonly used Machine Learning Algorithms (with Python and R Codes), https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/,Sep. 2017.

[39] S. A. Kaur Guneet, "Predict Chronic Kidney Disease using Data Mining Algorithms in Hadoop," international J. Adv. Comput. Eng. Netw. , vol. 5, no. 6, pp. 1–5, 2017.

[40] J. Joshi, R. Doshi, and J. Patel, "Diagnosis and Prognosis Breast Cancer Using Classification Rules," *Int. J. Eng. Res. Gen. Sci.*, vol. 2, no. 6, pp. 315–323, 2014, [Online]. Available: www.ijergs.org.

[41] V. Chaurasia and S. Pal, "Data mining techniques: To predict and resolve breast cancer survivability," *Int. J. Comput. Sci. Mob. Comput. IJCSMC*, vol. 3, p. 15, 2017.

[42] A. Vikas Chaurasia and I. Saurabh Pal, "Data Mining Approach to Detect Heart Dieses," *Int. J. Adv. Comput. Sci. Inf. Technol.*, vol. 2, no. 4, pp. 2296–1739, 2013, [Online]. Available: http://ssrn.com/abstract=2376653.

[43] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[44] D. S. P. Jaymin Patel, Prof.TejalUpadhyay, "Heart Disease Prediction Using Machine learning and Data Mining Technique," *IJCSC*, vol. 7, no. March, pp. 129–137, 2016, doi: 10.090592/IJCSC.2016.018.

[45] L. Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm," *Int. J. Biol. Med. Sci.*, vol. 3, no. 3, pp. 157–160, 2008.

[46] J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," *Int. J. Comput. Appl.*, vol. 17, no. 8, pp. 43–48, Mar. 2011, doi: 10.5120/2237-2860.

## AUTHORS

**Khalid Amen** is a System Engineering and Computer Science PhD student in the Electrical and Computer Engineering department, Oakland University, Rochester, MI, USA.

**Dr. Mohammed Zohdy** is a professor in the Electrical and Computer Engineering department, Oakland University, Rochester, MI, USA.

**Dr. Mohammed Mahmoud** is a professor in the Computer Science and Engineering department, Oakland University, Rochester, MI, USA.

# LEBANON UPRISING: A THOROUGH STUDY OF LEBANESE TWEETS

Reda Khalaf and Mireille Makary

Department of Computer Science and Information Technology,
Lebanese International University, Beirut, Lebanon

## *ABSTRACT*

*Recent studies showed a huge interest in social networks sentiment analysis such as Twitter, to study how the users feel about a certain topic. In this paper, we conducted a sentiment analysis study for the tweets in spoken Lebanese Arabic related to the Lebanon Uprising hash tag (#لبنان_ينتفض), which was trending upon a socio-economic revolution that started in October, using different machine learning algorithms. The dataset was manually labelled to measure the precision and recall metrics and to compare between the different algorithms. Furthermore, the work completed in this paper provides two more contributions. The first is related to building a Lebanese – Modern Standard Arabic (فصحة) mapping dictionary and the second is an attempt to detect sarcastic and funny emotions in the tweets using emojis. The results we obtained seem satisfactory especially considering that there was no previous or similar work done involving Lebanese Arabic tweets, to our knowledge.*

## *KEYWORDS*

*Lebanese Arabic tweets, sentiment analysis, machine learning, emotions, emojis*

## 1. INTRODUCTION

Nowadays, micro blogging services such as Facebook and Twitter are considered essential communication tools between people to share their opinions about a certain topic and spread information, and it can all be done in real-time [1]. As published on Statista website, by J. Clement, according to recent social media industry figures, Twitter currently ranks as one of the leading social networks worldwide based on active users. As of the fourth quarter of 2019, Twitter had 152 million monetizable daily active users worldwide. In Lebanon, and as shown by stat counter - GlobalStats1 for this year, Twitter was mostly used between October and November 2019 and then again between March and April 2020, however Facebook remains the most used social media platform by the Lebanese users. We chose to conduct our analysis using tweets since it was easier to collect the ones related to Lebanon Uprising topic, while on Facebook, it will be harder to detect the posts, the images that include text about the topic without using any hashtag. But it will definitely be interesting to compare between the two networks in further studies. October 17 was the date when a social-economic revolution started in Lebanon, and users became more active on social media, that could be the interpretation of having the peak of usage of Twitter between October and November. Given the fact that no previous study has been made to tweets in Lebanese Arabic dialect, we decided to conduct a sentiment analysis study of the spoken Lebanese Arabic tweets related to the Lebanon Uprising hashtag (#لبنان_ينتفض) which was the trending hashtag during the revolution.

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 225-233, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101119

The Arabic language is one of the top five spoken languages in the world [2]. Sentiment Analysis (SA) in Arabic could be a very challenging task since it is rich morphologically and there is always a difference between the formal written Arabic and the daily spoken one [3]. Sentiment analysis also known as opinion mining is a challenging natural language processing or text mining problem [4]. Most research studies treat sentiment analysis as a text classification problem where a particular text is classified as positive, negative or neutral opinion and this process can be automated through the use of machine learning algorithms.

From sentiment analysis, we can study emotions, which are closely related to sentiments, and which are usually subjective feelings and thoughts [5]. According to the study presented in [6], there are six primary emotions shared by people: love, joy, surprise, anger, sadness, and fear, which can be sub-divided into many secondary and tertiary emotions. These emotions can vary in intensity as well. When posting on social media, users frequently use emojis to express their emotions. And therefore, some studies covered the possibility of detecting a particular emotion through certain emojis. We will discuss them in more details in the upcoming sections. The remainder of the paper is divided as follows: the next section covers some of the related work completed in sentiment analysis for Modern Standard Arabic (MSA) tweets and some Jordanian or Saudi dialect tweets. Section 3 describes the experimental design which includes the dataset collected and the tools used, we also go through the manual annotation process for the tweets, the preprocessing steps we applied and then we go through the machine learning algorithms we used to train and predict the sentiments in the tweets. In section 4, we go through the experiments we conducted and compare between the results obtained. We report the accuracy, precision and recall metrics values. We discuss the hypothesis related to automatically detecting sarcastic and funny tweets based on emojis, we report the outcome of the experiments we applied. Since we collected tweets between two different periods, we compared between the users who were active in October and those active between May and August in attempt to study if new users became more involved or if the same users were still using the Lebanon Uprising hashtag. In the last section, we provide a conclusion of the experiments and we provide a future direction for the work.

## 2. RELATED WORK

SA for Arabic tweets has been an active field for quite some time now, especially considering that it is the native language for 22 countries [7]. Authors in [8] performed opinion mining for tweets targeting unemployment in Saudi Arabia and they faced the challenges related to Saudi dialects compared to MSA, they applied preprocessing techniques to raw data and then used supervised machine learning techniques to analyse sentiments. The classification obtained was satisfying. Another contribution in the field was presented in the Arabic sentiment analysis tool, a lexicon that maps Jordanian Dialect to MSA and a lexicon for emoticons [9]. In their study, the authors collected around 350,000 tweets. Through crowd sourcing, they were able to label more than 25,000 tweets, and then three different machine learning classifiers were used for sentiment analysis. The best accuracy achieved in the experiments the authors reported was obtained using SVM [10] and the score reported was 71.68% when compared to NB [11] and k- nearest neighbours (KNN) [12]. Abdullah et al. [13] performed a comprehensive study on sentiment analysis for Arabic tweets. They reported the challenges in the Arabic domain and not having enough studies that analyse people's opinion in Arabic language. The study demonstrated the need to perform more studies in different Arabic dialects. A hybrid method for sentiment analysis for Arabic tweets for Saudi dialect was also studied in [14]. They provided a two-way classification that classified the tweets as positive or negative, then a three-way classification that led to three classes: positive, negative and neutral and the four-way classification that added the "mixed" class to the three-way classes. The novelty of the work presented in this paper lies in the analysis conducted on tweets based on the spoken Lebanese dialect and in the use of emojis to detect emotions and not just opinion mining.

## 3. Experimental Design

In this section, we discuss the dataset, how it was annotated, the Lebanese – MSA dictionary built, the preprocessing steps and the machine learning classifiers chosen for the experiments.

### 3.1. Dataset

Abed Khooli collected 100K tweets with hashtag Lebanon Uprising (#لبنان_ينتفض) between October 18th and 21st, 2019 and as mentioned earlier that was following the socio-economic revolution that started on the 17th of October. They were collected in JSON format using workbench data2, an open source platform for data collection and they were posted on Kaggle3 which is another platform for sharing code and data related to data science work. Kaggle offers a wide range of public datasets and python notebooks that can be used for data analysis. The tweets in the Lebanon Uprising dataset were not all in Arabic language, so we had to filter them because our main focus and interest was in the tweets written in spoken Lebanese. Thus, after removing the duplicates, due to the retweets, cleaning the tweets, mostly those consisting of one word, we were left with 21,529 tweets. This dataset was manually labelled, and we will explain in the next section the platform we built for labelling, and it was used to train and build the machine learning models to predict sentiments. We will refer to this dataset as TDS (training dataset).

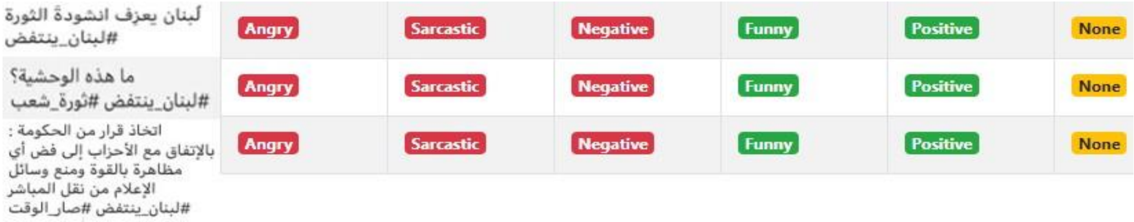
Then we started collecting tweets with the same hash tag starting May 2020, and until 8 August 2020 using workbench data as well and in the same JSON format. The same cleaning process was applied, and the total number of Lebanese tweets was 24,798 tweets, we will refer to this dataset as PDS (prediction dataset).

### 3.2. Manual Labelling of the TDS

In order to be able to build the machine learning models that could predict the sentiment for a Lebanese tweet, we had to train the machine learning classifiers. And in any supervised machine learning algorithm, the classifier has to be trained with labelled data so the accuracy can be measured, and parameters can be tuned to obtain the best model possible for the prediction. Since, to our knowledge, we could not find any labelled dataset using the Lebanese dialect, we decided to label the TDS manually. Hence, we built a web application that would allow the user to classify the tweet into one of the below categories:

1. Sarcastic
2. Angry
3. Negative
4. Neutral (none)
5. Funny
6. Positive

To note that the first 3 categories refer to a negative opinion, the last two refer to a positive opinion. But since we needed to test our hypothesis related to detecting sarcastic or funny emotion through emojis, we kept a flag for the tweets that belong to these two categories. Five users participated in the labelling process. Below is a figure that shows what the platform looked like.



لبنان يعزف انشودةَ الثورة
#لبنان_ينتفض

ما هذه الوحشية؟
#لبنان_ينتفض #ثورة_شعب

اتخاذ قرار من الحكومة :
بالإتفاق مع الأحزاب إلى فض أي
مظاهرة بالقوة ومنع وسائل
الإعلام من نقل المباشر
#لبنان_ينتفض #صار_الوقت

Figure 1. Manual labelling web application

## 3.3. Lebanese – MSA Mapping

One of the main steps of the pre processing included mapping words from the spoken Lebanese to MSA. We needed to make sure that the same word in MSA that could be written in different forms in the Lebanese Arabic is being considered the same when training the classifiers. For that reason, we built a dictionary that maps between the two and we tried to cover as many words included in the tweets as possible. We cannot claim that the work we did is fully complete, but it definitely covered most of the words we could find. And this step could be crucial to apply before we move to the stemming and stop words removal step. A glimpse of the content of the dictionary is shown in Table 1 below.

Table 1. Lebanese – MSA Mapping

| Lebanese word | MSA Word | English meaning |
|---|---|---|
| تحرفوا \ تحرفو\<br>تحرف \ بتحرف \ تحرفوا\ يحرفو | تحريف | twist |
| يأسو \ يأست \ ييأسوا\ ييأسن<br>\يأسون<br>\ييأست | يأس | depression |
| تعطو \يعطو\ يعطوا<br>\تعطوا | اعطاء | giving |

## 3.4. Preprocessing

The scripts we used were all written in python and therefore we made use of the available packages to preprocess the text such as removing all the links from the tweets, the emojis, the punctuations, the Arabic stopwords, and then stemming each word using Snowball4 stemmer that finds the stem or the root of the word after it was mapped to the MSA format.

## 3.5. Supervised Machine Learning Classifiers

We have selected five supervised machine learning (ML) algorithms to train and use for prediction of sentiments. These algorithms are usually used in case of text classification and for the implementation we used the classifiers implemented in scikit-learn5. The algorithms we tested were the Naïve Bayes – MultinomialNB6, the Support Vector Machines - SVM, the K-Nearest Neighbor – KNN, the SGD Classifier implemented in scikit-learn which is a linear classifier with the Stochastic Gradient Descent – SGD [15] training and the Logistic Regression [16].

## 4. EXPERIMENTS AND RESULTS

We considered the problem at hand as a two-way classification problem. The tweets in the training dataset, TDS, were classified as either positive or negative. Therefore, neutral tweets were discarded, and the angry and sarcastic tweets were considered as a part of the negative tweets. Their count is 8,943 tweets. The funny ones were counted as a part of the positive tweets and their total number was 12,586. We can see clearly that users were more positive than negative in the first few days of the revolution.

To start the experiments, we split the TDS into 85% for training and 15% for testing or as a cross validation step. For each of the machine learning algorithms, we used the grid search class in scikit-learn to look for the best combination of parameters that would lead to the best accuracy on both training and testing sets. We report here the best accuracy outcomes of the different classifiers. The results are detailed in Table 2 below.

Table 2. Accuracy for different machine learning classifiers

| **Machine Learning Classifier** | **Test Set Accuracy** |
|---|---|
| SVM | 74.11% |
| LogisticRegression | 73.65% |
| SGDClassifier | 73.77% |
| MultinomialNB | 73.40% |
| KNN | 66.93% |

As we can see, the SVM achieved the best accuracy score. The SGD Classifier, Logistic Regression and Mutlinomial NB classifiers had very similar accuracy scores on the test set. The KNN had the lowest accuracy value and that could be due to how the algorithms actually work and the combination of parameters that were more suited for the Lebanese dataset.

### 4.1. Precision and Recall

In addition to the accuracy measure, we computed the precision and recall metrics in an attempt to better understand how well each algorithm is performing. We define the precision metric as being the ratio of true positives (TP) over the sum of true and false (FP) positives. The recall metric is the ratio of the true positives over the sum of true positives and false negatives (FN).

Precision = TP/ (TP+FP)
Recall = TP / (TP + FN)

Accordingly, Table 3 shows how each algorithm performed in labelling the tweets as either positive or negative.

Table 3. Precision and recall metrics for different machine learning classifiers

| **Machine Learning Classifier** | **Precision** | **Recall** |
|---|---|---|
| SVM (C=1, gamma=1, kernel='rbf') | 0.759 | 0.822 |
| LogisticRegression (C=1) | 0.752 | 0.825 |
| SGDClassifier (alpha=0.0001, penalty=12) | 0.756 | 0.820 |
| MutlinomialNB (alpha=1) | **0.771** | 0.781 |
| KNN (neighbors=6, p=2) | 0.676 | **0.843** |

Based on the scores reported above, we can see that the best precision was achieved using the NB classifier with a value of 0.771 while the highest recall value was obtained using the KNN with a score of 0.843. The algorithms which provided a better accuracy than the NB and KNN had a close precision and recall scores with a difference less than 0.019 for the precision and less than 0.062 for the recall. In all cases, the numbers seem satisfactory when compared to the other Arabic SA studies described in the related work section.

## 4.2. From Sentiment to Emotions using Emojis

As mentioned in [5], emotions are related to sentiments, but they often express a more subjective opinion. Users tend sometimes to use emojis to express their emotions. A recent study exploited emojis for sarcasm detection [17]. The authors showed that the usage of Face with tongue out 😜 emoji is the highest among the sarcastic comments. The Face with tears of Joy 😂, Loud crying face, Grinning and Pouting face are the three specific emojis that are most frequently used with non-sarcastic comments.

So, we considered the study to do some statistics related to the use of emojis in the TDS and we found the results detailed in Table 4.

Table 4. Statistics related to emojis in tweets

| Label | Total Number of tweets | Tweets contain emojis | Angry Emoji | Sad Emoji | Happy Emoji | Other emojis |
|---|---|---|---|---|---|---|
| Angry | 2,341 | 338 | 92 | 56 | 123 | 166 |
| Sarcastic | 1,803 | 460 | 51 | 85 | 296 | 141 |
| Negative | 4,798 | 626 | 108 | 146 | 255 | 286 |
| Funny | 1,456 | 659 | 69 | 75 | 526 | 148 |
| Positive | 11,130 | 3,302 | 556 | 251 | 1,957 | 1,652 |

Looking at the numbers in Table 4, we can see that emojis were not used frequently in the tweets, however, those labelled as "Funny" had the highest percentage of tweets using emojis and that is ~45%. We looked for particular emojis like the Face with tears of Joy, the face with tongue out and winking eye since they seem to be very used according to [17]. We noticed that 138 tweets out of the 460 sarcastic ones contained the face with tears of joy and 243 out of the 659 funny ones contained the same emoji, that was the highest count for emojis in these two classes. Thus, we hypothesized that when the classifiers will predict negative tweets and in case these tweets contained the face with tears of joy, we will label the tweet as sarcastic, in case the tweet was classified as positive, and had the same emoji, it will be labelled as funny. Validating our hypothesis was done on the PDS, by predicting the sentiment and trying to detect sarcastic and funny ones after prediction.

## 4.3. Prediction on PDS

As described in section 3.1, the PDS consists of 24K+ tweets. It is worth noting that this number of tweets was collected over a duration of 4 months, while nearly the same number was collected in just 4 days at the beginning of the revolution. The same preprocessing steps applied to the TDS were applied to the tweets in PDS. We used the models with the best accuracy reported in Table 3 to label the tweets as either positive or negative. Among the ones predicted as negative, we detected the tweets that contained the face with tears of joy, and we labelled them as sarcastic, the ones predicted positive and that contained that same emoji, were considered funny. Table 5 shows the numbers of tweets predicted using each classifier.

Table 5. PDS sentiment prediction and emotion detection

| ML algorithm | Predicted Positive | Predicted Negative | Detected as Sarcastic | Detected as Funny |
|---|---|---|---|---|
| SVM | 9,100 | 15,698 | 627 | 68 |
| LogisticRegression | 7,446 | 17,352 | 644 | 51 |
| SGDClassifier | 8,638 | 16,160 | 634 | 61 |
| MultinomialNB | 9,955 | 14,843 | 629 | 66 |
| KNN | 15,928 | 8,870 | 585 | 110 |

We can see that in almost all the classifiers, except the KNN, the number of tweets predicted negative is higher than the number predicted positive. We can consider an accuracy of ~74% for the SVM, nearly 73% for the Logistic Regression, SGD Classifier and Multinomial NB and ~67% for the KNN, with a precision greater than 70% for all and a recall greater than 78% as measured on the test set in the TDS.

We did perform a manual verification for the tweets detected as sarcastic or funny. The number of true positive in each case is listed in Table 6.

Table 6. Sarcastic and Funny tweets Validation

| | Sarcastic | | | Funny | | |
|---|---|---|---|---|---|---|
| ML Algorithm | Predicted | True Positive - TP | Accuracy | Predicted | TP | Accuracy |
| SVM | 627 | 404 | 64.43% | 68 | 53 | 77.94% |
| LogisticRegression | 644 | 412 | 63.97% | 51 | 44 | 68.62% |
| SGDClassifier | 634 | 408 | 64.35% | 61 | 51 | 83.60% |
| MultinomialNB | 629 | 405 | 64.38% | 66 | 52 | 78.78% |
| KNN | 585 | 368 | 62.90% | 110 | 57 | 51.81% |

The accuracy in the funny tweets seems better than the one computed for the sarcastic ones because their number is much less than the sarcastic ones. The overall accuracy seems acceptable considering that we are only building our assumption on the existence of one particular emoji in the tweet. It might not be enough for other cases, but it could be something to build on for future work.

## 4.4. Variation in Users

The last factor we studied was the number of users who tweeted and re-tweeted in both periods October and between May and August. So, we compared the users in both TDS and PDS, once including the re-tweets and after removing the re-tweets. Therefore, the numbers are as shown in Table 7 below.

Table 7. Comparison between the number of users in TDS and PDS

| | Including retweets | Excluding retweets |
|---|---|---|
| **Nb. Of users in TDS** | 38,322 | 9,316 |
| **Nb. Of users in PDS** | 20,850 | 5,406 |
| **Common users between TDS and PDS** | 3,425 | 617 |

From the reported numbers in Table 7, we can see that the number of users tweeting and using the hashtag Lebanon Uprising is reduced, that could indicate a change of interest in the topic for the users. Only 5% of the users tweeted in October and kept tweeting between May and August using the same hashtag while new users started using Lebanon Uprising hash tag sometime between May and August.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we showed the detailed experiments used to build a Lebanese tweets dataset using the Lebanon Uprising hashtag, how we manually labelled the tweets in the dataset and built a Lebanese – MSA mapping dictionary. This dataset was used to train a set of supervised machine learning classifiers which were then used to predict the sentiments of Lebanese tweets collected over a different period of time. We showed that the performance of the different classifiers was very similar. We have also proposed a hypothesis relating emojis to emotions and in particular sarcasm and funny emotions. We tested our hypothesis on the newly collected tweets, and we obtained satisfactory results. We then compared between the number of users on Twitter tweeting and using Lebanon Uprising hashtag in October and then between May and August. We showed that new users were tweeting starting May and that only 5% of the users were common between the two timeframes. As a future direction of the work, one can expand the classification from simply opinion mining to emotions detection using not just emojis, but also benefiting from deep neural networks (DNN) which are known to provide satisfactory results in such tasks and words embedding to relate the semantics of the tweets rather than simply considering them as bag of words.

## REFERENCES

[1] Kouloumpis, Efthymios., Wilson, T., Moore, J. (2011) Twitter sentiment analysis: the good the bad and the OMG! ICWSM 11, 538–541

[2] Ibrahim, H.S., Abdou, S.M., Gheith, M. (2015) Sentiment analysis for modern standard Arabic and colloquial. arXiv preprint, arXiv:1505.03105

[3] Haifa, Kfir & Azmi, Aqil. (2015). Arabic tweets sentiment analysis - A hybrid scheme. Journal of Information Science. 42. 10.1177/0165551515610513.

[4] Liu, Bing (2010) Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau)

[5] Wiesław, Wolny (2016) Emotion analysis of twitter data that use emoticons and emoji ideogram 25th International Conference On Information Systems Development (ISD2016 Katowice)

[6] Parrott, W (2001). Emotions in social psychology: Essential readings. Psychology Press.

[7] Korayem, M., Crandall, D., Abdul-Mageed, M (2012) Subjectivity and sentiment analysis of Arabic: a survey. In: Hassanien, A.E., Salem, A.-B.M., Ramadan, R., Kim, T. (eds.) AMLTA 2012. CCIS, vol. 322, pp. 128–139. Springer, Heidelberg. doi:10.1007/978-3-642-35326-0 14

[8] Alwakid, Ghadah, Osman, Taha, Hughes-Roberts, Thomas (2017) Challenges in Sentiment Analysis for Arabic Social Networks 3rd International Conference on Arabic Computational Linguistics, ACLing 2017, Dubai, UAE

[9] Duwairi, Rehab & Marji, Raed & Sha'ban, Narmeen & Rushaidat, Sally. (2014). Sentiment Analysis in Arabic tweets. 2014 5th International Conference on Information and Communication Systems, ICICS 2014. 1-6. 10.1109/IACS.2014.6841964.

[10] Boser, Bernhard E., Guyon Isabelle M., & Vapnik, Vladimir N. (1992). A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory (COLT '92). Association for Computing Machinery, New York, NY, USA, 144– 152. DOI:https://doi.org/10.1145/130385.130401

[11] Lewis, David. (1998) Naive Bayes at forty: the independence assumption in information retrieval. In Machine Learning: ECML-98, Proceedings of the 10th European Conference on Machine Learning, Chemnitz, Germany (pp. 4–15). Berlin: Springer.

[12] Fix, Evelyn, and J. L. Hodges. (1989) Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. International Statistical Review / Revue Internationale De Statistique 57, no. 3: 238-47. Accessed August 26, 2020. doi:10.2307/1403797.

[13] Abdullah, Malak & Hadzikadic, Mirsad. (2017). Sentiment Analysis on Arabic Tweets: Challenges to Dissecting the Language. 191-202. 10.1007/978-3-319-58562-8_15.

[14] Al-Twairesh, Nora, Hend Al-Khalifa, AbdulMalik Alsalman, and Yousef Al-Ohali. (2018) Sentiment analysis of arabic tweets: Feature engineering and a hybrid approach. arXiv preprint arXiv:1805.08533.

[15] Ruder, Sebastian. (2016). An overview of gradient descent optimization algorithms. ArXiv Preprint ArXiv:1609.04747.

[16] LogisticRegression https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

[17] Subramanian, Jayashree & Sridharan, Varun & Shu, Kai & Liu, Huan. (2019). Exploiting Emojis for Sarcasm Detection.

## AUTHORS

**Reda Khalaf**: Master student in the Computer Science and Information Technology Department at the Lebanese International University

**Mireille Makary, PhD**: Lecturer in the Computer Science and Information Technology Department at the Lebanese International University with research focus on machine learning and information retrieval.

# SURVEY ON FEDERATED LEARNING TOWARDS PRIVACY PRESERVING AI

Sheela Raju Kurupathi[1] and Wolfgang Maass[1, 2]

[1]German Research Center for Artificial Intelligence, Saarbrücken, Germany
[2]Saarland University, Saarbrücken, Germany

## *ABSTRACT*

*One of the significant challenges of Artificial Intelligence (AI) and Machine learning models is to preserve data privacy and to ensure data security. Addressing this problem lead to the application of Federated Learning (FL) mechanism towards preserving data privacy. Preserving user privacy in the European Union (EU) has to abide by the General Data Protection Regulation (GDPR). Therefore, exploring the machine learning models for preserving data privacy has to take into consideration of GDPR. In this paper, we present in detail understanding of Federated Machine Learning, various federated architectures along with different privacy-preserving mechanisms. The main goal of this survey work is to highlight the existing privacy techniques and also propose applications of Federated Learning in Industries. Finally, we also depict how Federated Learning is an emerging area of future research that would bring a new era in AI and Machine learning.*

## *KEYWORDS*

*Federated Learning, Artificial Intelligence, Machine Learning, Privacy, Security, Distributed Learning.*

## 1. INTRODUCTION

Due to the emergence of AI and Machine learning over the past few decades, there has been significant progress in various domains like Robotics, Computer Vision and Gaming Applications. One of the major concerns is to preserve data privacy. Preserving the data privacy is of utmost importance in these days as the data is created in abundance every day. Data leaks on publicly available data and the private data of the companies lead to alarming increase towards data privacy. Utilizing the data which is isolated as data islands by maintaining specific privacy standards is very crucial for better data security. Misusing the personal data of the user may cause overhead to the user forcing him not to enclose his personal details. Even in the companies and industries, it is essential to protect data from data leaks as it would lead to grave consequences for the company. The data leaks, in turn, would affect the financial and commercial aspects of the company on a large scale leading to huge losses. One of the well-known standards for ensuring data privacy is the General Data Protection Regulation (GDPR) [1, 2] in the European Union. The GDPR was proposed in 2018 to ensure data privacy of every user, which in turn motivates to use AI and machine learning frameworks adapting to this standard while using data.

Many machine learning and AI models need sufficient data for training and to produce high-quality models. Although the models need to use user data if they need to build good prediction models for the user, there should be a way to ensure user privacy. Few organizations need to exchange data for working collaboratively for better performance of the companies, in turn,

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 235-253, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101120

ensuring the data privacy and confidentiality. In edge devices where users interact with different applications like in mobile phones, there is an ample amount of private data related to the user which is being exposed every day. To solve the problem of using data to train models, ensuring data privacy, we have a new approach known as Federated Learning (FL) [3]. The term Federated Learning (FL) was introduced by McMahan et al. [4] in 2016. Federated Learning is a collaborative Machine learning technique where the machine learning models are trained on edge devices (like mobiles) instead of a central server to ensure data privacy. The data is not exchanged between the devices. However, only the model updates (gradient updates) are sent to the server to build a global model using the aggregated gradients from all the computing edge devices. Thus, the server has no information about the raw data that the edge devices have been trained on, maximizing the data privacy of the users. Federated Learning has been evolving over the past few years due to the increasing demand for data privacy and security. It mitigates the risk of data privacy in comparison to centralized machine learning approaches. It also reduces the cost involved in traditional and centralized machine learning approaches.

The rest of the survey work is organized as follows: Section 2 details various related works in the area of Federated Learning. Section 3 details about Federated Learning, its working principle, training process, categorization of Federated Learning architectures along with various implementation frameworks, and Section 4 elaborates more about the privacy- preserving mechanisms in FL. Section 5 describes the application of Federated Learning in Industries along with its drawbacks, and in Section 6, we discuss in detail about the privacy- preserving aspect of Federated Learning. Section 7 concludes the survey work and suggests a few possible directions for the future area of research.

## 2. RELATED LITERATURE

As the data is vastly distributed over many devices, it is crucial for machine learning and AI models to access the data reliably for building efficient models. The goal of many research communities in the fields of Machine Learning, Artificial Intelligence, Cryptography and Distributed Systems has been to learn from the massively distributed data ensuring data security and privacy. Federated Learning is the recent trends for training the machine learning models in a decentralized way without having any information about the raw data except for the updated gradients from the client models. Federated Learning focusses on the edge and mobile computing [4, 5] devices and then extended its application to large scale production systems. Now industries are extensively using Federated Learning as part of their production systems for better product and profit generation on a large scale.

The data scattered everywhere as data islands need to be integrated on a large scale for useful application of AI models. It is challenging to integrate the data from these islands as it gives a cost overhead. Federated Learning has been the saviour for reducing the cost for data integration through execution of AI models on the data available on edge computing devices. Federated Learning is being used by Google in its Gboard mobile keyboard [6, 7, 8, 9, 10]. They also implemented a few of the features using Federated Learning in Android Messages [11] and Pixel phones [12]. Even Apple is using Federated Learning in iOS 13 [13], for various applications like the vocal classifier for “Hey Siri” [14] and QuickType keyboard. Other applications include Federated Learning for medical research [15] and the detection of hot words [16].

Currently, much of the research work is being focussed in FL, due to the privacy-preserving aspect of Federated Learning. Clifton and Vaidya proposed secure k-means [17], secure association mining rules [18], and a naive Bayes classifier [19] for vertically partitioned data. The authors of [20] implemented a privacy-preserving protocol using homomorphic encryption for applying linear regression on horizontally partitioned data. The authors of [21, 22] have proposed a linear regression approach for vertically partitioned data. FL directly solved the linear regression problem. The authors of [23] have approached the problem with Stochastic Gradient Descent (SGD) and also proposed privacy-preserving protocols for neural networks and logistic regression. The authors of [24] proposed a novel algorithm for association rules on horizontally partitioned data. Secure Support Vector Machines (SVM) algorithms have been implemented for horizontally partitioned data [25] and vertically partitioned data [26]. The authors of [27] proposed various secure protocols for multi-party linear regression and classification. The authors of [28] proposed efficient, secure multi-party gradient descent methods. All these works used Secure Multiparty Computation (SMC) [29, 30] for preserving privacy and ensuring security.

The authors of [31] proposed a secure logistic regression protocol based on homomorphic encryption. Shokri and Shmatikov [32] proposed training of neural networks on horizontally partitioned data with exchanges of updated parameters. The authors of [33] used homomorphic encryption to enhance the security of the entire system and preserve the privacy of gradients. With recent trends in machine learning and AI, privacy-preserving neural networks are also one of the research interest [34, 35, 36, 37]. Therefore, building a decentralized system with collaborative machine learning models and ensuring data privacy is one of the crucial aspects of many industries.

## 3. Federated Machine Learning

The term Federated Learning refers to a decentralized machine learning setting where all the participating clients train a shared global model without exposing the data to the central server. Only the model updates from each participating device are sent to the server. The model updates are then aggregated based on Federated Averaging mechanism [5] to obtain an efficient global model. Therefore, it is a collaborative machine learning where all the clients contribute the model updates to achieve a common learning objective. FL allows for smarter models, lower latency, and less power consumption, all while ensuring privacy. It is also used in distributed architectures where machine learning needs to be integrated into them.

### 3.1. Definitions

To understand the term Federated Learning, it is essential to know the terms distributed learning [38, 39], centralized and decentralized Federated Learning [5, 40, 41].

- **Distributed Machine Learning:** In distributed machine learning, we train a model on a large dataset. Here, the clients are computing nodes in a single cluster or datacenter. All the clients can access any part of the dataset. The data is distributed onto multiple computing nodes in a datacentre. Distributed learning aims at parallelization of computing power through the distribution of data or model.
- **Centralized Federated Learning:** It has a central server which is used to orchestrate the entire training process and coordinate all the participating nodes during the learning process. The central server is responsible for the selection of nodes initially before the training starts and the aggregation of the received model updates. The server may become a bottleneck here as all the selected nodes have to send updates to a single entity.
- **Fully Decentralized/ Peer-to-Peer Learning:** It has a peer-to-peer topology [42], where every participating client can talk to the other participating clients. It has a possibly dynamic connectivity graph structure without any central orchestration.
- **Decentralized Federated Learning:** In this Federated Learning setting, the computing nodes can coordinate between themselves to compute the global model. As the model

updates are exchanged only between interconnected nodes without the orchestration of the central server, this setting prevents single-point failures.

## 3.2. Federated Learning Life Cycle

Federated Learning ensures secure collaborative machine learning with the decentralization of data. It uses hub-and-spoke topology, with the hub representing a coordinating service provider and the spokes connecting to clients. Despite preserving privacy, it has many challenges like if the server fails, then the global update of the model would be difficult. For better performance of the FL systems, the federated system must be resilient to failures.

The clients participating in the FL process can be multiple clients or multiple organizations. Based on the participating client, we have two Federated Learning settings, namely Cross-device and Cross-silo Federated Learning [43]. In Cross-device Federated Learning, multiple clients like mobile devices are stateless and highly unreliable. Only a few of the clients would be available at any point in time, thus making the computation and communication difficult. In Cross-silo Federated Learning, the clients are different organizations (medical or financial domains) participate in the FL process. This setting is typically limited to a hundred organizations while Cross-device setting can have an extensively large number of clients. In both Cross-device and Cross-silo setting, the data is decentralized and local to each client. The server acts as a central authority for organizing the training process, and it never sees the raw data of the participating clients. In this paper, we mainly consider the cross-device federated setting for explaining the Federated Learning life cycle and the training process.

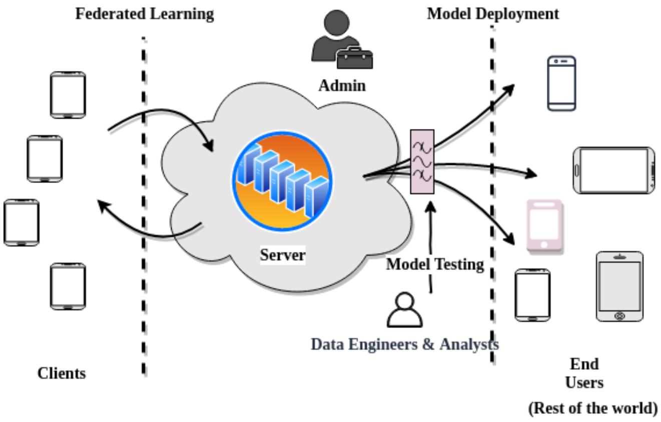


Figure 1. Life cycle of a Federated Learning (FL) model.

Initially, in the FL process, a model engineer develops a model for a particular application. In natural language processing, a domain expert may develop a prediction model for next word prediction to use in a virtual keyboard application. Figure 1 depicts the primary components and actors involved in the FL process. A typical workflow of the FL model can be realized, as shown in Figure 2. The life cycle of a Federated Learning (FL) model consists of six stages, as shown in Figure 2.

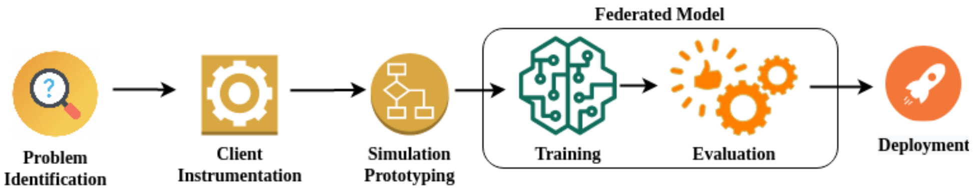


Figure 2. Stages in Life cycle of a Federated Learning (FL) model.

1. **Problem identification:** In this stage, the model engineer first identifies a problem that needs to be solved with Federated Learning.

2. **Client instrumentation:** The clients store the necessary training data locally. In a few cases, additional data or metadata might need to be maintained; for example, the labels for a supervised learning task.

3. **Simulation prototyping:** The model architectures are prototyped by the model engineer and then test's learning hyperparameters in a Federated Learning simulation using a proxy dataset.

4. **Federated model training:** Usually, all the federated training tasks are initiated to train different variations of the model. We could also use different optimization hyperparameters for further training.

5. **Federated model evaluation:** Once the Federated Learning tasks have been trained sufficiently, the models are then analyzed, and the best candidates are selected. The analysis depends on various metrics computed on standard datasets in the datacenter. Federated evaluation is carried out on local client data wherein the models are pushed to held-out clients.

6. **Deployment:** Once a good model is selected, it then goes through a standard model launch process, including live A/B testing, manual quality assurance and a staged rollout. The application owner sets the specific launch process for the selected model and is independent of how the model is trained.

### 3.3. Federated Learning Training Process

Federated Learning decouples the ability to do machine learning from the need to store the data in the central server or cloud. We could make use of local models to make predictions on mobile devices by bringing model training to the device as well. From Figure 3, the device first downloads the current model, improves it by learning from data on the phone, and then summarizes the changes as a small, focused update. Only this focused update is sent to the server through encrypted communication. Then immediately averaged with other user updates to improve the shared global model. Since all the training data remains on the device, no individual updates are stored in the server.

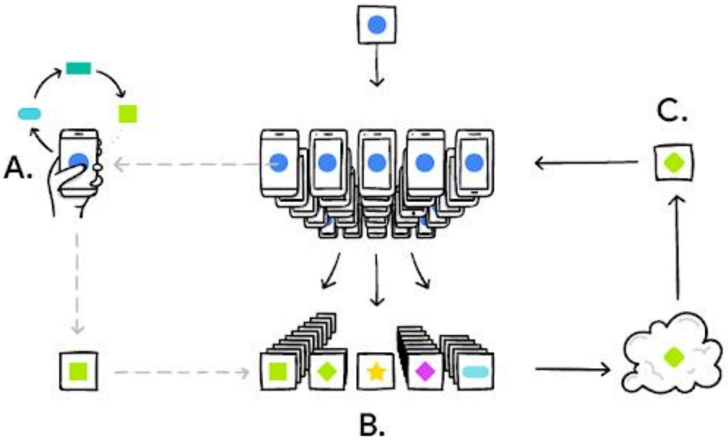


Figure 3. Process of Federated Learning (FL) model training.
(A) represents many users' updates are aggregated
(B) to form a consensus change (C) to the shared model, after which the procedure is repeated. [3]

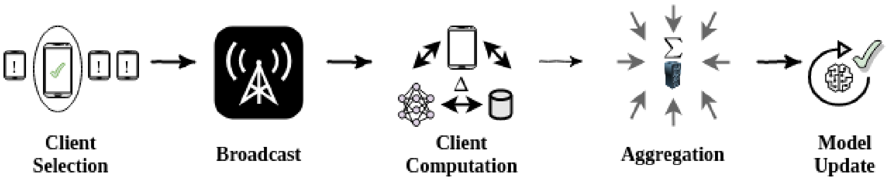


Figure 4. Stages in Federated Learning (FL) training process.

Federated Learning (FL) training process consists of five steps, as shown in Figure 4. A central server orchestrates the training process in FL setting, by iterative execution of the five steps shown in Figure 4 until the training process is stopped.

1. **Client selection:** First, the server samples from a set of participating clients meeting eligibility requirements. Mobile phones would only check in to the server if they are plugged in, and idle, to avoid impact on the device user.

2. **Broadcast:** In this step, the selected clients download the current model weights and a training program from the central server. For example, a training program can be a TensorFlow graph [44].

3. **Client computation:** In this step, each selected device locally computes a focused update to the model by executing the training program, like running SGD on the local data as in Federated Averaging algorithm.

4. **Aggregation:** The central server collects an aggregate of the device focused updates for efficiency. This step also includes other techniques like secure aggregation for added privacy, noise addition, a lossy compression of aggregates for communication efficiency and update clipping for differential privacy.

5. **Model update:** Finally, in this step, the server locally updates the shared model based on the aggregated update computed from all the participating clients in the current round. For

better performance of the central global machine learning model, FL relies on an iterative process of model updates.

## 3.4. Federated Learning Categorization

The data used for training the Federated Learning (FL) is non-identical as the data is on multiple devices. Based on how the data is distributed across multiple participating devices in the Federated Learning (FL) process, we classify FL into three different categories as Horizontal FL, Vertical FL and Federated Transfer Learning. The central authority to execute the final global update of the model based on the model updates from the clients plays a vital role in FL. Based on whether there is a central authority or not for coordination, FL can be classified as Centralized FL and Decentralized FL as already discussed. The clients participating in the FL process may be mobile devices or can be organizations. Few organizations need to collaborate to implement effective practical solutions which are profitable to the organizations involving as a whole. Therefore, we can classify FL into Cross-silo and Cross-device Federated Learning when the participating client is organization and mobile device, respectively. We will discuss in detail about FL categorization based on data partitioning.

Horizontal Federated Learning (HFL): In a Horizontal Federated Learning (HFL) system, only the central server can compromise the privacy of data participants. Horizontal Federated Learning (HFL), also known as sample-based Federated Learning (FL), is used in the scenarios in which datasets share the same feature space but different space in samples, as shown in Figure 5.

