Computer Networks & Communications

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

- 9th International Conference on Computer Networks & Communications (CCNET 2022), February 19~20, 2022, Dubai, UAE
- 2nd International Conference on NLP & Text Mining (NLTM 2022)
- 10th International Conference on Instrumentation and Control Systems (CICS 2022)

Published By

AIRCC Publishing Corporation

Volume Editors

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

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

ISSN: 2231 - 5403
ISBN: 978-1-925953-61-9
DOI: 10.5121/csit.2022.120301- 10.5121/csit.2022.120305

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

Preface

9th International Conference on Computer Networks & Communications (CCNET 2022), February 19~20, 2022, Dubai, UAE, 2nd International Conference on NLP & Text Mining (NLTM 2022) and 10th International Conference on Instrumentation and Control Systems (CICS 2022) was collocated with 9th International Conference on Computer Networks & Communications (CCNET 2022). 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 CCNET 2022, NLTM 2022 and CICS 2022 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, CCNET 2022, NLTM 2022 and CICS 2022 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 CCNET 2022, NLTM 2022 and CICS 2022.

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.

David C. Wyld,
Dhinaharan Nagamalai (Eds)

General Chair

Organization

David C. Wyld, Southeastern Louisiana University, USA
Dhinaharan Nagamalai (Eds) Wireilla Net Solutions, Australia

Program Committee Members

Abbasi, Islamic Azad University, Iran
Abdel-Badeeh M. Salem, Ain Shams University, Egypt
Abdelhadi Assir, Hassan 1st University, Morocco
Alia Karim AbdulHassan, University of Technology, Iraq
Ann Zeki Ablahd, Northern Technical University, Iraq
António Abreu, ISEL - Instituto Politécnico de Lisboa, Portugal
Anuj Singal, GJU S&T HISAR, India
Ayush Dogra, Ronin Institute, USA
Brahim Lejdel, University of El-Oued, Algeria
Dadmehr Rahbari, Tallinn University of Technology, Estonia
Diptiranjan Behera, The University of the West Indies, Jamaica
Dongping Tian, Baoji University of Arts and Sciences, China
Ekbal Rashid, RTC Institute of Technology, India
Everton Flemmings, iValley Nutraceuticals-President, Canada
Farid Kheiri, TVTO, Iran
Felix J. Garcia Clemente, University of Murcia, Spain
Grigorios N. Beligiannis, University of Patras, Greece
Hamid Ali Abed AL-Asadi, Iraq University College, Iraq
Hassan Hadi Saleh, University of Diyala, Iraq
Henok Yared Agizew, Mettu University, Ethiopia
Holger Kyas, University of Applied Sciences Berne, Switzerland
Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India
J. Naren, iNurture Education Solutions Private Limited, India
Khan Ferdous Wahid, Airbus, Germany
Ljubomir Lazic, Belgrade UNION University, Serbia
Loc Nguyen, Independent scholar, Vietnam
Luisa Maria Arvide Cambra, University of Almeria, Spain
Mehdi Gheisari, Islamic Azad University, Iran
Mehdi Nezhadnaderi, Islamic Azad University, Iran
Mounir Zrigui, University of Monastir, Tunisia
Mu-Song Chen, Da-Yeh University, Taiwan
Nadia Abd-Alsabour, Cairo University, Egypt
Rajeev Kanth, Savonia University, Finland
Rajini Kanth, SNIST, India
Ramakrishnan, Drmgr Educational And Research Institute, India
Ramana Murthy, Osmania University, India
Ramgopal Kashyap, Amity University Chhattisgarh, India
Richa Purohit, DY Patil International University, India
Sadique Shaikh, AIMSR, India
Samira Hazmoune, University of 20 Août 1955-Skikda, Algeria
Seppo Sirkemaa, University of Turku, Finland
Shashikant Patil, ViMEET Khalapur Raigad MS, India
Siarry Patrick, Universite Paris-Est Creteil, France
SILMI Souhila, USTHB University, Algeria

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Artificial Intelligence Community (AIC)

Soft Computing Community (SCC)

Digital Signal & Image Processing Community (DSIPC)