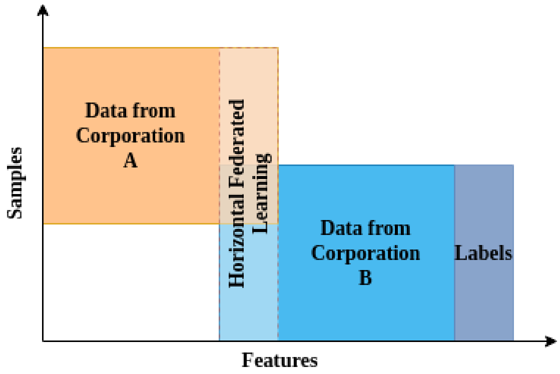


Figure 5. Horizontal Federated Learning (HFL) data partitioning.

For example, two regional banks that differ in user groups have a small intersection of users. However, as the business is very similar, they have the same feature spaces. The authors of [45] proposed a collaboratively deep-learning setting wherein participants train independently and share only subsets of parameter updates. Google proposed a Horizontal Federated Learning (HFL) solution for Android phone model updates [46]. In this framework, a single user using an Android phone updates the model parameters locally and then uploads the parameters to the Android cloud. Thus, jointly training the centralized model together with other data owners. A secure aggregation was used to protect the privacy of aggregated user updates, as shown in [47]. The authors of [48] use homomorphic encryption for model parameter aggregation to provide security.

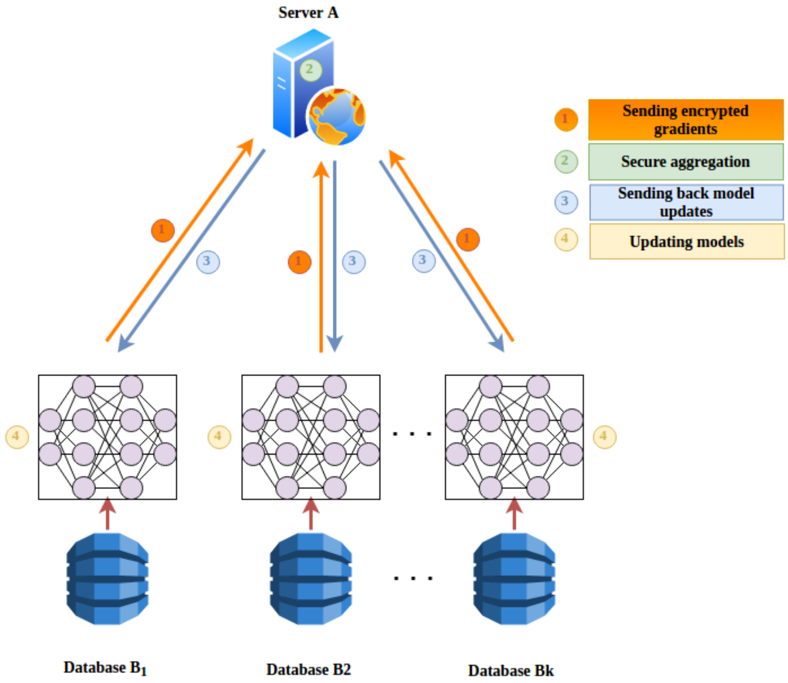


Figure 6. Horizontal Federated Learning (HFL) architecture.

Sample architecture for a horizontal Federated Learning (FL) system is shown in Figure 6. In this system, k participants with the same data structure collaboratively learn a machine-learning model using a parameter or cloud server. It assumes that there is no leakage of information from any participants to the server [48]. The training process of the HFL system usually contains the following four steps.

- **Step 1:** Initially, all the participants locally compute training gradients and then mask selected gradients with differential privacy [49], encryption [48], or secret sharing [47] techniques. Later these masked results are sent to the server.
- **Step 2:** The server then performs secure aggregation without learning any information about any participating client.
- **Step 3:** The server sends the aggregated results to all the participants.
- **Step 4:** Participants update their respective model with the decrypted gradients.

All the steps go through iterations until the loss function converges, thus completing the entire training process.

**Vertical Federated Learning (VFL):** Vertical Federated Learning (VFL) or feature-based Federated Learning (FL) is applicable to the cases in which two datasets share the same sample ID space but differ in feature space. For example, two different companies, like the bank and the other is an e-commerce company in the same city. Their user sets contain most of the residents of the area, and the intersection of their user space is enormous. However, their feature spaces are very different. Vertically Federated Learning (VFL) aggregates these different features and computes the gradients and training loss in a privacy-preserving manner. It finally builds a model with data from both parties collaboratively. At the end of the learning phase, each party holds only those model parameters associated with its features. Finally, at inference time, the two parties need to collaborate to generate output.

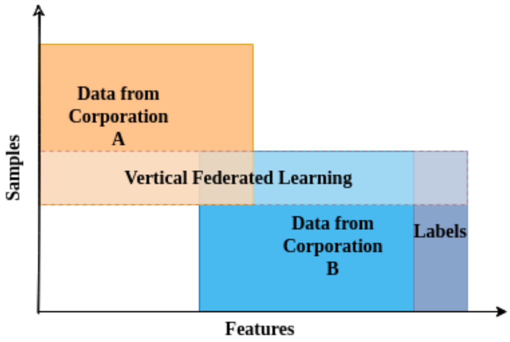


Figure 7. Vertical Federated Learning (VFL) data partitioning.

Many privacy-preserving machine learning algorithms have been proposed for vertically partitioned data, including association rule mining [50], cooperative statistical analysis [51], secure linear regression [52, 53], gradient descent [54] and classification [55]. Two companies A and B would like to train a machine-learning model jointly, and their business systems each have their data. For data privacy and security reasons, companies A and B cannot directly exchange data. During the training process, a third-party collaborator C is involved to ensure the privacy and confidentiality of the data. This Federated Learning (FL) system consists of two parts, as shown in Figure 8.

**Part 1. Encrypted entity alignment:** As the user groups of the two companies, A and B are different; the system uses the encryption-based user ID alignment techniques [56, 57] to confirm the standard users of both parties without A and B exposing their data. During the entity alignment, the system does not expose users that do not overlap with each other.

**Part 2. Encrypted model training:** Once the common entities are determined, we can use these common entities' data to train the machine-learning model.

The training process of VFL can be divided into the following four steps, as shown in Figure 8.

- **Step 1:** Initially, Collaborator C creates encryption pairs and sends a public key to A and B.
- **Step 2:** Both A and B encrypt and exchange the intermediate results for gradient and loss calculations.
- **Step 3:** Companies A and B compute the encrypted gradients and add a mask, respectively. Company B also computes an encrypted loss. Both A and B send encrypted values to C.
- **Step 4:** C decrypts and send the decrypted gradients and loss back to A and B. Then A and B unmask the gradients and update the model parameters accordingly.

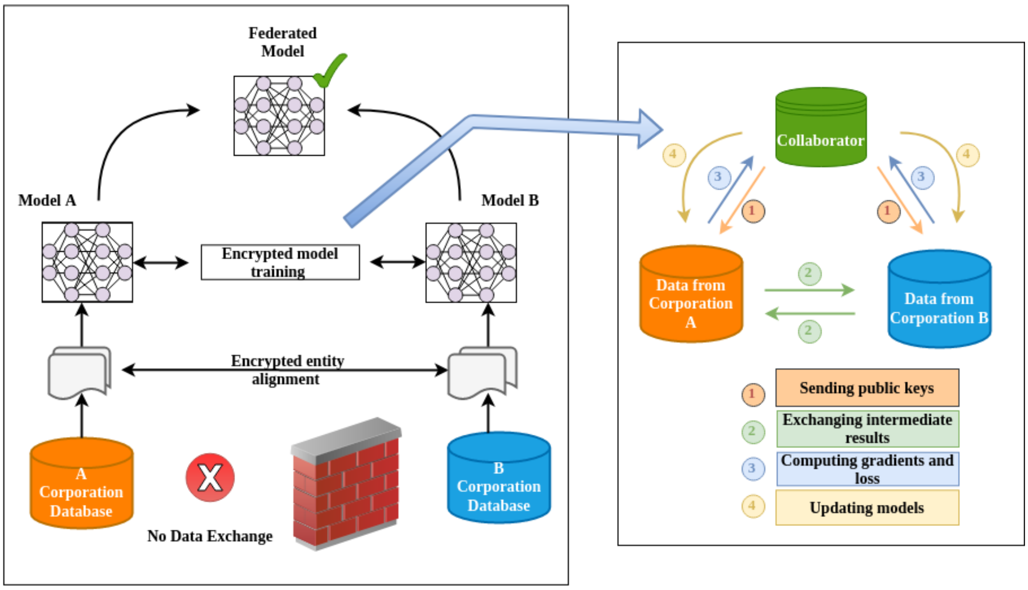


Figure 8. Vertical Federated Learning (VFL) architecture.

**Federated Transfer Learning (FTL):** Federated transfer learning is used in scenarios in which two datasets differ in both sample and also in feature space. FTL is a vital extension to the existing Federated Learning (FL) systems as it deals with the problems outside the scope of existing Federated Learning (FL) algorithms. For example, if one is a bank located in Russia and the other is an e-commerce company located in Ireland. A small portion of the feature space overlaps from both parties due to geographical restrictions. The architecture of VFL works only for the overlapping dataset. To extend it to the entire sample space, we introduce transfer learning. Typically, transfer learning involves learning a common representation between the features of parties A and B. It minimizes the errors in predicting the labels for the target-domain. At inference time, it still requires both parties to compute the prediction results. Thus, transfer-learning [58] techniques can be applied to provide solutions for the entire sample and feature space under a federated setting. Generally, a common representation is learnt between the two feature spaces using limited common sample sets and then later applied to obtain predictions for only one-side feature samples.

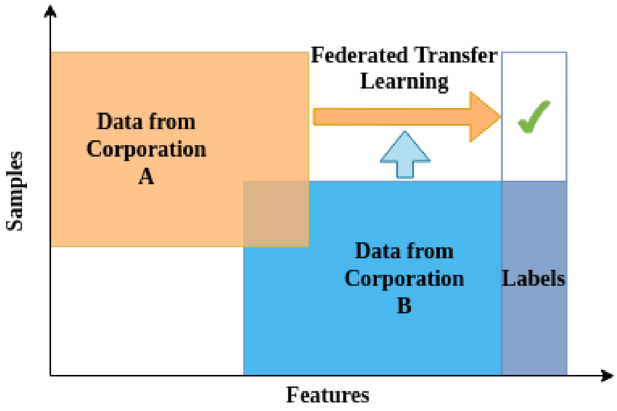


Figure 9. Federated Transfer Learning (FTL) data partitioning.

### 3.5. Federated Learning Implementation Frameworks

Federated Learning is difficult to implement and deploy in real life due to the heterogeneity in edge computing devices. These devices may have different programming languages, frameworks, and hardware configurations. There are many federated frameworks available to simulate FL algorithms. Few of the available tools and frameworks are TensorFlow Federated [59], PySyft [60], Federated AI Technology Enabler [61], PaddleFL [62], Leaf [63] and Clara Training Framework [64]. TensorFlow Federated (TFF) introduced by Google is an extensible, powerful framework for implementing Federated Learning (FL) research by simulating Federated Learning (FL) computations on realistic proxy datasets. It has Federated Core (FC) API is used for expressing new algorithms, and Federated Learning (FL) API can be used for implemented federated models. PySyft is another open-source library built for Federate Learning (FL) and preserving privacy. It was developed by the OpenMined community which combines these different tools for building secure and private machine learning models. It is built as an extension of well known DL libraries, such as PyTorch, Keras and Tensorflow. Using these popular deep learning frameworks, we can immediately begin to build privacy-preserving applications without having to learn a new Deep Learning framework. Thus, Federated Learning (FL) and other tools could be easily adopted in any application domain for preserving privacy. Therefore, these frameworks are designed to simulate FL in a server environment. However, they do not allow experimentation in a distributed mobile setting for a large number of clients. Another framework Leaf includes a set of open-source federated datasets, an evaluation framework, and a set of reference implementations using for practical federated environments.

## 4. PRIVACY MECHANISMS IN FEDERATED LEARNING

Privacy is one of the crucial properties of Federated Learning (FL). Therefore, it requires analysis and security models to provide privacy guarantees. In this section, we briefly review various privacy techniques for Federated Learning (FL).

**Secure Multiparty Computation (SMC):** SMC security models involve multiple parties and provide security proof in a well-defined simulation framework to guarantee that each party knows nothing except its input and output. Here, the parties have zero knowledge about other parties. Zero-knowledge is highly desirable, but this desired property usually requires highly complicated computation protocols and may not be achieved efficiently. In certain exceptional scenarios, disclosure of partial knowledge can be considered acceptable if security guarantees are provided. Therefore, it is possible to build a security model with SMC under lower security requirements in exchange for efficiency.

**Differential Privacy:** Differential Privacy involves adding noise to the data, or using generalisation methods to hide certain sensitive attributes until the third party cannot distinguish the individual, thereby making the data impossible to be restored to protect user privacy. The DP method is lossy as machine learning models are built after noise is injected, which can reduce much performance in prediction accuracy.

- ***Local Differential Privacy:*** Differential privacy can be achieved without requiring trust in a centralised server by having each client apply a differentially private transformation to their data before sharing it with the server.
- ***Distributed Differential Privacy:*** Here, the clients first compute and encode a minimal, focused report, and then send the encoded reports to a secure computation function, whose output is available to the central server. The output already satisfies differential privacy requirements by the time the central server can inspect it. The encoding is done

to help maintain privacy on the clients. This privacy-preserving technique can be implemented via secure aggregations and secure shuffling.

- ***Hybrid Differential Privacy:*** This combines multiple trust models by partitioning users by their trust model preferences. There are two options before the advent of HDP like most-trusting and the least trusting model.

**Homomorphic Encryption (HE):** Homomorphic encryption is adopted to protect user data privacy through an exchange of parameters under the encryption mechanism. Unlike differential privacy protection, the data and the model itself are not transmitted, nor can they be guessed by the other party's data. Homomorphic encryption (HE) schemes allow certain mathematical operations to be performed directly on ciphertexts, without any prior decryption. Homomorphic encryption is a powerful tool for enabling Multiparty Computation (MPC) by enabling a participant to compute functions on values, keeping the values hidden. Different variations of HE exist, ranging from general Fully Homomorphic Encryption (FHE) [65] to the more efficient levelled variants [66, 67, 68, 69]. There are also partially homomorphic schemes allowing either homomorphic multiplication or addition.

**Secure Aggregations:** Secure aggregation is functionality for n number of clients and a server. It enables each client to submit a tensor value such that the server learns just an aggregate function of the clients' values, generally the sum. The server learns just an unordered collection of the messages from all clients. The server cannot link any message to its sender beyond the information in the message itself. There are many research literature exploring secure aggregation in both the single-server setting using threshold homomorphic encryption [70, 71, 72], pairwise additive masking [73, 74, 75], and generic secure multi-party computation [76]. It is also used in the multiple non-colluding servers setting [77, 78, 79]. Secure aggregation can also be implemented using trusted execution environments as in [80].

**Secure Shuffling:** Secure shuffling can be considered as an instance of Secure Aggregation where the values are multiset-singletons, and the aggregation operation is multiset-sum. It is mostly the case that very different implementations provide the best performance for secure shuffling and secure aggregation. Secure shufflers have been studied in the context of secure multi-party computation [81, 82] and also in trusted computing [83].

**SecureBoost:** SecureBoost is a novel gradient-tree boosting algorithm in the setting of Federated Learning (FL). It consists of two main steps. First, it aligns the data under the privacy constraint. Second, it collaboratively learns a shared gradient-tree boosting model while keeping all the training data secure over multiple private parties. SecureBoost is beneficial as it provides the same level of accuracy in comparison to non-privacy-preserving approach while at the same time, reveal zero information of each private data provider. The SecureBoost framework is as accurate as other non-federated gradient tree-boosting algorithms that bring the data into one place and is theoretically proven.

**Private Information Retrieval (PIR):** PIR is functionality for one client and one server. It enables the client to download an entry from a server-hosted database such that the server gains no information about which entry the client has requested. MPC approaches to PIR can be put into two main categories: computational PIR (cPIR), in which a single party can execute the entire server-side of the protocol [84], and information theoretic PIR (itPIR), in which multiple non-colluding parties are required to execute the server-side of the protocol [85]. Computational PIR has a very high computational cost [86], while the non-colluding parties setting has been complex to achieve in industrial scenarios. Recently, the results on PIR have shown dramatic reductions in the computational cost through the use of lattice-based cryptosystems [87, 88, 89].

It shows how to construct communication-efficient PIR on a single-server by leveraging side information available at the user [90]. Research works propose to leverage local client state to speed up PIR. Patel et al. [91] showed a practical hybrid PIR scheme on a single server was implemented and validated. Corrigan-Gibbs and Kogan [92] present protocols for PIR with sublinear online time by working in an offline/online model. During an offline phase, clients fetch information from the server(s) independent on the future query to be executed.

## 5. Applications in Industries and Limitations

Federated Learning (FL) is not only a technology standard but also a business model for many industries. When we consider the effects of big data, the first thing is to aggregate the data, compute the models through a remote processor, and then download the results for further use. In such cases, cloud computing comes into demand. With the increasing importance for data privacy and data security and a high relationship between a company's profits and its data, the cloud computing model has been challenged. However, the business model of Federated Learning (FL) has provided a new paradigm for many applications of big data. When the isolated data by each institution fails to produce an ideal model, the mechanism of Federated Learning (FL) makes it possible for many institutions and enterprises to share a united global model without the exchange of data.

However, Federated Learning (FL) could make equitable rules for profits allocation using blockchain techniques. We believe that the establishment of the business model for data alliance and the technical mechanism for Federated Learning (FL) should be implemented together. Various standards for Federated Learning (FL) in many fields need to be put into use for the betterment of industries and enterprises. Industries could use Federated Learning mechanism for the resilience management where the computing or manufacturing devices are likely to fail due to the quality fails or manufacturing defects at later stages. It could affect the profits of industries on a large scale, especially in the pandemic situations wherein there should be manual intervention for the devices for parameter settings. It could be an area of application for Federated mechanism where the devices run efficiently even in case of failures and thus optimizing the profits on a global scale. Cross-silo Federated Learning applications can be seen in various domains including finance risk prediction for reinsurance [93], electronic health records mining [94], pharmaceuticals discovery [95], medical data segmentation [96], and smart manufacturing [97]. Commercial data platforms incorporating Federated Learning (FL) are in progress in various technology companies and smaller start-up companies.

Even though there are significant practical privacy improvements of Federated Learning over centralizing all the training data, there is still no formal guarantee of privacy in this baseline Federated Learning (FL) model. The significant challenges of the Federated Learning (FL) setting are non-Independent and Identically Distributed (IID) data, unbalanced, massively distributed, and limited communication. Each user generates quite different data, and thus the data is non-IID data. Due to the massive number of participating clients in the federated process, some of the users produce significantly more data than others, making it unbalanced. It is massively distributed, and therefore there are more mobile device owners than the average training samples on each device. Due to unstable, unreliable and asymmetric mobile network connections between the clients and the server, there is limited communication.

## 6. Discussion

Federated Learning (FL) embodies basic principles of focused data collection and minimization and can reduce many of the systemic privacy risks. Although there are many existing privacy-preserving techniques in a Federated Learning (FL) setting, they still do not offer much support to complete privacy of the system. An actively malicious adversary controlling the central server could lead to a large number of fake client devices as a "Sybil attack" [98]. Adversarial attacks include data poisoning [99], model evasion attacks [99, 100] and model update poisoning [101, 102], which degrade the model performance. Hitaj et al. [103] devised an attack based on GANs, which shows that record-level differential privacy is generally ineffective in Federated Learning (FL) systems. Balcan et al. [104] introduced a concept to add statistical noise only to a subset of data, but the resulting privacy in this scenario is dependent on the number of statistical queries required to learn the dataset. It is interesting to investigate various approaches to facilitate federated privacy mechanisms, which could then be integrated into many business models for scaling the profits. Therefore, Federated Learning (FL) mechanism motivates the development of novel and highly trusted models taking Federated Learning's unique computational model as the baseline.

## 7. CONCLUSIONS AND FUTURE SCOPE

The emphasis on data privacy and security with the isolation of data has become the next challenges for AI, but Federated Learning (FL) has emerged with the solution. It could establish a united model for multiple enterprises and institutions while local data is protected so that enterprises could work together on data security. Thus, Federated Learning (FL) provides a platform to build a cross-enterprise, cross-data, and cross-domain ecosphere for AI, Machine learning and big data. This paper generally introduces the basic working of Federated Learning (FL), various architectures, privacy-preserving techniques of Federated Learning (FL), and discusses its potential in industrial applications. In the near future, Federated Learning (FL) would break the barriers between industries and establish a new community, wherein the data and knowledge could be shared. It ensures the safety and the benefits would be equally distributed based on the contribution of each participant. Finally, the essence and need of AI would be brought to every corner of our lives through Federated Learning (FL).

### ACKNOWLEDGEMENTS

We thank DFKI SPAICER project for giving an opportunity and encouraging to do this survey work.

### REFERENCES

[1] Albrecht, J. P. (2016), How the gdpr will change the world. Eur. Data Prot. L. Rev. 2:287.

[2] Regulation, P. (2016), The general data protection regulation. European Commission. [Online]. Available: https://eur-lex. europa. eu/legal-content/EN/TXT

[3] H. B. McMahan & Daniel Ramage. (2017), "Federated learning: Collaborative machine learning without centralized training data", Google AI Blog. [Online]. Available: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

[4] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017), "Communication-efficient learning of deep networks from decentralized data", *In Artificial Intelligence and Statistics,* ppl. 1273-1282.

[5] H. B. McMahan, Eider Moore, Daniel Ramage, Seth Hampson, & Blaise Aguera y Arcas. (2017), "Communication-efficient learning of deep networks from decentralized data", *In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ppl 1273–1282.

[6] Sundar Pichai. (2019), Privacy Should Not Be a Luxury Good, *New York Times.*

[7] Swaroop Ramaswamy, Rajiv Mathews, Kanishka Rao, & Françoise Beaufays. (2019), "Federated learning for emoji prediction in a mobile keyboard", [Online]. Available: http://arxiv.org/abs/1906.04329

[8] Andrew Hard, Kanishka Rao, Rajiv Mathews, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, & Daniel Ramage. (2018), "Federated learning for mobile keyboard prediction", [Online]. Available: http://arxiv.org/abs/1811.03604

[9] Mingqing Chen, Rajiv Mathews, Tom Ouyang, & Françoise Beaufays. (2019), "Federated learning of out-of-vocabulary words". [Online]. Available: http://arxiv.org/abs/1903.10635

[10] Timothy Yang, Galen Andrew, Hubert Eichner, Haicheng Sun, Wei Li, Nicholas Kong, Daniel Ramage, & Françoise Beaufays. (2018), "Applied federated learning: Improving Google keyboard query suggestions", [Online]. Available: http://arxiv.org/abs/1812.02903

[11] support.google. (2019), Your chats stay private while Messages improves suggestions. [Online]. Available: https://support.google.com/messages/answer/9327902.

[12] ai.google. (2018), Under the hood of the Pixel 2: How AI is supercharging hardware. [Online]. Available: https://ai.google/stories/ai-in-hardware/

[13] Apple. (2019), Private Federated Learning, *NeurIPS 2019 Expo Talk Abstract.* [Online]. Available: ExpoConferences/2019/schedule?talk_id=40.

[14] Apple.(2019), Designing for privacy, *Apple WWDC.* [Online]. Available: https://developer.apple.com/videos/play/wwdc2019/708

[15] Walter de Brouwer. (2019), The federated future is ready for shipping. [Online]. Available: https://medium.com/@_doc_ai/the-federated-future-is-ready-for-shipping-d17ff40f43e3

[16] David Leroy, Alice Coucke, Thibaut Lavril, Thibault Gisselbrecht, & Joseph Dureau. (2018), "Federated learning for keyword spotting", [Online]. Available: http://arxiv.org/abs/1810.05512

[17] Jaideep Vaidya & Chris Clifton. (2003), "Privacy-preserving K-means clustering over vertically partitioned data", *In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03). ACM*, ppl 206–215. [Online]. Available: https://doi.org/10.1145/956750.956776

[18] Jaideep Vaidya & Chris Clifton. (2002), "Privacy preserving association rule mining in vertically partitioned data". *In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02). ACM,* ppl 639–644. [Online]. Available: https://doi.org/10.1145/775047.775142

[19] Jaideep Vaidya & Chris Clifton. (2004), "Privacy preserving naïve Bayes classifier for vertically partitioned data", *In Proceedings of the 4th SIAM Conference on Data Mining*, ppl 330–334.

[20] Murat Kantarcioglu & Chris Clifton. (2004), "Privacy-preserving distributed mining of association rules on horizontally partitioned data", *IEEE Trans. on Knowl. and Data Eng*, ppl 1026–1037. [Online]. Available: https://doi.org/10.1109/TKDE.2004.45

[21] Adrià Gascón, Phillipp Schoppmann, Borja Balle, Mariana Raykova, Jack Doerner, Samee Zahur, & David Evans. (2016), "Secure linear regression on vertically partitioned datasets", *IACR Cryptology*, 892.

[22] Irene Giacomelli, Somesh Jha, Marc Joye, C. David Page, & Kyonghwan Yoon. (2017), "Privacy-preserving ridge regression with only linearly-homomorphic encryption", *IACR Cryptology*, 979. [Online]. Available: https://eprint.iacr.org/2017/979

[23] Payman Mohassel & Yupeng Zhang. (2017), "SecureML: A system for scalable privacy-preserving machine learning". *IACR Cryptology*.

[24] Murat Kantarcioglu & Chris Clifton. (2004), "Privacy-preserving distributed mining of association rules on horizontally partitioned data", *IEEE Trans. on Knowl. and Data Eng*, ppl 1026–1037. [Online]. Available: https://doi.org/10.1109/TKDE.2004.45

[25] Hwanjo Yu, Xiaoqian Jiang, & Jaideep Vaidya. (2006), "Privacy-preserving SVM using nonlinear kernels on horizontally partitioned data", *In Proceedings of the 2006 ACM Symposium on Applied Computing (SAC'06)*, ppl 603–610. [Online]. Available: https://doi.org/10.1145/1141277.1141415

[26] Hwanjo Yu, Jaideep Vaidya, & Xiaoqian Jiang. (2006), "Privacy-preserving SVM classification on vertically partitioned data", *In Proceedings of the 10th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD'06),* ppl 647–656. [Online]. Available: https://doi.org/10.1007/11731139_74

[27] Wenliang Du, Yunghsiang Sam Han, & Shigang Chen. (2004), "Privacy-preserving multivariate statistical analysis: Linear regression and classification", *In SDM*, Vol. 4, ppl 222–233.

[28] Li Wan, Wee Keong Ng, Shuguo Han, & Vincent C. S. Lee. (2007), "Privacy-preservation for gradient descent methods", *In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07)*, ppl 775–783. [Online]. Available: https://doi.org/10.1145/1281192.1281275

[29] O. Goldreich, S. Micali, & A. Wigderson. (1987), "How to play any mental game", *In Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC'87),* ppl 218–229. [Online]. Available: https://doi.org/10.1145/28395.28420

[30] Andrew C. Yao. (1982), "Protocols for secure computations", *In Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (SFCS'82),* IEEE Computer Society, ppl 160–164. [Online]. Available: http://dl.acm.org/citation.cfm?id=1382436.1382751

[31] Yoshinori Aono, Takuya Hayashi, Le Trieu Phong, & Lihua Wang. (2016), "Scalable and secure logistic regression via homomorphic encryption", *In Proceedings of the 6th ACM Conference on Data and Application Security and Privacy (CODASPY'16),* ppl 142–144. [Online]. Available: https://doi.org/10.1145/2857705.2857731

[32] Reza Shokri & Vitaly Shmatikov. (2015). "Privacy-preserving deep learning", *In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS'15)* , ppl 1310–1321. [Online]. Available: https://doi.org/10.1145/2810103.2813687

[33] Le Trieu Phong, Yoshinori Aono, Takuya Hayashi, Lihua Wang, & Shiho Moriai. (2018). Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Trans. Information Forensics and Security* (2018), ppl 1333–1345.

[34] Florian Bourse, Michele Minelli, Matthias Minihold, & Pascal Paillier. (2017), "Fast homomorphic evaluation of deep discretized neural networks", *IACR Cryptology*, 1114.

[35] Nathan Dowlin, Ran Gilad-Bachrach, Kim Laine, Kristin Lauter, Michael Naehrig, & John Wernsing. (2016), "CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy". [Online]. Available: https://www.microsoft.com/en-us/research/publication/cryptonets-applying-neural-networks-to-encrypted-data-with-high-throughput-and-accuracy/

[36] Ehsan Hesamifard, Hassan Takabi, & Mehdi Ghasemi. (2017), "CryptoDL: Deep neural networks over encrypted data", *CoRR,* [Online]. Available: http://arxiv.org/abs/1711.05189

[37] Bita Darvish Rouhani, M. Sadegh Riazi, & Farinaz Koushanfar. (2017), "DeepSecure: Scalable provably-secure deep learning", *CoRR*. [Online]. Available: http://arxiv.org/abs/1705.08963

[38] Joost Verbraeken, Matthijs Wolting, Jonathan Katzy, & Jeroen Klop. (2020), " A survey on Distributed Machine Learning", ACM Computing Surveys, Vol. 53, [Online]. Available: https://doi.org/10.1145/3377454

[39] Diego Peteiro-Barral & Bertha Guijarro-Berdiñas. (2013), "A survey of methods for distributed machine learning", *Progress in Artificial Intelligence*, Vol. 2, No. 1, ppl 1–11.

[40] Abhijit Guha Roy, S. Siddiqui, S. Pölsterl, Nassir Navab, & Christian Wachinger. (2019), "BrainTorrent: A Peer-to-Peer Environment for Decentralized Federated Learning," arXiv: 1905.06731

[41] Chenghao Hu, Jingyan Jiang, & Zhi Wang, (2019), "Decentralized Federated Learning: A Segmented Gossip Approach", *1st International Workshop on Federated Machine Learning for User Privacy and Data Confidentiality (FML'19).*

[42] Róbert Ormándi, István Hegedűs, & Márk Jelasity. (2013), Gossip learning with linear models on fully distributed data. Concurrency and Computation: Practice and Experience, Vol. 25, No. 4, ppl 556–571.

[43] Peter Kairouz et. al., (2019), "Advances and Open Problems in Federated Learning", [Online]. Available: http://arxiv.org/abs/1912.04977

[44] Martín Abadi et. al., (2015), Large-scale machine learning on heterogeneous systems. [Online]. Available: https://www.tensorflow.org/

[45] Reza Shokri & Vitaly Shmatikov. (2015), "Privacy-preserving deep learning", *In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS'15),* ppl 1310–1321. [Online]. Available: https://doi.org/10.1145/2810103.2813687

[46] H. Brendan McMahan, Eider Moore, Daniel Ramage, & Blaise Agüera y Arcas. (2016), "Federated learning of deep networks using model averaging", CoRR, [Online]. Available: http://arxiv.org/abs/1602.05629

[47] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, & Karn Seth. (2017), "Practical secure aggregation for privacy-preserving machine learning", *In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS'17),* ppl 1175–1191. [Online]. Available: https://doi.org/10.1145/3133956.3133982

[48] Le Trieu Phong, Yoshinori Aono, Takuya Hayashi, Lihua Wang, & Shiho Moriai. (2018), "Privacy-preserving deep learning via additively homomorphic encryption", *IEEE Trans. Information Forensics and Security,* Vol. 13, No. 5, ppl 1333–1345.

[49] Reza Shokri & Vitaly Shmatikov. (2015), "Privacy-preserving deep learning", *In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS'15),* ppl 1310–1321. [Online]. Available: https://doi.org/10.1145/2810103.2813687

[50] Jaideep Vaidya & Chris Clifton. (2002), "Privacy preserving association rule mining in vertically partitioned data", *In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02),* ppl 639–644, [Online]. Available: https://doi.org/10.1145/775047.775142

[51] W. Du & M. Atallah. (2001), "Privacy-preserving cooperative statistical analysis", *In Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC'01),* IEEE 102. [Online]. Available: http://dl.acm.org/citation.cfm?id=872016.872181

[52] Adrià Gascón, Phillipp Schoppmann, Borja Balle, Mariana Raykova, Jack Doerner, Samee Zahur, & David Evans. (2016), "Secure linear regression on vertically partitioned datasets", *IACR Cryptology,* 892.

[53] Ashish P. Sanil, Alan F. Karr, Xiaodong Lin, & Jerome P. Reiter. (2004), "Privacy preserving regression modelling via distributed computation", *In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04),* ppl 677–682. [Online]. Available: https://doi.org/10.1145/1014052.1014139

[54] Li Wan, Wee Keong Ng, Shuguo Han, & Vincent C. S. Lee. (2007), "Privacy-preservation for gradient descent methods", I*n Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07),* ppl 775–783. [Online]. Available: https://doi.org/10.1145/1281192.1281275

[55] Wenliang Du, Yunghsiang Sam Han, & Shigang Chen. (2004), Privacy-preserving multivariate statistical analysis: Linear regression and classification. *In SDM,* Vol. 4. ppl 222–233.

[56] Gang Liang & Sudarshan S. Chawathe. (2004), "Privacy-preserving inter-database operations", *In International Conference on Intelligence and Security Informatics,* ppl 66–82.

[57] Amit P. Sheth & James A. Larson. (1990). "Federated database systems for managing distributed, heterogeneous, and autonomous databases", *ACM Comput. Surv.,* Vol. 22, No. 3, ppl 183–236. [Online]. Available: https://doi.org/10.1145/96602.96604

[58] Sinno Jialin Pan & Qiang Yang. (2010), "A survey on transfer learning", *IEEE Trans. Knowl. Data Eng.,* Vol. 22, No. 10, ppl 1345–1359. [Online]. Available: https://doi.org/10.1109/TKDE.2009.191

[59] The TFF Authors. (2019), TensorFlow Federated. [Online]. Available: https://www.tensorflow.org/federated

[60] Theo Ryffel, Andrew Trask, Morten Dahl, Bobby Wagner, Jason Mancuso, Daniel Rueckert, & Jonathan Passerat-Palmbach. (2018), A generic framework for privacy preserving deep learning.

[61] The FATE Authors. (2019) Federated AI technology enabler, [Online]. Available: https://www.fedai.org/

[62] The PaddleFL Authors. (2019) PaddleFL. [Online]. Available: https://github.com/PaddlePaddle/PaddleFL

[63] The Leaf Authors. (2019), Leaf. [Online]. Available: https://leaf.cmu.edu/

[64] The Clara Training Framework Authors. (2019), NVIDIA Clara. [Online]. Available: https://developer.nvidia.com/clara.

[65] Craig Gentry et al. (2009), Fully homomorphic encryption using ideal lattices. (2009), *In Stoc,* Vol. 9, ppl 169–178.

[66] Zvika Brakerski. Fully homomorphic encryption without modulus switching from classical gapsvp. (2012), *CRYPTO,* Vol. 7417, ppl 868–886.

[67] Junfeng Fan & Frederik Vercauteren. (2012), Somewhat practical fully homomorphic encryption, *IACR Cryptology*.

[68] Zvika Brakerski, Craig Gentry, & Vinod Vaikuntanathan. (2012), "(leveled) fully homomorphic encryption without bootstrapping", *In ITCS,* ppl 309–325.

[69] Jean-Sébastien Coron, Tancrède Lepoint, & Mehdi Tibouchi. (2014), "Scale-invariant fully homomorphic encryption over the integers", *In Public Key Cryptography,* Vol. 8383, ppl 311–328.

[70] Elaine Shi, HTH Chan, Eleanor Rieffel, Richard Chow, & Dawn Song , (2011), "Privacy-preserving aggregation of time-series data", *In Annual Network & Distributed System Security Symposium (NDSS).*

[71] Shai Halevi, Yehuda Lindell, & Benny Pinkas. (2011), "Secure computation on the web: Computing without simultaneous interaction", *In Annual Cryptology Conference*, ppl 132–150.

[72] T-H Hubert Chan, Elaine Shi, & Dawn Song. (2012), "Privacy-preserving stream aggregation with fault tolerance", *In International Conference on Financial Cryptography and Data Security*, ppl 200–214.

[73] Gergely Ács & Claude Castelluccia. (2011), "I have a DREAM!: DIfferentially PrivatE smart Metering", *In Proceedings of the 13th International Conference on Information Hiding, IH'11,* ppl 118–132. [Online]. Available: http://dl.acm.org/citation.cfm?id=2042445.2042457

[74] Slawomir Goryczka & Li Xiong. (2017), "A comprehensive comparison of multiparty secure additions with differential privacy", *IEEE Trans. Dependable Sec. Comput.*, Vol. 14, No. 5, ppl 463–477. [Online]. Available: https://doi.org/10.1109/TDSC.2015.2484326

[75] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, & Karn Seth, (2017), "Practical secure aggregation for privacy-preserving machine learning. *In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security"*, ppl 1175–1191.

[76] Martin Burkhart, Mario Strasser, Dilip Many, & Xenofontas Dimitropoulos. (2010), "SEPIA: Privacy-preserving aggregation of multi-domain network events and statistics", Network, Vol. 1.

[77] Peter Bogetoft, Dan Lund Christensen, Ivan Damgård, Martin Geisler, Thomas P. Jakobsen, Mikkel Krøigaard, Janus Dam Nielsen, Jesper Buus Nielsen, Kurt Nielsen, Jakob Pagter, Michael I. Schwartzbach, & Tomas Toft. (2009), "Secure multiparty computation goes live", *In Financial Cryptography*, Vol. 5628, ppl 325–343.

[78] Toshinori Araki, Jun Furukawa, Yehuda Lindell, Ariel Nof, & Kazuma Ohara. (2016), "High-throughput semi-honest secure three-party computation with an honest majority", *In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ppl 805–817.

[79] Henry Corrigan-Gibbs & Dan Boneh. (2017), "Prio: Private, robust, and scalable computation of aggregate statistics", *In 14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17)*, ppl 259–282.

[80] David Lie & Petros Maniatis. (2017), "Glimmers: Resolving the privacy/trust quagmire", *In Proceedings of the 16$^{th}$ Workshop on Hot Topics in Operating Systems*, ppl 94–99.

[81] Albert Kwon, David Lazar, Srinivas Devadas, & Bryan Ford. Riffle. (2016), *Proceedings on Privacy Enhancing Technologies*, Vol. 2, ppl 115–134.

[82] David Chaum. (1981), "Untraceable electronic mail, return addresses, and digital pseudonyms", *Communications of the ACM*, Vol. 24, No. 2.

[83] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnes, & Bernhard Seefeld. Prochlo (2017), "Strong privacy for analytics in the crowd", *In Proceedings of the 26th Symposium on Operating Systems Principles, SOSP '17,* ppl 441–459, [Online]. Available: http://doi.acm.org/10.1145/3132747.3132769

[84] Eyal Kushilevitz & Rafail Ostrovsky. (1997), "Replication is not needed: Single database, computationally-private information retrieval", *In Proc. of the 38th Annu. IEEE Symp. on Foundations of Computer Science*, ppl 364–373.

[85] Benny Chor, Eyal Kushilevitz, Oded Goldreich, & Madhu Sudan. (1998), "Private information retrieval", J. ACM, Vol .45, No. 6, ppl 965–981. [Online]. Available: http://doi.acm.org/10.1145/293347.293350

[86] Radu Sion & Bogdan Carbunar. (2007), "On the computational practicality of private information retrieval", *In Proceedings of the Network and Distributed Systems Security Symposium.*

[87] Femi Olumofin & Ian Goldberg. (2011), "Revisiting the computational practicality of private information retrieval", *In International Conference on Financial Cryptography and Data Security*, ppl 158–172.

[88] Sebastian Angel, Hao Chen, Kim Laine, & Srinath T. V. Setty. (2018), "PIR with compressed queries and amortized query processing", *In IEEE Symposium on Security and Privacy*, ppl 962–979.

[89] Craig Gentry & Shai Halevi. (2019), "Compressible FHE with applications to PIR", *In TCC*, Vol. 11892, ppl 438–464.

[90] S. Kadhe, B. Garcia, A. Heidarzadeh, S. E. Rouayheb, & A. Sprintson. (2017), "Private information retrieval with side information: The single server case", *In 55th Annual Allerton Conference on*

*Communication, Control, and Computing (Allerton)*, ppl 1099–1106. [Online]. Available: 10.1109/ALLERTON.2017.8262860

[91] Sarvar Patel, Giuseppe Persiano, & Kevin Yeo. (2018), "Private stateful information retrieval", *In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS '18*, ppl 1002–1019. [Online]. Available: http://doi.acm.org/10.1145/3243734.3243821

[92] Henry Corrigan-Gibbs & Dmitry Kogan. (2019), "Private information retrieval with sublinear online time", *IACR Cryptology.*

[93] WeBank. (2019), WeBank & Swiss re signed cooperation MOU, [Online]. Available: https://finance.yahoo.com/news/webank-swiss-signed-cooperation-mou-112300218.html.

[94] FeatureCloud. (2019), FeatureCloud: Our vision. [Online]. Available: https://featurecloud.eu/

[95] EU CORDIS. (2019), "Machine learning ledger orchestration for drug discovery", [Online]. Available: https://cordis.europa.eu/project/rcn/223634/factsheet/en?WT.mc_id=RSS-Feed&WT.rss_f=project&WT.rss_a=223634&WT.rss_ev=a

[96] ai.intel. (2019), Federated learning for medical imaging. [Online]. Available: https://www.intel.ai/federated-learning-for-medical-imaging/

[97] Musketeer. (2019), Musketeer: About, 2019. [Online]. Available: http://musketeer.eu/project/.

[98] John R. Douceur. (2002), "The sybil attack", *In Revised Papers from the First International Workshop on Peer-to-Peer Systems, IPTPS '01*, ppl 251–260, [Online]. Available: http://dl.acm.org/citation.cfm?id=646334.687813

[99] Battista Biggio, Blaine Nelson, & Pavel Laskov. (2012), "Poisoning attacks against support vector machines", *In Proceedings of the 29th International Coference on International Conference on Machine Learning, ICML'12*, ppl 1467–1474. [Online]. Available: http://dl.acm.org/citation.cfm?id=3042573.3042761

[100] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, & RobFergus. (2013), "Intriguing properties of neural networks", *ICLR.*

[101] Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, & Vitaly Shmatikov. (2018), "How to backdoor federated learning", [Online]. Available: http://arxiv.org/abs/1807.00459

[102] Arjun Nitin Bhagoji, Supriyo Chakraborty, Prateek Mittal, & Seraphin Calo. (2019), "Analyzing federated learning through an adversarial lens", *In Proceedings of the 36th International Conference on Machine Learning,* ppl 634–643.

[103] Peter Bogetoft, Dan Lund Christensen, Ivan Damgård, Martin Geisler, Thomas P. Jakobsen, Mikkel Kroigaard, Janus Dam Nielsen, Jesper Buus Nielsen, Kurt Nielsen, Jakob Pagter, Michael I. Schwartzbach, & Tomas Toft. (2009), "Secure multiparty computation goes live", *In Financial Cryptography*, Vol. 5628, ppl 325–343.

[104] Nitin Agrawal, Ali Shahin Shamsabadi, Matt J. Kusner, & Adrià Gascón. (2019), "QUOTIENT: two-party secure neural network training and prediction", *In Proceedings of the ACM Conference on Computer and Communication Security (CCS).*

## AUTHORS

Sheela Raju Kurupathi, Deep Learning Researcher, DFKI, Germany.

# A CASE STUDY ON MAINTAINABILITY OF OPEN SOURCE SOFTWARE SYSTEM JABREF

Denim Deshmukh, Ravi Theja Kataray and Tallari Rohith Girikshith

Blekinge Tekniska Högskola, Sweden

## *ABSTRACT*

*Maintainability is a major aspect of any software project; maintainability refers to the ease by which software can adapt to changes. There are various factors that affect the effort required for maintenance, in this paper we conducted a study to observe the extent up to which a metric could affect the maintainability of a software. We have considered various versions of JabRef and studied how maintainability of various packages had changed across versions. This is done using a framework called Goal Question Metric(GQM) which provides a systematic procedure to study various attributes of entities. Data of the attributes are collected using various Object-Oriented code metric tools which provide numerical data to compare the attributes between the versions. The data collected is visualized to answer the questions formulated which indeed tends to achieve the goal to identify the modules that are hard to maintain.*

## *KEYWORDS*

*Size, Structure, Complexity, Maintainability, Understandability & Goal Question Metric (GQM) approach.*

## 1. INTRODUCTION

This Software quality plays an important role while it comes to the development of the software product as it provides a brief ideology about the software product for the developers or the end users for further changes or modification that would be needed as the technology advances rapidly in the future. Hence, assessing the software quality is done using the measurement units like code size, coupling, maintainability, cohesion and structure of the product. Since software maintenance is critical, prospective developers should consider the maintainability aspect as a high priority during software system development [4]. In this study we have observed various metrics that assess various attributes over the Object-Oriented (OO) system JabRef. The software maintainability being an external attribute and is complicated and vital to calculate to know as it is affected by various factors. Maintainability is an important quality attribute, but it is difficult to estimate, because it involves making predictions about future changes that will occur in a software module, once it has been deployed [3]. As the Object-oriented based software systems are used more widely and often in all the software systems, they are different from the non-OO system due some programming concepts like inheritance, encapsulation. Hence, we cannot apply the well-known software metrics used to predict the non-OO software [3]. Many various studies have been done to provide the OO metrics that are reliable and effective to measure the maintainability of the product. In this paper a detailed analysis of various versions on an object-oriented project namely JabRef is conducted to assess the packages which are high to maintain.

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 255-273, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101121

### 1.1. Summery

This paper is divided into various sections Firstly analysis will be carried out by a Goal Question Metric (GQM) framework in which Goals to be achieved will be mentioned and various questions will be formulated questioning about the attributes of the entities and the metrics which will be best suited to answer the question will be mentioned, suitable scales for measuring metrics will be selected, tools which support object oriented metrics will be used to collect data of various versions, visualization methods like bar graphs , flow charts, tables will be uses to analyse the collected data. Later results of the analysis will be discussed along with the change log and related or similar work done on analysis of maintainability on various versions of JabRef.

## 2. RESEARCH METHODOLOGY

### 2.1. Study Type

Being an empirical study, Case study would be the suitable methodology. Case study is an empirical method aimed at investigating contemporary phenomena in their context [1]. This study is flexible in nature and could be conducted in real life context which makes it suitable for case study. The objective of this study is to explore for the least maintainable package which requires more effort and focus when compared to others. Such an objective can be achieved by analysing the various factors and metrics in a software.

### 2.2. System Considered for the Study

In this study we have considered an open source software called JabRef a cross-platform citation and reference management tool. The study is done on 10 recent/stable releases which are 5.0, 4.3, 4.2, 4.1, 4.0, 3.8, 3.7, 3.6, 3.5, and 3.4 versions. JabRef is developed in an object-oriented approach using Java Language. Source code, previous releases and related data was collected from its GitHub repository [7].

### 2.3. Study Design

Design for the case study and analyses in phases as following steps:

- Define a Goal Question Metric based on the template.
- Define Study type and provide justification for the specific study type.
- Define metrics to measure and justify the use of specific metrics.
- Define Question, Entities, attributes and relevant metric to give answer and result.
- Collect the data with the tool and analyse the data for results.
- Answer the question and provide analysis and overall evaluation.
- Discuss the reflection on the project and related works.

### 2.4. Data Collection

#### 2.4.1. Quantitative Data

The qualitative data is collected by using three metrics extraction tools which are metrics reloaded, lizard and maintainability index.

*Metrics Reloaded*: Metrics Reloaded is a Plugin to eclipse IDE and able to extract various metrics such as Ca, Ce, CBO, LOC, LCOM, DIT and WMC form Java source code. This tool can extract metrics at package level and support various suits such as Martin Suit and CK suit etc. This tool can extract various metrics and have an option to extract metrics in CSV (Comma-Separated Values) or XML format.

*Lizard*: Lizard is a metric extraction tool written in python. This tool is an open source tool and available in form of a python library but could extract metrics from the source code written in various programming languages. It is used to extract metrics such as Cyclomatic Complexity, Average cyclomatic complexity and Line of Code etc. This tool can be used at package, project as well as class level. It is a command line tool and supports CSV, text file-based output.

*Maintainability Index*: It is an extension of the lizard tool used to extract Maintainability Index at package, project and class level and support CSV output like the Lizard tool.

### 2.4.2. Quantitative Data

Change logs are analysed to find major changes in the project and to relate the report [14] with the data extraction.

## 3. GQM TREE

Goal-Question-Metric (GQM) approach to process metrics provides a framework for deriving measures from organization or business goals. According to Basili's GQM process, the first phase is to develop project goals [5][6].Then in the second phase we develop questions that define the goals as completely as possible in a quantifiable way. Then in the third phase we specify the metrics to be collected to answer the questions we defined. In the fourth phase we define data collection and in the last phase we collect data, validate it and provide feedback for corrective measure [5][6].

Table1 GQM Tree

| **Goal** | |
|---|---|
| Goal 1: To measure maintainability of various packages. | |
| Q1 | How maintainable the packages are with respect to size and structure? |
| Q2 | How does complexity of the packages change the maintainability of the product? |
| Q3 | To what extent does the understandability of each package provide the ease of maintenance of the product? |
| Q4 | How does the cohesive nature of packages influence the maintainability of the product? |
| Q5 | How does the maintainability change with instability in terms of coupling? |

Table2 Metrics Table

| The metrics that are considered in this study for the evaluation of the JabRef systems are as follows:**S.no** | **Metrics** | **Metrics Full Name** |
|---|---|---|
| M1 | DIT | Depth of Inheritance Tree |
| M2 | LOC | Line of Code |
| M3 | LCOM | Lack of Cohesion of Method |
| M4 | v(G) | Cyclomatic complexity |
| M5 | Ca | Afferent Coupling |
| M6 | Ce | Efferent coupling |
| M7 | MI | Maintainability Index |
| M8 | CR | Comment Ratio |
| M9 | DIT | Coupling between object classes |
| M10 | LOC | Weighted methods per class |

## Justification

1) DIT: This metric gives information about the inheritance within a class by measuring the number of nodes between the root node and given node within a class hierarchy, which indirectly describes the structure of the code. This metric was obtained in class level which was aggregated to package level by taking the mean values of all the classes in a specific package.

2) LOC: This is the simplest and most reliable metric to measure the size of a package.

3) LCOM: This metric measures the cohesion between methods of a class. Cohesion is the interdependence of methods in a class, which may affect the maintainability of a package and describes the structure of Source code. Hence this metric is considered to measure size and structure of a package. This metric was obtained in method level which was aggregated to package level by taking the mean values of all the methods of a class in a specific package.

4) Cyclomatic complexity v(G): This metric is essentially used to measure and denote the complexity of code. It measures the number of linearly independent paths in the program. It is also a basis for calculating the maintainability index. This metric can be obtained on class level and averaged on package level.

5) Ca: This metric measures the number of classes in other packages that are dependent on the classes in the package. This metric was obtained on package level. This metric is obtained on package level and shows the dependency of other packages on the package in scope. It is also known as incoming dependency. Afferent coupling will have a significant effect on maintainability.

6) Ce: This metric measures the dependency of the class in scope on the other classes of the package. This metric is obtained on package level and shows the dependency on other

packages in scope. It is also known as outgoing dependency. Efferent coupling has a significant effect on maintainability.

7) MI: This metric measures the maintainability and give value between zero to hundred, zero means very hard to maintain and hundred means easy to maintain. This metric measure maintainability using Cyclomatic complexity, Line of code and Halstead Volume. This provide an insight to the ease of maintenance with respect to complexity

8) CR: This metric defines the ratio of commented lines to lines of code of that particular package which helps to know exactly how many lines of code (LOC) are exactly available per package hence would be useful while estimating the efforts required during the maintenance of that particular package.

9) CBO: Also referred to as coupling between objects usually defines the coupling between the classes for each class to which they are coupled with. This indeed helps us to know the coupling between the packages and helps further while the product is modified or analysed for maintenance.

10) WMC: Also referred to as a weighted method per class usually describes the number of methods present per class which helps to know exactly how much effort will be required to maintain that particular class in maintenance phase which indeed helps to prevent excess cost and time.

## 3.1. Entities, Attribute and the Metrics

Table 3: Entities, Attribute and Metrics of Question Table

| Q1: How maintainable the packages are with respect to size and structure? | |
|---|---|
| Entity | Package |
| Attribute1 | Size |
| Attribute2 | Structure |
| Attribute Type | Internal |
| Metric1 | LOC |
| Metrics2 | LCOM, DIT, CBO |

| Q2: How does complexity of the packages change the maintainability of the product? | |
|---|---|
| Entity | Package |
| Attribute | Complexity |
| Attribute Type | Internal |
| Metric | CBO, WMC, v(G), DIT |

| Q3: To what extent does the understandability of each package provide the ease of maintenance of the product? | |
|---|---|
| Entity | Package |
| Attribute | Understandability |
| Attribute Type | Internal |
| Metric | Ca, Ce, CR, WMC |

| Q4: How does cohesive nature of packages influence the maintainability of the product? | |
|---|---|
| Entity | Package |
| Attribute | Cohesion |
| attribute Type | Internal |
| Metric | LCOM |

| Q5: How does the maintainability change with instability in terms of coupling? | |
|---|---|
| Entity | Package |
| Attribute | Coupling |
| Attribute Type | Internal |
| Metric | Ca, Ce |

**Justification**: Supportive aspects like why the certain metrics are selected for that internal attribute.

1) Understandability: It is an important attribute in software development process as it plays a vital role in maintaining the software product as understanding the models can help in modifying, analysing the system for later requirements and advancements which indeed save a lot of cost and time for avoiding the implementation errors[19]. Hence understandability is affected with various factors hence out of those many factors Comment ratio (CR), efferent coupling (Ce)and afferent coupling (Ca) are considered [16] in this study.

2) Size and Structure: Size of a package could be measured using lines of code(LOC) and number of functional points(FP), but by using these metrics one cannot draw any conclusion regarding the maintainability of the packages as there are other metrics related to the structure of the code which should be considered to provide a better picture on the overall maintenance of the packages. CK metric suite [13] provides various metrics which could depict the structure of the code like Depth Inheritance Tree(DIT)[11][12], Lack of Cohesion between Methods (LCOM)[10].

3) Complexity: Maintainability is highly dependent on the complexity of the software. Complexity is one of the most important factors affecting the overall maintainability of a software. Software complexity is described as the degree of difficulty in analysing, maintaining and modifying software[3].The cyclomatic complexity is a measurement of independent paths in a program and Average cyclomatic complexity for a package means that every class on average in a program has that much cyclomatic complexity. Similarly, the Weighted method per class provides us with an estimate for complexity for methods [4]. Coupling between object classes is the count of other classes being used which provide help in visualizing dependencies. The depth of inheritance provides a visualization ability to understand the abstraction level in the program and give a prospective of depth of abstraction in the program. Hence v(G), DIT, CBO and WMC are used to calculate complexity.

## 3.2. Scale Type

This section briefly discusses the scale type and its utilization in this study and in-detail justification is provided for selecting the scale type for the suitable metrics. Two different types of scales were considered in this paper.

*Ratio*: As this scale is the best suitable for the ratio values and the measurement of the metrics like CBO and comment ratio are taken under this scale type as their measurement values start from 0(zero) and increase gradually at equal intervals.

*Absolute*: This scale is usually associated with the number of the entity which is in the scope to measure. Hence, the measurement of the metrics like WMC, Ca and Ce are taken under this scale type as their measurement values are exact counted entities.

**Justification**

CBO: The best scale for measuring CBO is the absolute scale as it is the measure of the number of classes coupled to a class.

CR: The ratio scale is selected to represent the comment ratio as this is the best suitable scale to represent the ratio values better than any other scale types.

WMC: Ratio scale is taken into consideration for the weighted method per class as it is the best suitable for the arithmetic analysis and most of the measurement values have been ratio of complexity and number of elements which are indeed taken for ratio scale type.

DIT: According to the definition of DIT it is the count of number of nodes between root node to leaf node which could be measured using absolute scale.

LOC: Absolute scale is the best suitable scale to measure LOC as an exact count of the source line of code can be calculated at package and class level.

LCOM: Ratio scale is the best suitable scale to measure this metric as representing a link between methods and local variables is done for more than one class.

v(G): Cyclomatic complexity measures the number of linearly independent paths in a program. For calculating absolute values and representing on a scale Absolute scale is best to be used in this case.

Ca: The best scale fit for Afferent coupling is absolute since it represents the absolute numerical value of the number of packages on whom the package in scope is dependent.

Ce: The best scale fit for Efferent coupling is absolute since it represents the absolute numerical value of the number of packages who are dependent on the package in scope.

## 4. RESULT

The Study was formed according to the Goal Question Metrics tree and analysed to find the answer to the question for the goal. All 10 versions of JabRef were analysed and compared to form the results. An overview of JabRefprojects according to size with respect to number of classes and Line of code for every version is given in the table below.

Table 4. Overview Table

| Version | LOC | Number of Children(NOC) |
|---|---|---|
| 3.4 | 101246 | 183 |
| 3.5 | 102112 | 184 |
| 3.6 | 106703 | 239 |
| 3.7 | 113227 | 245 |
| 3.8 | 115359 | 250 |
| 4.0 | 123062 | 334 |
| 4.1 | 125653 | 340 |
| 4.2 | 126238 | 345 |
| 4.3 | 126563 | 373 |
| 4.5 | 119085 | 331 |

Based on the goal and analysing the software there were 10 main packages or called modules were found in the main directory and the analysis is done on these packages. These packages were significant and persistent throughout all the versions.

- net.sf.jabref.cli as cli
- net.sf.jabref.logic as logic
- net.sf.jabref.migration as migration
- net.sf.jabref.model as model
- net.sf.jabref.preferences as preferences
- net.sf.jabref.pdfimport as pdfimport
- net.sf.jabref.gui as gui
- net.sf.jabref.collab as collab
- net.sf.jabref.shared as shared

Some packages were there in the project which were either very small or merged in some other packages. There packages are listed below

- net.sf.jabref.architecture as architecture
- net.sf.jabref.external as external
- net.sf.jabref.sql as sql
- net.sf.jabref.util as util
- net.sf.jabref.exporter as exporter
- net.sf.jabref.specialfields as specialfields
- net.sf.jabref.bst as bst

## 4.1. Analysis of Size and Structure

In this section Q1 of GQM is answered providing a detailed analysis on the effect of size and structure on the maintainability of the product using DIT, LOC, LCOM metrics. The data is collected using Metrics Reloaded plugin installed to IntelliJ IDEA

DIT metric values obtained are depicted below in a tabular form:

Table 5. DIT table

| DIT | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **V 3.4** | **V3.5** | **V3.6** | **V 3.7** | **V 3.8** | **V 4** | **V 4.1** | **V 4.2** | **V 4.3** | **V 5** |
| Architecture | | | | | | 0 | 0 | 0 | 0 | 0 |
| Cli | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Logic | 1.1 | 1.1 | 1.11 | 1.15 | 1.14 | 1.15 | 1.15 | 1.16 | 1.16 | 1.15 |
| Migration | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Model | 1.13 | 1.12 | | 1.16 | 1.17 | 1.14 | 1.14 | 1.16 | 1.15 | 1.17 |
| Preferences | | | 1.43 | 1.3 | 1.3 | 1.27 | 1.27 | 1.27 | 1.27 | |
| Styletester | | | | | | | | | | 1 |
| Pdfimport | 2.25 | 2.25 | 2.25 | 2.25 | 2.25 | 1 | 1 | 1 | 1 | |
| Gui | 2.56 | 2.57 | 2.48 | 2.46 | 2.47 | 2.12 | 2.11 | 2.08 | 2.08 | 1.29 |
| Collab | 1.58 | 1.58 | 1.58 | 1.61 | 1.61 | 1.33 | | | | |
| Shared | | | 1.46 | 1.53 | 1.53 | 1.53 | 1.53 | | | |

Maintenance of a product or code is directly proportional to DIT value i.e., maintainability decreases with increase in DIT value from our observations package “model” has a gradually increasing DIT value and could be classified as high to maintain. “GUI” package has gradually decreasing values over versions, which indicates low maintenance and frequently updated versions. “pdfimport” package had a major update from V3.8 to V4.0 decreasing DIT values indicates low maintenance of the package. Rest of the packages are classified into moderate level maintenance and low maintenance categories depending on their mean values across the versions which is discussed in later sections of the paper.

Maintenance is directly proportional to the product size in terms of Line of Code(LOC). The metric value obtained are depicted below in tabular form:

Table 6. LOC table

| LOC | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSION** | **V 3.4** | **V3.5** | **V3.6** | **V3.7** | **V3.8** | **V 4** | **V4.1** | **V4.2** | **V 4.3** | **V 5** |
| Architecture | | | | | | 9 | 9 | 9 | 9 | 9 |
| Cli | 714 | 727 | 1050 | 1060 | 1062 | 1073 | 1073 | 946 | 946 | 885 |
| Logic | 25861 | 25717 | 35308 | 34295 | 34625 | 36205 | 37066 | 39369 | 39429 | 40609 |
| Migration | 324 | 324 | 344 | 343 | 412 | 488 | 530 | 650 | 669 | 564 |
| Model | 5463 | 5567 | 6108 | 11946 | 11749 | 12374 | 12346 | 12665 | 12678 | 12843 |
| Preferences | | | 1435 | 1750 | 1750 | 1953 | 1961 | 2088 | 2088 | 2336 |
| Styletester | | | | | | | | | | 561 |
| Pdfimport | 485 | 485 | 450 | 474 | 474 | 478 | 478 | 473 | 473 | |
| Gui | 41282 | 42265 | 48735 | 52236 | 53165 | 56494 | 58590 | 58365 | 58452 | 47818 |
| Collab | 1784 | 1784 | 1574 | 1493 | 1482 | 56594 | | | | |
| Shared | | | 1208 | 1659 | 1661 | 1662 | 1681 | | | |

In “Model” package LOC has increases drastically from V3.6 to V3.7 and has a considerable change in LOC on further packages, increasing the maintenance of the package as LOC is directly proportional to maintenance of the product.

Cohesion is the inter-relatedness among class members and methods of a class. Cohesion has a negative effect on the complexity of the code, increasing the maintenance of the product. The metric to measure cohesion is Lack of Cohesion between Methods (LCOM). Greater LCOM values indicate very poor cohesion between methods of a class. LCOM metric is inversely proportional to maintenance of the code. The following table depicts LCOM values obtained by using Metric reloaded tool.

Table7. LCOM Table

| LCOM | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSION** | **V 3.4** | **V3.5** | **V3.6** | **V3.7** | **V3.8** | **V 4** | **V 4.1** | **V 4.2** | **V4.3** | **V 5** |
| Architecture | | | | | | 0 | 0 | 0 | 0 | 0 |
| Cli | 2.25 | 2.25 | 1.56 | 1.56 | 1.56 | 1.56 | 1.56 | 1.62 | 1.62 | 1.86 |
| Logic | 1.77 | 1.78 | 2.03 | 1.94 | 1.94 | 1.99 | 2.04 | 2.07 | 2.09 | 2.13 |
| Migration | 2 | 2 | 2 | 2 | 1.67 | 2.33 | 2.33 | 2 | 1.2 | 1.2 |
| Model | 1.91 | 1.94 | 1.8 | 2.22 | 2.03 | 2.25 | 2.25 | 2.28 | 2.31 | 2.49 |
| Preferences | | | 2 | Y | 2.36 | 2.33 | 2.33 | 2.33 | 2.33 | 2.35 |
| Styletester | | | | | | | | | | 2.5 |
| Pdfimport | 1.75 | 1.75 | 1.75 | 2 | 2 | 2 | 2 | 2 | 2 | |
| Gui | 1.79 | 1.77 | 1.81 | 1.77 | 1.78 | 1.79 | 1.78 | 1.81 | 1.81 | 1.98 |
| Collab | 1.84 | 1.84 | 1.84 | 1.89 | 1.89 | 1.89 | | | | |
| Shared | | | 2.14 | 1.95 | 1.95 | 1.95 | 1.95 | | | |

“cli” package is frequently updated over versions and has no pattern which depicts instability of the package. ”Migration” package had a constant value from V3.4 to V3.7 then the value was gradually increasing up to 4.2 and a sudden fall in LCOM value was noticed from V4.2 to V5.0 indicating low maintenance required to handle the package.

## 4.2. Analysis on the Understandability

Here with these metrics we have successfully discussed and found the suitable answers for the Q3 and the following data to prove those results was collected from the MetricsReloaded plugin installed in the IntelliJ IDE.

The results collected from the tools for the Ca metrics is shown in the table below:

Table8. Ca Table

| Ca | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSIONS** | **V 3.4** | **V3.5** | **V3.6** | **V 3.7** | **V3.8** | **V 4** | **V 4.1** | **V 4.2** | **V 4.3** | **V 5** |
| Architecture | | | | | | 0 | 0 | 0 | 0 | 0 |
| Cli | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 |
| Logic | 23 | 10 | 12 | 12 | 7 | 6 | 6 | 6 | 6 | 3 |
| Migration | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 5 | 5 | 5 |
| Model | 17 | 32 | 31 | 43 | 54 | 55 | 54 | 59 | 59 | 31 |
| Preferences | | | 64 | 63 | 65 | 69 | 70 | 65 | 65 | 79 |
| Styletester | | | | | | | | | | 0 |
| Pdfimport | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | |
| Gui | 282 | 280 | 271 | 267 | 272 | 351 | 358 | 352 | 352 | 202 |
| Collab | 13 | 13 | 13 | 13 | 13 | 13 | | | | |
| Shared | | | 6 | 11 | 11 | 11 | 11 | | | |

Afferent coupling is measured and shown in the table. With increase in Afferent coupling the understandability decreases, since the dependency of the class on other packages is measured by afferent coupling which means increase in value of afferent coupling means more dependency on other packages. This dependency on other packages makes the program hard to understand. From the table three packages “model”, “preferences” and “gui” are having high value of afferent coupling. The model package shows an increase of afferent coupling value from version 3.4 to version 3.8 and then approximate constant value till version 4.3 and then a decrease to a relatively low value which show the scope for maintainability. The package preferences have also relatively high value from version 3.6 to version 5.0. The package logic stands different and has very high afferent coupling value relative to all other packages, it shows a significant increment from version 3.8 to version 4.0 and also a drop from version from 4.3 to version 5.0 but still the value is relatively very high. This shows the package “gui” is hard to understand thus hard to maintain.

Efferent coupling is measured and shown in table with increase in efferent coupling the understandability decreases. The dependency of other packages on the class in scope is measured by efferent coupling. This increase in dependency of other packages on this class in scope makes understandability low. The increase in efferent coupling reduces the ease of maintenance. As we can observe the graph for the package GUI the initial versions had high maintenance as the Ce values of the package was large and later versions the Ce values have reduced, and the maintenance of the package reduced. In the package cli package we observe that the initial versions have exceptionally low Ce values and had low maintenance and in the later versions has slightly increased and resulted in the high maintenance of the package. In the package migration we also observe that the initial versions have exceptionally low values of Ce and had low maintenance of the package and in further versions the values have slightly increased making the package high maintenance when compared to the initial versions.

Table9. Ce Table

| **Ce** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSIONS** | **V3.4** | **V3.5** | **V3.6** | **V3.7** | **V3.8** | **V 4** | **V4.1** | **V4.2** | **V4.3** | **V5** |
| Architecture | | | | | | 0 | 0 | 0 | 0 | 0 |
| Cli | 24 | 24 | 47 | 50 | 50 | 50 | 50 | 46 | 46 | 43 |
| Logic | 9 | 8 | 8 | 8 | 4 | 4 | 4 | 4 | 4 | 3 |
| Migration | 12 | 12 | 17 | 17 | 24 | 25 | 25 | 37 | 37 | 33 |
| Model | 6 | 7 | 7 | 8 | 8 | 6 | 6 | 6 | 6 | 2 |
| Preferences | | | 25 | 38 | 38 | 44 | 44 | 47 | 47 | 70 |
| Styletester | | | | | | | | | | 4 |
| Pdfimport | 20 | 20 | 22 | 27 | 27 | 28 | 28 | 26 | 26 | |
| Gui | 459 | 457 | 421 | 437 | 411 | 442 | 446 | 432 | 432 | 229 |
| Collab | 85 | 85 | 89 | 85 | 85 | 87 | | | | |
| Shared | | | 14 | 22 | 22 | 22 | 22 | | | |

It was also observed for the CR metrics that the maintenance is inversely proportional to the metrics CR. The results that were collected for the CR metrics for the packages of 10 different versions are as follows:

Table10 CR Table

| **Comment Ratio** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSIONS** | **V 3.4** | **V3.5** | **V3.6** | **V3.7** | **V3.8** | **V4** | **V4.1** | **V4.2** | **V4.3** | **V5** |
| Cli | 0.067 | 0.083 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.054 | 0.054 | 0.055 |
| Logic | 0.286 | 0.291 | 0.202 | 0.191 | 0.191 | 0.185 | 0.187 | 0.184 | 0.183 | 0.176 |
| Migration | 0.256 | 0.256 | 0.209 | 0.209 | 0.199 | 0.178 | 0.181 | 0.153 | 0.153 | 0.125 |
| Model | 0.3 | 0.299 | 0.265 | 0.256 | 0.252 | 0.246 | 0.247 | 0.244 | 0.244 | 0.237 |
| Preferences | | | 0.132 | 0.11 | 0.106 | 0.103 | 0.105 | 0.099 | 0.099 | 0.096 |
| Pdfimport | 0.169 | 0.169 | 0.095 | 0.09 | 0.09 | 0.089 | 0.089 | 0.088 | 0.088 | |
| Gui | 0.177 | 0.178 | 0.127 | 0.125 | 0.124 | 0.119 | 0.117 | 0.115 | 0.115 | 0.089 |
| Collab | 0.237 | 0.237 | 0.128 | 0.138 | 0.137 | 0.138 | | | | |
| Shared | | | 0.236 | 0.203 | 0.203 | 0.203 | 0.201 | | | |

According to the above results generated and the conclusions drawn from ”logic” package the outputs that had been collected and observed the variations keenly in this package declares us that the Code to comment ratio in the version 3.4 and 3.5 we can observe a negligible increase of the comment ratio and later it had an exceptional decrease in the 3.6 version and had gradually decreased in the further versions but an except case where a negligible increase of the ratio between 4.0 and 4.1 versions. Hence, the decrease in the values are only in the initial versions and thus 3.6 version the package is more difficult to maintain than the 3.4 and 3.5 versions and 4.1 versions is easy to maintain than the 4.0 version. According to the above results generated and the conclusions drawn from ”GUI” package the outputs that had been collected and observed the variations keenly in this package declares us that the Code to comment ratio in the versions 3.4 and 3.5 it has negligibly increased and has remarkably decreased in the next version i.e. 3.6 version and has gradually decreased for further versions hence 4.0 version has a decreased value than the difficult to maintain than the previous version 3.8 and the further versions are also equally maintained for this package. According to the above results generated and the conclusions drawn from ”pdfimport” package the outputs that had been collected and observed the variations keenly in this package declares us that the Code to comment ratio in the versions 3.4 and 3.5 versions is constant and has exceptionally decreased in 3.6 version and since has slightly decreased in further versions. Hence there is a decrease in the values in 3.5 and 3.6 versions thus 3.6 is more difficult to maintain than the initial versions and the further versions have not shown any noticeable change and thus are considered to be equally maintained versions for these packages.

## 4.3. Analysis on Complexity

The Q2 of GQM is discussed in this part where the relation of how the complexity effect the maintainability of the software project. Complexity is a crucial factor in determining the ease of maintenance of the software project. Ease of maintaining of software depend on various factor but complexity is one of the important factors in determining the maintainability. Metrics as v(G), CBO and WMC are used to estimate complexity and maintainability relation.

Table11. v(G) Table

| **COMPLEXITY** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSION** | **V 3.4** | **V3.5** | **V3.6** | **V3.7** | **V3.8** | **V4** | **V4.1** | **V4.2** | **V4.3** | **V5** |
| Architecture | | | | | | 0 | 0 | 0 | 0 | 0 |
| Cli | 2.6 | 2.6 | 2.8 | 2.8 | 2.8 | 2.7 | 2.7 | 2.5 | 2.5 | 2.4 |
| Logic | 3.2 | 3 | 3.2 | 3.4 | 3.3 | 3.2 | 3.2 | 3 | 3 | 2.9 |
| Migration | | 4.7 | 4.6 | 4.6 | 4.3 | 3.6 | 3.6 | 3 | 2.9 | 2.6 |
| model | 2.3 | 2.2 | 2.1 | 2.1 | 2.1 | 2 | 2 | 1.9 | 1.9 | 2 |
| Preferences | | | 1.7 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| Styletester | | | | | | | | | | 1 |
| Pdfimport | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 | 2.8 | 2.8 | |
| Gui | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.3 | 2.3 | 2.2 | 2.2 | 1.8 |
| Collab | 2.7 | 2.7 | 2.7 | 2.8 | 2.8 | 2.8 | | | | |
| Shared | | | 1.9 | 1.8 | 1.8 | 1.8 | 1.8 | | | |

Cyclomatic complexity is measured with the Lizard tool and collected values are shown in the table. With increase in complexity maintenance also increases. The module “cli”,“pdfimport”, migration and logic are the modules which are showing relative high complexity. Here the logic module is having highest complexity among other modules and migration also have high complexity in initial versions but have as shown in table migration module have steady drop in complexity from first version 3.4 to last version 5.0 which show the scope of maintainability. The module logic does not show any significant drop in complexity and maintain high complexity value thought all versions which is from version 3.4 to 5.0 which show that module is hard to maintain. The module “cli” and “pdfimport” are also showing steady relatively high complexity and are also belong to the group which show these are hard to maintain based on their complexity nature.

The maintainability index which is calculated based on cyclomatic complexity and number of lines of code (LOC) and Halstead volume has can be used to see a relation between the complexity and maintenance and to validate the result. The maintainability index shows that the lower the value of maintainability index the harder is to maintain the project and higher the value means highly maintainable. There is a very similar patter between the maintainability index and the complexity, and both shows the similar result of relation between the complexity and maintainability. The result observed is that cyclomatic complexity is inversely proportional to maintainability of packages.

Table12. MI Table

| **Maintainability Index** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSIONS** | **V3.4** | **V3.5** | **V3.6** | **V3.7** | **V 3.8** | **V4** | **V4.1** | **V4.2** | **V4.3** | **V5** |
| architecture | | | | | | 0 | 0 | 0 | 0 | 0 |
| Cli | 43.83 | 43.83 | 43.41 | 43.37 | 42.98 | 44.49 | 44.49 | 44.96 | 44.96 | 44.34 |
| Logic | 45.04 | 45.33 | 45.1 | 44.83 | 44.96 | 45.62 | 45.8 | 46.03 | 46.1 | 46.21 |
| Migration | 27.7 | 27.7 | 27.09 | 26.73 | 27.93 | 33.9 | 32.73 | 37 | 37.11 | 38.03 |
| Model | 42.93 | 43.36 | 44.39 | 45.91 | 45.89 | 46.8 | 46.79 | 47.25 | 47.23 | 48.76 |
| preferences | | | 40.55 | 41.47 | 41.28 | 39.5 | 39.48 | 38.44 | 38.44 | 36.7 |
| Styletester | | | | | | | | | | 53.5 |
| Pdfimport | 39.57 | 39.57 | 39.57 | 38.91 | 38.91 | 36.55 | 36.55 | 36.64 | 36.64 | |
| Gui | 37.75 | 37.6 | 37.91 | 37.89 | 37.8 | 41.4 | 41.67 | 41.88 | 41.85 | 47.65 |
| Collab | 42.23 | 42.23 | 42.28 | 41.99 | 42.12 | 42.11 | | | | |
| Shared | | | 48.31 | 48.01 | 49.33 | 49.32 | 49.22 | | | |

Maintenance is directly proportional to the WMC metrics i.e. as the WMC increase it becomes easy to maintain the package.

Table13. WMC Table

| **WMC** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSIONS** | **V 3.4** | **V3.5** | **V3.6** | **V 3.7** | **V 3.8** | **V 4** | **V 4.1** | **V 4.2** | **V 4.3** | **V 5** |
| Architecture | | | | | | 0 | 0 | 0 | 0 | 0 |
| Cli | 113 | 113 | 148 | 148 | 149 | 152 | 152 | 136 | 139 | 18.43 |
| Logic | 3278 | 3194 | 5265 | 5153 | 5187 | 5299 | 5416 | 5677 | 5686 | 5844 |
| Migration | 41 | 41 | 44 | 44 | 51 | 62 | 69 | 91 | 91 | 76 |
| Model | 742 | 759 | 760 | 1593 | 1570 | 1786 | 1784 | 1827 | 1831 | 1836 |
| Preferences | | | 123 | 187 | 191 | 218 | 218 | 236 | 236 | 291 |
| Styletester | | | | | | | | | | 6 |
| Pdfimport | 55 | 55 | 54 | 56 | 56 | 57 | 57 | 58 | 58 | |
| Gui | 4723 | 4830 | 5770 | 6356 | 6463 | 6851 | 7048 | 6979 | 6985 | |
| Collab | 199 | 199 | 199 | 183 | 183 | 183 | | | | |
| Shared | | | 137 | 200 | 200 | 200 | 201 | | | |

According to the results generated and observations we here conclude that the package like ”GUI” the WMC metrics values have gradually decreased over the versions 3.4 to 5.0 of the JabRef system which makes it a low maintenance of the package. For the package ”Migration” the outputs that had been collected and observed the variations keenly in this package declares us that the WMC in the versions 3.4-3.5 and 3.6-3.7 remains constant and increase between v3.5-v3.6 and later there is an exceptional decrease between 3.7 and 3.8 and later it increases from version 3.8 to 4.0. And again, decreases between 4.1 and 4.2 and remains constant between 4.2 and 4.3 and again decreases in the version 5 making the migration package low maintenance. For the package ”Model” the data collected and observed shows that the values were decreasing overall from the version 3.4 and 5.0 but had an exceptional decrease in the versions 3.6 and 3.7 and hence that makes the package model low maintenance. For the package ”Logic” the outputs that had been collected and observed the variations keenly in this package declares us that the WMC in the versions 3.4 and 3.5 there is a gradual decrease and then from the versions 3.5 and 3.6 there has been a remarkable increase and for later version it has been increasing since then making the package high maintenance package.

It was also observed for the CBO metrics that the maintenance is directly proportional to the metrics CBO.

The results that were collected for the CBO metrics for the packages of 10 different versions are as follows:

Table14. CBO Table

| CBO | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **VERSIONS** | **V 3.4** | **V3.5** | **V3.6** | **V3.7** | **V3.8** | **V 4** | **V41** | **V 4.2** | **V4.3** | **V 5** |
| Architecture | | | | | | 0 | 0 | 0 | 0 | |
| Cli | 0.75 | 0.75 | 0.44 | 0.44 | 0.44 | 0.44 | 0.44 | 0.38 | 0.38 | 0.43 |
| Logic | 2.62 | 2.51 | 3.18 | 3.25 | 3.31 | 3.4 | 3.44 | 3.38 | 3.39 | 3.53 |
| Migration | | 0 | 0 | 0 | 0 | 0.33 | 0.33 | 0.2 | 0.2 | 0.2 |
| Model | 4.49 | 1.81 | 2.08 | 3.06 | 2.84 | 2.9 | 2.91 | 2.86 | 2.84 | 3.55 |
| Preferences | 6.36 | | 0.75 | 0.73 | 0.73 | 0.67 | 0.67 | 0.67 | 0.67 | |
| Styletester | | | | | | | | | | 0 |
| Pdfimport | 8.23 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | |
| Gui | 10.1 | 4.44 | 4.62 | 5.19 | 5.15 | 4.99 | 5 | 4.92 | 4.93 | 4.84 |
| Collab | | 0.37 | 0.37 | 0.28 | 0.28 | 0.28 | | | | |
| Shared | 11.97 | | 1.14 | 1.6 | 1.6 | 1.6 | 1.6 | | | |

According to the above results generated and the conclusions drawn from it were that for “cli” package the outputs that had been collected and observed the variations keenly in this package declares us that the CBO in the versions 3.4 and 3.5 have the same constant outputs whereas, the later versions from the 3.6 to 4.1 had a massive decrease in the outputs values but were constant in all the versions further which was equal to initial versions i.e. 3.4 and 3.5 respectively. Hence as the CBO metrics value decreased from versions 3.4-5.0 it is concluded as a low maintenance package for the latest version when compared to the previous versions. According to the above results generated and the conclusions drawn from the package “gui” the outputs that had been collected and observed the variations keenly in this package declares us that the CBO for all the version is initially increasing from the version 3.4 to 3.7 but negligible decrease between 3.4 and 3.5 and has further decreased from version 3.7 to 5 but negligible increase between the 4.0-4.1 and 4.2-4.3 respectively as the values is frequently changing and the has noticeably increased from 3.4 version to 5.0 version hence it is high maintenance package. According to the above results generated and the conclusions drawn from the ”migration” package the outputs that had been collected and observed the variations keenly in this package declares us that for most of the versions the CBO has been 0 and the has rose to 0.33 from 4 version and then has been decreased to 0.2 from version 4.2 and further. Hence, considering the frequent change and considering the initial version and final version and values had increased and thus it is considered a highmaintenance package.

## 4.4. Analysis on Cohesion

In this part we are answering Q4 of GQM, how does cohesion effect maintainability. LCOM of CK metric suite is considered for analysing cohesion and maintainability relation. The data is collected using Metric-Reloaded Plugin with IntelliJ IDE. Observation and Reflection Cohesion is directly proportional to maintainability of a package i.e. with increase with cohesion package are easy to maintain. LCOM shows the lack of cohesion between methods of a class. The greater value of LCOM means packages are hard to maintain.

package is frequently updated over versions and has no specific pattern which shows instability of package. Cohesion in the migration package has constant value from version 3.4 to version3.7. In the later versions the package becomes a bit unstable as we go from version 3.7 to version 4.2 and suddenly fall in LCOM value from version 4.2 to 5.0 which shows low maintenance.

## 4.5. Overall Evaluation

In this part we are answering Q4 of GQM, how does cohesion effect maintainability. LCOM of for overall evaluation we have processed the collected data and summarized it for more possible evaluation. In this section for every package all the values of the metrics are processed by mean and the mean is done over time to find a single value for all packages and the values are normalized to find comparative results. The data is processed according to the direct proportionality or inverse proportionality as required by values and as the result have been found in the above subsections.

Table15. Normalised Table

| PACKAGE | NORMALIZED METRIC VALUES | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | DIT | LCOM | LOC | CBO | v(G) | Ca | Ce | MI | CR | WMC |
| Cli | 0 | 0 | 0.0095 | 0.0998 | 0.5927 | 0.017 | 0.0945 | 0.3076 | 1 | 0.974 |
| Logic | 0.1121 | 0.3131 | 0.6705 | 0.6535 | 0.7734 | 0.0304 | 0.0038 | 0.2538 | 0.094 | 0.687 |
| Migration | 0 | 0.175 | 0 | 0.0514 | 1 | 0.0097 | 0.0482 | 1 | 0.131 | 1 |
| Model | 0.1218 | 0.5368 | 0.1932 | 0.599 | 0.3831 | 0.1456 | 0.0053 | 0.2384 | 0.221 | 0.851 |
| Preferences | 0.2466 | 0.7872 | 0.0283 | 0.1426 | 0.1897 | 0.2259 | 0.0972 | 0.5153 | 0.54 | 0.948 |
| Styletester | 0 | 1 | 0.0018 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pdfimport | 0.5682 | 0.2324 | 0.0001 | 0.102 | 0.6224 | 0.02 | 0.0506 | 0.5846 | 0.536 | 0.712 |
| Gui | 1 | 0.0907 | 1 | 1 | 0.4807 | 1 | 1 | 0.4692 | 0.2 | 0.576 |
| Collab | 0.4487 | 0.1644 | 0.2012 | 0.0645 | 0.6325 | 0.0435 | 0.1987 | 0.3923 | 0.2 | 0.523 |
| Shared | 0.4222 | 0.3263 | 0.2163 | 0.3078 | 0.2963 | 0.0334 | 0.0397 | 0.1384 | 0.0826 | 0.505 |

All the packages are classified into three difficulty levels on ordinal scale as high maintainability, moderate maintainability, low maintainability packages. Mean values of a metric for a specific package is calculated and normalized according to the dependency of the metric on maintainability. The above table depicts the normalized values of various metrics. Depending upon these values packages are classified into 3 categories, then the mode represents the difficulty of maintenance as form difficulty level of high to low if the mode have sufficientvalue to be in high difficulty level it is placed in high else it is considered for medium level of difficulty and still if not fit in the difficulty level it is finally in low difficulty level.

Table16. Maintainability Level Table

| PACKAGE | HIGH | MEDIUM | LOW | FINAL MAINTENANCE LEVEL |
|---|---|---|---|---|
| Cli | 2 | 1 | 7 | MEDIUM |
| Logic | 3 | 0 | 7 | HIGH |
| Migration | 3 | 0 | 7 | HIGH |
| Model | 1 | 3 | 6 | MEDIUM |
| Preferences | 2 | 2 | 6 | MEDIUM |
| Styletester | 1 | 0 | 9 | LOW |
| Pdfimport | 1 | 4 | 5 | MEDIUM |
| gui | 5 | 3 | 2 | HIGH |
| Collab | 0 | 4 | 6 | MEDIUM |
| Shared | 0 | 2 | 8 | LOW |

Table17. Packages Maintainability Classification Table

| Package | Maintenance | Reason |
|---|---|---|
| cli | Medium | - |
| Logic | High | LOC,v(G),WMC,CBO |
| migration | High | v(G),WMC |
| Model | Medium | LCOM,CBO,WMC |
| preferences | Medium | - |
| Styletester | Low | - |
| pdfimport | Medium | DIT,CR,v(G) |
| gui | High | DIT,LOC,CBO,Ca,Ce,WMC |
| Collab | Medium | v(G) |
| Shared | Low | - |

## 4.6. Change log and Timeline

The timeline for how the JabRef versions is released and duration between them is given in the figure below.