9th International Conference on Computer Networks & Communications (CCNET 2022)

2nd International Conference on NLP & Text Mining (NLTM 2022)

10th International Conference on Instrumentation and Control Systems (CICS 2022)

ANALYZING AND PERSONALIZING THE LEARNING PERFORMANCE FOR SPECIAL NEEDS STUDENTS USING MACHINE LEARNING AND DATA ANALYTICS

Aaron Fei[1] and Yu Sun[2]

[1]St. Margaret' s Episcopal High School, 31641 La Novia Ave, San Juan Capistrano, CA 92675, USA
[2]California State Polytechnic University, Pomona, CA, 91768, Irvine, CA 92620, USA

ABSTRACT

Recognizing the fact that autistic kids usually have troubles socially and focusing on academic studies, this research attempts to give a more insightful perspective on the ethnic way of helping autistic kids through technologies [4]. The core idea of this paper is to find a way of helping the autistic kids to maximize their potential instead of accommodating this society using assistive tools. Holding the responsibility of sharing the advantages in this society, this application is built to the end of using a general value to connect level of focus to level of reward. This solution is achieved by three steps: Designing a text box with different variables that evaluates focus level, calculating the level of reward based on achievements on the variables, and the game begins with different hardness according to the level of reward. The results show that the designed application increases the focus level of the kids and their willingness to communicate surprisingly.

KEYWORDS

Python, Computer Science, Autism

1. INTRODUCTION

There are 1 in 44 children who are struggling with autism in the United States [5]. Autistic kids usually struggle in focusing, communicating, and socializing with others [6]. The goal is to design an application to help autistic kids to focus. Since autistic children are born with a certain level of disadvantage in society, sharing our advantage with them through the present technology is demanded in present days [14]. To this end, this application is attempting to give an intriguing solution to help the autistic kids through games.

Nowadays, there are many assistive technologies for autistic kids [7]. For example, there are augmentative and alternative communication (AAC), non-verbal people to robots built to increase social skills in children on the spectrum, and picture boards. Even though all of them may have their intention of helping autistic kids in a variety of ways, it is also true that there are a few problems with these assistive technologies [8].

Admittedly, augmentative and alternative communication is a high technology tool for autistic kids to use to improve their social performance. It is true that it has the potential that autistic kids

David C. Wyld et al. (Eds): CCNET, NLTM, CICS - 2022
pp. 01-09, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120301

will spend more time in the process of learning how to use it than using it to improve their skills. Also, it is too expensive to produce widely to allow all the autistic to use it.

Secondly, even though non-verbal people can train the autistic kids according to their own personality, it should be heavily doubted by the practicality of this robot for all autistic kids since the huge population.

Last but not least, paint boards may seem to be a solution to the large population of autistic kids, but its problem is that it never treats the autistic kids equally according to their own circumstances. It is obvious that it is incorrect to use one tool to help all of the kids with completely different demands and personality.

My method is to find a way to encourage their focus by giving awards to evoke their internal motivation to strive [9]. Since the level of focus can be evaluated differently under different situations, I define it as the number of characters that the kid enters in a certain amount of time. So, this tool itself should be used for writing. However, this tool is designed as modularized that could eventually link different applications that evaluate focus levels differently and different games according to individual's demand in future development. This makes my tool easily usable to different individuals, which makes it better than a picture board. It is better than AAC since it is easily usable. Moreover, it is cheaper and allows more autistic kids to use it.

I will give this application to a number of different autistic kids as users to let them use it for a while. Then, I will give them a survey that rates certain aspects of it from 1 to 10. After they answer the questions, I will see if the result is significant enough that I can conclude this application will help them in focusing and studying using the reward system.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Identifying the Problem and the Righteous Way to Solve it

Going back to the basics of this problem, these children were born with autism. Then, what should be our proper approach to help them? The moral problem is equality or equity. If we help the autistic kids who were born wealthy, then what about those kids who were born normal but poor. Who should we help to achieve real justice? Taking this problem into consideration, I apply the idea from John Rawls that we should bring the exceeding advantage from us to those people who were born demanding help. Instead of forcing the autistic kids to accommodate the normal way of socializing or giving them assistance to help them communicate, which might eventually be considered as deviant to others, I regard designing a tool that allows the autistic kids to use it according to their free will is more vital [10].