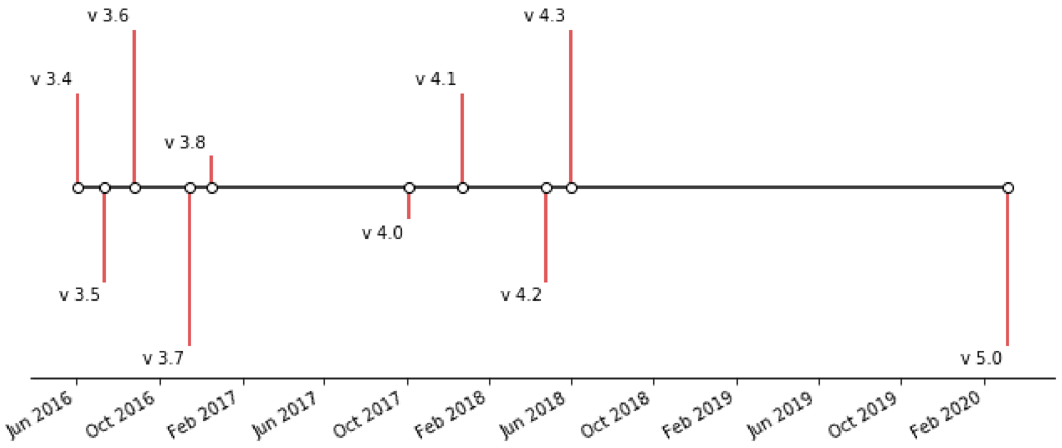


Fig 1. Version Release Timeline

Table18. Change log Table

| Version | Changes | Fixed | Removed |
|---|---|---|---|
| V 3.4 | 18 | 31 | 6 |
| V 3.5 | 8 | 15 | 0 |
| V 3.6 | 33 | 44 | 7 |
| V 3.7 | 48 | 45 | 5 |
| V 3.8 | 17 | 15 | 0 |
| V 4.0 | 11 | 22 | 1 |
| V 4.1 | 26 | 28 | 0 |
| V 4.2 | 28 | 17 | 1 |
| V 4.3 | 9 | 6 | 1 |
| V 5.0 | 8 | 25 | 2 |

We have observed all the 10 versions of the JabRef system namely 3.4, 3.5, 3.6, 3.7, 3.8, 4.0, 4.1, 4.2, 4.3, 5.0 that initially in the 3.4 version there many packages like “Cli”, “logic”, ” migration”, “model”, “pdfimport”, “gui”, “collab”, “specialfields”, “event”, “external”, “bst”, exporter, “sql”, “util”, “importer” and as the new versions were introduced these packages were either removed or were merged in the other packages i.e. the “bst” and “specialfields” packages were removed and the packages like importer, exporter, external, collab were merged into the GUI package and che packages namely “shared”, “event”, “util”, “sql” were merged into “model” package. Hence as these small packages were present all in few versions and were again merged in other packages, we have neglected these packages during the analysis done by the metrics. From change log the extracted data for the issue changes in version and issue fixed and issue removed are shown in table below.

## 4.7. Conclusion of Results and Future Work

Proportionality of an object-oriented metric to the maintainability of a product. The relation of the metric and maintainability is discussed in the above sub-section of analysis and the inverse or direct proportionality is also discussed. We could use the study to obtain the maintainability level for various Object oriented  software system by considering the goal and the questions covering our goal and the metrics which are helpful in answering the question in scope to fulfil the goal and a similar approach can be applied to obtain the result to be found.

As in this paper we majorly focused on factors such as complexity, size, structure and understandability to determine the Maintainability. The Future work for the study is to do a study on various Object-Oriented open source system independent of the language and to analyse the result to while increasing the scope of attribute and similarly increasing the scope of metrics.

### ACKNOWLEDGEMENTS

The authors would like to thank everyone, just everyone!

# REFERENCES

[1] Robson, Colin. (2002). Real World Research : A Resource for Social Scientists and Practitioner-Researchers / C. Robson.

[2] Laing, Victor Coleman, Charles: Principal Components of Orthogonal Object-Oriented Metrics. White Paper Analyzing Results of NASA Object-Oriented Data. SATC, NASA, 2001.

[3] Horst Zuse. 1991. Software complexity: measures and methods. Walter de Gruyter Co., USA.

[4] Bansal, M., Agrawal, C.P., 2014. Critical Analysis of Object Oriented Metrics in Software Development, in: 2014 Fourth International Conference on Advanced Computing Communication Technologies. IEEE, pp. 197–201. doi:10.1109/ACCT.2014.106.

[5] V.R. Basili and D. Weiss (1984), A Methodology for Collecting Valid Software Engineering Data, IEEE Trans. Software Engineering, vol. 10, pp.728-738.

[6] V.R. Basili and H.D. Rombach (1988), The Tame Project: Towards Improvement-Oriented Software Environments, IEEE Trans. Software Engineering, vol.14, pp.758-773.

[7] https://github.com/JabRef/jabref

[8] Yadav, A. and Khan, R.A., 2011, September. Class cohesion complexity metric (C 3 M). In 2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011) (pp. 363-366). IEEE.

[9] Kaur, K. and Singh, H., 2011, February. Towards a Valid Metric for Class Cohesion at Design Level. In 2011 Second International Conference on Emerging Applications of Information Technology pp. 351-354. IEEE.

[10] Al Dallal, J., 2012. Theoretical analysis for the impact of including special methods in lack-of-cohesion computation. Procedia Technology, 1, pp.167-171.

[11] Dubey, S.K., Rana, A., 2011. Assessment of maintainability metrics for object-oriented software system. ACM SIGSOFT Softw. Eng. Notes 36, pp. 1–7. doi:10.1145/2,020,976.2020983.

[12] Sandesh Ganjare, Koustubh Kulkarni, Dinesh B Hanchate et al. Measuring Structural Code Quality Using Metrics. Inventi Rapid: Soft Engineering, 2015(3):1-7, 2015.

[13] S R Chidamber and C F Kemerer. A Metrics Suite for Object Oriented Design. IEEE International Conference on Data Mining, 150-159, 2008.

[14] Simon, Martin, Linus W. Dietz, Tobias Diez and Oliver Kopp. “Analyzing the Importance of JabRef Features from the User Perspective.” ZEUS (2019).

[15] Al-Jamimi, H.A., Ahmed, M., 2012. Prediction of software maintainability using fuzzy logic, in: 2012 IEEE International Conference on Computer Science and Automation Engineering. IEEE, pp. 702–705. doi:10.1109/ICSESS.2012.6269563.

[16] Wirotyakun, A., Netisopakul, P., 2012. Improving software maintenance size metrics A case study: Automated report generation system for particle monitoring in Hard Disk Drive Industry, in: 2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE). IEEE, pp. 334–339. doi:10.1109/JCSSE.2012.6261975.

[17] Kaur, A., Kaur, K., Pathak, K., 2014. Software maintainability prediction by data mining of software code metrics, in: 2014 International Conference on Data Mining and Intelligent Computing (ICDMIC). IEEE, pp. 1–6. doi:10.1109/ICDMIC.2014.6954262.

[18] Dubey, S.K., Rana, A., 2011. Assessment of maintainability metrics for object-oriented software system. ACM SIGSOFT Softw. Eng. Notes 36, pp. 1–7. doi:10.1145/2,020,976.2020983.

[19] Saifan, Ahmad Alsghaier, Hiba Khateeb, Khaled. (2017). Evaluating the Understandability of Android Applications. International Journal of Software Innovation. 6. 10.4018/IJSI.2018010104.

[20] Jehad Al Dallal,Object-oriented class maintainability prediction using internal quality attributes,Information and Software Technology,Volume 55, Issue 11,2013,Pages 2028-2048,ISSN 0950-5849, https://doi.org/10.1016/j.infsof.2013.07.005.

## AUTHORS

**Denim Deshmukh**

Denim Deshmukh is currently completing his master's degree in Software Engineering from Blekinge TekniskaHögskola, Sweden. He is looking forward for contributing to the field of Software Engineering.

**Ravi ThejaKataray**

Ravi ThejaKataray currently pursuing a Master of Science in the field of Software Engineering from Blekinge TekniskaHögskola, Sweden. Striving hard to work in a team for a cause of gradually developing software methodology.

**Rohith Girikshith**

Rohith Girikshith is currently pursuing a Master of Science in the field of Software Engineering from Blekinge TekniskaHögskola, Sweden. He is doing his research work in the field of Software Engineering.

# LOCAL SELF-ATTENTION BASED CONNECTIONIST TEMPORAL CLASSIFICATION FOR SPEECH RECOGNITION

Deng Huizhen and Zhang zhaogong

Computer Science Institution, Heilongjiang University of China, China

## *ABSTRACT*

*Connectionist temporal classification (CTC) has been successfully applied to end-to-end speech recognition tasks, but its main body recurrent neural network makes parallelization very difficult. Since the attention mechanism has shown very good performance on a series of tasks such as machine translation, handwriting synthesis, and image caption generation for loop sequence generators conditioned on input data. This paper applies the attention mechanism to CTC, and proposes a connectionist temporal classification based on the local self-attention mechanism, in which the cyclic neural network module in the traditional CTC model is replaced by the self-attention module. It shows that it is attractive and competitive in end-to-end speech recognition. The proposed mechanism is based on local self-attention, which uses a sliding mechanism to obtain acoustic features locally. This mechanism effectively models long-term scenarios by stacking multiple sliders to obtain a larger receiving field to achieve online decoding. Moreover, the CTC training joint cross-entropy criterion makes the model converge better. We have completed experiments on the AISHELL-1 dataset. The experiments show that the basic model has a lower character error rate than the existing state-of-the-art models, and the model after cross entropy has been further improved.*

## *KEYWORDS*

*Connectionist temporal classification, self-attention mechanism, cross entropy, Speech Recognition*

## 1. INTRODUCTION

End-to-end speech recognition is a recently proposed method, which does not require a pre-defined alignment between speech frames and characters to directly transcribe speech into text [1-9]. The latest work on end-to-end speech recognition can be divided into two main methods: based on connectionist temporal classification (CTC) [10,1-3] and attention-based encoder-decoder [4-6]. Both of these methods solve the problem of variable-length input and output sequences. In the traditional deep neural network hidden Markov model hybrid system, the deep neural network is used to generate each frame of sound data, and its distribution is re-expressed as the transmission probability of the Hidden Markov Model (HMM). Then, model training can be performed by using frame-level cross entropy (CE) criteria, using sequence discrimination training methods such as maximum mutual information (MMI) [8]. For this model, its problem is that the frame-level training target must be inferred from the alignment determined by the HMM. Different traditional speech recognition methods, the end-to-end model learns the mapping of acoustic frames to characters for the final target of interest, and tries to correct the sub-optimal problems caused by the irrelevant training process. Among them, the key idea of CTC is to use intermediate label representation, it allows duplicate labels and uses blank labels to identify

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 275-284, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101122

labels which are not output. The CTC loss can be effectively calculated by the forward and backward algorithm, which can predict the target of each frame, and provided that the conditions between the targets are independent of each other.

Recently, self-attention mechanism [11, 12] was proposed, which uses the entire sequence to model feature interactions at any distance in time. It is used in the encoder, decoder and feedforward context to accelerate the translation speed, and provides the latest translation results, sentiment analysis [13] and other tasks. The success of self-attention in these tasks inspired the initial work of self-attention in speech recognition. So the attention-based encoder decoder model appeared. Although it was first applied to machine translation, its versatility also made it useful for speech recognition tasks [14-17]. The attention-based encoder decoder model directly learns the mapping from the acoustic frame to the character sequence. At each output time step, the model sends out a label based on the history of the input and target labels. Since the attention model does not use any conditional independence assumptions, it exhibits a lower character error rate (CER) than CTC without using an external language model. However, some speech recognition tasks in real environments, the model shows poor results because the estimated alignment in the attention mechanism is easily damaged by noise and other details. Another problem is that it is difficult to learn the model from scratch due to the misalignment of long input sequences.

In order to overcome the above problems, this paper proposes a novel end-to-end speech recognition method, which uses a local self-attention model based on CTC training criteria to improve performance and accelerate learning. The key of our method is to use a shared encoder representation trained by CTC and self-attention model targets at the same time. We believe that the weakness of the attention model is due to the lack of left-to-right constraints used in DNN-HMM and CTC, which makes it is difficult to properly align the training encoder network under noisy data or long input sequences. Our proposed method improves the performance by correcting the CTC loss based on the forward and backward algorithm plus the cross-entropy loss function to assist the alignment problem of CTC training. In addition, combining the characteristics of attention and the defects of CTC, and inspired by time-delayed neural networks, this paper proposes a mechanism based on local self-attention that uses a sliding mechanism to obtain acoustic features locally, and stacked a larger receiving field to effectively model long-term scenarios to achieve online decoding.

## 2. RELATED WORKS

Recently, there have been some works applying the self-attention mechanism to speech recognition, and good results have been obtained compared with traditional hybrid speech models [18, 19]. Different from these, this paper introduces the self-attention mechanism into the CTC-based model and proposes a sliding mechanism similar to the convolutional neural network to achieve online decoding. Different from the block jumping mechanism in [19], this article divides the entire pronunciation into several overlapping blocks as input, and the slider has an asymmetric context. We use a sliding window at each layer to limit the scope of self-attention. They all use sliding windows to model the local dependencies between inputs, without any modification to the self-attention network structure. The sliding chunk mechanism only uses sliding windows to limit the range of attention, and stacks multiple self-attention sliders for long-term dependencies are modeled. Existing work [20] believes that only using CTC to train the model, sometimes the training failed to converge, or the cross-entropy loss function is used to pre-train the model, and then the CTC training model is used on this model, and there is still a problem of instability of the model. Therefore, this paper proposes to use CTC training and cross-entropy loss function at the same time to make the model converge better.

## 3. CONNECTIONIST TEMPORAL CLASSIFICATION

With CTC as the acoustic model sequence of the loss function, CTC can automatically learn the alignment between the input speech frame sequence and its label sequence (such as phonemes or characters) without using frame-level alignment information. Only one input sequence and one output sequence are needed for training. CTC cares about whether the predicted output sequence is close to the real sequence, and does not care whether each result in the predicted output sequence is exactly aligned with the input sequence at the time point. The CTC modeling unit is a phoneme or a word, and its main idea is to introduce Blank (-) tags, delete blank tags and merge duplicate tags to obtain a unique corresponding sequence. For a piece of speech, the last output of the CTC is a sequence of spikes, the position of the spike corresponds to the Label of the modeling unit, and the other positions are Blank.

For alphabet L, the size after adding the blank label (-) introduced by CTC is $L' = L \cup \{'-'\}$ ,Input an acoustic sequence x of length T, $x = (x_1, ..., x_T)$ , The corresponding output length is U tag sequence l, $l = (l_1, ..., l_U)$ , and U≤T.One way to match the input x neural network output $\pi = (\pi_1, ..., \pi_T)$ ,It is defined as a sequence above $L'$, which is $\pi \in L'^T$ .There is such a transformation function B, for all possible output paths after B transformed into label l, namely $l = B(\pi)$ , This transformation removes the blank tags in the path and merges the consecutively repeated tags to obtain a unique corresponding sequence, for example, B(_,a,a,_b,_c,c,_)=abc,B(a,_,b,b ,_,_c,c,_)=abc. Therefore, for a given input x, the probability of its output sequence l after the neural network is the sum of the probabilities of all possible paths, the formula is as follows:

$$p(l \mid x) = \sum_{\pi \in B^{-1}(l)} p(\pi \mid x) \qquad (1)$$

which

$$p(\pi \mid x) = \prod_{t=1}^{T} y_{\pi}^{t} \qquad (2)$$

In order to facilitate the calculation of the gradient, the log objective function is generally minimized:

$$\ell_{ctc}(x) = -\log\, p(l \mid x) \qquad (3)$$

Since there are many possible paths and the amount of calculation is too large, CTC uses a forward-backward algorithm to calculate the loss, and uses a cluster search algorithm to decode.

## 4. MODEL

In order to improve the parallel computing power and performance of the model, this paper proposes an end-to-end model that does not use recurrent neural networks, a CTC model based on local self-attention, which uses a self-attention mechanism to replace the original CTC The recurrent neural network structure in the model. A slider mechanism similar to the convolutional neural network is proposed. The input acoustic feature length is 25% as the slider length and the features are stacked in chronological order. The experiment proves that this ratio has a certain effect.

## 4.1. Model Structure

For alphabet L, given an input sequence x of length T, $x = (x_1,...,x_T)$ , x has T*d dimensions, defining an output sequence $y = (y_1,..., y_U)$ .In speech recognition, x is an acoustic feature, L is a collection of characters or phonemes, and the output sequence y is the corresponding real label on the alphabet. For the model structure of this article, the structure of the recurrent neural network replaced by the self-attention mechanism is shown in Figure 1, and the self-attention mechanism module is shown in Figure 2:

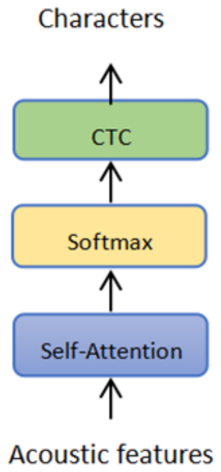


Figure 1. Basic model

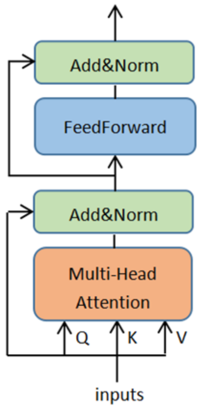


Figure 2.Self-attention block

This paper proposes a sliding mechanism similar to the convolutional neural network, which scans the acoustic features locally according to the 25% ratio of the input length as the slider length, as shown in Figure 3. The specific operation will be described in detail in section 4.1.2.

### 4.1.1. Multi-Head attention

Self-attention is a mechanism that associates different positions in the input sequence to calculate the input representation. Specifically, it has three inputs, namely query, key and value. The output of a query will be calculated as a weighted sum of values, where the weight of each value is calculated by the design function of the query and the corresponding key. Here, we use zoomed dot product attention, which is an effective self-attention mechanism, which has been demonstrated in [11]. As shown in Figure 2, Q represents the query, K is the key and V is the value, and the dimensions of the three variables are the input acoustic feature length multiplied by the model dimension, then the output of self-attention is:

$$Attention(Q,K,V) = soft\max(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

The function of the scaling factor is to prevent the softmax function from entering an area with a very small gradient.On the basis of single-head attention, we adopt a multi-head attention mechanism, which calculates the dot product attention of h zooms, where h represents the number of heads. The original paper maps d_model (model dimensions) h times, each time three dimensions are obtained, d_q, d_k, d_v, and the attention value of each head is calculated in parallel, then the output of multi-head attention is:

$$MultiHead\,(Q,K,V) = Concat(head_1,...,head_h)W^O \quad (5)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (6)$$

The dimension of the mapping matrix $W_i^Q$ is the model dimension multiplied by the query dimension, Similarly, the dimensions of $W_i^k$ and $W_i^V$ are the model dimension multiplied by the key dimension, the model dimension multiplied by the value dimension, and the query dimension equals the key dimension equals the value dimension. The dimension of $W^O$ is the model dimension multiplied by the model dimension, and the above dimension is expressed by the formula:

$$d_q = d_k = d_v = d_{\mathrm{mod}\,el} / h$$

In addition, there is a position feedforward network layer after the multi-head attention layer, which contains two linear transformations and a RELU activation function:

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (7)$$

The dimension of the mapping matrix $W_1$ is the model dimension multiplied by the feedforward network layer dimension, the dimension of $W_2$ is the feedforward network layer dimension multiplied by the model dimension, and the bias vector $b_1 b_2$ is learned, and the dimension is consistent with the model dimension. In these two After each sublayer, a layer normalization operation is connected, LayerNorm(x + Sublayer(x)), where Sublayer(x) is a function implemented by the sublayer itself.

### 4.1.2. Sliding Chunk mechanism

Since this article uses the self-attention mechanism, due to the characteristics of this mechanism, we must use the entire feature sequence as the input of the model to calculate the attention weight, and CTC originally did not achieve real-time output text, combined with the characteristics of attention and the shortcomings of CTC, and Inspired by the time-delay network, this paper proposes a local self-attention mechanism that uses a slider mechanism to obtain acoustic features locally. This mechanism effectively models long-term models by stacking multiple sliders to obtain a larger receiving field. Scenario to achieve online decoding function. In addition, regarding the size design of the slider, according to the input acoustic feature size, a fixed-length slider is taken by multiplying the acoustic feature length by a ratio of 25%. As shown in Figure 3, the fixed-length slider flows along the time axis of the feature sequence. Moreover, stacking multiple self-attention blocks makes it possible to model a longer time context without causing excessive performance degradation.

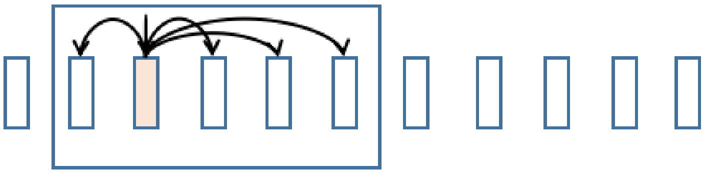


Figure 3. Local self-Attention block

The attention module maps the input $x_t$ into three vectors: $q_t, k_t v_t$ represents the query, key and value, respectively, and the output $h_t$ is the weighted sum of the value $v_t$ (varying with time), where the weight is determined by the dot product of the query and the key (through the softmax activation function Standardization) decision. For the single head example, it can be defined by the following formula:

$$h_t = \sum_{\tau=t-L}^{t+R} c_{t,\tau} v_t \tag{8}$$

Where ct(τ) = exp(qt · kτ)/Zt, Zt represents the normalization operation, which guarantees $\sum_{\tau} c_{t,\tau} = 1$. In a fixed-length slider, L and R represent the frames to the left and right of the current time t, respectively number. For the long example, its formula is:

$$h_{i,t} = \sum_{\tau=t-L}^{t+R} \alpha_{i,\tau} s_\tau \tag{9}$$

Where $\alpha_{i,\tau} = Attention(s_\tau, K, V)$ and K, V are the τ-th vector in the slider, hi, t represents the i-th head in the multi-head attention layer at time t, and $s_\tau$, K, V represent the τ vectors in the slider. The slider length of each block is L+R+1.

## 4.2. Loss Function

Existing work believes that only the CTC training model is used, and sometimes the training fails to converge, or the cross-entropy loss function is used to pre-train the model. If the CTC training model is used on this model, there is still the problem of instability and model instability. Therefore, this paper proposes to use CTC and cross entropy at the same time to make the model converge better. Then the loss function after using CTC and CE jointly is as follows:

$$\ell_{joint}(x) = \ell_{ctc}(x) + \ell_{ce}(x) \tag{10}$$

$$\ell_{ce}(x) = -\sum_{i=2}^{K}(1 - p(y_1 \mid x))t_i \log p(y_i \mid x) \tag{11}$$

The CTC loss function is shown in formula (3), $p(y_1 | x)$ represents the probability of the CTC blank label of the softmax output layer, and the probability of 1 minus the blank label is used as the normalization factor of the cross-entropy loss function, $t_i$ (i=2, ..., K) represents the target label at the frame level. This normalization factor plays an important role. At the beginning of training, the prediction of the acoustic model is like random guessing, and then CTC and cross-entropy loss function both play an important role in guiding the training. In the training process, the CTC loss often produces a shape peak distribution, each output target has only a few peaks, and the rest of the time is likely to predict blank labels. Therefore, the standardized cross-entropy loss function will help produce accurate alignment for the output target without affecting the allocation of blank labels. As a result, the proposed joint CTC-CE training will be more stable and help alleviate the delay problem.

## 5. Experiment

### 5.1. Dataset

The experiment mainly performed speech recognition tasks on Chinese (Mandarin). The open source speech corpus AISHELL-1 [21] is used for Mandarin speech recognition, and all speech files are sampled at 16 K Hz and 16 bits. The training set contains 150 hours of speech recorded by 340 speakers; the development set contains 20 hours of speech recorded by 40 speakers; and the test set contains 10 hours of speech recorded by 20 speakers. And the speakers in the training set, development set and test set do not overlap.

### 5.2. Experiment Set

As mentioned earlier, this article uses a connectionist-based temporal classification (CTC) speech recognition system. When doing the experiment, we used the 40-dimensional Mel filter library coefficient feature calculated on a 25ms window with a 10ms displacement. Each feature is rescaled so that the mean and unit variance of each audio sample is zero. When the processed frame is at time t, the number of left and right frames is asymmetrical, and these features are stacked in time through the local self-attention module, and finally down-sampled to a 30ms frame rate. This paper selects 4231 characters (including "blank" characters) as the model unit. During training, all audio samples are sorted by length and modeled using PyTorch [22], and Kaldi [23] is used for data preparation. Like the attention mechanism [2], this paper uses 6 self-attention models as encoders, in which the dimensions of query, key and value are all 128, the number of attention heads is 8, the model dimension and feedforward network layer The dimension of is 1024. Due to the introduction of local self-attention (see section 4.1.2), this article does not use the position coding formula in the original text of attention to reduce the amount of calculation. After that, an initial learning rate of 0.001 is used to train the network to reduce the joint loss function of CTC and CE.

### 5.3. Results and Discussion

Table 1. The CERs (%) of the development and test sets of AISHELL-1.

| Model | Dev. | Test. |
|---|---|---|
| Baseline | 7.36 | 8.51 |
| Baseline+CE | 7.29 | 8.42 |
| BN | 8.35 | 9.71 |
| ABN-U | 7.40 | 8.40 |

As shown in Table 1, we compared the character error rates of the four models on the AISHELL-1 dataset. Dev. and Test. respectively represent development dataset and test dataset.The Baseline model represents the model proposed in this article, and the second is the joint training with cross entropy. Model, BN is the basic model in [24], it is an acoustic model based on cyclic neural network and using CTC training, ABN-U is the sentence-level attention batch normalization model in [24]. From the data in the table, it can be seen that on the development dataset, the model proposed in this article is relatively better. On the test set, the model performs better than the basic model after adding cross entropy, and the effect is similar to ABN-U.

## 6. CONCLUSION

In this work, we propose a local self-attention encoder, which replaces the recurrent neural network with a self-attention module. The performance of the self-attention encoder is better than the BN (CTC original model under the same data set) model. Use local self-attention mechanism (slider mechanism) to realize online decoding function. Moreover, this paper also proposes a CTC joint cross-entropy criterion training model to improve model stability and facilitate better convergence of the model. The results show that the CTC joint cross-entropy criterion training method has greatly improved the basic model. During decoding, we observed that the model can predict characters with similar pronunciations well. In future work, we will explore how to optimize the parameter estimation problem of language models and unknown distribution data.

## REFERENCES

[1] Alex Graves and Navdeep Jaitly, “Towards end-to-end speech recognition with recurrent neural networks,” in Proceedings of the 31st International Conference on Machine Learning(ICML-14), 2014, pp. 1764–1772.

[2] Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, et al., “Deep speech: Scaling up endto-end speech recognition,” arXiv preprint arXiv:1412.5567,2014.

[3] Yajie Miao, Mohammad Gowayyed, and Florian Metze,“EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding,” in 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU).IEEE, 2015, pp. 167–174.

[4] Jan Chorowski, Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, “End-to-end continuous speech recognition using attention-based recurrent NN: First results,” arXiv preprint arXiv:1412.1602, 2014.

[5] Jan K Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk,Kyunghyun Cho, and Yoshua Bengio, “Attention-based models for speech recognition,” in Advances in Neural Information Processing Systems, 2015, pp. 577–585.

[6] William Chan, Navdeep Jaitly, Quoc V Le, and Oriol Vinyals,“Listen, attend and spell,” arXiv preprint arXiv:1508.01211,2015.

[7] Dzmitry Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, and Yoshua Bengio, “End-to-end attentionbased large vocabulary speech recognition,” arXiv preprint arXiv:1508.04395, 2015.

[8] Liang Lu, Xingxing Zhang, and Steve Renals, “On training the recurrent neural network encoder-decoder for large vocabulary end-to-end speech recognition,” in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016, pp. 5060–5064.

[9] William Chan and Ian Lane, “On online attention-based speech recognition and joint mandarin character-pinyin training,” Interspeech 2016, pp. 3404–3408, 2016.

[10] A. Graves, A. Fernandez, F. Gomez, and J. Schmidhuber, “Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks,” in Proceedings of the 23rd International Conference on Machine Learning. IEEE, 2006.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in Advances in Neural Information Processing Systems, 2017, pp. 6000–6010.

[12] J. Cheng, L. Dong, and M. Lapata, “Long short-term memorynetworks for machine reading,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2016, pp. 551–561.

[13] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang,“DiSAN: Directional self-attention network for RNN/CNNfree language understanding,” in Proc. AAAI Conf. Artificial Intell. (AAAI), 2018.

[14] Y. Zhang, W. Chan, and N. Jaitly, “Very deep convolutional networks for end-to-end speech recognition,” in Proc. IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP). IEEE,2017, pp. 4845–4849.

[15] S. Kim, T. Hori, and S. Watanabe, “Joint CTC-attention based end-to-end speech recognition using multi-task learning,” in Proc. IEEE Int. Conf. Acoustics Speech Signal Process.(ICASSP). IEEE, 2017, pp. 4835–4839.

[16] C.-C. Chiu, T.N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen,Z. Chen, A. Kannan, R.J. Weiss, K. Rao, K. Gonina,et al., “State-of-the-art speech recognition with sequence-tosequence models,” in Proc. IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP). IEEE, 2018, pp. 4774–4778.

[17] A. Zeyer, K. Irie, R. Schluter, and H. Ney, “Improved training of end-to-end attention models for speech recognition,” in Proc. Ann. Conf. Int. Speech Communication Assoc. (INTERSPEECH), 2018.

[18] L. Dong, S. Xu, and B. Xu, “Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 5884–5888.

[19] L. Dong, F. Wang, and B. Xu, “Self-attention aligner: A latencycontrol end-to-end model for asr using self-attention network and chunk-hopping.” arXiv: Computation and Language, 2019.

[20] H. Sak, A. Senior, K. Rao, O. Irsoy, A. Graves, F. Beaufays, and J. Schalkwyk, “Learning acoustic frame labeling for speech recognition with recurrent neural networks,” in Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. IEEE, 2015, pp. 4280–4284.

[21] H.Bu, J. Du, X. Na, B. Wu, and H. Zheng, “Aishell-1: An open-source mandarin speech corpus and a speech recognition baseline,” in O-COCOSDA2017.

[22] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” 2017.

[23] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz et al., “The kaldi speech recognition toolkit,” IEEE Signal Processing Society, Tech. Rep., 2011.

[24] Fenglin Ding, Wu Guo, Lirong Dai, Jun Du,“Attentive batch normalization for lstm-based acoustic modeling of speech recognition，″CoRR abs/2001.00129，2020

## AUTHORS

**Huizhen Deng** was born in Anhui, China, in 1995.She received the B.S. degree in Qindao College of Qingdao Technological University, China, in 2018.She is currently pursuing the M.S. degree in computer science and technology in Heilongjiang University.

**Zhaogong Zhang**, professor, Ph.D, postdoctoral, master supervisor, Department of Software and Theory, School of Computer Science and Technology, Heilongjiang University. Member of the Big Data Division of Heilongjiang Institute of Industry and Application. He received a bachelor's degree in basic mathematics, a master's degree in basic mathematics and a doctorate degree in computer software and theory from Harbin Institute of Technology in the Department of Mathematics of Hei longjiang University in 1985, 1991 and 2003, respectively. His main research interests include bioinformatics, data mining, statistical genetics, big data, cloud computing, etc.

# APPLICABILITY OF DEEP NEURAL NETWORKS ON THE TASK OF DOCUMENT RETRIEVAL

M. Shoaib Malik[1, 2] and Dagmar Waltemath[1]

[1]Medical Informatics Laboratory, Institute for Community Medicine,
University Medicine Greifswald, Germany
[2]Department of Computer Science, Air University, Islamabad, Pakistan

## *ABSTRACT*

*A Deep Neural Network (DNN) can be used to learn higher-level and more abstract representations of a particular input. DNNs have successfully been applied to analysis tasks including image processing, unsupervised feature learning, and natural language processing. DNNs furthermore can improve computing performance when compared to shallower networks, for example in pattern recognition tasks in machine learning. Recent usage of DNNs in search engines for the Web have impacted that technology in industrial scale applications. One example for such an application is deepgif - a search engine for Graphics Interchange Format (GIF) images that is based on a convolutional neural network and takes natural language text as query. In this study, we developed a tool and compared the performance of feed-forward neural networks and deep architectures of recurrent neural network using the case of document retrieval. This study first discusses two architectural setups used to build the models and then provide a detailed comparison of their performance. The goal is to identify the architecture that is most suited for the task of document retrieval.*

## *KEYWORDS*

*Deep Neural Network, Machine Learning, Document Retrieval, Feed-Forward Neural Network, Recurrent Neural Network*

## 1. INTRODUCTION

Textual documents are an integral part of our everyday life. With an increased amount of text, and thus documents, becoming available on the world wide web, the chance of finding the information that best meets a user's need is significantly decreasing (Skovajsova, 2010). The availability of cheap and effective storage media has resulted in an enormous rise in the size of textual databases which are widely used in traditional library science environments, in business applications (e.g., manuals, newsletters, and electronic data interchanges), and in scientific applications (e.g., electronic community systems and scientific databases) (Chen, 1995). This has required greater efforts in the retrieval of relevant information, especially from large scaled databases. Various methods have been proposed over the years to deal with the large amounts of data in textual document space. Models that are built on neural networks usually cluster or classify documents into groups with similar documents belonging to the same group. Many document retrieval systems are built on keyword searches. However, these systems do not consider the relations among the passages of text within a document (Treeratpituk and Callan, 2006) (Wei and Croft, 2006).

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 285-304, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101123

The task of document retrieval is to find documents of unstructured or semi-structured nature that satisfies the information need of a user from within a large collection. The goal of document retrieval systems is thus to retrieve documents that are relevant to the user's information need. Document retrieval systems must hence be able to retrieve desired information about a subject rather than to retrieve documents that look similar to a given query (Baeza-Yates and Ribeiro-Neto, 1999). To accomplish that task, document retrieval systems apply specific concepts to represent the query and documents, and to assign relevant documents to the query (Skovajsova, 2010). The issue of predicting relevance of a document to the user's information need is usually based on a ranking algorithm where documents appearing at the top of the list are considered to be more relevant than those at the bottom (Baeza-Yates and Ribeiro-Neto, 1999).

Many Web-based tools retrieve information through general-purpose search engines like Google, Bing and Yahoo or through specialized search engines such as PubMed (for biological and medical publications). While early search engines ranked their results based on content of the document, more modern search engines evaluate the semantics of the document. Search engines like Google and Yahoo consider page reputation as one of the major criteria of relevance ranking (Deepak and Deepika, 2012).

Several approaches to document retrieval have been proposed over the years including the boolean model (van Rijsbergen, 1979) (Baeza-Yates and Ribeiro-Neto, 1999) (Cordon et al., 2002) (Herrera-Viedma, 2001), vector space model (Manning et al., 2008) (Hotho et al., 2005) (Lan et al., 2005) (Scheir and Lindstaedt, 2006), document based language model (Ogilvie and Callan, 2003) (Wang et al., 2005), and the models built using neural networks (Cheung and Cannons, 2002).

The Boolean model is one of the first models of information retrieval for document extraction. It uses a term-document matrix where every cell in the Boolean matrix is filled with 0 or 1 based on whether a word appears in the document or not. The vectors for the terms in the query are put together using a Boolean operation such as AND, OR, NOT. For a collection with 1 million documents with each 1000 words, roughly 6GB are required to store such a collection (assuming an average of 6 bytes/word). For this reason, the documents are stored in an inverted index (Zobel and Moffart, 2006) with variable sized posting lists.

The vector space model is commonly used in information retrieval systems. It works on the Term Frequency- Inverse Document Frequency (TF-IDF) weights of the terms, which is a common weighting scheme in information retrieval. Opposed to results returned by Boolean model, which are not ranked in any presumed order of importance, the retrieved documents can easily be ranked in decreasing order of the query-document similarity for vector space models (Salton et al., 1983). The most common similarity metric used is the cosine similarity (Turney and Pantel, 2010) between the query and the document. A lot of variants have been proposed over the years for better retrieval output.

The document based language model is the most direct way to estimate a language model from a large collection and assumes an underlying multinomial model. It estimates the probability of each document in the collection generated the query independently. This approach is not very good at estimating novel terms. For this reason, smoothing (Chen and Goodman, 1999) is applied that compensates for data sparseness by stealing a little probability mass from the seen terms and adding it to the unseen terms.

Neural networks provide a convenient knowledge representation for document retrieval applications in which nodes typically represent objects such as user query, keywords or documents. Such models are usually a combination of a feed-forward and spreading activation neural network. The feed-forward model learns the keywords against the query whereas the spreading activation model learns the relevant documents against the input keywords as depicted in figure 1 (Skovajsova, 2010) (Chen, 1995) (Mokris and Skovajsova, 2005). Neural networks have been used in different context but mostly on sentiment analysis. Le and Mikolov (2014) presented a framework that learns continuous distributed vector representations for a paragraph from a document. This framework works in a similar manner as learning word embedding described in Mikolov et al. (2010). It maps every paragraph to a unique vector just like every word and predicts the next word in context using the concatenated sentence and current word representation. Similar concept is also described in Lin et al. (2015) for document modeling on sentiment analysis.

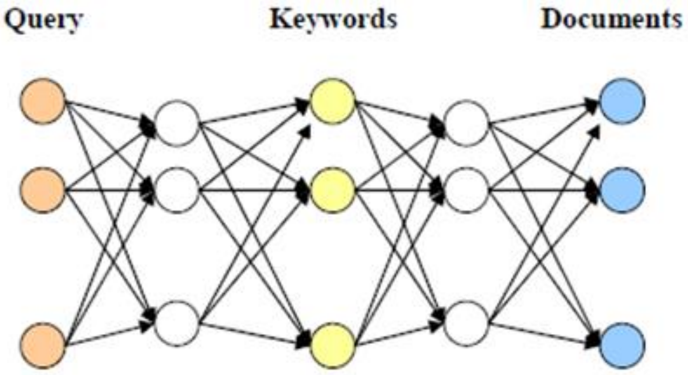


Figure 1: Cascade neural network model

The primary objective of this study is to develop and then evaluate the performance of deep learning based information retrieval systems, namely feed-forward based system and recurrent neural network based system. In particular, we will investigate feed-forward and recurrent neural networks for this task. Recent results have shown that recurrent models outperform feed forward model on the task of language modeling. This is mainly because the deep architecture of recurrent neural network allows to store context information in the hidden layer to better inform the current prediction. Our goal is to see if storing the context information for longer period of time is beneficial as well when building document model instead of a language model. Moreover, we also evaluate how change in the number of words in the query affects the retrieval output of models trained on both these architectures.

## 2. METHODS

This study investigates two different neural network architectures: the feed-forward (Bengio et al., 2003) and recurrent (Jain and L.R., 1999) neural network architectures.

### 2.1. Feed Forward Neural Network

The feed-forward neural network document model (Svozil et al., 1997) is an n-gram model where the posterior probability distribution of topic and document is computed for given n words. All collections of this study group documents into topics. Hence the output layer is factorized into a topic and a document layer. The size of the output topic layer is equal to the number of topics C and the size of the output document layer is equal to the number documents D in the collection. The feed-forward neural networks were introduced as an alternative to widely used back-off language models and have been reported to perform better in (Bengio et al., 2003), (Schwenk and Gauvain, 2004), (Gauvain et al., 2005), and (Emami and Jelinek, 2004) when used on the application of language modeling.

## 2.2. Recurrent Neural Network

When building a feed-forward document model, the network reads a subset of words (i.e., n-gram) at each time-step and then predicts the probability distribution of the topic and document that the input n-gram belongs to (Lin et al., 2015). The feed-forward neural network ignores the contextual information in texts and remains unsatisfactory for capturing the semantics of the words (Lai et al., 2015). It is an open question if capturing the semantics of the text by a deep recurrent neural network is of any value when learning topics along with documents in the output layer. Deep recurrent neural networks are created by stacking multiple hidden layers on top of each other, with the output sequence of one layer forming the input sequence for the next.

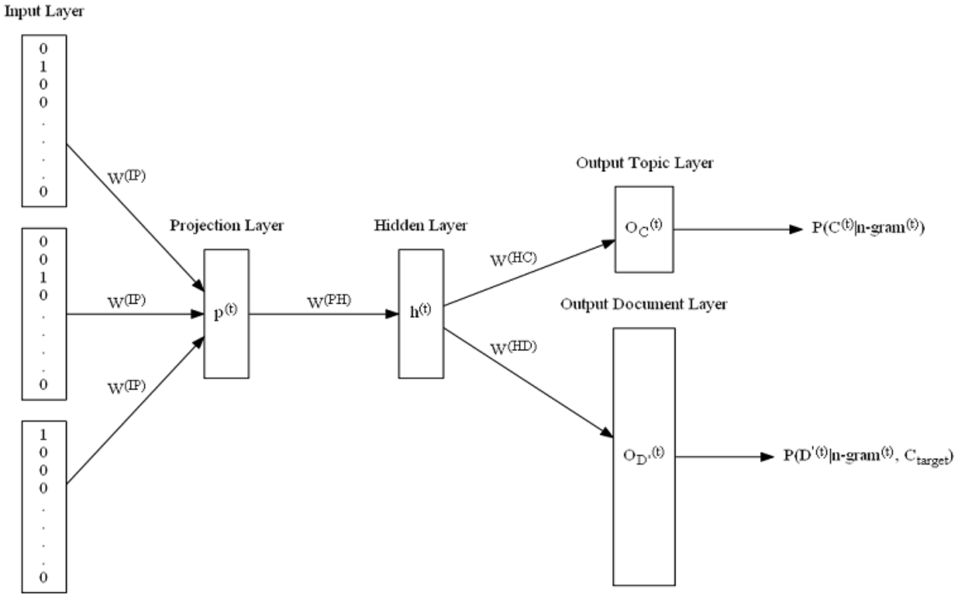


Figure 2: Architecture of feed-forward neural network

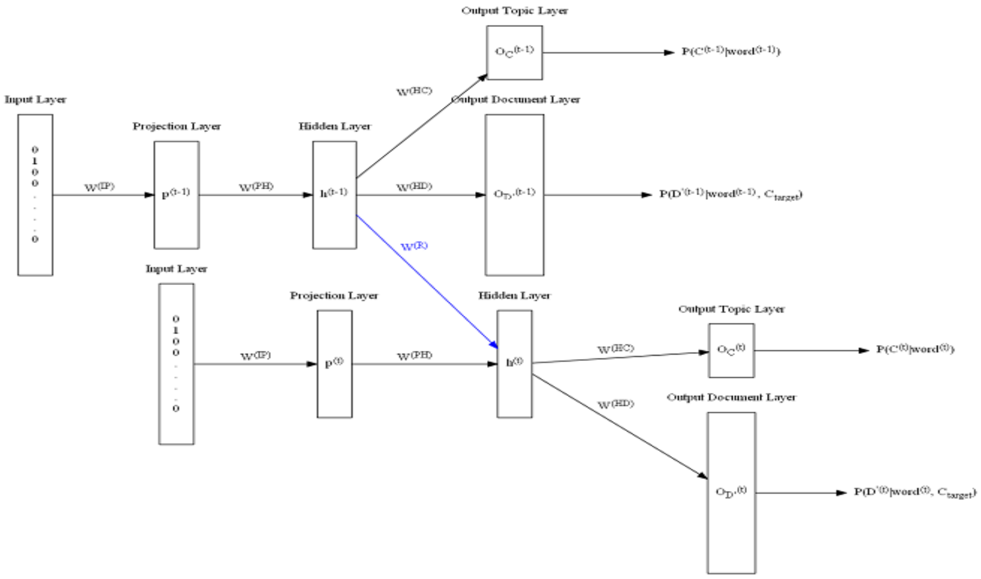


Figure 3: Architecture of recurrent neural network

## 3. IMPLEMENTATION DETAILS

The Neural Network Document Retrieval (NNDR) toolkit, which implements feed-forward neural network and deep architecture of recurrent neural network. is implemented using CUDA C/C++ (Nickolls et al., 2008) and Java (Arnold et al., 2000). This chapter discusses the translation of functions in respective forward pass and learning algorithms to CUDA pseudocode along with some of the considerations that were considered during the implementation of the toolkit in order to train the networks.

### 3.1. CUDA

CUDA is a parallel computing platform and application programming interface model created by NVIDIA for applications running on Graphics Processing Unit (GPU). Nowadays, hundreds of industry-leading scientific computing applications are already GPU-accelerated making use of multiple cores of GPU and fast arithmetic operation capability with greater floating-point performance from CUDA. From the programmer's perspective, the CUDA architecture is divided into threads, blocks and grids. A grid is organized as a 2D array of blocks and a block is organized as 3D array of threads. The high performance comes from the concurrent execution of multiple threads resulting in reduced latency.

### 3.2. Data Parallelism

With data parallelism, we take our documents and we process a subset of these documents in a batch. This means that we use the same model for each mini-batch but feed it with different document. The two dimensionalities of the CUDA grid can be taken advantage of in order to achieve data parallelism with each mini-batch or document being processed in the separate row of the grid. Data parallelism uses the same weights in forward pass of the network to give out topic and document probabilities as output for each mini-batch. However, in the backward pass, the weight gradients need to be synchronized from all the mini-batches and then averaged. If we do not process documents in a batch and rather process each document individually, such an approach would undo learning that it did with document D-1, D-2, etc. and would not converge to the optimal parameters. It would just hop back and forth because it does not consider all the topics at once. To avoid this, documents are processed in batch and the batch size is set to the number of topics and only one document is processed from each topic. It is made sure that the number of words processed in batch is equal for all documents. In the next iteration, the next document from same topic is chosen and similar steps described in preceding text are performed.

### 3.3. Text Processing

The raw text in the collections is pre-processed to reduce the problem's dimensionality and to ensure the completeness, consistency, and interpretability of the data. The following steps are performed.

1. Substitute TAB, NEWLINE and RETURN characters by SPACE.
2. Turn all letters to lowercase.
3. Substitute multiple SPACES by a single SPACE.
4. Remove the 524 SMART stopwords.
5. The title of each document is simply added in the beginning of the document's text.
6. Apply Porter's Stemmer (Porter, 1980) to the remaining words to reduce the words to their morphological root, so that the number of different terms in the documents is reduced.

7. Split the training dataset into training and validation dataset using heldout technique. The validation data is used to control the learning rate.

## 3.4. Vocabular Truncation

Just like hidden and output layer, the computation done between input and projection layer is also very computationally expensive - even more so than hidden and output layer in some cases, which limits its application to real world problems. Taking an example of language modeling, most of the models trained in today's research are trained on millions of words. It would take impractically long to train these models with very large vocabulary. We took the same measures as most of the researchers working in the field of natural language processing do - merge all infrequent words into a special <unk> class that represents the probability of all rarely seen words in the collection. Although this approach improves the speed of the training, it suffers from a loss of accuracy (Mikolov et al., 2010).

## 3.5. Out of Vocabulary Words

Out of Vocabulary words are unknown words that appear in the test data but not in the training data. Since it is practically impossible to train a system on all words that exist in natural language, some steps must be taken to deal with this problem. Although, most common approach used in textual applications of natural language processing for models trained on neural networks is to assign an <unk> tag to some of the most infrequent words in the collection, we went an extra mile. From the piece of text that was selected to be trained from a document, we replaced a word at a random location with an <unk> tag. This way, we tried to make sure that the models that were trained were also able to generalize well for the words that are not seen in the training data.

## 3.6. Variable Learning Rate

A standard refinement to gradient descent is to use a variable learning rate that is updated after each training epoch. We used perplexity as the evaluation criterion when training the models. Perplexity is a measure of the average branching factor of the topics and documents when predicting them from the input n-grams or words. The learning rate is varied according to changes in validation perplexities across epochs. If the previous learning rate decreased the validation perplexities across epochs, then the learning rate is left unchanged. If the previous learning rate did not decrease the validation perplexities across epochs, the learning rate is reduced to half after each subsequent epoch (Blackwood, 2005). The default value of initial learning rate for models trained on feed-forward neural network is 0.1 and it is 0.2 for models trained on recurrent neural network.

The document retrieval system was trained and evaluated on 3 collections. Every collection consists of a set of pre-classified documents where every document in the collection belongs to a particular topic. The datasets are the 20-Newsgroup collection, the Reuters-21578 collection, and the Cade collection. All collections were obtained from (Cardoso-Cachopo, 2007).

Table 1: Document distribution and vocabulary sizes of all the collections.

| Collection | No. of Training Documents | No. of Validation Documents | No. of Test Documents | No. of Topics | Vocabulary Size |
|---|---|---|---|---|---|
| 20-Newsgroups | 8951 | 2231 | 7528 | 20 | 13926 |
| Reuters-21578 | 4901 | 1200 | 2568 | 52 | 8862 |
| Cade | 13597 | 3393 | 13661 | 12 | 18947 |

Every document in the collections is available as a single running text. Since the documents are not available as sentences, the number of words trained from each document in batch is equal to the number of words W in the shortest document. From the rest of the documents except the shortest document, a random chunk of text is taken whose length is equal to W.

The 20-Newsgroups collection is a set of newsgroup documents, which are nearly evenly partitioned across 20 different newsgroups. The collection has become a popular dataset for experiments in text applications of machine learning techniques, such as text classification and text clustering.

The Reuters-21578 collection is also one of the widely used collections in text classification. All the documents contained in this collection appeared on the Reuters newswire in 1987 and were manually classified by personnel from Reuters Ltd. and Carnegie Group, Inc. in 1987.

The documents in the Cade collection correspond to a subset of web pages extracted from the Cade Web Directory, which points to Brazilian web pages classified by human experts.

## 4. EVALUATION METRICS

A document retrieval system assigns higher relevance to documents that are more similar to the query. We compared the performance of different models that we developed and trained. This comparison can be carried out either by looking at the ranked retrieval results, or by adopting a performance measure as an indicator to derive the perfection in prediction, in particular as a function of the number of topics and topic imbalance.

As far as the collections used for this study are concerned, every test document is available as a single running text and hence a substring of document is chosen as a query to the networks. The substring is selected from the beginning of the document as we found the words in beginning of the document to be most informative regarding the document and the topic. When evaluating the model trained on a particular architecture for a particular collection, for each input word or n-gram, the topic with maximum probability from output topic layer is recorded. Then, the probabilities of all the documents from that topic are computed in output document layer and are added to the output vector whose length is equal to the number of training documents in the collection. In other words, we estimate the probability that each document generated the query.

### 4.1. Mean Average Precision

Precision (Powers, 2007) is the fraction of retrieved documents that are relevant to the query. A document is considered retrieved if its probability in the output vector is greater than 0. In recent years, other measures have become more common, one of which is Mean Average Precision (MAP) (Manning et al., 2008), which provides a single-figure measure of quality across recall levels and has been shown to have especially good discrimination and stability. For a single query, average precision (Manning et al., 2008) is the average of the precision values obtained for the set of top k documents. For our system, the number of relevant documents is equal to the number of documents belonging to target class Ctarget. For evaluation purposes, we evaluate only the top 20 documents that are retrieved (at most) and hence the number of relevant documents is the lower bound on 20 and number of documents belonging to Ctarget (i.e., n=min(20,|Ctarget|)). At the end when all queries from the query set Q have been run, mean average precision can be calculated as the mean of the average precision scores for each query.

## 4.2. Matthews Correlation Coefficient

The Matthews correlation coefficient (MCC) (Matthews, 1975) is used in machine learning as a measure of the quality of classifications. MCC formulation was originally reported for binary classification that works on a 2 x 2 contingency table taking into account the true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) and is generally regarded as a balanced measure which can be used even for imbalanced collection. The MCC is in essence a correlation coefficient between the observed and predicted classifications returning a single value in range [-1, +1] where +1 is perfect classification with all zeros in the contingency table except the diagonal, -1 is the extreme mis-classification case with all zeros in the diagonal of the contingency table, and 0 corresponds to prediction at random. A totally random prediction can occur when for all the queries, the model retrieves documents of one particular topic only or when for queries belonging a particular topic, the number of predictions is equal for all the topics.

Since all the collections used for this study are divided into more than 2 topics, the definition of MCC reported for the multi-class case (Jurman et al., 2012) was used. For multi-class case, the MCC formulation works on a N x N contingency table C where N is the number of topics/classes in the collection. For evaluation purposes, the number of true positives for a particular topic is equal to the number of successful retrievals for that topic. If a retrieval is unsuccessful for a query belonging to a topic i, the entry ij in the contingency table is incremented by 1 where j is the topic that the highest ranked document in the output vector belongs to.

## 4.3. Mean Rank

Ranking (Baeza-Yates and Ribeiro-Neto, 1999) of the documents can be computed by sorting the documents in decreasing order of probability from output vector. The rank is the highest index of document from the target topic Ctarget in the output vector.

# 5. EXPERIMENTS AND RESULTS

This section discusses the results obtained by running experiments on the collections in order to evaluate the retrieval outputs of both the architectures. To evaluate the performance of the feed-forward neural network, 1-gram, 3-gram, and 5-gram models were built. A 1-gram model takes one word in the input whereas 3-gram and 5-gram models take a sequence of 3 and 5 words in the input respectively. After each network run, the window in the document slides by 1 and the next sequence of 1, 3, or 5 words is read from the document. For recurrent neural network, models were built using bptt parameter of 1, 3, and 5 for each collection to see how storing the context information affects the retrieval output. Recurrent models were trained using an initial learning rate of 0.2 whereas feed-forward models were trained using an initial learning rate of 0.1. To avoid over-fitting of the training documents, the regularization parameter used for models of both architectures was fixed at 10-6. It shall be noted that for models built on feed-forward neural network, the projection layer size is 200 for each word in input n-gram.

Table 2: Network setups used to build the models.

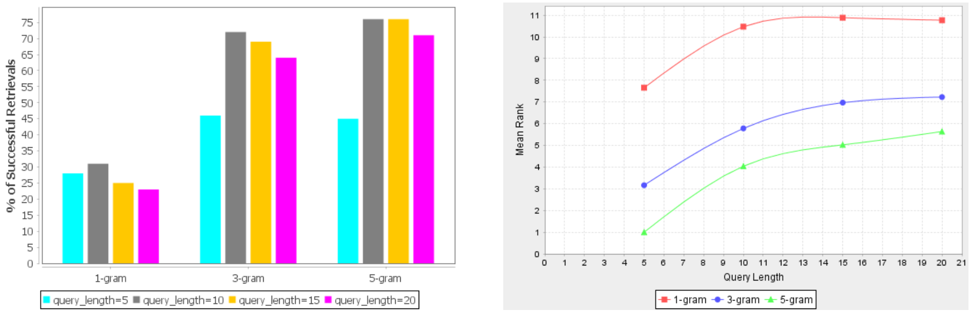
| | Collection | | | | | |
|---|---|---|---|---|---|---|
| | 20-Newsgroup | | Reuters-21578 | | Cade | |
| | Projection Layer Size | Hidden Layer Size | Projection Layer Size | Hidden Layer Size | Projection Layer Size | Hidden Layer Size |
| FFNN | 200 | 800 | 200 | 600 | 200 | 800 |
| RNN | 200 | 800 | 200 | 600 | 200 | 800 |

### 5.1. Retrieval Effectiveness for FFNN

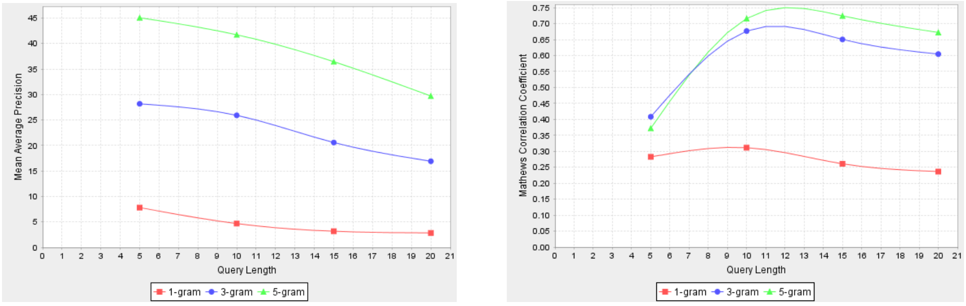
Figure 4, 5, and 6 show the percentage of successful retrievals along with the results returned by the evaluation metrics for the respective collections trained on feed-forward model. A retrieval is considered to be successful if at least one of the relevant documents is retrieved for a given query. It is seen that the mean rank and the mean average precision for the Reuters-21578 and 20-Newsgroup collection is significantly better than the Cade collection which is expected as the number of documents and vocabulary size of the Reuters-21578 and 20-Newsgroup collection is lower than that of Cade collection. A larger vocabulary and greater number of documents in the Cade collection makes it a little hard for the network to capture the n-gram to topic and n-gram to document relationship.

For higher order n-grams, the models trained on all the collections performed better than lower order n-grams, which had been expected. The larger the n-gram on which we train the model, the more coherent the training documents. For 1-gram model, there is no coherent relation between words whereas the 5-gram model has some local word-to-word coherence which is reflected in the results. For 5-gram model that was trained for each collection, it is seen that with an increase in the length of query from 5 words to 10 words, the retrieval performance is significantly decreased. However, for the 20-Newsgroup collection, MAP becomes better for query length of 15 and 20. The reason for best performance in case of shortest query of 5 words for a 5-gram model is that average precision for a query will be 1 whenever a correct prediction is made in the output topic layer for the input 5-gram. However, for queries where incorrect prediction is made in the output topic layer, the average precision will be 0. Under such circumstances, the percentage of successful retrievals gives the same estimate as the MAP score which is evident from the results shown for respective collections.

For the Reuters-21578 and Cade collection, we were able to achieve maximum retrieval performance for shortest query length of 5. For longer queries, performance of the Reuters-21578 collection became poorer while that of the Cade collection remained constant. Figure 7 shows the skewness of training data for all collections. Since the training dataset for the Reuters-21578 collection is very skewed, the inability of models trained on these two collections to perform better for longer queries can be explained with the help of an example model trained on a skewed collection.

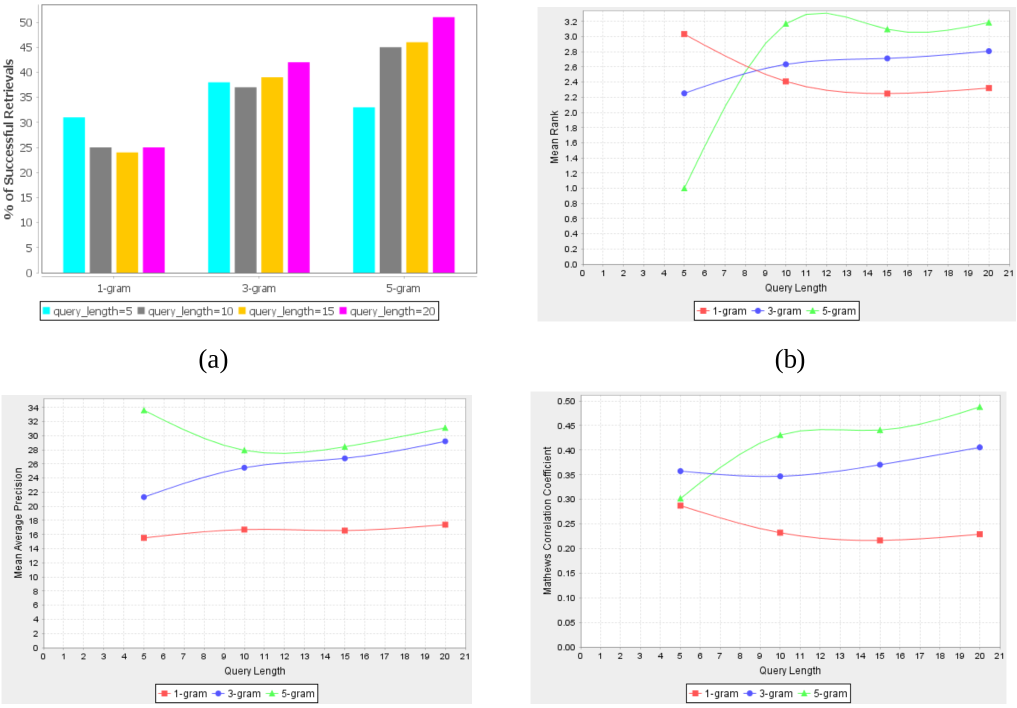


(a) (b)



(c) (d)

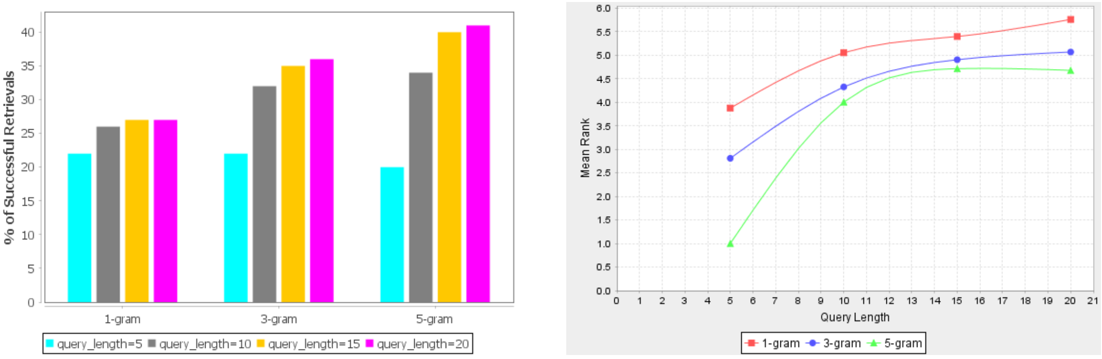
Figure 4: (a) percentage of successful retrievals, (b) mean rank, (c) mean average precision, and (d) matthews correlation of the Reuters-21578 collection trained on feed-forward neural network



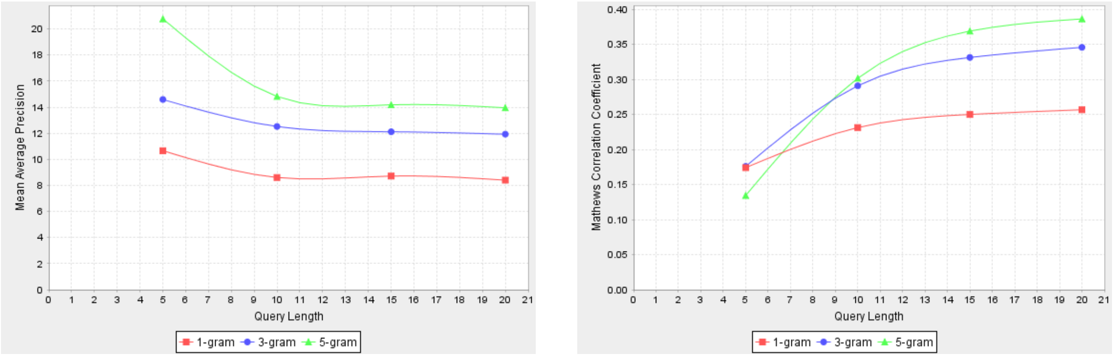
(a) (b)

(c) (d)

Figure 5: (a) percentage of successful retrievals, (b) mean rank, (c) mean average precision, and (d) matthews correlation of the Reuters-21578 collection trained on feed-forward neural network



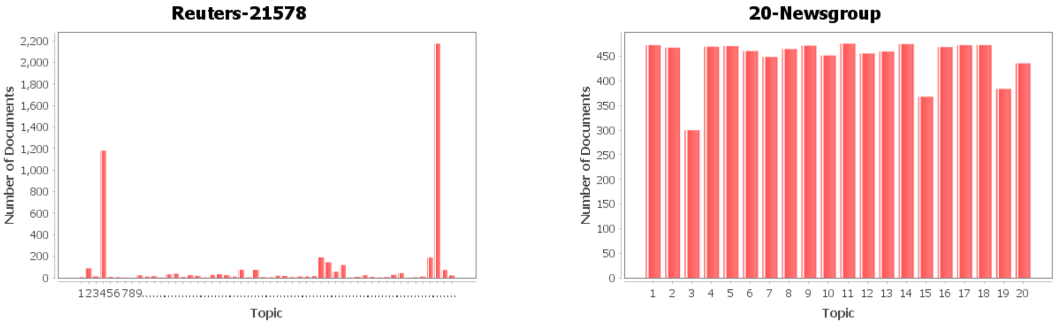
(a) (b)



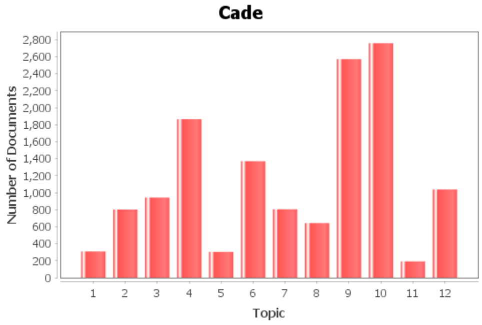
(c) (d)

Figure 6: (a) percentage of successful retrievals, (b) mean rank, (c) mean average precision, and (d) matthews correlation of the Cade collection trained on feed-forward neural network

Figure 8 shows snapshot of network runs for a 1-gram model with the state of the output vector after the softmax layers are computed for both topics and documents for each input word. Let us suppose we have a test document “university of saarland germany”, and the collection consists of 7 documents, divided into 3 topics: d0, d1, d2 belong to t0; d3, d4, d5 belong to t1; d6 belongs to t2. Further suppose that the test document belongs to t1 (i.e. Ctarget = t1). For evaluation purposes, let us consider the first 5 documents (at most) that are retrieved. Let us say our model predicts the first two words of the test document belonging to t0 with a probability of 0.6 and 0.5, receptively. The state of the output vector after sorting and averaging is d0 = 0:4; d1 = 0:3; d2 = 0:3. It can be seen that Ctarget is never maximized by the model and hence no document from that topic is retrieved and as of yet the average precision for that query is 0. Let us increase the query length to 3 and suppose that for the next input word saarland, Ctarget is maximized with probability of 0.4 and we are able to retrieve the relevant documents, d3, d4 and d5, for this input word. Now, the state of the output vector after sorting and averaging is d0 = 0:27; d1 = 0:2; d2 = 0:2; d4 = 0:13; d5 = 0:1; d3 = 0:08 and rank is 4. At this stage, the average precision for the query is 0.22. Obviously if Ctarget is not or seldom maximized with a shorter queries, the chances of the highest ranked document from Ctarget to be higher in the output vector for longer queries also decrease which is why rank is poor for longer queries. If we now increase the query length to 4, provide the next word in the test document (i.e., germany) to the model and assume that this time the model predicts this word belonging to topic t2. Since there is only one document belonging to t2, the model will predict the probability of document d6 to be 1.0. The state of the output vector after sorting and averaging at this stage is d6 = 0:25; d0 = 0:2; d1 = 0:15; d2 = 0:15; d4 = 0:1; d5 = 0:08; d3 = 0:06 and the average precision for this query is decreased to 0.07.



(a) (b)

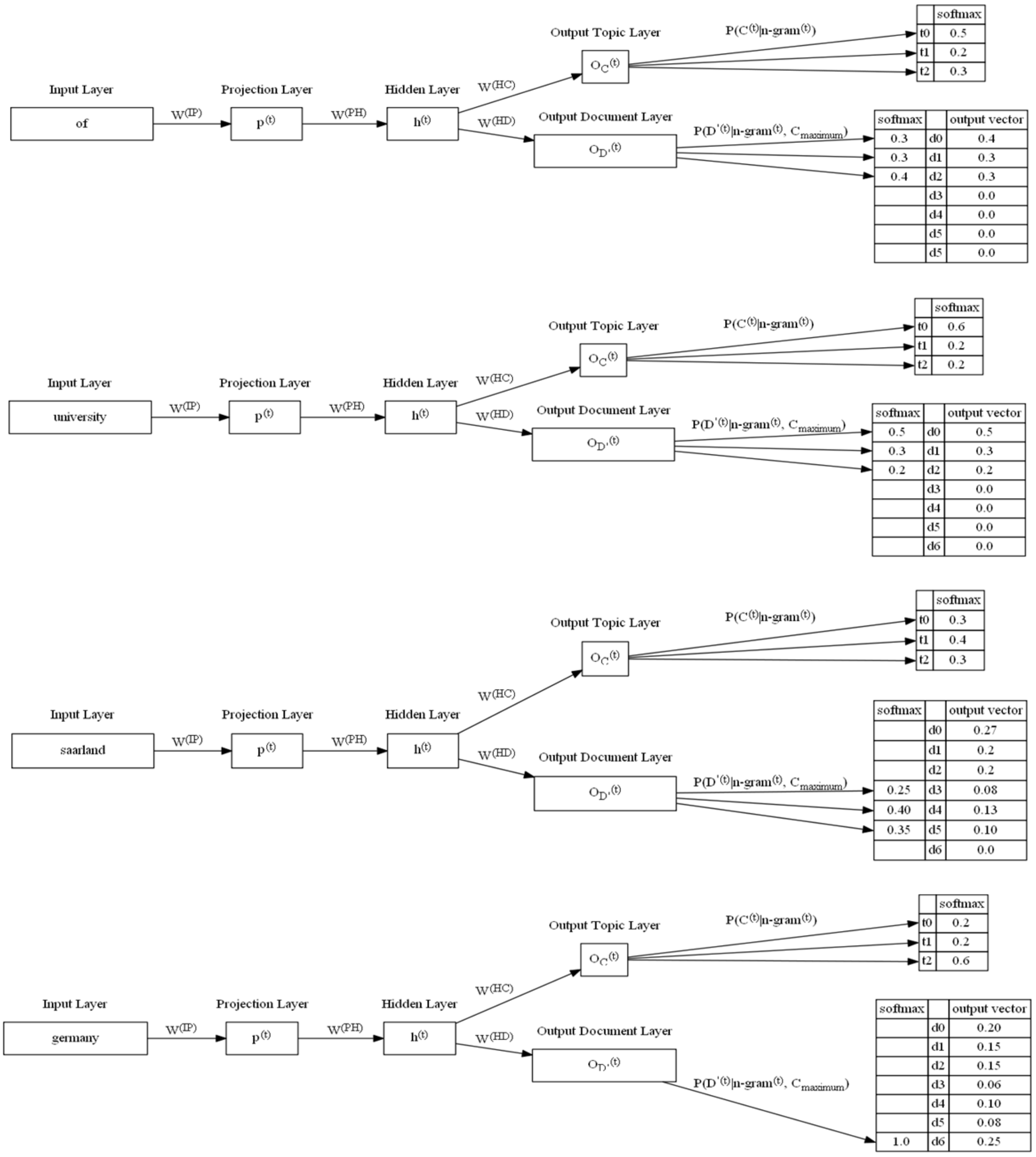


(c)

Figure 7: Skewness of the training data for (a) Reuters-21578, (b) 20-Newsgroup, and (c) Cade collection

We clearly see how a wrong prediction in the output topic layer can result in a drastic decrease in the performance on a model trained on a very skewed dataset. From our example, the document d6 has the highest probability in the output vector and consequently has the highest rank and has pushed the relevant documents d3, d4, and d5 further down in the output vector. Because of this, the chances of these relevant documents to be among the top few documents to be evaluated becomes very low. Thus, the performance of the system trained on skewed datasets gets poorer with an increase in the query length. For the Reuters-21578 collection, the number of successful retrievals increases when the number of words in the query is altered from 5 words to 10 words. After that the number of successful retrievals exhibit a continuous decrease for longer queries which is also depicted in figure 9. For topic 4 (i.e., acq) of that collection, which has 1181 documents out of a total of 4901 in the training dataset, increasing the length of the query has resulted in lesser number of successful retrievals when increasing the query length from 10 to 20 words. Conversely for query length of 15 or 20 words for the 20-Newsgroup collection, the increase in the length of the query has resulted in more relevant documents to be among the top 20 documents that are evaluated as compared to 10 words in the query. A non-skewed collection like 20-Newsgroup is less prone to the wrong predictions when increasing the length of query. For almost all 20 topics of the 20-Newsgroup collection, increasing the length of the query has resulted in more successful retrievals as shown in figure 10. For topics that show slight decrease in number of successful retrievals with increase in query length, wrong predictions for few of the n-grams in the query does not greatly affect the overall performance of the model for that query because of non-skewness of the dataset.

A striking observation can be made when looking at the MAP and MCC metrics. For all the collections trained on forward neural network, there is an inverse relationship between the MAP and the MCC value with respect to change in the length of query. For the Reuters-21578 and Cade collection, very high MCC values are recorded for longer queries along with low MAP scores. The inverse relationship between the MCC values and the MAP scores indicate that for the models trained on feed-forward neural network for the Reuters-21578 and Cade collection, the confidence in the relatively better retrieval output is very low for shorter queries. The high MCC values for low precision scores also suggest that for most of the queries, the models were able to maximize the topic Ctarget at least once and were able retrieve the relevant documents but most probably these documents were low in the output vector because of the skewness of the training dataset. For the 20-Newsgroup collection, different results are observed. For this collection, high MCC values are recorded for longer queries along with high precision scores suggesting that the confidence in relatively better retrieval output for longer queries is greater as compared to shorter queries.



| | softmax |
|---|---|
| t0 | 0.5 |
| t1 | 0.2 |
| t2 | 0.3 |

| softmax | | output vector |
|---|---|---|
| 0.3 | d0 | 0.4 |
| 0.3 | d1 | 0.3 |
| 0.4 | d2 | 0.3 |
| | d3 | 0.0 |
| | d4 | 0.0 |
| | d5 | 0.0 |
| | d5 | 0.0 |

| | softmax |
|---|---|
| t0 | 0.6 |
| t1 | 0.2 |
| t2 | 0.2 |

| softmax | | output vector |
|---|---|---|
| 0.5 | d0 | 0.5 |
| 0.3 | d1 | 0.3 |
| 0.2 | d2 | 0.2 |
| | d3 | 0.0 |
| | d4 | 0.0 |
| | d5 | 0.0 |
| | d6 | 0.0 |

| | softmax |
|---|---|
| t0 | 0.3 |
| t1 | 0.4 |
| t2 | 0.3 |

| softmax | | output vector |
|---|---|---|
| | d0 | 0.27 |
| | d1 | 0.2 |
| | d2 | 0.2 |
| 0.25 | d3 | 0.08 |
| 0.40 | d4 | 0.13 |
| 0.35 | d5 | 0.10 |
| | d6 | 0.0 |

| | softmax |
|---|---|
| t0 | 0.2 |
| t1 | 0.2 |
| t2 | 0.6 |

| softmax | | output vector |
|---|---|---|
| | d0 | 0.20 |
| | d1 | 0.15 |
| | d2 | 0.15 |
| | d3 | 0.06 |
| | d4 | 0.10 |
| | d5 | 0.08 |
| 1.0 | d6 | 0.25 |

Figure 8: Example to demonstrate the effect on evaluation metrics with alteration in query length for collections trained on feed-forward model

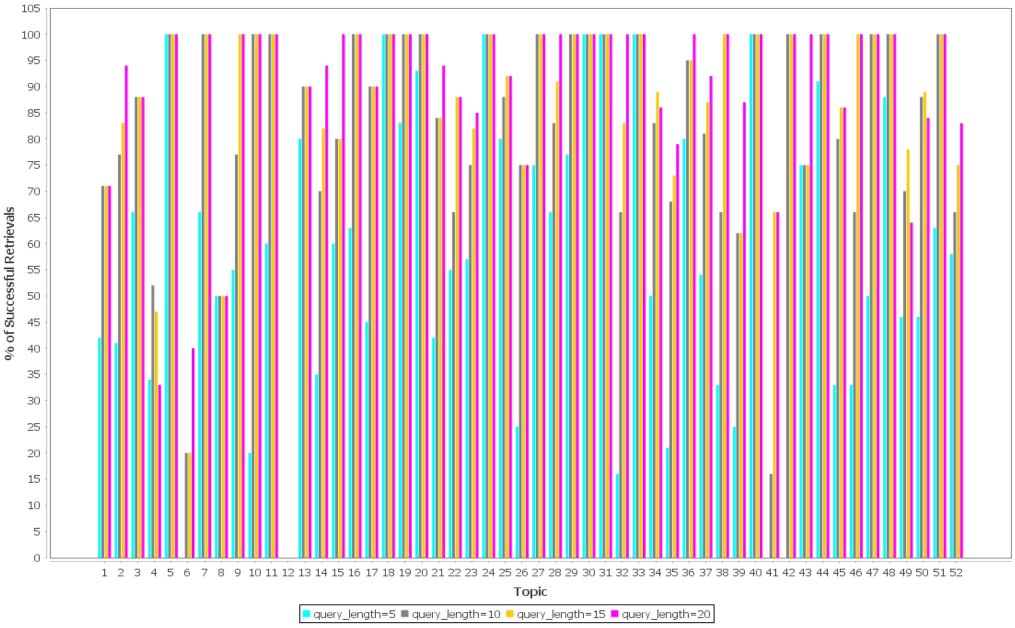


Figure 9: Percentage of successful retrievals for each topic of Reuters-21578 collection with increase in length of query for 5-gram model

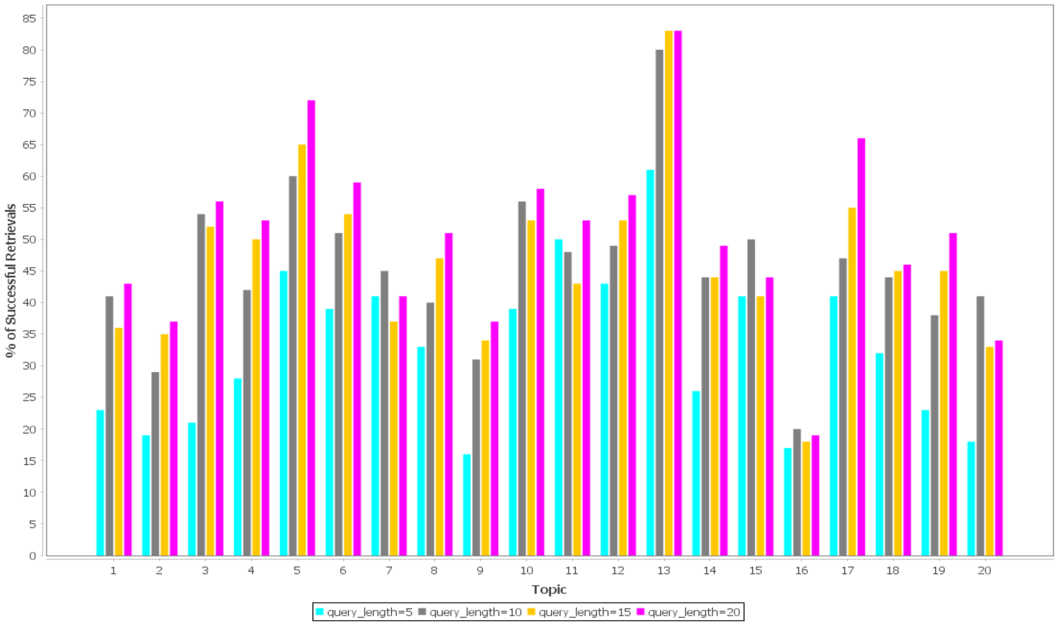


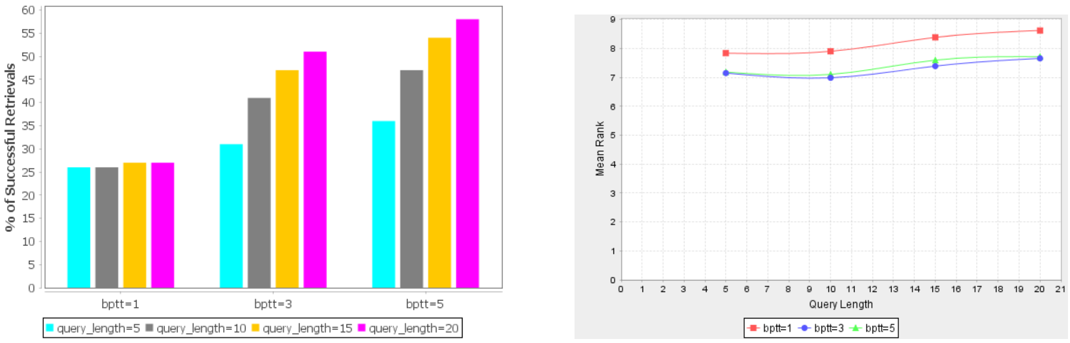
Figure 10: Percentage of successful retrievals for each topic of 20-Newsgroup collection with increase in length of query for 5-gram model

## 5.2. Retrieval Effectiveness for RNN

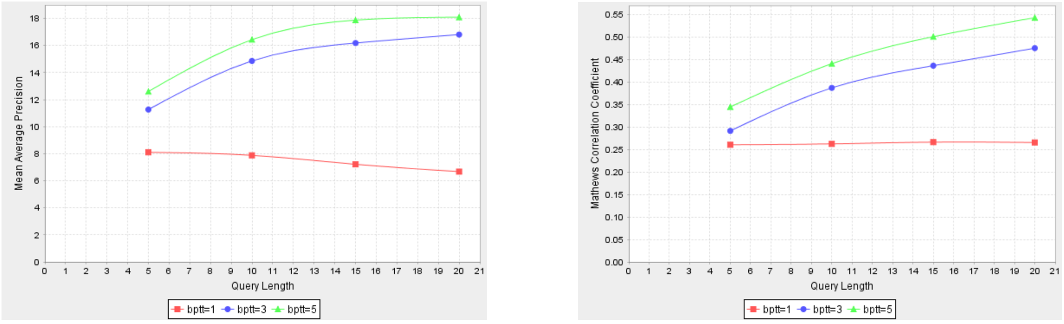
In the previous passage, we described the effect on retrieval output when changing the number of words in query for models trained on feed-forward neural network. On the other hand, when we ran the queries on the models trained on recurrent neural network, we saw slightly different results for Reuters-21578. In contrast to results obtained from the feed-forward neural network model trained on this collection, we saw an improvement in retrieval output when we altered the length of query from 5 words to 20 words for higher order bptt parameter. The improvement in performance reflects the ability of recurrent neural networks to capture the context in the passage of text. The effectiveness of recurrent neural networks can only be seen by allowing the network to see relatively greater number of words in input query as compared to feed-forward network.

The most significant observation made when evaluating retrieval output for recurrent models trained on all the collections was the shift in performance with an increase in the bptt parameter. For the Reuters-21578 collection, the network was able to learn and generalize well across sequences of words in a query for a bptt parameter of 3 and 5. For the Cade collection, storing a context of up to 3 words gave the best performance whereas for 20-Newsgroup collection, the performance actually degraded by storing any context while training. This degradation in performance can be explained by the fact that many times local context does not provide the most useful predictive clues, which instead are provided by long distance dependencies for which long short-term memory neural networks (Hochreiter and Schmidhuber, 1997) are used. Although recurrent neural networks are able to connect past information in order to better inform about the current prediction, this is not the case every time. In cases where the gap between the relevant information and the place that information is needed is small, recurrent neural network can learn to use that past information. But there are also cases where a larger context is needed in order to predict the current word and where the gap between the relevant information in context and the point where that information is needed increases. Long-term memory based neural networks are a special kind of recurrent networks, that successfully cater to the problem of remembering information for longer periods of time and work well where there is an increased gap between the relevant information and the point where it is required.

It is seen that the recurrent models trained on only the Reuters-21578 collection, having a vocabulary size of 8862 words, performed better when we increase the number of words in the query. For the 20-Newsgroup and Cade collection, we observed degradation in performance for longer queries. A larger vocabulary of 20-Newsgroup and Cade collection has more tendency to have long distance dependencies in the text because of which the models were not able to generalize well for longer queries. One of the reasons of using the truncated back propagation through time algorithm is that the algorithm suffers from vanishing gradient problem. Whenever the gradient of the error function of the neural network is propagated back through time, it gets scaled by a certain factor which is either greater than one or smaller than one. As a result, the gradient either blows up or decays exponentially over time. Thus, the gradient either dominates the next weight adaptation step or effectively gets lost.