2.2. Difference Between Individuals

Taking an individual's difference into consideration, I recognize the problem that this application should not simply divide human beings into autistic and unautistic because this is not treating each individual as themselves. To this end, I regard a modularized design is vital that allows more kids with different demands in games and focus areas to use. Even though the application itself is not designed with different choices of games and assistance tools for focus, it generalizes the idea of using a general value to connect the reward and assistance tool to approach this problem. This could be used for future research.

2.3. Games available and reward level

There are very few attractive games that will provide their codes to the public. To find a game that is both satisfying having a variable that could change accordingly with level of focus and public is not easy. In addition, to determine the level of reward according to focus to make the game not too easy or too hard is a vital process to make the reward attractive.

3. SOLUTION

This application is designed in three major steps. They are input, transfer, and reward. This first step is to give the user an interactive surface to ask for their goal. According to this input value, a timer is going to be set. All the inputs will be calculated after the user clicks. Secondly, the user's performance is going to be evaluated by calculating wpm and checking the time requirement. Through comparison between the time requirement that the user inputs initially and wpm's calculation, the reward is going to be determined as negative reward, positive reward, or normal reward. Lastly, different rewards are going to determine the level of hardness for tetris that changes the speed of blocks falling downward.

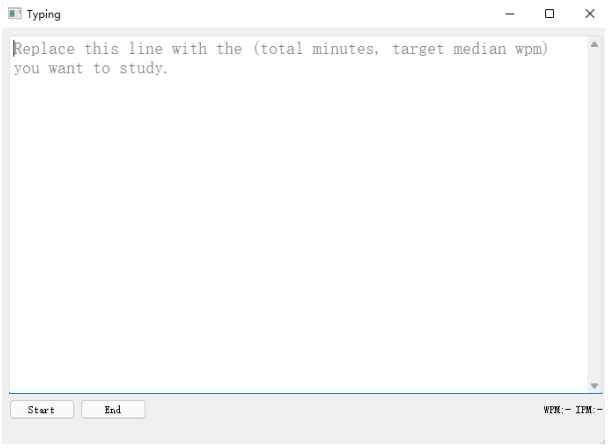

Figure 1. The user interactive surface

The figure above demonstrates the user interactive surface of the text box that constantly updates the user's level of focus in the format of WPM and IPM. Both start and end button were used to confirm the goal of focus level and the concluding time for the application to decide the level of reward.

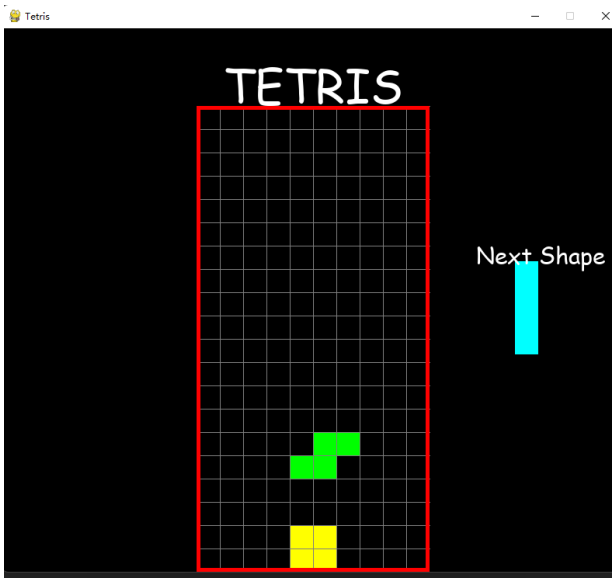

Figure 2. Screenshot of the reward

The above figure demonstrates the reward that is Tetris. The game has almost exactly the same rules as the official game of tetris. To the end of giving users rewards based on their performance during writing, the speed of blocks that were falling has a negative relationship with the achievements on wpm and time requirements.

```
def startTest(self):

    text = self.textEdit.toPlainText()
    print(len(text.split()))
    break_text = text.split()

    if len(break_text) > 2:
        self.textEdit.clear()
        self.textEdit.setPlaceholderText('YOU HAVE ENTERED TOO MANY ARGUMENTS !!!')
        return
    if break_text[0].isnumeric():
        if break_text[1].isnumeric():
            self.timeLimit = int(break_text[0])
            self.targetWPM = int(break_text[1])
        else:
            self.textEdit.clear()
            self.textEdit.setPlaceholderText('WPM must NUMBERS ONLY !!!')
            return

    else:
        self.textEdit.clear()
        self.textEdit.setPlaceholderText('Time limit MUST BE NUMBERS ONLY !!!')
        return
```