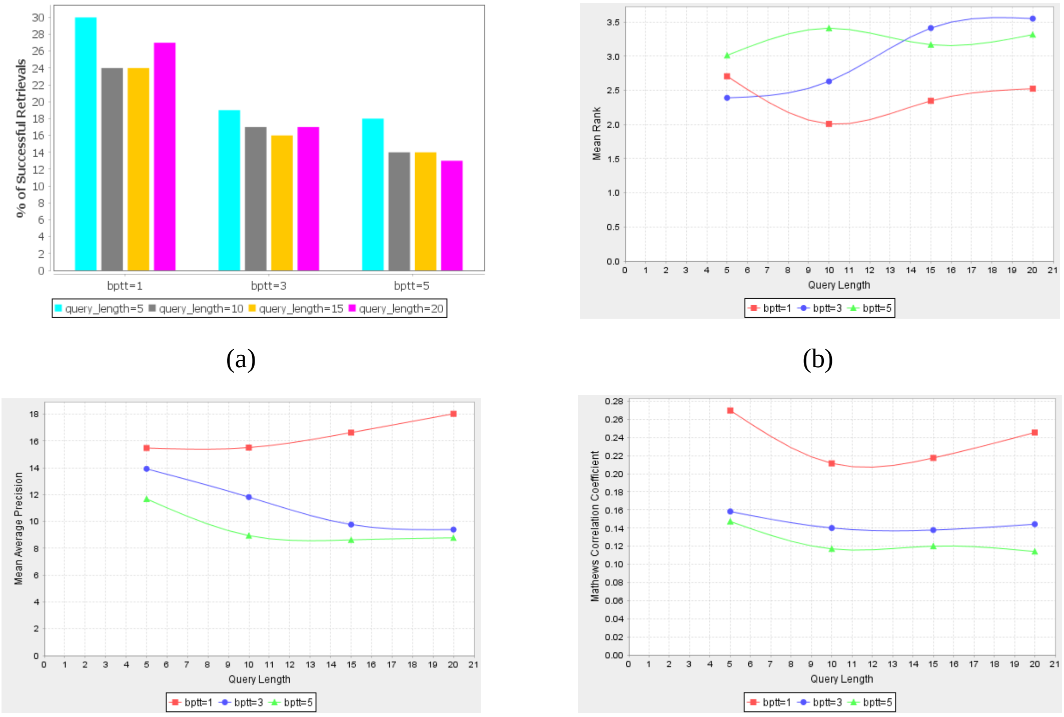


(a) (b)



(c) (d)

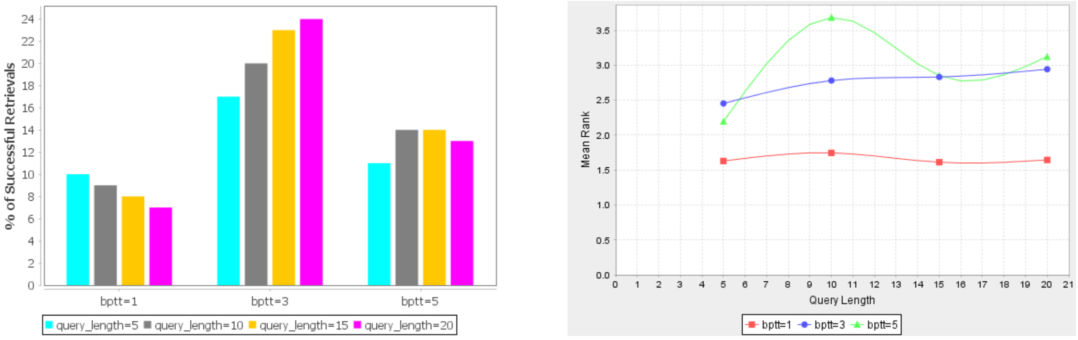
Figure 11: (a) percentage of successful retrievals, (b) mean rank, (c) mean average precision, and (d) matthews correlation of the Reuters-21578 collection trained on recurrent neural network



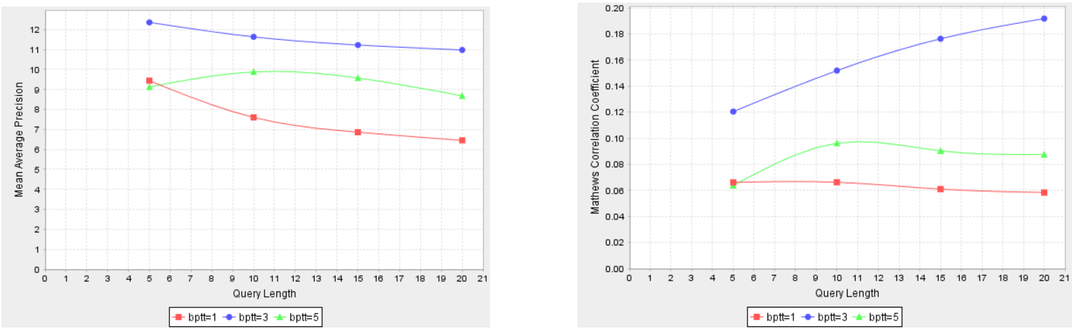
(a) (b)

(c) (d)

Figure 12: (a) percentage of successful retrievals, (b) mean rank, (c) mean average precision, and (d) matthews correlation of the Reuters-21578 collection trained on recurrent neural network



(a) (b)



(c) (d)

Figure 13: (a) percentage of successful retrievals, (b) mean rank, (c) mean average precision, and (d) matthews correlation of the Cade collection trained on recurrent neural network

## 5.3. Best Performance

In theory, the recurrent neural networks consider the long-term dependencies when modeling text in natural language. In practice, however, learning long term dependencies with gradient descent on a document modeling application when the number of words are small is a difficult task. For Reuters-21578 and Cade collection, average number of words trained in batch from a document were 10. For 20-Newsgroup collection, that number was 20. The work on recurrent neural networks described in (Mikolov et al., 2010) does not address this problem. That work focuses more to democratize the use of recurrent neural networks for the application of language modeling by making them relatively fast to train in comparison to old techniques.

Table 3 shows a comparison of the mean average precision scores for models trained on both feed-forward and recurrent neural network for all the collections. The comparison is shown for a query length of 5 since it is very unlikely for a user to query a search engine with 10 or more words. The best performance on a collection is highlighted in bold and it can be seen that a 5-gram feed-forward model gives the best performance on all the collections. The fact that long term dependencies are still difficult to learn in case of document modeling would argue in favor of using n-gram sequences as an input to the neural network.

The main advantages of using a recurrent neural network over feed-forward neural network would be the greater representational power of recurrent neural networks and their ability to perform intelligent smoothing by considering syntactic and semantic features but we have seen in this study that it does not lead to very good results on the application of document modeling where we train very few words from a document. Although n-gram based feed-forward neural network does not solve the problem of n-gram context that expresses the semantic character of text but by comparison this approach works better in determining the topic of the document from input n-gram. The representation by n-grams assumes that high probability is assigned for those input words that co-occur and low probability is assigned for those input words that do not co-occur without caring too much about where they appear in the document. The architectural setup of feedforward neural network is relatively simple and modeling systems for document retrieval task is also fairly easy. For this reason, their application is easily verified and are much more suitable for text document retrieval for predefined document set structures.

Table 3: MAP scores for query length of 5.

| | FFNN | | | RNN | | |
|---|---|---|---|---|---|---|
| | 1-gram | 3-gram | **5-gram** | bptt=1 | bptt-3 | bptt=5 |
| 20-Newsgroup | 15.52 | 21.30 | **33.59** | 15.47 | 13.92 | 11.68 |
| Reuters-21578 | 7.82 | 28.15 | **45.05** | 8.10 | 11.26 | 12.59 |
| Cade | 10.65 | 14.58 | **20.77** | 9.43 | 12.36 | 9.12 |

## 6. CONCLUSION

We described the implementation of feed-forward and recurrent neural networks and have reported the performance of both networks with respect to retrieval output on altering the length of the query. We also showed how a wrong prediction in the output topic layer can lead to significant decrease in performance in case of a skewed dataset. As a first step, we showed how altering the length of the query affects the retrieval output of feed-forward and recurrent models. Moreover, a comprehensive analysis of feed-forward and recurrent neural network architectures was provided. We saw that backpropagation through time algorithm does not improve the performance of retrieval as compared to standard backpropagation on the application of document retrieval mainly due to small number of words trained from the documents in batch.

The results show that the retrieval system returns best MAP scores for a 5-gram feed-forward model. If our target topic is maximized with lesser number of words in a query, the chances of irrelevant documents to be in the top 20 documents gets minimized since there is less opportunity for the model to maximize a non-target topic. More relevant documents seen in the output vector also increase the chances of highest ranked document from the target topic to be higher in the output vector as well. Although the number of successful retrievals is low for shorter queries, but whenever a target topic is maximized for a given 5-gram, most of the documents from that topic appear to be among the top 20 documents. This also explains why MAP scores are better when the model is presented with fewer number of words in query. In other words, the decrease in performance with longer queries can also be explained with the fact that our design assumes of user foreseeing the exact words and phrases belonging to that topic and only to that topic. Consequently, longer word phrases lead to smaller chances of its subset belonging to the same topic.

Moreover, we showed that using higher order bptt parameter to store context information does improve the performance of the retrieval system in some cases but in other cases local context is unable to provide effective clues in prediction, which instead are provided by long distance context for which long short-term memory neural networks are used. We saw that learning context information with backpropagation through time algorithm in case of recurrent neural network does not outperform the standard backpropagation algorithm of feed-forward neural network since it is difficult to capture those dependencies with relatively fewer words from document as input in the batch.

## 7. FUTURE WORK

In future, it will be interesting to investigate how our approach compares with keyword-based models presented in (Skovajsova, 2010) and with recent research on vector based and Boolean models on the task of document retrieval. It will be interesting to train the models on a two sub-system network with non-factorized output layer, the architecture of which is somewhat similar to the cascade neural network presented in (Skovajsova, 2010). The inputs to the networks are n-gram(s) or word(s), and our implementation learns the input to topic and input to document relationship in a single pass with the error backpropagating from both the output layers to the hidden layer. An alternative to this approach would be to just learn input to topic relationship in the first subsystem for all the documents in the collection and then later learn the input to document relationship in the second subsystem.

## REFERENCES

1. Arnold, K., Gosling, J., and Holmes, D. (2000). The Java Programming Language. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.
2. Baeza-Yates and Ribeiro-Neto (1999). Modern Information Retrieval. Addison Wesley/ ACM Press.
3. Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. (2003). A Neural Probabilistic Language Model. Journal of Machine Learning Research, 3:1137-1155.
4. Blackwood, G. W. (2005). Neural Network-based Language Model for Conversational Telephone Speech Recognition. PhD thesis, St. Catherine's College.
5. Bullinaria, J. A. (2015). Neural Computation.
6. Cardoso-Cachopo, A. (2007). Improving Methods for Single-label Text Categorization. PhD thesis, Instituto Superior Tecnico, Universidade Tecnica de Lisboa.
7. Chen, H. (1995). Machine Learning for Information Retrieval: Neural Networks, Symbolic Learning, and Genetic Algorithms. Journal Of The Americal Society For Information Science, 46(3):194-216.
8. Chen, S. F. and Goodman, J. (1999). An Empirical Study of Smoothing Techniques for Language Modeling. Computer Speech and Language, 13:359-364.
9. Cheung, V. and Cannons, K. (2002). An Introduction to Neural Networks.
10. Cordon, O., Moya, F., and Zarco, C. (2002). A New Evolutionary Algorithm Combining
11. Simulated Annealing and Genetic Programming for Relevance Feedback in Fuzzy Information Retrieval Systems. Soft Computing - A Fusion of Foundations, Methodologies and Applications, 6(5):308-319.
12. Deepak, G. and Deepika, S. (2012). Information Retrieval on the Web and its Evaluation. International Journal of Computer Applications, 40(3):70-78.
13. Dunne, R. A. and A., C. N. (1997). On the pairing of the softmax activation and crossentropy penalty functions and the derivation of the softmax activation function. In 8th Australian Conference on Neural Networks, Melbourne, Australia, pages 181-185.
14. Emami, A. and Jelinek, F. (2004). Exact Training of a Neural Syntactic Language Model. In ICASSP, pages 245-248.
15. Gauvain, J., Adda, G., Adda-Decker, M., Allauzen, M., Gendner, V., Lamel, L., and H., S. (2005). Where Are We In Transcribing BN French? In Eurospeech, pages 1665-1668.
16. Goodman (2001). A Bit of Progress in Language Modeling Extended Version. Machine Learning and Applied Statistics Group.
17. Herrera-Viedma (2001). Modelling the Retrieval Process for an Information Retrieval System using an Ordinal Fuzzy Linguistic Approach. Journal of the American Society for Information Science and Technology, 52(6):460-475.
18. Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8):1735-1780.
19. Hotho, A., Nurnberger, A., and Paas, G. (2005). Brief Survey of Text Mining. LDV-forum, 20(1):19-62.
20. Jain, L. C. and L.R., M. (1999). Recurrent Neural Networks: Design and Applications. CRC Press, Inc. Boca Raton, FL, USA.
21. Jurman, G., Riccadonna, S., and Furlanello, C. (2012). A Comparison of MCC and CEN Error Measures in Multi-Class Prediction. PLoS ONE, 7(8).
22. Lai, S., Xu, L., Liu, K., and Zhao, J. (2015). Recurrent Convolutional Neural Networks for Text Classification. In Proceeding AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pages 2267-2273.
23. Lan, M., Tan, C., Low, H., and Sung, S. (2005). A Comprehensive Comparative Study on Term Weighting Schemes for Text Categorization with Support Vector Machines. In Special interest Tracks and Posters of the 14th international Conference on World Wide Web, Chiba, Japan, pages 1032-1033.

24. Le, Q. and Mikolov, T. (2014). Distributed Representations of Sentences and Documents. Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 32:1188-1196.
25. Lin, R., Liu, S., Yang, M., Li, M., Zhou, M., and Li, S. (2015). Hierarchical Recurrent Neural Network for Document Modeling. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, pages 899-907.
26. Manning, C. D., Raghavan, P., and Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
27. Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure, 405(2):442-451.
28. Mikolov, T., Karafiat, M., Burget, L., Cernocky, J. H., and Khudanpur, S. (2010). Recurrent neural network based language model. INTERSPEECH, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan.
29. Mokris, I. and Skovajsova, L. (2005). Neural Network Model Of System For Information Retrieval From Text Documents In Slovak Language. Acta Electrotechnica et Informatica, 5(3).
30. Nickolls, J., Buck, I., Garland, M., and Skadron, K. (2008). Scalable Parallel Programming with CUDA. ACM Queue, 6(2):40-53.
31. Ogilvie, P. and Callan, J. (2003). Language Models and Structured Document Retrieval. Proceedings of the First Workshop of the Initiative for the Evaluation of XML Retrieval (INEX), Delos.
32. Porter (1980). An algorithm for suffix stripping. Program, 14(2):130-137.
33. Powers, D. (2007). Evaluation: From Precision, Recall and FFactor to ROC, Informedness, Markedness & Correlation. Human Communication Science SummerFest.
34. Rumelhart, D., Hinton, G. E., and R.J., W. (1986). Learning representations by backpropagating errors. Nature, 323:533-536.
35. Salton, G., Edward, A. F., and Wu, H. (1983). Extended Boolean Information Retrieval. Communications of the ACM, 26(11):1022-1036.
36. Scheir, P. and Lindstaedt, S. N. (2006). A Network Model Approach to Document Retrieval taking into account Domain Knowledge. Lernen-Wissendeckung-Adaptivitat, pages 154-158.
37. Schwenk, H. and Gauvain, J. (2004). Neural Network Language Models for Conversational Speech Recognition. In ICSLP, pages 1215-1218.
38. Skovajsova, L. (2010). Text Document Retrieval by Feed-forward Neural Networks. Information Sciences and Technologies Bulletin of the ACM Slovakia, 2(2):70-78.
39. Svozil, D., Kvasnicka, V., and Pospichal, J. (1997). Introduction to multilayer feed-forward neural networks. Chemometrics and Intelligent Laboratory Systems, pages 43-62.
40. Treeratpituk, P. and Callan, J. (2006). An Experimental Study on Automatically Labeling Hierarchical Clusters using Statistical Features. Annual ACM Conference on Research and Development in Information Retrieval, pages 707-708.
41. Turney, P. and Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. Journal of Artificial Intelligence Research, 37(1):141-188.
42. van Rijsbergen (1979). Information Retrieval (2nd edition).
43. Wang, S., Schuurmans, D., Peng, F., and Zhao, Y. (2005). Combining Statistical Language Models via the Latent Maximum Entropy Principle. Machine Learning, 60(1-3):229-250.
44. Wei, X. and Croft, B. (2006). LDA-based Document Models for ad-hoc Retrieval. Proceedings of the 29th Annual International ACM SIGIR Conference on Research and
45. Development in Information Retrieval, Seattle, Washington, USA, pages 178-185.
46. Werbos (1990). Backpropagation Through Time: What It Does and How to Do It. Proceedings of the IEEE, 78(10):1550-1560.
47. Zobel, J. and Moffart, A. (2006). Inverted Files for Text Search Engines. ACM Computing Surveys (CSUR), 38(2).

# THE DESIGN AND IMPLEMENTATION OF LANGUAGE LEARNING CHATBOT WITH XAI USING ONTOLOGY AND TRANSFER LEARNING

Nuobei SHI, Qin Zeng and Raymond Lee

Division of Science and Technology, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China

## ABSTRACT

*In this paper, we proposed a transfer learning-based English language learning chatbot, whose output generated by GPT-2 can be explained by corresponding ontology graph rooted by fine-tuning dataset. We design three levels for systematically English learning, including phonetics level for speech recognition and pronunciation correction, semantic level for specific domain conversation, and the simulation of "free-style conversation" in English - the highest level of language chatbot communication as 'free-style conversation agent'. For academic contribution, we implement the ontology graph to explain the performance of free-style conversation, following the concept of XAI (Explainable Artificial Intelligence) to visualize the connections of neural network in bionics, and explain the output sentence from language model. From implementation perspective, our Language Learning agent integrated the mini-program in WeChat as front-end, and fine-tuned GPT-2 model of transfer learning as back-end to interpret the responses by ontology graph.*

*All of our source codes have uploaded to GitHub: https://github.com/p930203110/EnglishLanguageRobot.*

## KEYWORDS

*NLP-based Chatbot, Explainable Artificial Intelligence (XAI), Ontology graph, GPT-2, Transfer Learning*

## 1. INTRODUCTION

*Language chatbot* has widely used in customer services or personal assistants for task-orientated, interactive chats in special domains and knowledge base for question-answer systems. All have comprised of automatic speech recognition (ASR), natural language understanding (NLU), dialogue management (DM), natural language generation (NLG), speech synthesis (SS). Figure 1 shows the system flow of a typical chatbot system.

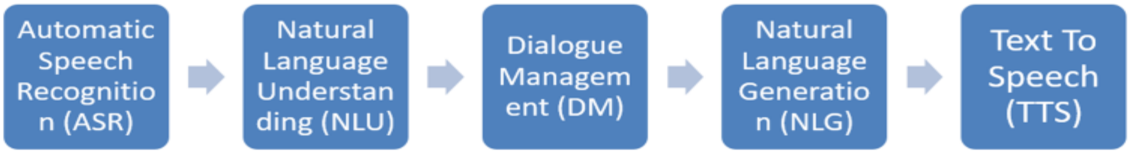

Automatic Speech Recognitio n (ASR)
Natural Language Understan ding (NLU)
Dialogue Managem ent (DM)
Natural Language Generatio n (NLG)
Text To Speech (TTS)

Figure 1. Flowchart of a typical chatbot system

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 305-323, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101124

Researches on rule-based matching chatbot were incited since the first chatbot was invented and tried the Turing Test in 1950s. To build a chatbot in such pattern require tremendous amount of human dialogues as knowledgebase. Moreover, this kind of simple chatbot for daily conversation was incapable to extract information to transform into knowledge and even generate new knowledge like agent with AI technology nowadays. Traditional chatbot with sufficient corpus can correspond to suitable responses for human questions in both grammar and matching rates due to responses are natural conversations produced by human. Additional matching words signified better selected responses. Thus, AI-based NLP technology challenges nowadays are machines' capability to generate responses rather than by patterns recognition which is the focus our language learning chatbot.

Neural network as language model in Natural Language Processing (NLP) supports machine to generate appropriate responses in recent years. Recurrent Neural Networks (RNN) with popular framework like TensorFlow and Keras are the mainstream for Language Model generation. In late 2018, Google published a basic language model called Bidirectional Encoder Representations from Transformers (BERT) with outstanding performance in 11 common NLP tasks which concentrated on Encoder scheme. Few months later, Open AI released another transformer based on unsupervised learning with pre-trained model focusing on Decoder scheme. By using unsupervised learning as pre-training scheme, the bi-directional transfer learning model can be served as a promising Language Model framework in NLP. With the pre-trained language model, our relatively small dataset can achieve better performance than traditional language models. Based on GPT-2 with fine-tuned model [1], our language agent has fluent and syntactic response as a virtual AI English tutor for industrial usage.

No matter of how excellent performance of these models, the essential neural networks are always in needed of big data as data source. Human minds make inferences that go far beyond the data available. The reverse-engineering of human learning and cognitive development helps the engineering of humanlike machine learning system [2]. Neural Network outputs are the mathematical computation results of neurons layers. It always considered as a black box, but the basic concept in bionics is inspired by human thinking and learning processes. We use the way of human learning and reasoning to explain the output of the neural networks, which also witness the development of search engines. That related to another question: How do human get information and knowledge?

Information system is the basis to build a knowledgebase, from websites to search engines, then to the ontology graph to retrieve the simplified output and make it more accurate. Due to the relation of keywords, the ranking done by search engines are more suitable for human justifications. In this paper, we use ontology, also called knowledge graph to simulate the connection of neural networks. The ontology graph is the tree of real-world concepts in different areas acquired by raw data, which focus on the relation between different nodes of the ontology graph. Just like neural networks, the tree also has the characteristic of synapses, which can inspire the relation extraction in ontology graph as memory in human brain. The interaction of agent with human also use natural language rather than query language of database or mathematical distance computation for similarity. Facing the barrier of machine can understand the natural language without computation, we use ontology graph to explain the humanlike neural networks. To some extent, the graph has ability to reason and generate new knowledge when it has sufficient knowledgeable and capable ontology graph that can "absorb" and "generate" new knowledge.

From the implementation perspective, English learning chatbot is constructed with Question-Answer-type of conversation as fundamental interactions between human and machine, in order to construct a humanlike English learning system. In general, such Question-Answer system with knowledge base is better than the system without database, such as Information Retrieval-based Question-Answer (IRQA) by crawler or search engine and chatbot with rule-based distance matching. The knowledge base is divided into two parts, the task-oriented knowledge aims at special-domain knowledge base like expert system. However, the chatbot for daily chats need open-domain knowledge to answer unpredicted questions. For example, the customer service chatbot like Ali Xiaomi [3], which is the typical example of E-commerce online support staff to substitute human online customer service. The more specific domain, the more suitable for chatbot to predict and set personality problems from users. The opposite is open-domain KBQA such as Siri for Apple, Xiao Ice for Microsoft, the interaction form provides a 24 hours personal assistant for users including database and APIs to search engine and other apps within one terminal to answer questions of open-domain knowledge.

In our English language learning chatbot system, we use unstructured data, English text, from daily dialogue to construct knowledge base with dictionary and graphs (ontology graphs) from fine-tuned dataset. With Python’s AI ecosystem development platform, researchers will obtain more ideas between neural network and cognition to find a highly accurate answers from massive unstructured data.

From the implementation perspective, we propose a mini-program in WeChat for real-world usage, with the fine-tuned GPT-2 model [1] and speech recognition service from Google, whose three levels systematically English Learning method provide an efficient way in natural language learning. Simultaneously, the ontology graph visualized on Neo4j, graph database, to explain the generated response from agent.

The main contributions and originality of this paper include:

1. Following the Explainable Artificial Intelligence (XAI) concept to explain the output natural language from our Neural Network model.
2. The introduction of GPT-2 framework with dialogue format [1] as a language model for our system, different from the original GPT-2 used in text generation with reminders, to extend the usage of transfer learning into our language learning chatbot.
3. The successful integration of transfer learning substitute to traditional seq2seq model with ontology graph for the fine-tuning of dataset.
4. The creative idea in users’ convenience to develop an AI NLP-based English agent into mini-program into WeChat as intelligent mobile English learning chatbot tutor.
5. The successful design and implementation of a Webchat-based mini-program for real-world use for English learning.

This paper is presented as follows: Section II is the literature review to review the contemporary chatbot system from technology companies such as Microsoft, Alibaba and Hugging face in respect of function design and technology component to analyse existing idea and optimize our idea for agent. Also, the research direction and related work for XAI and our practical method of using ontology graph with NLP. Section III states the framework and methodology in theoretical and practical of agent, which include THREE levels for English learning system for users, connectionism in language model with GPT-2 and ontology. Section IV presents the implementation of agent at mini-program and the testing of constructed system in real-world to analyse NLP raw data and interpreter, the ontology graph. Last section is to evaluate the performance combined with ontology for conclusion.

## 2. LITERATURE REVIEW

In recent decades, industrial market launched numerous chatbots, dialogue systems and online customer services. All of them are related products of Question-Answer (QA) systems. These systems have different technological backgrounds to control responses from systems. Furthermore, some researchers published structured dataset to fine-tune and evaluate the performance of systems to define the language model. Due to the particularity of our English language learning agent, we encountered situations that users require unstructured raw data as source text with the help of AI ecosystem to develop a multistep function-oriented learning method to language agent. This is a process from data pre-processing to the end of user interface to implement our idea for agent. Although AI technology provide lots of framework to generate conversation responses, we always adhere to research direction and related work of XAI and Ontology Graph (OG) with NLP.

### 2.1. An Overview of Chatbot

### 2.1.1. A Knowledge-Grounded Neural Conversation Model (seq2seq RNN) [4]

Since the origin of seq2seq model generated by RNN, neural network based chatbots had engaged in both industrial and academic communities. In 2018, Microsoft extended their industrial conversation system to make responses from the system to avoid brief and illogical contents as compared with human responses.

In respect of design, Microsoft already possessed functions to give simple responses in open domain. Figure 2 shows the extension exist only in the branch of Encoder to add facts into response. Both versatility and scalability in open-domain and external information knowledge of textual and structured are combined in this system, which has the recommendation system function for restaurant but not task-oriented.

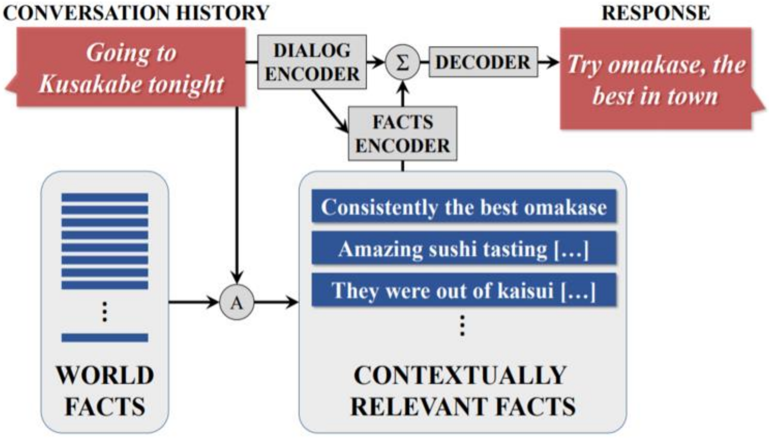

CONVERSATION HISTORY
Going to Kusakabe tonight
DIALOG ENCODER
Σ
DECODER
RESPONSE
Try omakase, the best in town
FACTS ENCODER
A
WORLD FACTS
Consistently the best omakase
Amazing sushi tasting […]
They were out of kaisui […]
CONTEXTUALLY RELEVANT FACTS

Figure 2. Architecture of Knowledge-Grounded Model

From the implementation aspects, different with the slot-filling to grounded content using rule-based scheme, the system uses another seq2seq neural network model fed by dataset from Twitter and Foursquare tips, the same technique with original language model. After that, the conversation provides with more meaningful and logical contents in responses, which only infuse knowledge information into the trained data-driven neural model.

The progress of knowledge grounded supported our NLP part to extract the Triple with useful information to consist knowledge. The meaning of XAI in our research is to explain the output response. This review shows the dataset can be absorbed by neural networks and generate output by sort or upgrade these data to information or knowledge, in which thinking resemble human to give response. In other words, abstract information can be visualized on Ontology and explained the response in relations.

### 2.1.2. AliMe Chat: A Sequence to Sequence and Re-rank based Chatbot Engine (rule-based IR+seq2seq) [5]

In order to improve response quality to obtain the most matching response sentence within the restriction of neural network, AliMe, a commercial chatbot specialize in E-commerce industry, integrate both traditional information retrieval based and Seq2Seq neural models.

From the design perspective, AliMe is applied to E-commerce as a substitute of human service in Taobao. So, for the function of AliMe, it should search the most similar question to obtain response from the QA database to reply customers. If the client posed a new question to system, it should be intelligent enough to answer the question. At optimizer part, whether every response is generated or searched, both will be selected again by Seq2Seq neural network to obtain the most suitable response to human users. Also, AliMe is not limited in a task-oriented service for Taobao. According to survey by Alibaba, AliMe received 5% of questions within E-commerce span. It intended to upgrade to an open domain chatbot for questions expansion.

Before the transformation of AliMe to an open domain chatbot, most of responses depended on 9,164,834 QA pairs in database. After that, Seq2Seq model of GRU with SoftMax and Beam Search algorithms provided better response from neural network in both open and special domain. Also, the Seq2Seq model used second time to re-rank candidate responses.

### 2.1.3. The Design and Implementation of Xiao Ice, an Empathetic Social Chatbot （IR+seq2seq+KG） [3]

Compared with chatbots mentioned before, Xiao Ice as shown in Figure 3 has the most complex chatbot structure built. The key-points of Xiao Ice is comprised of Intelligent Quotient (IQ) and Emotional Quotient (EQ) into the system design.

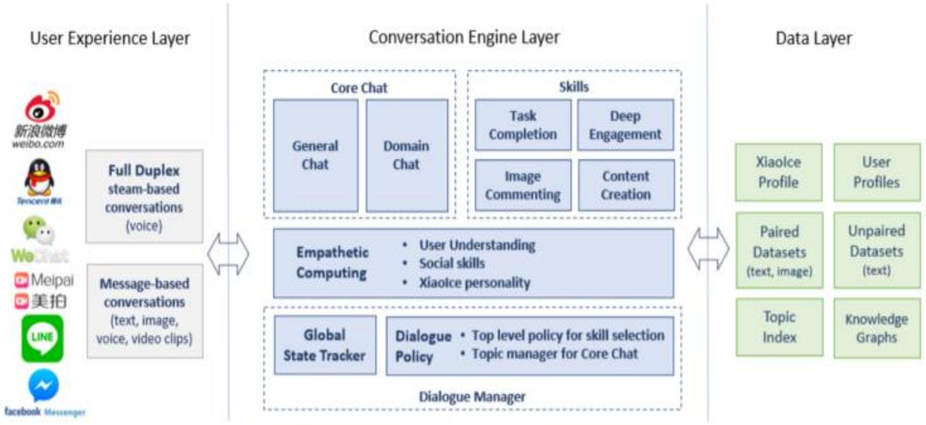
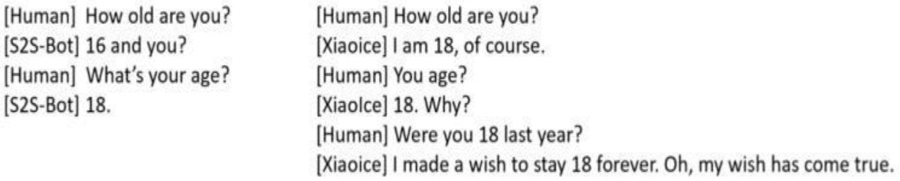


Figure 3. Xiao Ice System Architecture [3]

In the architecture above, Xiao Ice can adapt to input with different format, which not limited in voices and texts but processed images and videos. EQ and IQ show on topics adjustment with users' preferences in special and open domains. Considering at an English learning scenario, our agent defines the highest two layers for topic discussion and open domain of 'free-style conversation' to improve conversational skills aims to one field.



[Human] How old are you?
[S2S-Bot] 16 and you?
[Human] What's your age?
[S2S-Bot] 18.

[Human] How old are you?
[Xiaoice] I am 18, of course.
[Human] You age?
[XiaoIce] 18. Why?
[Human] Were you 18 last year?
[Xiaoice] I made a wish to stay 18 forever. Oh, my wish has come true.

Figure 4. Examples of inconsistent responses generated using a seq2seq model [3]

The truth is, chatbot with emotion provided more natural and humanlike responses.

From the implementation perspective, the core of Xiao Ice is based on the language model using RNN to create responses. As mentioned, Xiao Ice integrates three methods to create responses. First, its chatbot has divided into retrieval-based and generated-based models. Xiao Ice has both candidate generator and candidate ranker. For the generator, it uses rule-based matching with real-world conversation collected and stored by natural language. Second, it uses deep learning model trained by paired dataset to simulate human dialogue to build human-like system. Third, it uses query on knowledge graph to get related entities. The candidate ranker also corresponding to the generator but including the semantic computation in NLP and empathy matching of Xiao Ice personality.

For our language learning agent, in order to simulate the IELTS test, we choose fine-tuned GPT-2 model with daily dialogue dataset to shape the tutor with different personalities and background. Our ontology graph acts as an interpreter for language model rather than graph database as compared with Xiao Ice.

### 2.1.4. Transfer Transfo: A Transfer Learning Approach for Neural Network Based Conversational Agents (transfer learning with GPT-2) [1]

The above chatbots discussed are basically adopting the Seq2Seq model, it gets good performance for the generation of responses. Since neural network is a data-driven model, it means that the performance are heavily relies on the amount and quality of the *big data*. Thus, based on attention mechanism, transfer learning using self-attention can used to find the relations and sequence dependency among words in the conversation.

The outstanding feature of transfer learning architecture in Google BERT and Open AI GPT-2 includes: 1) the proficient at encoder; and the proficient at decoder for response generation. GPT-2 uses self-attention method, which is an attention mechanism relating different positions of a single sequence to compute a representation of the sequence [6]. It extracts the relations of the sequence to rewrite the sequence. The process is similar to the extraction of information from text and make comprehension for further processing.

GPT-2 is more specialized in language generation according to self-attention scheme because it already absorbed 40G pure text to learn semantics and syntax of natural language. Fine-tuning is to use dataset with suitable format for special task to personalize the original GPT-2 model to task-oriented language model. Transfer Transfo we used as chatbot in our agent is a language system combining Transfer learning-based training scheme and a high-capacity Transformer model. [6] By using the persona-chat dataset to fine-tune the model, its utterance changes from long-text to dialogue format. Persona-chat in real-world helps to shape communicator's background to further define the topic and better understand user's input sentence. It shows the priority in the different AI language tutor with different personality resemble Speaking module in IELTS.

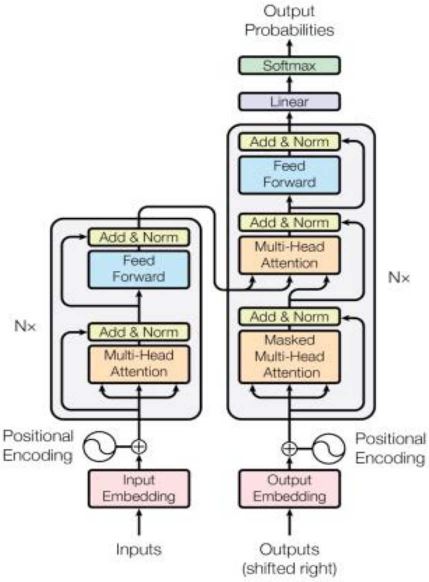


Figure 5. Transfer Learning-Model Architecture [6]

This paper shows the resulting fine-tuned model of significant improvements over the current state-of-the-art end-to-end conversational models like memory augmented seq2seq and information-retrieval models [6]. This is why we chose transfer learning rather than seq2seq model to apply XAI. The transfer learning achieved better and advanced performance, whereas self-attention is corresponded to our academic research in connectionism with ontology to explain the neural network are associated with bionics.

## 2.2. Explainable Artificial Intelligence (XAI)

Deep learning has widely applied in industrial projects. However, the explainable segment is the main barrier for artificial intelligence future development of. If the AI output model is unexplainable or uninterpreted, the output validity is unreliable. It is undeniable that AI is influential in prediction, classification and in decision making specifically the computation process of AI models is parameters, we only input big data for an output, but are unable to explain the recommended output response to human users. Just like the input format feed to neural model is word embedding of vector rather than natural language.

Neural network is a *black box* resembles to the construct of human brain. So far, we do not know its computation process. The process of human cognition is from concept to practice. It has axons and synapses to activate the storage of another neuron which is the human memory. The definition of classes, attributes (properties) are concepts and relationships as memory in human brain. The way to store these relationships is similar to human brain to transform raw data to knowledge, machine also requires special data format which is different with natural language to compute machinal thinking to reasoning new relations. So, we use triple <entity, relation, entity> as a basic format to store computer memory and visualize the connections by ontology. XAI is an interpreter to solve the question that the machine can understand the input data and get the output data human readable that means machine has ability to thinking like human. In NLP, it is Natural language Understanding (NLU). The NLU improvement is the progress of encoder part corresponding to end to end model. Natural Language Generation is the verification to get expected responds from human.

IBM delivered a speech on the Explainable AI (XAI) on April 2020 to visualize the implementation of the *true* Human-Computer Interaction (HCI). It meant the directional approach did not restrict to reveal the black-box algorithm but used user-centred approach to connect user needs and technical advancement.

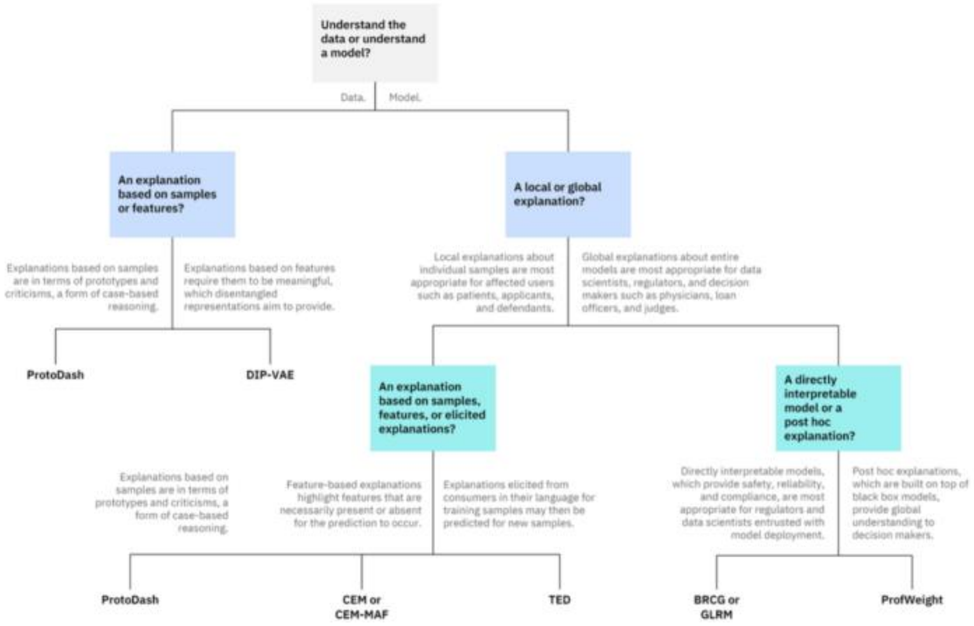

Understand the data or understand a model?
Data.
Model.
An explanation based on samples or features?
A local or global explanation?
Explanations based on samples are in terms of prototypes and criticisms, a form of case-based reasoning.
Explanations based on features require them to be meaningful, which disentangled representations aim to provide.
Local explanations about individual samples are most appropriate for affected users such as patients, applicants, and defendants.
Global explanations about entire models are most appropriate for data scientists, regulators, and decision makers such as physicians, loan officers, and judges.
ProtoDash
DIP-VAE
An explanation based on samples, features, or elicited explanations?
A directly interpretable model or a post hoc explanation?
Explanations based on samples are in terms of prototypes and criticisms, a form of case-based reasoning.
Feature-based explanations highlight features that are necessarily present or absent for the prediction to occur.
Explanations elicited from consumers in their language for training samples may then be predicted for new samples.
Directly interpretable models, which provide safety, reliability, and compliance, are most appropriate for regulators and data scientists entrusted with model deployment.
Post hoc explanations, which are built on top of black box models, provide global understanding to decision makers.
ProtoDash
CEM or CEM-MAF
TED
BRCG or GLRM
ProfWeight

Figure 6. The taxonomy of AI Explainable methods introduced in AIX360 [7]

When we use AI technology on NLP, whether such algorithm can be trust and explainable is an important factor [7], which is also the core meaning of XAI. Thus, the algorithm operation should be able to understand and analysed by human. The following 5 covers the related XAI application. [8]

a) Transparency
   Focus on readable format by human
b) Causality
   Data-driven model provide both accurate inferences with decision background
c) Bias
   Black-box cannot calculate model bias against complicated computation
d) Fairness
   AI system always operate in a fair manner to users.
e) Safety
   AI system output are understandable and trustworthiness in regulatory sectors.

XAI transparency and compliance should be taken into account in association with a given related prediction [8]. The first approach is required more attention to deep learning and neural network considering it powerful prediction ability representing the deep explanation focusing on neural network layers and structures such as computation and parameters. The second approach is interpretable models such as casual models like linear regression, Bayes, logistic regression, decision tree are models in statistics which can be explained during calculation and reasoning. On the contrary, random forest, which consist of lots of decision trees has higher accuracy but are uninterpretable. Third approach is model induction which can infer any explainable model from black-box model. It put forward high-level requirements to explain every model for users against actual features. In this paper, we use the last method to interpret transfer learning-based dialogue system of end-to-end decoder part to obtain the response.