Figure 3. Code of the level of focus

This is a piece of the code that asks the user for the level of focus in a period of literary writing on the computer they are aiming for. Before they start writing, they must input the time limit and wpm goal into the text box; Then, they can click the button to start. In the figure, the user was asked to enter the correct input in order for the application to run successfully by checking if

there are only two arguments that are both numbers. If not, they must re enter until the application could run successfully.

```
def resetTest(self):
    self.allInputNum = []
    self.validAlphaNum = []
    self.inputNum = 0
    self.textEdit.setPlainText("")
    self.textEdit.setDisabled(False)
    self.label.setText("WPM:-")
    self.label_2.setText("IPM:-")
    print("wpms: ", self.wpms, "\nipms: ", self.ipms)
    wpm_median = statistics.median(self.wpms)
    ipm_median = statistics.median(self.ipms)
    print("wpm med: ", wpm_median, "\nipm med: ", ipm_median)
    duration = (time.time() - self.startTime) / 60
```

Figure 4. Code of variables

The figure above shows variables that were recorded during the user's input. Wpm is defined as word per minute and Ipm is defined as input per minute. Normally, other applications that assess the level of focus will use the mean of wpms. However, this application uses median to assess a more constant focus. For example, if the user did not write anything in 9 minutes, but copied and past 500 words in the last minute, the mean will demonstrate that he has a 50 wpm, which obviously did not truly demonstrate the user's level of focus. Indeed, a median is used to prevent any false interpretation on the user's focus level in this application.

```
if duration < self.timeLimit:
    if (wpm_median > self.targetWPM):
        print("You are quick but somewut productive")
        if not ( games):
            tt.win = pg.display.set_mode((tt.s_width, tt.s_height))
            tt.pygame.display.set_caption('Tetris')
            tt.main_menu(0.05)
        else:
            print("dir: ", dir)
            os.system('python '+dir+ " 0.05")
    else:
        if not games:
            print("You failed.")
            tt.win = pg.display.set_mode((tt.s_width, tt.s_height))
            tt.pygame.display.set_caption('Tetris')
            tt.main_menu(0.01)
        else:
            print("dir: ", dir)
            os.system('python '+ dir + " 0.01")
```

Figure 5 (1). Transfer the variables of level of focus to variables for rewards

```
if duration > self.timeLimit:
    if (wpm_median > self.targetWPM):
        if not games:
            print("Wow a really good student.")
            init_speed = duration* 0.03+ 0.05
            tt.win = pg.display.set_mode((tt.s_width, tt.s_height))
            tt.pygame.display.set_caption('Tetris')
            tt.main_menu(init_speed)
        else:
            init_speed = duration * 0.03 + 0.05
            print("dir: ", dir)
            os.system('python '+dir + " "+ str(init_speed))
    else:
        if not games:
            print("you have met the time requirement but your wpm is too slow")
            init_speed = duration / 10 * 0.002 + 0.05
            tt.win = pg.display.set_mode((tt.s_width, tt.s_height))
            tt.pygame.display.set_caption('Tetris')
            tt.main_menu(init_speed)
        else:
            init_speed = duration / 10 * 0.002 + 0.05
            print("dir: "dir)
            os.system('python '+dir + " "+ str(init_speed))
```

Figure 5 (2). Transfer the variables of level of focus to variables for rewards

The above figures demonstrate statements that transfer the variables of level of focus to variables for rewards. Through the comparison between median wpm and goal wpm, and time written and time requirement goal, the users will have four different results. Giving more weight on the median wpm than time requirement, taking that time requirements could be easily met without achieving productivity, the level of reward will be negative if the median wpm does not meet the goal. If the wpm goal is met, according to the achievement on time requirement, positive and normal rewards will be given accordingly.