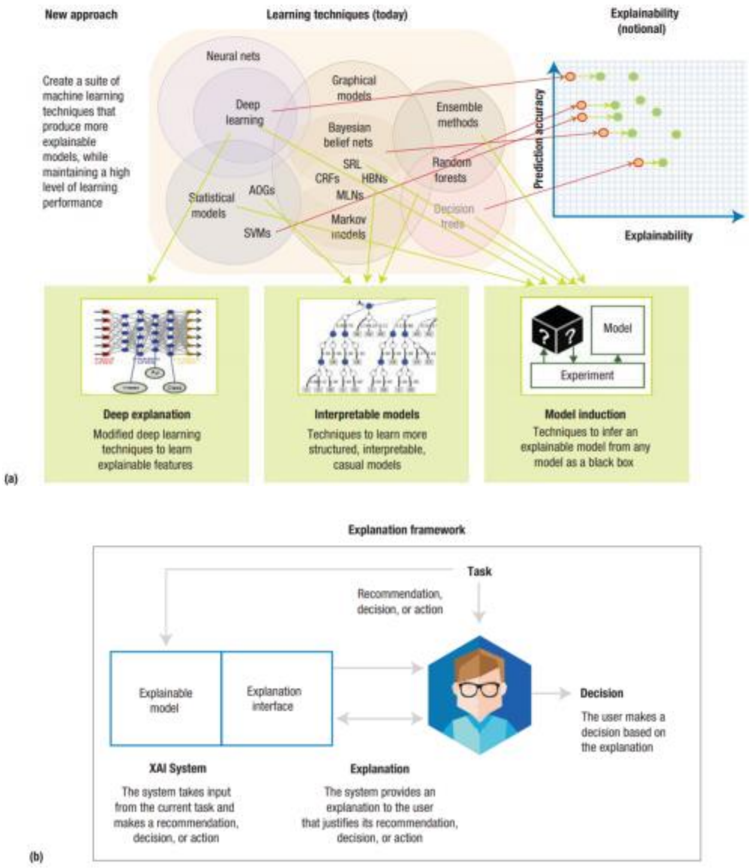

New approach
Learning techniques (today)
Explainability (notional)
Create a suite of machine learning techniques that produce more explainable models, while maintaining a high level of learning performance
Neural nets
Deep learning
Graphical models
Bayesian belief nets
Ensemble methods
SRL
CRFs
HBNs
MLNs
Random forests
Statistical models
AOGs
SVMs
Markov models
Decision trees
Prediction accuracy
Explainability
Deep explanation
Modified deep learning techniques to learn explainable features
Interpretable models
Techniques to learn more structured, interpretable, casual models
Model induction
Techniques to infer an explainable model from any model as a black box
Model
Experiment
(a)
Explanation framework
Task
Recommendation, decision, or action
Explainable model
Explanation interface
Decision
The user makes a decision based on the explanation
XAI System
The system takes input from the current task and makes a recommendation, decision, or action
Explanation
The system provides an explanation to the user that justifies its recommendation, decision, or action
(b)

Figure 7. An overview of explainable models of AI (XAI).[8]

In this article [8], Hani Hagras explained the Fuzzy Logic Systems (FLSs) and human understandable AI, FLSs tried to mimic human thinking and research on the approximate way to thinking rather than limit human brain resemble to neural network. It has upgraded to the philosophy to build numerous of if-then reasoning rules to describe given human behaviour in human-readable way. The rules are the highest reasoning format for inference called OWL in ontology. Knowledge reasoning is a developing area in ontology and natural language generation. From existed knowledge to generate new knowledge based on logical rules to make up the contents of dialogue system is to be solved.

## 2.3. Ontology

Ontology could be seen as a visualization of knowledge base or the update of search engine, whose previous version is semantic Web. It represents the fundamental form of knowledge representation with graph. We are familiar with database, which used as the container to store different type of data. The knowledge is the senior of data. It should be extract from raw data such as domain texts or constructed database to save as the computer or human readable format, such as the triple of RDF, OWL. The hierarchy is shown in Figure 8.

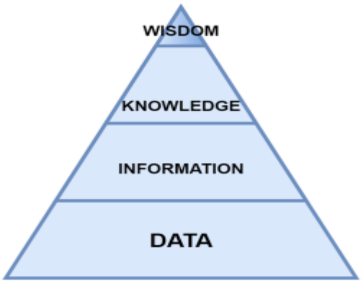


Figure 8. Pyramid of Knowledge

Ontology more like a tree to consist of the concepts, categories and relations systematically for the development of knowledge-based system, which responsible for question concerning about what entity exist and how such entity may be grouped and related within hierarchy and subdivided according to difference and similarities. [9] It is readable for both human and computer. Because the natural words are not transformed to the computer bits but use natural language processing tools to separate the objects sentence to triple format, which refreshed rapidly and interact between computers conveniently. Once we extract the triple as basic constructed data, then the graph database like neo4j, Orient DB, Amazon Neptune, will show the RDF/OWL triple to visualize the relationships of entities. The relations give ambiguous help to the inference of OWL to generate new knowledge.

Ontology has three levels, after the development of decade, another name from Google called knowledge graph, which also divide it into two level. For traditional ontology, the top-level corresponding to open-domain ontology graph (OG) and domain ontology match special-domain knowledge graph for specific industrial applications, for example finance, medicine. Domain ontology can extend to top-level ontology with grounded knowledge for concepts, because it built from bottom to top as open-domain to including as much as possible knowledge concepts for interoperable, information retrieval, automatic reasoning and other specific natural language processing tasks about common sense. In opposite, domain knowledge graph build from top to bottom, which always digs the deeper relationship to enlarge the domain ontology of different entities.

With the development of cognition and knowledge representation, knowledge engineering tools also update from handcraft to graph database. For academic, researchers hope the system really equipped with inference ability with OWL format, the tool proposed by Stanford University that protégé has updated to 5.5 but also half-manually to type the RDF/OWL triple. Protégé concentrate on special domain ontology edition owing to the handcraft, which is so hard to build open-domain knowledge. However, protégé supports various of ontology languages such as RDF, RDFs, OWL, XML, UML, etc. Figure 9 shows most ontology language and corresponding levels in semantic Web.

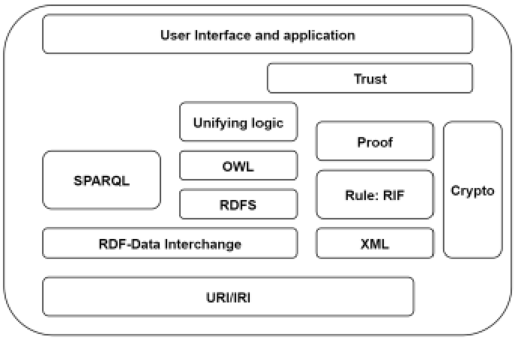
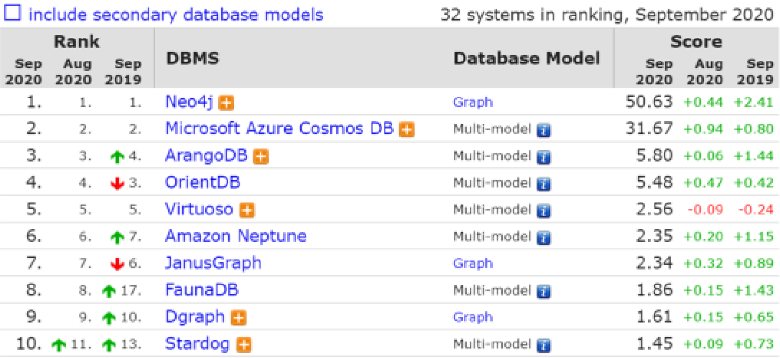


Figure 9. Semantic Web Framework

Now, knowledge engineering developers already built magnitude open-domain knowledge base for their purpose. Ontology designed to optimize the rank and contents of search engine also for commercial usage. It will re-rank the search output more intelligent through the keywords. Compared with traditional search engine, the relation can improve the search efficient and display related information.

The ontology storage system also develops with larger magnitude for commercial usage. For the vision and operation in construction, neo4j substitute other graph database or storage system to be the most welcomed graph database to visualize and analysis by Cypher, a custom-made SPARQL for neo4j. Figure 10 is the rank of storage system in September 2020.



☐ include secondary database models 32 systems in ranking, September 2020

| Rank |  |  | DBMS | Database Model | Score |  |  |
|---|---|---|---|---|---|---|---|
| Sep 2020 | Aug 2020 | Sep 2019 |  |  | Sep 2020 | Aug 2020 | Sep 2019 |
| 1. | 1. | 1. | Neo4j | Graph | 50.63 | +0.44 | +2.41 |
| 2. | 2. | 2. | Microsoft Azure Cosmos DB | Multi-model | 31.67 | +0.94 | +0.80 |
| 3. | 3. | ↑ 4. | ArangoDB | Multi-model | 5.80 | +0.06 | +1.44 |
| 4. | 4. | ↓ 3. | OrientDB | Multi-model | 5.48 | +0.47 | +0.42 |
| 5. | 5. | 5. | Virtuoso | Multi-model | 2.56 | -0.09 | -0.24 |
| 6. | 6. | ↑ 7. | Amazon Neptune | Multi-model | 2.35 | +0.20 | +1.15 |
| 7. | 7. | ↓ 6. | JanusGraph | Graph | 2.34 | +0.32 | +0.89 |
| 8. | 8. | ↑ 17. | FaunaDB | Multi-model | 1.86 | +0.15 | +1.43 |
| 9. | 9. | ↑ 10. | Dgraph | Graph | 1.61 | +0.15 | +0.65 |
| 10. | ↑ 11. | ↑ 13. | Stardog | Multi-model | 1.45 | +0.09 | +0.73 |

Figure 10. DB-Engines Ranking of Graph DBMS, top 10, date Sep 2020 [10]

The rank of DB-Engines graph system shows the popularity degree of different type storage system. Considering the running speed to generate graph relations and storage format extracted from raw data, neo4j always the top-1 causing it industrial magnitude and relative high refresh speed, more important is the API with python, java and other frequently used language.

## 3. Methodology

To start with, we need to define some questions for our English Language Learning agent first. We prefer to define it as an QA-based dialogue system rather than for task-oriented execution but with grounded knowledge and special scenario. Combing with XAI, our system should concern three aspects and make direction for theoretical research. [11]

The first aspect is artificial intelligence (AI). Once AI mentioned, from the aspect of understanding of human things, it has two layers, the content and the methodology. In general, we divide AI into two mainstreams, Computer Vision (CV) and natural language processing (NLP). The other hand, methodologies to research AI is Symbolism, connectionism, and behaviourism.

Symbolism is an intelligence that use mathematical logical thinking to simulate the human thinking. For example, the expert system. Connectionism belongs to bionics for human brain, in general, neural network is a typical model. Behaviourism focus on the prediction of human behaviours, such as AlphaGo of Reinforcement learning and genetic algorithm, in which researchers think human get adaptivity from the interaction with external environments. In our research, rule-based system is symbolism but generate-based with word2vec is connectionism by neural network.

The second aspect is natural language, which is truly the real-world human communication language. NLP is relative to computer. It hopes computer can understand human language.

The third aspect is understanding. The branch of NLP in understanding is Natural Language Understanding. For human, there are a thousand Hamlet in a thousand people's eyes, which means the understanding related to similar life experience, common topics, context, knowledge base of individual and semantic and pragmatic of sentence more than dictionary. So, our direction to integrate as more as possible about what human thinking factors to applied in human intelligence for machine.

For XAI, readable and interpretable for human is explainable AI techniques. If the machine concerns these factors like human, no matter the computation method, the output and parameters can be interpreted by human trustily. In practical, our system research following the connectionism. By the reviews about self-attention mechanism, the ontology with connection are similar to the relations of attention mechanism.

In order to develop a complete system, we first define the direction as an AI English tutor with three levels. That part is pre-set at the User Interface (UI) based on mini-program in WeChat developer. It means WeChat user can log in and use it directly. As a platform, mini-program need whole NLP architecture to process the learning tasks. Our architecture has five parts. The first and last belongs Voice Recognition, which part we use service from Google, Audio to Words. NLU and NLG use the model of fine-tuned use daily dialogue from thesis Transfer Transfo: A Transfer Learning Approach for Neural Network Based Conversational Agents. [1] based on Open AI GPT-2.

As shown in Figure 11, the central part, Dialogue Management (DM), we use ontology graph for visualization on neo4j platform, where Ontology will explain the relations of entities in fine-tune data and generated response from language model. We will display the techniques we used in our system below, that Spacy for NLP raw data processing, neo4j as graph database to store ontology. Open AI GPT-2 as the original model to specific used into dialogue generation.

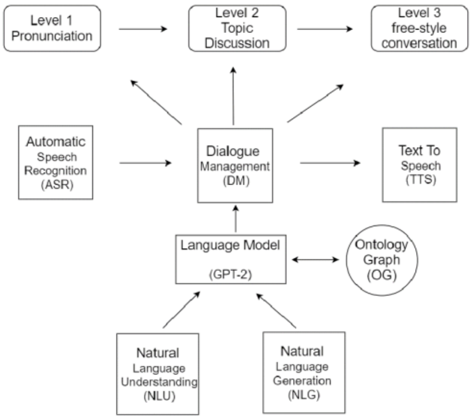


Figure 11. Architecture of this paper

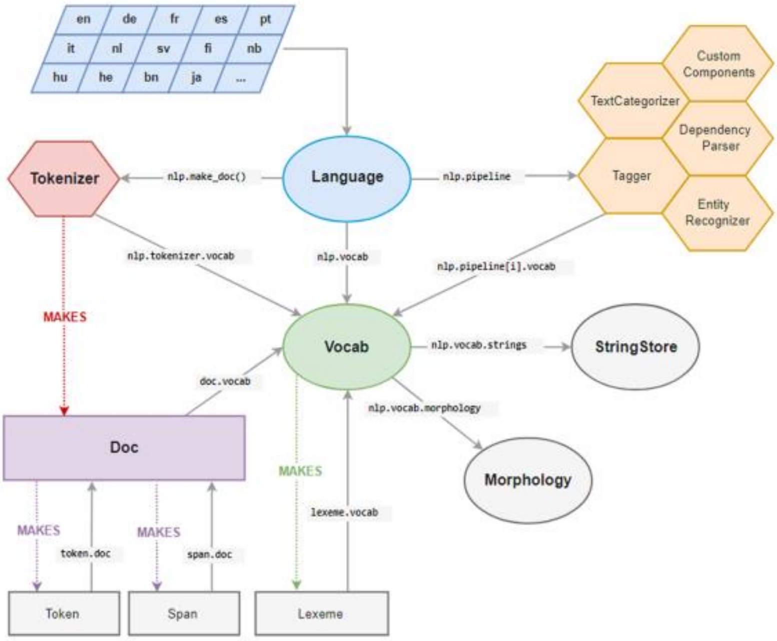
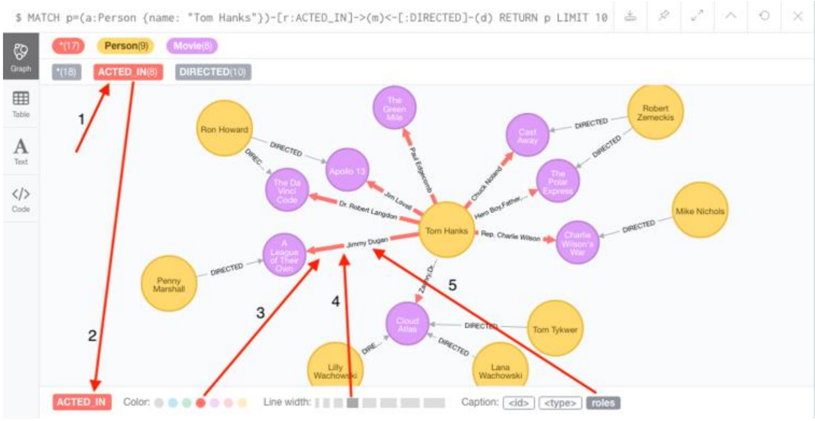


Figure 12. Architecture of Spacy [13]



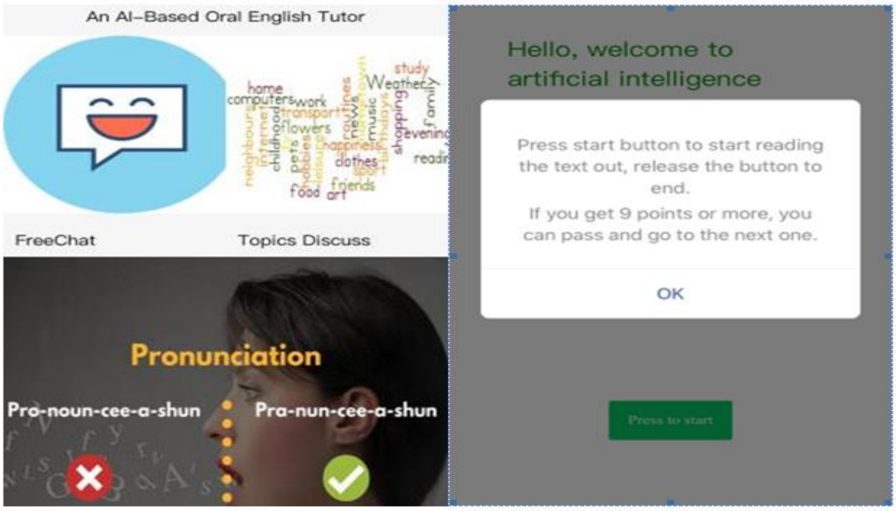
```
$ MATCH p=(a:Person {name: "Tom Hanks"})-[r:ACTED_IN]->(m)<-[:DIRECTED]-(d) RETURN p LIMIT 10
```

Figure 13. Demo of Ontology in Neo4j [14]

# 4. SYSTEM IMPLEMENTATION

## 4.1. Chatbot at Mini-program in WeChat

Before we explain the output sentence, following XAI, we integrate the language model, the GPT-2 model-based Transfer Transfo [1] with our THREE-level functions for language learning. Level 1 is pronunciation correction, where match the difference between standard pronunciation and human interaction. Level 2 is Topic discussion by pre-define the topics into UI.



Press start button to start reading the text out, release the button to end.
If you get 9 points or more, you can pass and go to the next one.

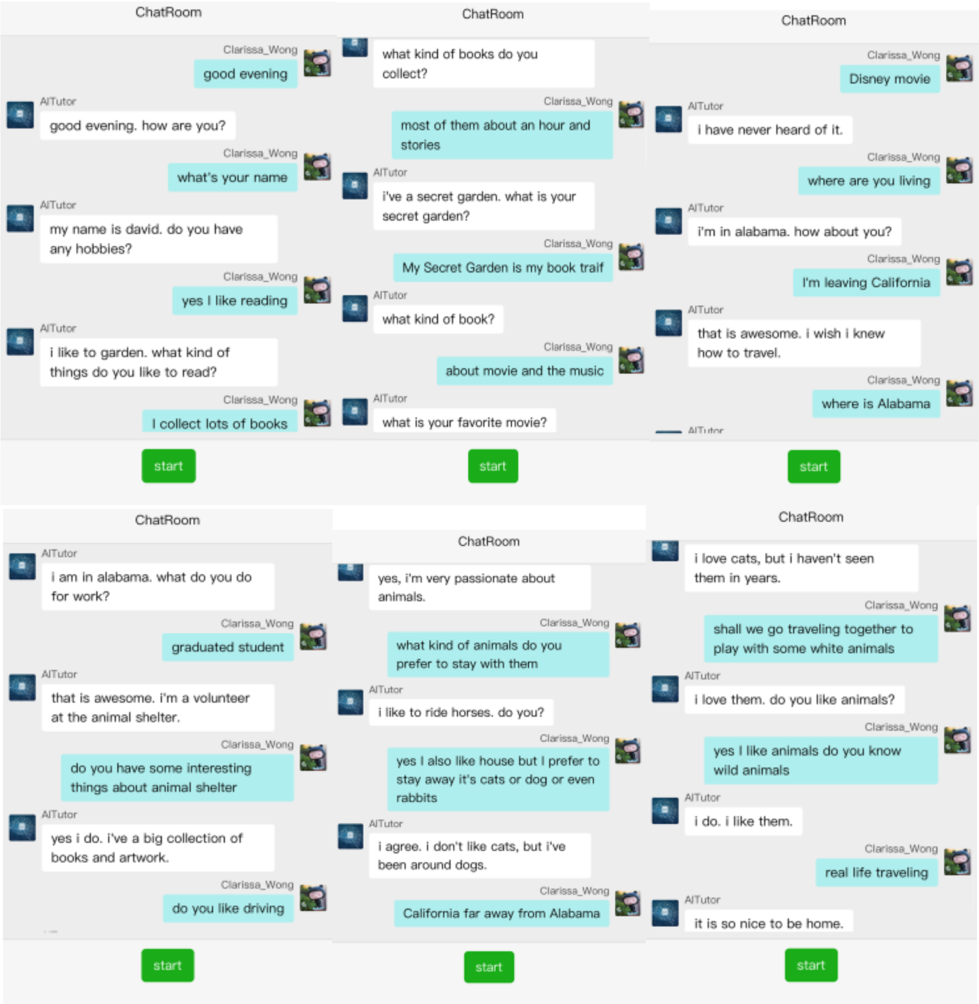
Figure 14. UI of AI English tutor develop by ourselves.

Level 3 is free-style conversation, which also the key-point part to be explained by Ontology. The Free-style conversation module has no restriction, which use the fine-tuned GPT-2 model to generate response. The system, from the perspective of industrial usage like this:

Step1: Record the user’s voice and store it in a local temporary file.
Step2: Upload temporary file to server to analysis and convert it to text.
Step3: Use the text as input string to our language model to get the generated response back to front-end (mini-program).
Step4: Show the response text in the Chat Room.

## 4.2. Ontology Graph

Different with protégé which only could be create the entities and relations by hand. Neo4j take in so many kinds of data format. So, we cooperate neo4j with python to create the nodes and links automatically. The dataset of Open AI GPT-2 is json format, so we first transform the json data to txt data as text for entities and relations extraction. Then we use Spacy web model to extract the SPO triple as our knowledge. Every entity contains a property or an extended range to expand knowledge to get more information based on web. So, we choose DBpedia as our entity links. Because it is the constructed data from Wikipedia, which already an OWL database. It will help our ontology graph link to the internet semantic web to enlarge our database and extend the entities. After we have extracted SPO triple, we should build the node and relations for the neo4j according to it.



Clarissa_Wong: good evening

AITutor: good evening. how are you?

Clarissa_Wong: what's your name

AITutor: my name is david. do you have any hobbies?

Clarissa_Wong: yes I like reading

AITutor: i like to garden. what kind of things do you like to read?

Clarissa_Wong: I collect lots of books

what kind of books do you collect?

Clarissa_Wong: most of them about an hour and stories

AITutor: i've a secret garden. what is your secret garden?

Clarissa_Wong: My Secret Garden is my book traif

AITutor: what kind of book?

Clarissa_Wong: about movie and the music

AITutor: what is your favorite movie?

Clarissa_Wong: Disney movie

AITutor: i have never heard of it.

Clarissa_Wong: where are you living

AITutor: i'm in alabama. how about you?

Clarissa_Wong: I'm leaving California

AITutor: that is awesome. i wish i knew how to travel.

Clarissa_Wong: where is Alabama

AITutor: i am in alabama. what do you do for work?

Clarissa_Wong: graduated student

AITutor: that is awesome. i'm a volunteer at the animal shelter.

Clarissa_Wong: do you have some interesting things about animal shelter

AITutor: yes i do. i've a big collection of books and artwork.

Clarissa_Wong: do you like driving

yes, i'm very passionate about animals.

Clarissa_Wong: what kind of animals do you prefer to stay with them

AITutor: i like to ride horses. do you?

Clarissa_Wong: yes I also like house but I prefer to stay away it's cats or dog or even rabbits

AITutor: i agree. i don't like cats, but i've been around dogs.

Clarissa_Wong: California far away from Alabama

i love cats, but i haven't seen them in years.

Clarissa_Wong: shall we go traveling together to play with some white animals

AITutor: i love them. do you like animals?

Clarissa_Wong: yes I like animals do you know wild animals

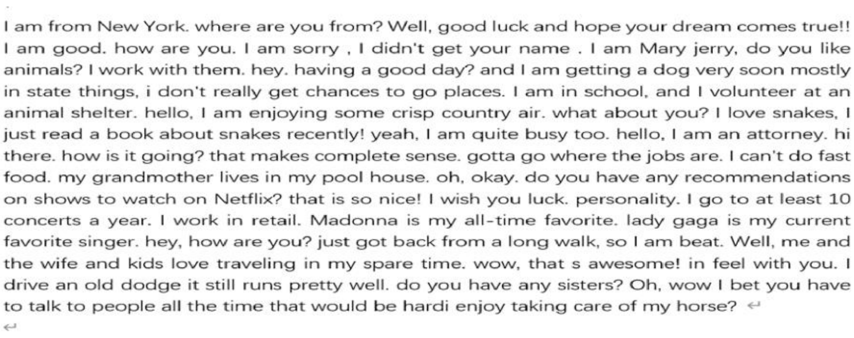
AITutor: i do. i like them.

Clarissa_Wong: real life traveling

AITutor: it is so nice to be home.

Figure 15. Test of Level 3 free-style conversation.

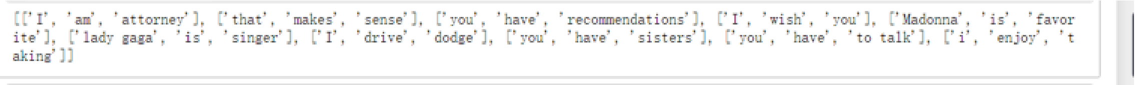
The conversation is the Level 3 free-style conversation, machine with back-end GPT-2 can generate any possible response as feedback. With this system, first three sentences are general speaking about hobbies in music and reading. When we talk about work, the contents from general to special domain about animals. We find that the special noun 'animal shelter' in part of training data has relatively high IDF (inverse-document frequency), so the TF-IDF score is higher. We find the most similar paragraph with dialogue content of the response we show in Figure 16.



I am from New York. where are you from? Well, good luck and hope your dream comes true!! I am good. how are you. I am sorry , I didn't get your name . I am Mary jerry, do you like animals? I work with them. hey. having a good day? and I am getting a dog very soon mostly in state things, i don't really get chances to go places. I am in school, and I volunteer at an animal shelter. hello, I am enjoying some crisp country air. what about you? I love snakes, I just read a book about snakes recently! yeah, I am quite busy too. hello, I am an attorney. hi there. how is it going? that makes complete sense. gotta go where the jobs are. I can't do fast food. my grandmother lives in my pool house. oh, okay. do you have any recommendations on shows to watch on Netflix? that is so nice! I wish you luck. personality. I go to at least 10 concerts a year. I work in retail. Madonna is my all-time favorite. lady gaga is my current favorite singer. hey, how are you? just got back from a long walk, so I am beat. Well, me and the wife and kids love traveling in my spare time. wow, that s awesome! in feel with you. I drive an old dodge it still runs pretty well. do you have any sisters? Oh, wow I bet you have to talk to people all the time that would be hardi enjoy taking care of my horse? ↵
↵

Figure 16. Sample text to Ontology.

We choose the matching sample text from the fine-tune dataset of GPT-2 model and transform the format of .json to the natural language, which will be the input to be extracted triple with Subject-Prrdicate-Object form.

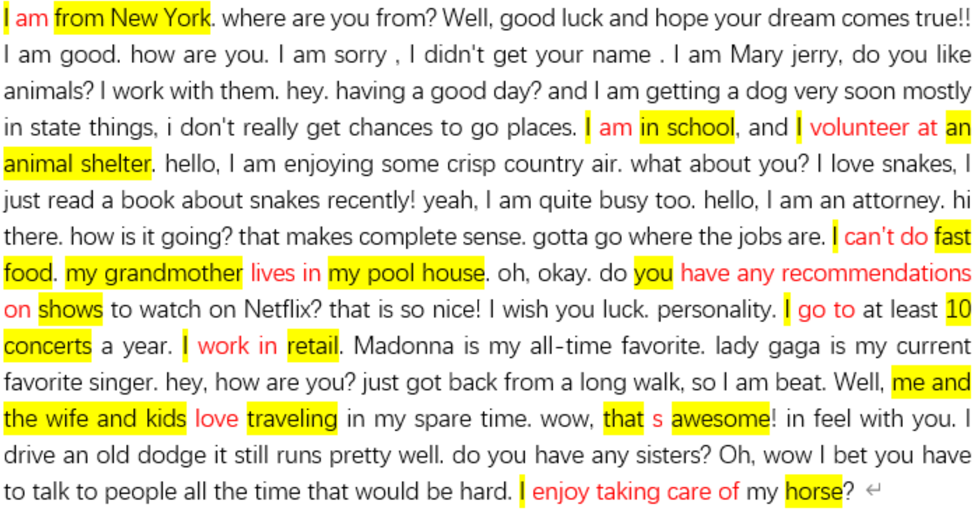


```
[['I', 'am', 'attorney'], ['that', 'makes', 'sense'], ['you', 'have', 'recommendations'], ['I', 'wish', 'you'], ['Madonna', 'is', 'favor
ite'], ['lady gaga', 'is', 'singer'], ['I', 'drive', 'dodge'], ['you', 'have', 'sisters'], ['you', 'have', 'to talk'], ['i', 'enjoy', 't
aking']]
```

Figure 17. Extracted Triple.

In order extract the text, textacy is another package which is built on Spacy for SPO triple. After the Spacy deal with its pipeline tasks of the model to pre-processing the data for triple extraction, such as tokenizer, Tagger, lemmatization, coreference resolution, textacy will extract the SPO tiple and print all of them as list. The result display at figure17, which is the output of text figure 16 [12]. Compared the original text and SPO triple, even though the text following the S-P-O syntax, there are also amount of entities cannot be distinguished by Name Entity Recognition in Spacy. With my experience of protégé of Stanford University to build ontology graph by hand, we decide to add the absent triple which sentence included in the demo paragraph with the Cypher commands in neo4j. The syntax of English has five basic sentence patterns: S-P-O, S-P-O-O, S-P-O-Object complement, S-intransitive verb, subject-linking verb-predicative (SVP). The handcraft triple will obey these five basic rules. Go through all triples corresponding to original sentence, we add the DBpedia link to the entities, which existing in the DBpedia with 'url' link to extend knowledge base.

After the automatic recognition of SPO triple, with 17 relationship types link to different two entities, the most subjects are 'I', due to the training dataset is daily dialogue with special personality in US. Even though the 40G pre-trained dataset of Open AI GPT-2, it only equips the transfer learning model with fundamental grammar to generate suitable response with correct syntax. However, the contents and the length of sentence are controlled by fine-tuning part of Transfer Transfo with daily dialogue dataset.



I am from New York. where are you from? Well, good luck and hope your dream comes true!! I am good. how are you. I am sorry , I didn't get your name . I am Mary jerry, do you like animals? I work with them. hey. having a good day? and I am getting a dog very soon mostly in state things, i don't really get chances to go places. I am in school, and I volunteer at an animal shelter. hello, I am enjoying some crisp country air. what about you? I love snakes, I just read a book about snakes recently! yeah, I am quite busy too. hello, I am an attorney. hi there. how is it going? that makes complete sense. gotta go where the jobs are. I can't do fast food. my grandmother lives in my pool house. oh, okay. do you have any recommendations on shows to watch on Netflix? that is so nice! I wish you luck. personality. I go to at least 10 concerts a year. I work in retail. Madonna is my all-time favorite. lady gaga is my current favorite singer. hey, how are you? just got back from a long walk, so I am beat. Well, me and the wife and kids love traveling in my spare time. wow, that s awesome! in feel with you. I drive an old dodge it still runs pretty well. do you have any sisters? Oh, wow I bet you have to talk to people all the time that would be hard. I enjoy taking care of my horse? ↵

Figure 18. the SPO sentence unextracted by Spacy.

The coloured text is the part of text, which match with the SPO but not recognized by Spacy. With my experience of protégé of Stanford University to build ontology graph by hand, we decide to add the absent triple which sentence included in the demo paragraph with the Cypher commands in neo4j. After we extract the triple from text manually, the neo4j will display more 12 relationships with almost same subject entities. Because the dialogue always uses personal pronoun.

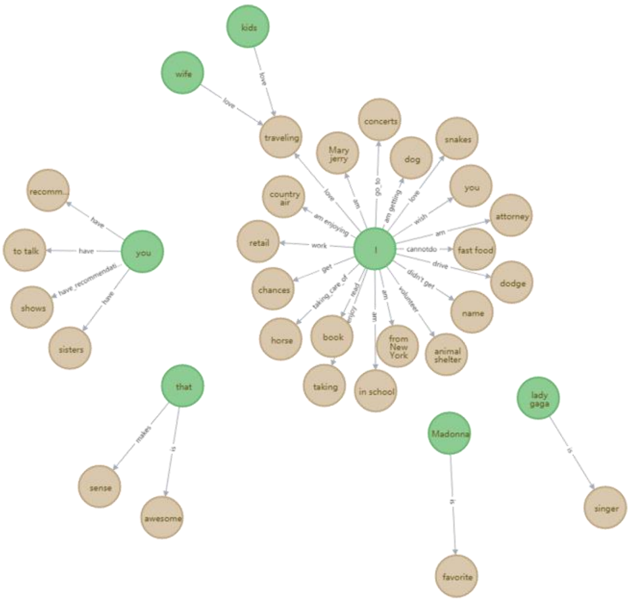


Figure 19. ontology graph with whole triples

In training dataset sample, it shows the paragraph we extracted 30 relationships also contains the content of dialogue. Such as the animals of horse, volunteer at animal shelter.

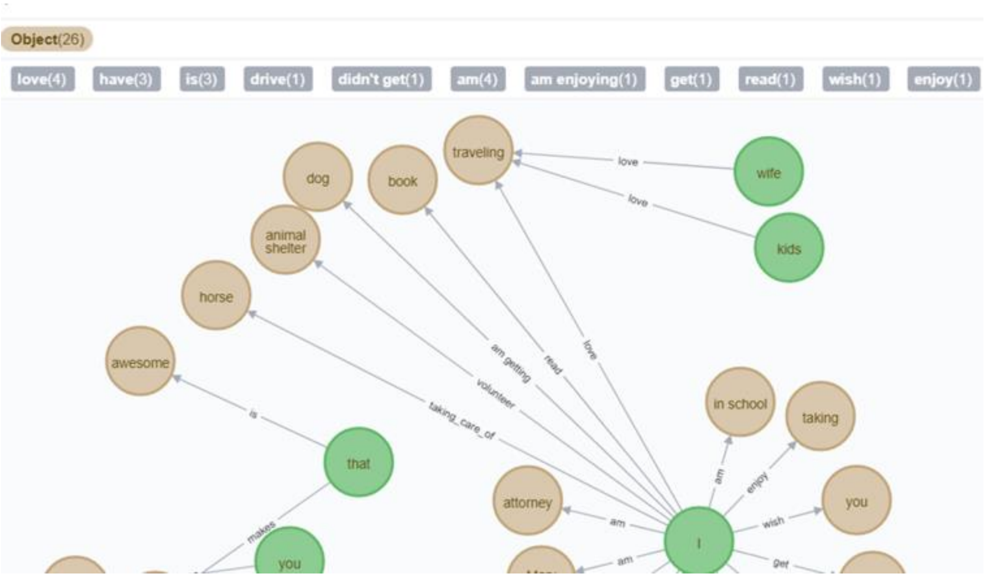


Figure 20. ontology graph visualizes the generated response

In order to prove the output sentence from system is explainable, we use some extract related triple in this demo paragraph. Due to our dialogue based on whole training data, the 200M json format data for trained. In general, only the rule-based pattern recognition can match the high-similarity sentence in database as response. However, the transfer learning and Transfer Transfo model are trained on end-to-end based neural network, which consider as a black-box.

Table 1. comparison of system generated and SPO triple

| system generated | training data triple at ontology graph |
|---|---|
| I'm a volunteer of the animal shelter | I volunteer at animal shelter |
| I like to ride horses | I taking care of horse. |
| I 've been around dogs | I am getting dog |
| I have a big collection of books and artwork.<br>( the reminder from user has collect) | I read book |
| that is awesome | that is awesome |
| … | … |

The specialization is that the Subject entities (the brown bubble) is too little compared with Objects (green bubble). From the aspect of our research direction connectionism, neural network is the bionic research, our ontology belongs to philosophy but the construct resembles the connection in humans' brain. As we see, the green bubbles represent different entities, but due to the same object, it will activate by the objects. At the beginning of this paper, we mentioned that we use ontology graph simulate the neural network. The partly visualization of training data shows the partly explain the output sentence and simulate the relationship in human's brain.

The size of ontology graph and language model restricted by the hardware. If the ontology extended with the whole training data, to some extent, with the help of TF-IDF for special words, we can explain more about the black-box.

## 5. CONCLUSIONS

In this paper, we design and implement an English Learning chatbot with the theory of explainable Artificial Intelligence using Ontology Graph (OG) and Transfer learning. We apply connectionism both in neural network and ontology into the simulation of human brain. In practical, our system techniques refer to Natural Language Processing (NLP) especially in Natural Language Understanding (NLU) and Natural Language Generation (NLG), Ontology, Transfer Learning, Search Engine and WeChat developer of mini-program. Due to the project is an industrial project, at experiment, we test several times with the three level to make sure that the system can be used to English Spoken training in a silent environment. From research aspect, our idea that use Ontology Graph to explain the output natural language sentences make a little progress. It means the neural network model except the feature that Open AI GPT-2 generate text according to the reminder written in algorithm, besides, the content of output sentence can be explain and visualize at the ontology graph with the same training dataset. The larger ontology graph contains more training data, the better and detailed explainable of output sentence.

## ACKNOWLEDGEMENTS

The authors would like to thank for UIC DST for the provision of computer equipment and facilities. This project is supported by UIC research grant R202008.

## REFERENCES

[1] Thomas Wolf, Victor Sanh, Julien Chaumond & Clement Delangue, "Transfer Transfo: A Transfer Learning Approach for Neural Network Based Conversational Agents," in USA, Association for the Advancement of Artificial Intelligence, arXiv:1901.08149v2, 2019

[2] Joshua B. Tenenbaum, et al , (2011) How to Grow a Mind: Statistics, Structure, and Abstraction," Science 11 Mar 2011 : 1279-1285.

[3] Li Zhou, Jianfeng Gao, Di Li, Heung-Yeung Shum, (2019) "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot." arXiv:1812.08989v2 [cs.HC]

[4] Marjan Ghazvininejad et al., (2018) "A Knowledge-Grounded Neural Conversation Model," arXiv:1702.01932v2 [cs.CL]

[5] MinghuiQiu et al., "AliMe Chat: A Sequence to Sequence and Rerank based Chatbot Engine," (2017) Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 498–503

[6] AshishVaswani et al, "Attention Is All You Need," (2017) 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, arXiv:1706.03762v5 [cs.CL]

[7] Q. Vera Liao et al, (2020) Introduction to Explainable AI, CHI 2020, April 25–30, Honolulu, HI, USA

[8] Hani Hagras, (2018) "Toward Human Understandable, Explainable AI," Computer, 51 (9). 28 - 36.

[9] Raymond S. T. Lee, (2020) Artificial Intelligence in Daily Life, Springer.

[10] DB-Engines, accessed in 2020 https://db-engines.com/en/ranking/graph+dbms

[11] Jiaxuan Li (2017) TensorFlow 技术解析与实战, Posts and Telecom Press.

[12] Raman Kishore, (2019), "https://medium.com/analytics-vidhya/a-knowledge-graph-implementation-tutorial-for-beginners-3c53e8802377", access in 2020

[13] Spacy, https://spacy.io/usage/spacy-101#architecture, access in 2020

[14] Neo4j, https://neo4j.com/developer/neo4j-browser/, access in 2020

[15] Yuli Vasiliev (2020) Natural Language Processing with Python and Spacy, No Starch Press, San Francisco.

## AUTHORS

**Clarissa N. B. Shi** is graduated from Beijing Normal University - Hong Kong Baptist University United International College (UIC) awarded M.Sc. (Data Science) with distinction from Hong Kong Baptist University in 2020. Her research interests covering Ontology graph and Chatbot in Natural Language Processing, Neural Network with Mathematical and Explainable Artificial Intelligence (XAI).

**Qin Zeng** is graduate from Beijing Normal University-Hong Kong Baptist University United International College (UIC) awarded his M.Sc. (Data Science) from Hong Kong Baptist University in 2020. Qin Zeng had worked in the software industry for decade before obtaining the master degree and has a very rich experience in Linux environment, cross-platform, multi-language development and deployment. His research interests covering natural language processing, recurrent neural networks and intelligent mobile applications in WeChat platform.

**Dr. Raymond S. T. Lee** (M'98) attained his B.Sc. (Physics) from Hong Kong University in 1989, M.Sc. (IT) and PhD (Computer Science) from Hong Kong Polytechnic University in 1997 and 2000 respectively. Dr Lee had worked at the Department of Computing of Hong Kong Polytechnic University as Associate Professor 1998 - 2005. During the past 20 years, Dr. Lee has published over 100 publications and the author of 8 textbooks and research monographs covering the fields of artificial intelligence, quantum finance, e-commerce, pattern recognition, intelligent agents and chaotic neural networks. Dr. Lee is now the Associate Professor in Beijing Normal University-Hong Kong Baptist University United International College (UIC) working in the field of quantum finance, quantum anharmonic oscillators, chaotic neural oscillators, fuzzy-neuro financial systems, chaotic neural networks and severe weather modelling and prediction.

# COMPARATIVE STUDY ON EYE GAZE ESTIMATION IN VISIBLE AND IR SPECTRUM

Susmitha Mohan and Manoj Phirke

Imaging and Robotics Lab, HCL Technologies, Bangalore, India

## *ABSTRACT*

*Eye gaze estimation aims to find the point of gaze which is basically," where we look". Estimating the gaze point plays an important role in many applications with varying usage. Gaze estimation is used in automotive industry to ensure safety. In the field of retail shopping and online marketing gaze estimation is used to analyse the consumer's interest and focus. Gaze estimation is also used for psychological tests and in healthcare for diagnosing some of the neurological disorders. This also has a significant role to play in the field to entertainment. There are multiple ways by which eye gaze estimation can be done. This paper is about a comparative study done on two of the popular methods for gaze estimation using eye features. An infra-red camera is used to capture data for this study. Method 1 tracks corneal reflection centre w.r.t the pupil centre and method 2 tracks the pupil centre w.r.t the eye centre to estimate gaze. There are advantages and disadvantages with both the methods which has been looked into. Choosing the right method for gaze estimation hence depends on the type of application, precision required and many other factors including environmental conditions. This paper can act as a reference for researchers working in the same field*

## *KEYWORDS*

*eye gaze, pupil, iris, cornea, corneal reflection, polynomial curve fitting, infra-red.*

## 1. INTRODUCTION

Eye gaze estimation tells a lot about the attention and the focus of the person. This information is useful in various ways in various fields and there are different ways by which gaze estimation is performed in accordance with the performance requirement. It is important for the web designers to understand where people gaze most on the screen and where not. This information helps them to place the right buttons at the right place. In healthcare the information about eye gaze variation helps to diagnose some of the neurological disorders and some of the developmental disorders like autism. In automobile industry gaze estimation of the driver helps to find out if he is attentive enough on the road or not. Gaze estimation is also used in the field of augmented or virtual reality. Vision aided applications also make use of eye gaze estimation to help the user.

There are different ways by which gaze estimation can be done. Some of the methods uses wearable gadgets whereas some doesn't. Analysing the movement of eye features is the underlying principle in all of the gaze estimation systems. Analysing the movement of pupil w.r.t the eye and the movement of corneal reflection w.r.t the pupil has gained a lot of attraction because of its simplicity and easiness. This movement of the eye features has been interpreted mathematically in various ways. Some of the methods like polynomial fitting and Gaussian fitting are widely used.

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD, SPM - 2020
pp. 325-338, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101125

With the increasing popularity and usability of gaze estimation, the accuracy and efficiency of gaze estimation is also important. The amount of accuracy and efficiency required can be different from one application to another application. Also, the minimum gaze angle to be measured varies with the kind of application it is used for. For example, in the case of a graphical user interface (GUI) developer the gaze angle measurement has to be more accurate when compared to the gaze estimation of a driver. Since, for a driver it may not be so important to understand precisely at which point on the road or outside environment he is gazing at. Whereas for a GUI developer it is important to understand precisely at which point on the screen the person is looking at. The type of environment and the lighting also plays a significant role in the selection of sensor and method for gaze estimation. The lighting variation inside a car can be large when compared to an indoor environment and hence the technology suiting the GUI developer for gaze estimation might not fit for driver gaze estimation. In case of person wearing spectacles or goggles gaze estimation is challenging and sometimes impossible when the eyes are fully occluded. For accurate estimation of eye gaze and for gaze estimation under challenging scenarios more sophisticated devices and techniques are used.

There is intense research happening over decades to improve the eye gaze estimation performance. Recent trends and developments happening in gaze estimation is present in [3], [4], [5]. In [1] visible light is used to obtain the corneal reflection instead of IR lighting. An efficient method of using geometric transforms in homography normalization (HN) method when corneal reflections are lesser than 4 in number is present in [6]. The possibility of accurately detecting and tracking human gaze based on the head pose information extracted by an RGB-D device is present [7]. A mapping function based on artificial neural networks is used in [8]. How well Gaussian process adapt to the non-linearity in eye movement over polynomial regression is presented in [2]. In [9] Fourier descriptors are used to describe the appearance-based features of eyes compactly. Shape and intensity based deformable eye pupil-centre detection and movement decision algorithms is present in [10]. An effective way of estimating eye gaze from the elliptical features of one iris is described in [11]. An investigation on the effect of gaze position on pupil size estimation by three common eye-tracking systems is presented in [12]. A low-cost system which uses web camera and open source Computer Vision Library Open CV is proposed in [13]. In [14] and [15] review of different eye tracking systems for diagnostic interpretation is presented. A review of eye gaze estimation techniques and applications for consumer products, which has progressed in diverse ways over the past two decades is presented in [16]. In [17] a new hardware friendly, convolutional neural network (CNN) model with minimal computational requirements is introduced and assessed for efficient appearance-based gaze estimation.

In this paper two of the popular gaze estimation methods are compared. Method 1 will be referred to as ’corneal reflection-based gaze estimation’ and method 2 will be referred to as ’pupil-based gaze estimation’ in the rest of the paper. Pupil based gaze estimation doesn’t need an infra-red (IR) camera for gaze estimation whereas corneal reflection-based method needs an IR camera to function. But for comparison purpose data captured from an IR camera is used for both pupil based and corneal reflection-based analysis as both the eye features are visible in IR data. This paper is hence a comparison of what can be achieved with an IR camera and without an IR camera. Both the methods have advantages and disadvantages which will be discussed in detail in further sections

## 2. EYE FEATURES AND CORNEAL REFLECTION

For pupil-based gaze estimation method features of the eye like pupil centre and eye corner points are used for gaze estimation. For corneal reflection-based method features of the eye like iris, pupil centre and the bright spot formed by the IR illuminators on the eye are used for gaze estimation. Figure 1 shows the eye features and corneal reflection.

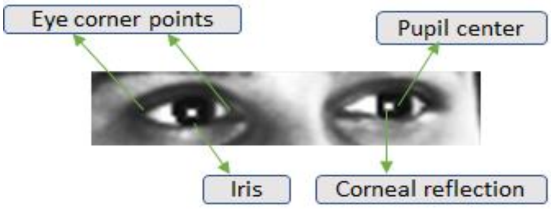


Figure 1. Eye features and corneal reflection.

## 3. CAMERA SETUP AND DATA CAPTURE

An IR camera is placed in between the person and a fixed pattern. The fixed pattern has markings in both vertical and horizontal directions to gaze at. To compare the pupil movement and corneal reflection movement images of eyes are captured using the IR camera looking at the frontal face of the person. The camera set up consists of an IR lens, set of IR illuminators that emits IR radiation within a specific bandwidth and an IR filter that allows only the radiation falling in this specific bandwidth to pass through. This set up minimizes image quality variations caused by surrounding illumination changes. Figure 2 shows the fixed pattern and the camera setup used for capturing data.

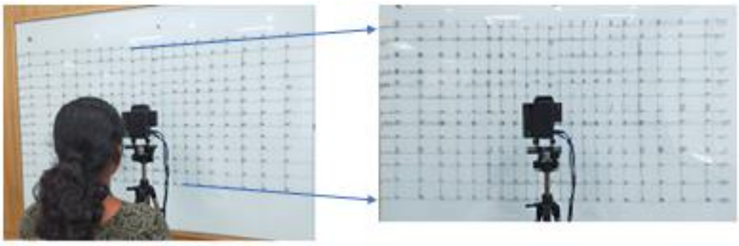


Figure 2. Fixed pattern to gaze at Images of eyes at different gaze angle are captured using the IR camera.

When the gaze changes from one point to the next point on the fixed pattern only the eyes are moved without any head movement. Sets of images are captured with eye gaze varying in horizontal direction from left to right with an offset of 5 degree and the vertical gaze angle kept constant. On successful completion of one set of horizontal data capture the vertical gaze is changed to 5 degree up or down and next set of horizontal eye gaze data is captured. The maximum variation in eye gaze in horizontal direction is from -45 degree to +45 degree and in the vertical direction is from -30 degree to +30 degree using the pattern shown in figure 2.

## 4. GAZE ESTIMATION METHODS

Two of the commonly used gaze estimation methods are chosen and compared in this paper. In corneal reflection-based method the movement of corneal reflection w.r.t the pupil centre is analysed and in pupil-based method the movement of pupil centre w.r.t the eye centre is analysed. Corneal reflection formation requires an IR camera along with IR illuminators, whereas for pupil movement analysis any normal camera would suffice. But for accurate and efficient comparison data set used for both the methods are captured using an IR camera as the images includes all the features required for both the methods. This study can also be considered as a comparative study of what can be achieved with an IR camera and what can be achieved with a normal camera. From the face images captured using the IR camera, eye region is detected and used for both the methods. Any state-of-the art method can be used to detect eye region, eye features and corneal reflection point. Eye features include corner points of eye, boundary of iris, pupil and pupil centre.

### 4.1. Corneal reflection-based gaze estimation

Near-infrared light directed towards the eyes forms bright reflections on the cornea called corneal reflection. As there are two sets of illuminators placed on either side of the lens, two bright spots are formed on the cornea. Centre of these two bright spots is referred to as corneal reflection centre in the whole paper and this spot is considered for the analysis. With the changes in eye gaze the corneal reflection centre moves around the pupil in some specific pattern. This specific pattern helps to estimate gaze angle. To find the relation between corneal reflection movement and gaze angle, measurements shown in (1), (2) and (3) are used.

$$x = |\text{PupilCenter\_x} - \text{CornealReflectionCenter\_x}| \quad (1)$$

$$y = |\text{PupilCenter\_y} - \text{CornealReflectionCenter\_y}| \quad (2)$$

$$R = \text{Radius of iris} \quad (3)$$

Figure 3 has x,y and R marked

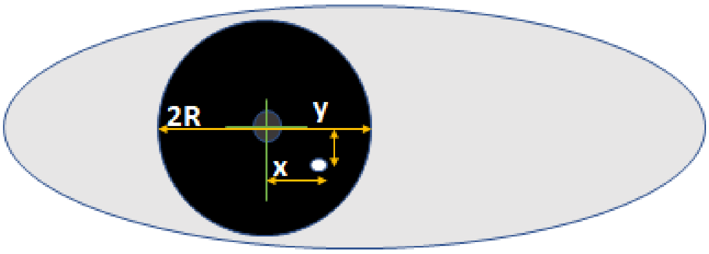
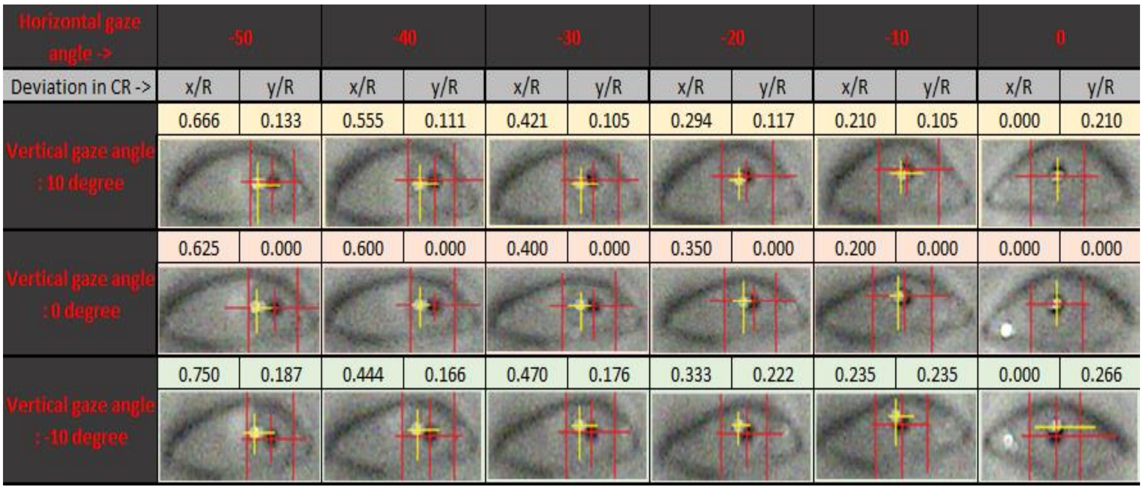


Figure 3. Measurements for gaze estimation from corneal reflection

#### 4.1.1. Horizontal gaze estimation using corneal reflection

To estimate horizontal gaze angle from corneal reflection eye gaze is shifted only in the horizontal direction from left to right and vertical gaze angle kept constant. It is observed that when camera is aligned with the centre of face and the gaze is at the centre of the fixed pattern then the corneal reflection centre overlaps with the centre of pupil. When the gaze shifts towards left from the centre of the pattern then the corneal reflection moves towards right from the pupil centre. And when the gaze shifts towards right from the centre of the pattern then the corneal reflection moves towards left from the pupil centre. The corneal reflection moves farther from the pupil centre when the gaze shifts more towards left or right from the centre of the pattern. But this relation is not linear. To find the relation between corneal reflection position and horizontal gaze angle ’x/R’ is calculated for different horizontal gaze angles with vertical gaze angle kept constant. Instead of ’x’ and ’y’, ’x/R’ and ’y/R is used in the analysis to remove the dependency on image resolution and distance from the camera. The values obtained for ’x/R’ for varying horizontal angles are as shown in Figure 4.

It is clear from the values that the relation is not linear as the difference between subsequent values is not constant. To find the non-linear relation connecting corneal reflection position with eye gaze angle, ’x/R’ is plotted against horizontal eye gaze and different curve fittings are tried on these calibration points.



| Horizontal gaze angle -> | -50 | | -40 | | -30 | | -20 | | -10 | | 0 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Deviation in CR -> | x/R | y/R | x/R | y/R | x/R | y/R | x/R | y/R | x/R | y/R | x/R | y/R |
| Vertical gaze angle : 10 degree | 0.666 | 0.133 | 0.555 | 0.111 | 0.421 | 0.105 | 0.294 | 0.117 | 0.210 | 0.105 | 0.000 | 0.210 |
| Vertical gaze angle :0 degree | 0.625 | 0.000 | 0.600 | 0.000 | 0.400 | 0.000 | 0.350 | 0.000 | 0.200 | 0.000 | 0.000 | 0.000 |
| Vertical gaze angle : -10 degree | 0.750 | 0.187 | 0.444 | 0.166 | 0.470 | 0.176 | 0.333 | 0.222 | 0.235 | 0.235 | 0.000 | 0.266 |

Figure 4. x/R values for different gaze angles

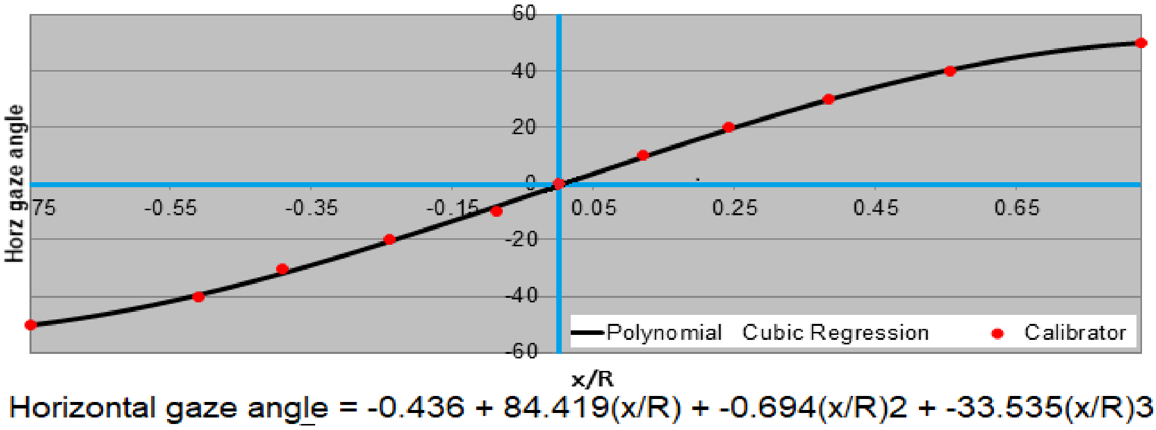


Figure 5. x/R vs gaze angle plot

From figure 5 it can be observed that a third order polynomial is fitting the calibration points effectively. For easy comparative analysis polynomial regression is preferred in this paper over Gaussian or other methods. As the relation between horizontal gaze angle and corneal reflection position is established as a polynomial function, it is possible to calculate the gaze angle from the iris dimensions, pupil and corneal reflection centre position.

$$\text{HorzGazeAngle} = -0.436 + 84.419(x/R) + -0.694(x/R)\ 2 + -33.535(x/R)\ 3 \qquad (4)$$

For a given person the relation represented in 4 is true for varying distance between camera and eyes. This relation is also true at varying vertical eye gaze angle. Figure 6 has ’x/R’ values at different vertical gaze angle and horizontal gaze angle kept constant (30 degree).

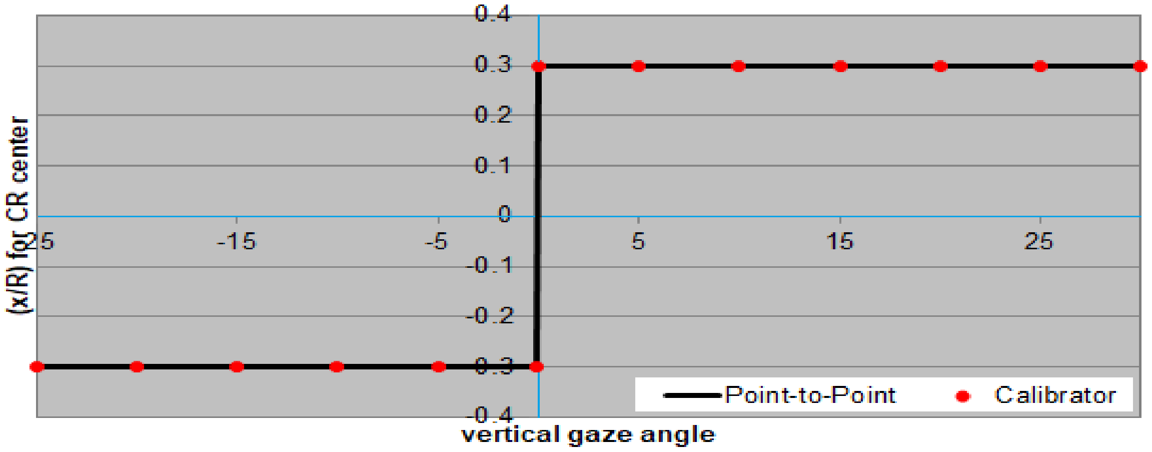
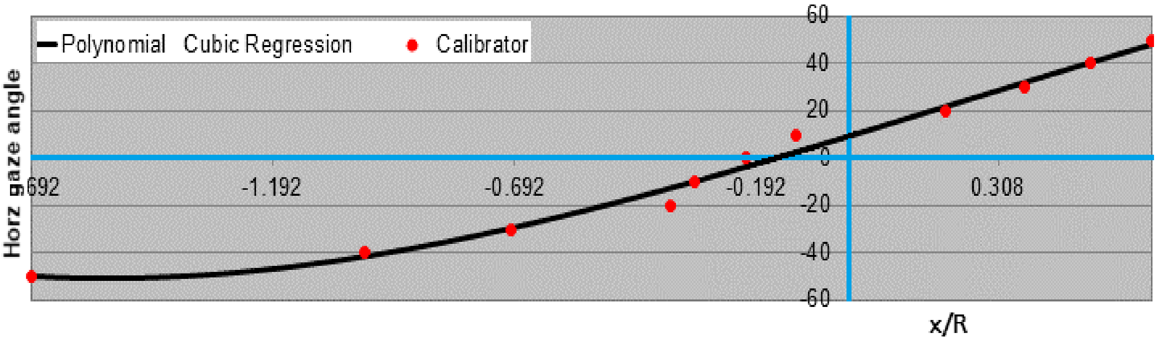


Figure 6. x/R at varying vertical gaze angle

As observed in figure 6 ’x/R’ is a constant at different vertical gaze angle with horizontal gaze angle kept constant.

When camera is not aligned with the centre of both eyes, then there is a shift in the polynomial curve from zero. The curve plotted in Figure 7 is with the camera shifted towards right by 10 degrees.



Horizontal gaze angle = 9.54 + 62.08(x/R) + 4.21(x/R)2 + -6.93(x/R)3

Figure 7. Shift in curve fitting with the shift in camera alignment

It is important that any gaze estimation method should be generic enough to work on multiple people. To check how generic is curve fitting method for gaze estimation, eye images with horizontal gaze varying from -50 to +50 degree and vertical gaze angle kept as zero is collected from different people and a third order polynomial curve is fitted on these calibration points. Figure 8 has values from 3 different people marked in three different colours. The error in curve fitting is calculated and is found to be minimal. Error values of each coefficient of curve fitting has values as shown in table I.

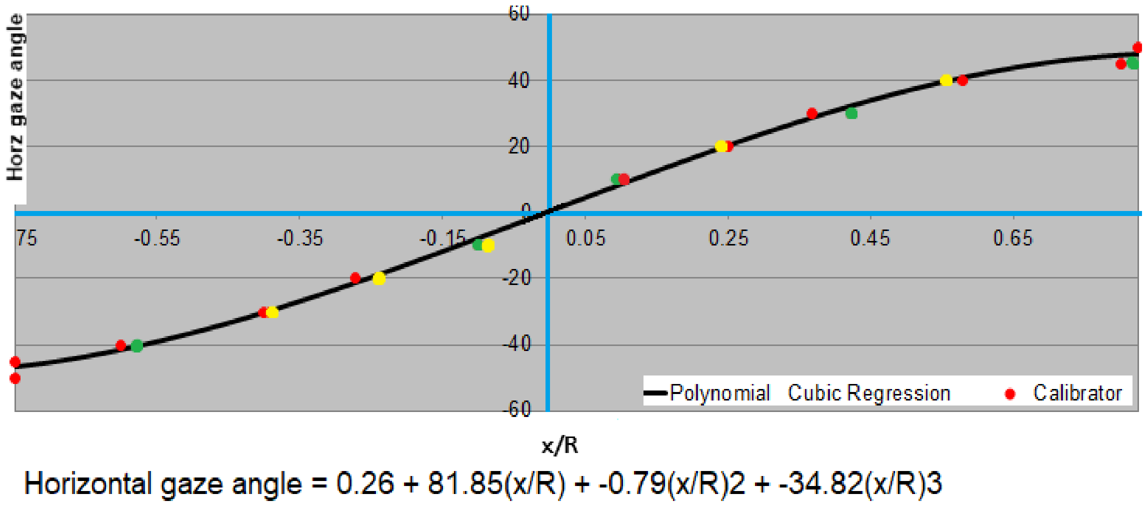


Figure 8. x/R values from 3 different people

Table 1: Error values of polynomial coefficients

| Coeff | Value | Error |
|---|---|---|
| a | 0.26 | 0.638 |
| b | 81.85 | 2.192 |
| c | -0.79 | 2.024 |
| d | -34.82 | 4.576 |

The polynomial curve in figure 8 is used to estimate gaze for person 1,2 and 3 and the error in gaze estimation is obtained as 1.33 deg, 1.61 deg and 1.11 deg respectively. When ’y/R’ is plotted against varying horizontal eye gaze angles (with vertical gaze angle kept constant) then the plot as shown in Figure 9 is obtained. As observed from Figure 9 ’y/R’ is constant for different horizontal eye gaze angle.

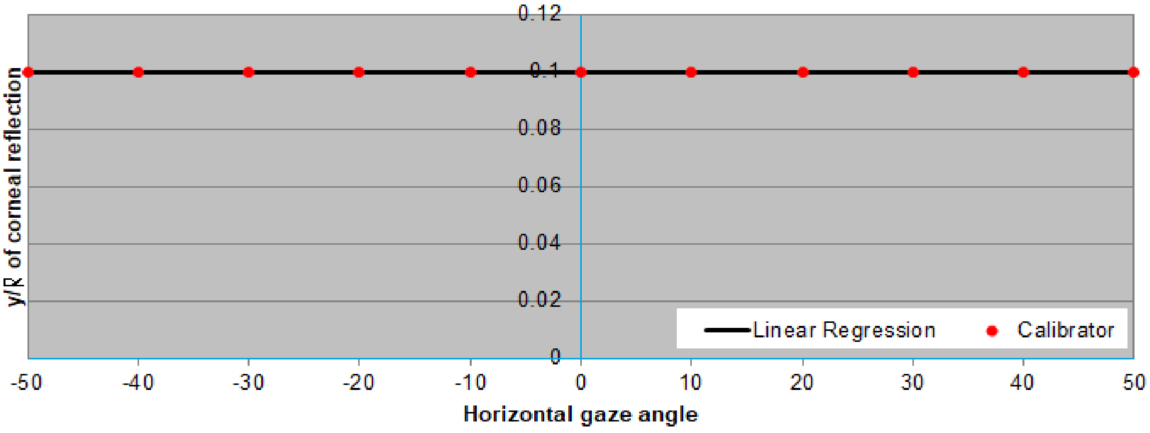


Figure 9. Horizontal gaze vs y/R

### 4.1.2. Vertical gaze estimation using corneal reflection

Similar experiments which are done for horizontal eye gaze estimation has been done for vertical eye gaze estimation also. To find the relation between vertical gaze angle and corneal reflection position, data is captured with eye gaze shifting in vertical direction from -30 degree to 30 degree and horizontal eye gaze kept constant. Following are the key observations from this experiment:

1. It is observed that when camera is aligned with the centre of both eyes and the gaze is at the centre of the fixed pattern then the corneal reflection centre overlaps with the centre of pupil. When the gaze shifts upwards from the centre of the pattern then the corneal reflection moves downwards from the pupil centre. And when the gaze shifts downwards from the centre of the pattern then the corneal reflection moves upwards from the pupil centre. The corneal reflection moves farther from the pupil centre when the gaze shifts more upwards or downwards from the centre of the pattern.

2. ’y/R’ varies with vertical eye gaze as a linear equation. When ’y/R’ is plotted against varying vertical eye gaze angles (with horizontal gaze angle kept constant) then the plot as shown in Figure 10 is obtained.

3. Vertical eye gaze can be estimated from the following eye coordinates: PupilCenter y, CornealReflectionCenter y, radius of iris.

4. Relation between y/R and vertical gaze angle holds true with varying distance between the camera and face for a given person.

5. Relation between ’y/R’ and vertical gaze angles measured at different horizontal gaze angle shows similar pattern (linear relation)

6. ’y/R’ values collected from multiple people is mixed and plotted and the curve fitting error is found to be minimal. 7. With the shift in camera little up from the eye horizon there is a shift in curve from zero.

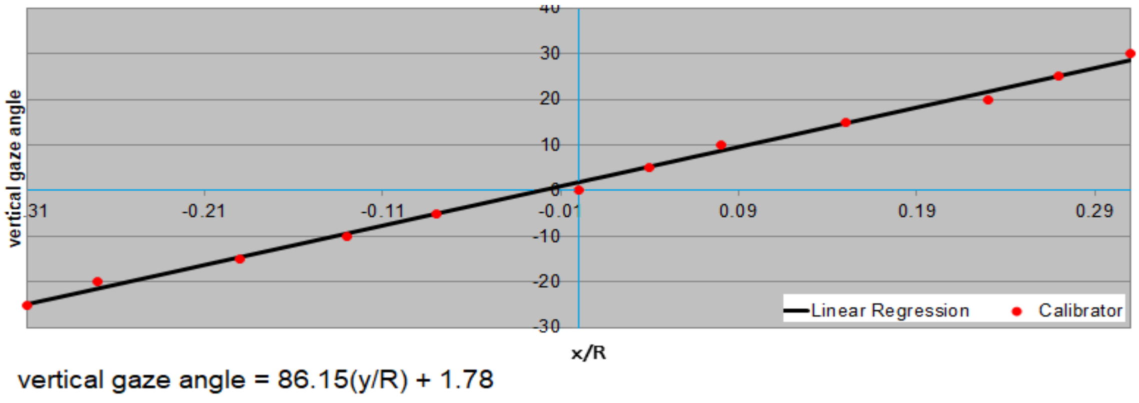


Figure 10. y/R against vertical gaze change

## 4.2. Pupil centre-based gaze estimation

In this method the centre of pupil is tracked relative to the centre of eye for estimating gaze. For accurate and efficient comparison, the same set of images used for the corneal reflection movement analysis are used for pupil-based method as well. Measurements shown in 5 and 6 are used for gaze estimation using pupil centre position.

$$x = |PupilCenterx - EyeCenterx| \quad (5)$$
$$R = DistanceBetweenEyeCornerPoints/2 \quad (6)$$

### 4.2.1. Horizontal gaze estimation using pupil centre

To estimate horizontal gaze angle from pupil centre position, images of eye region are captured with gaze shifting horizontally from left to right with no head movement and the vertical gaze angle kept constant. The movement of pupil centre w.r.t the changes in gaze angle also shows some pattern but is different from the pattern formed by corneal reflection centre. When the gaze is at the centre of the pattern the pupil centre falls at the centre of eye. When the gaze shifts towards left or right the pupil centre moves in the same direction. This is unlike the corneal reflection position which moves in the opposite direction of gaze. But like corneal reflection movement, the movement of pupil centre with gaze change is not linear. To understand this non-linear relation ’x/R’ is plotted against horizontal gaze angle and the plot as shown in figure 12 is obtained. It can be observed from figure 12 that a third order polynomial is fitting the calibration points effectively.

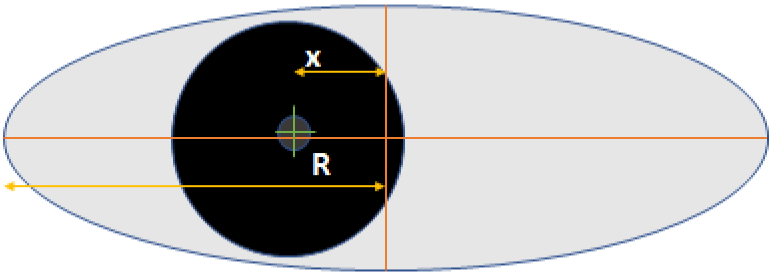
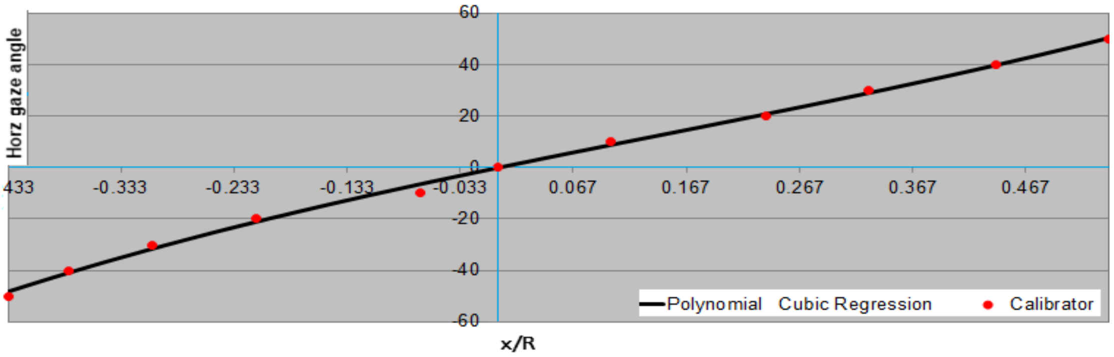


Figure 11. Measurements for gaze estimation from pupil centre



Horizontal gaze angle = -0.17 + 90.83(x/R) + -23.97(x/R)2 + 53.60(x/R)3

Figure 12. Pupil centre deviation with horizontal gaze change

From the polynomial curve it is clear that the horizontal gaze angle can be calculated from pupil centre position and width of eye. This relation between gaze angle and pupil position also holds true for varying distance between camera and eyes. This is also true at varying vertical eye gaze angle. Figure 13 has ’x/R’ values at different vertical gaze angle and horizontal gaze angle equals to 30 degree. ’x/R’ is a constant with varying vertical gaze angle.

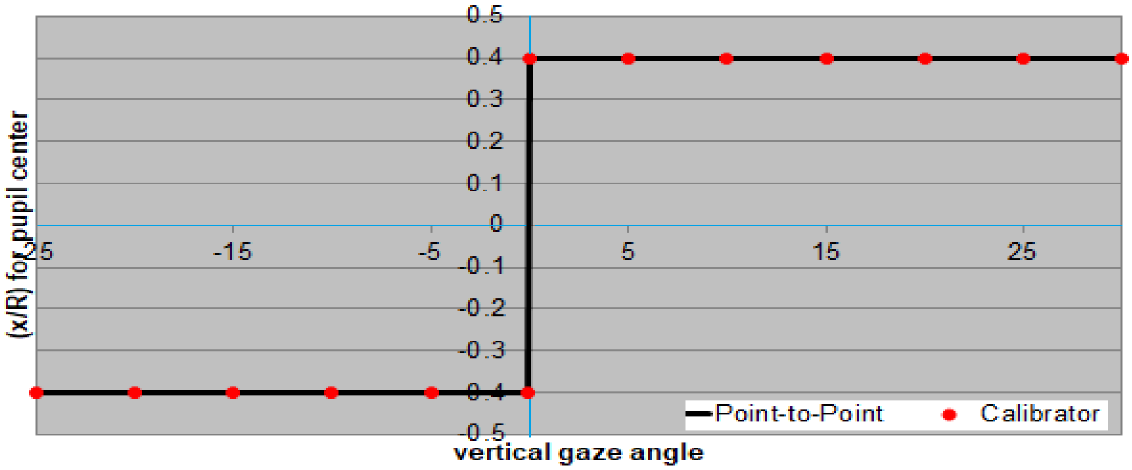


Figure 13. x/R at varying vertical gaze angle

When the camera is not aligned with the centre of face, then there is a shift in the polynomial curve from zero.

To check how generic is pupil centre-based gaze estimation method 'x/R' values calculated from multiple people are mixed and a third order polynomial curve is fitted on these calibration points. Figure 14 has values from 3 different people marked in three different colours. The error in curve fitting is calculated and is found to be more when compared to the error obtained from the corneal reflection-based method. Table II has values of coefficients and its error values.

The polynomial curve in figure 14 is used to estimate gaze for person 1,2 and 3 and the error in gaze estimation is obtained as 3.72 deg, 4.61 deg and 4.88 deg respectively. The higher error in the pupil-based approach is because of the variation in eye shape from person to person.

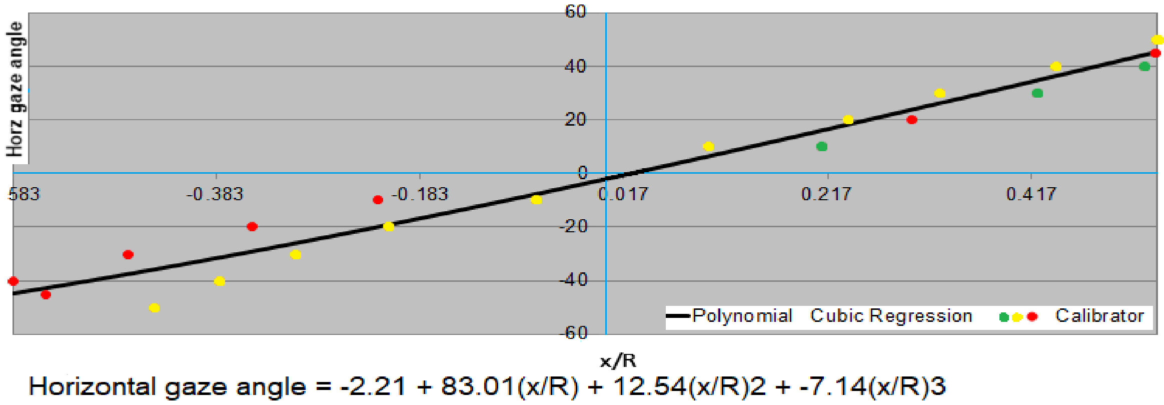


Figure 14. Pupil movement in different people

Table 2: Error values of polynomial coefficients

| Coeff | Value | Error |
|---|---|---|
| a | -2.21 | 2.67 |
| b | 83.01 | 10.41 |
| c | 12.54 | 14.43 |
| d | -7.14 | 43.20 |

### 4.2.2. Vertical gaze estimation using pupil centre:

To estimate vertical gaze angle from pupil centre position, images of eye region are captured with gaze shifting vertically and the horizontal gaze angle kept constant. Figure 15 is for vertical gaze angle equals 0, +10 and -10 degree and horizontal gaze angle kept as ’zero’. In all the three cases the pupil centre is lying above the horizontal line connecting the eye corner points.

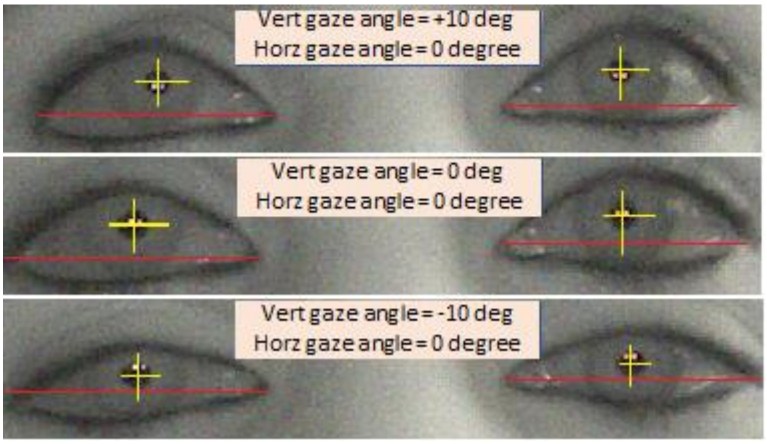


Figure 15. Pupil movement with vertical gaze change

In the case of Corneal reflection-based method, the target point is moving around the centre of pupil in a defined pattern and the pupil centre point can be considered as a reference point. All calculations to estimate eye gaze is independent of the width/height of the eye.

Whereas in the case of pupil centre-based method there is no fixed reference to estimate vertical gaze angle and hence the estimation becomes difficult. The variation in eye shape from person to person makes the gaze estimation challenging with pupil- based method.

## 5. MINIMUM GAZE ANGLE MEASURABLE

In certain applications it is important to correctly differentiate the gaze change while the eye gaze is shifting between two points which are lying really closer. For example, to predict at which point on the infotainment cluster the driver is looking at, high precision gaze estimation techniques are required. The difficulty in gaze estimation increases when the need for precise and high-resolution gaze estimation is required. A comparative study is done on similar lines for pupil based and corneal reflection-based gaze estimation to find out which method is more efficient for precise high-resolution gaze estimation. In order to perform this comparison a fixed pattern in which the points are closely marked is placed in front of the person at 50cm distance from eyes and the gaze is varied horizontally between these closely marked points from left to right. Images of resolution 1280 x 964 are captured using the IR camera. Deviation in corneal reflection centre from pupil centre and pupil centre from eye centre has been measured in pixels for each of the horizontal gaze angle. Table III contains the values, with the gaze changing from one point to next point with an offset of 1 degree.

Table 3: Minimum gaze angle measurable

| **Horz gaze angle (degree)** | **CR deviation (pixels)** | **Pupil deviation (pixels)** |
|---|---|---|
| 1 | 1 | 3 |
| 2 | 2 | 3 |
| 3 | 2 | 3 |
| 4 | 2 | 4 |
| 5 | 2 | 5 |
| 6 | 3 | 5 |
| 7 | 3 | 5 |
| 8 | 3 | 7 |
| 9 | 4 | 7 |
| 10 | 4 | 8 |
| 11 | 4 | 8 |
| 12 | 5 | 9 |
| 13 | 5 | 10 |
| 14 | 5 | 10 |
| 15 | 5 | 10 |
| 16 | 5 | 11 |
| 17 | 6 | 11 |
| 18 | 6 | 11 |
| 19 | 6 | 13 |
| 20 | 6 | 13 |

From the values shown in table III it can be inferred that the minimum gaze angle measurable with corneal reflection-based method is 4 degree and with pupil-based method is 3 degree for the given image resolution and distance from the camera. The resolution at which the gaze angle can be measured will always depend on the resolution of image captured. And for a given camera the resolution of gaze estimation depends on the distance of the camera from the eyes. With the increase in image resolution or with the camera placed more closer to the eyes, the gaze estimation resolution can be improved.

## 6. CONCLUSIONS

Gaze estimation method, its accuracy and efficiency are highly subjective to the use case it is meant for. Choosing the right method hence depends on many parameters. In this paper a comparative study of two methods are done under various scenarios. The pupil-based method has few limitations but it can work with any camera which can capture image of eye region clearly. Pupil based method doesn’t need any IR illuminators and uses only eye features for gaze estimation. The corneal reflection method has few advantages, but it requires infra-red illuminators to illuminate the eyes for forming the bright spot on cornea. On the other hand, using IR camera and IR illuminators is useful in low light or dark scenarios like in driver monitoring systems. From the analysis done on the resolution of gaze estimation it is observed that pupil-based method has better gaze resolution than corneal reflection-based method with the given camera setup.

## ACKNOWLEDGEMENTS

The authors would like to thank people who supported in data collection.

## REFERENCES

[1] J. Sigut and S. Sidha, ”Iris Center Corneal Reflection Method for Gaze Tracking Using Visible Light,” in IEEE Transactions on Biomedical Engineering, vol. 58, no. 2, pp. 411-419, Feb. 2011.

[2] Sesma-Sanchez, Laura. (2016). Gaussian processes as an alternative to polynomial gaze estimation functions. 10.1145/2857491.2857509.

[3] Chennamma, Hr & Yuan, Xiaohui. (2013). A Survey on Eye-Gaze Tracking Techniques. Indian Journal of Computer Science and Engineering. 4.

[4] Morimoto, Carlos& Mimica, Marcio. (2005). Eye gaze tracking techniques for interactive applications. Computer Vision and Image Understanding. 98. 4-24. 10.1016/j.cviu.2004.07.010.

[5] D. W. Hansen, Q. Ji, ”In the eye of the beholder: A survey of models for eyes and gaze”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 3, pp. 478-500, Mar. 2010.

[6] Chunfei Ma, Seung-Jin Baek, Kang-A Choi, Sung-Jea Ko, ”Improved remote gaze estimation using corneal reflection-adaptive geometric transforms,” Opt. Eng. 53(5) 053112 (14 May 2014).

[7] Cazzato, D.; Leo, M.; Distante, C. An Investigation on the Feasibility of Uncalibrated and Unconstrained Gaze Tracking for Human Assistive Applications by Using Head Pose Estimation. Sensors 2014, 14, 8363- 8379.

[8] Gneo, M., Schmid, M., Conforto, S. et al. A free geometry modelindependent neural eye-gaze tracking system. J NeuroEngineering Rehabil 9, 82 (2012). https://doi.org/10.1186/1743-0003-9-82.

[9] Lin, Y., Lin, R., Lin, Y. et al. Real-time eye-gaze estimation using a low-resolution webcam. Multimed Tools Appl 65, 543–568 (2013).

[10] Ince, I.F., Kim, J.W. A 2D eye gaze estimation system with lowresolution webcam images. EURASIP J. Adv. Signal Process. 2011, 40 (2011).

[11] Wen Zhang, Tai-Ning Zhang, Sheng-Jiang Chang, ”Eye gaze estimation from the elliptical features of one iris,” Opt. Eng. 50(4) 047003 (1 April 2011).

[12] Brisson, J., Mainville, M., Mailloux, D. et al. Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. Behav Res 45, 1322–1331 (2013).

[13] M. S. Mounica, M. Manvita, C. Jyotsna and J. Amudha, "Low Cost Eye Gaze Tracker Using Web Camera," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 79-85, doi: 10.1109/ICCMC.2019.8819645.

[14] Brunyé, T.T., Drew, T., Weaver, D.L. et al, "A review of eye tracking for understanding and improving diagnostic interpretation", Cogn. Research 4, 7 (2019).

[15] Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. (2018). Eye-tracking technology in medical education: a systematic review. Medical Teacher, 40(1), 62–69.

[16] A. Kar and P. Corcoran, "A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms," in IEEE Access, vol. 5, pp. 16495-16519, 2017, doi: 10.1109/ACCESS.2017.2735633.

[17] J. Lemley, A. Kar, A. Drimbarean and P. Corcoran, "Convolutional Neural Network Implementation for Eye-Gaze Estimation on Low-Quality Consumer Imaging Systems," in IEEE Transactions on Consumer Electronics, vol. 65, no. 2, pp. 179-187, May 2019, doi: 10.1109/TCE.2019.2899869.

## AUTHORS

**Susmitha Mohan** is a Lead engineer who has experience in developing various imaging and machine learning based solutions for automotive, aerospace and other domains. Worked in algorithm development from scratch, integrating components, testing and validating components as well as solutions, designing and modularization of algorithms. Worked closely with cross-functional teams like testing and validation, hardware porting, on road testing, data collection etc.

**Manoj** is an Imaging/Computer vision architect with wide experience in end to end product development including video/image processing, enhancement, view transformations, segmentation, object detection, recognition, tracking etc. Developed and managed solutions across multiple verticals like avionics, automotive, printing and document imaging, security and surveillance, manufacturing. Co-authored 3 USA patents granted for Automotive Driver Assistance Systems.

# AUTHOR INDEX