4. EXPERIMENT

To prove our solution, multiple experiments are run, each with multiple trials. Each experiment alters one part: algorithm parameters, volunteer autistic kids groups from 7 - 12, and training data sizes. Through these experiments, we will be able to determine if an algorithm depends on certain factors to produce the best results. Each algorithm is then tested and scored to find the most accurate prediction model. And we will also decide if the solution has an effective influence on autistic kids based on different group-level.

4.1. Experiment 1

For the first experiment, the goal is to verify at what age group our application works the best for children with autism. Through sampling 5 groups of children of different ages and letting them use the application from the same time, the result is collected by surveying the parents and counting how many children had significantly improved. Experiments have shown that all children in different age groups show a certain level of interests. Age group 8 and 10 showed the largest interest among all.

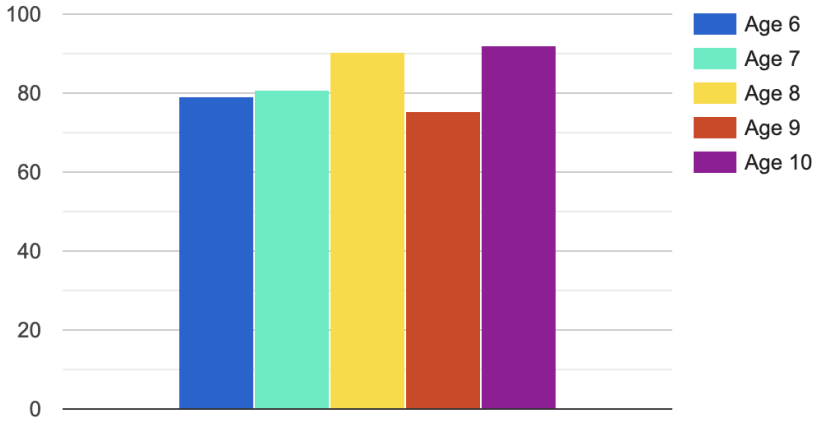

Figure 6. Result of Experiment 1

4.2. Experiment 2

The second experiment is to track to what extent the game was able to increase the desire to communicate in children with autism since this is also a part of autistic kid's disadvantages. Surprisingly, this produces a positive result that this paper did not expect initially. By selecting a group of students of the same age and giving a survey to their parents to see if they demonstrate an increase in willingness to communicate after shortly using the application, this experiment demonstrates more than 50% of kids showed an increase in willingness to communicate.

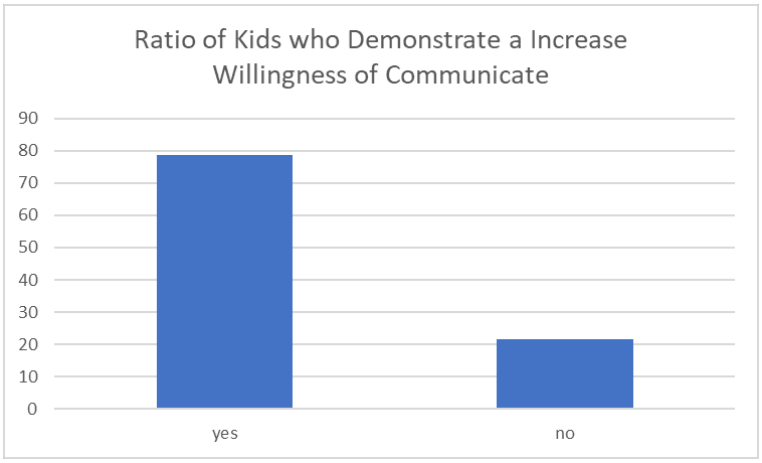

Figure 7. Result of Experiment 2

4.3. Experiment 3

The third experiment evaluates the original goal of this paper: the level of increase in focus level. By using the same method of giving surveys to the parents after the last question of assessing the willingness of communication, this question was asked if the kid showed an increase in ability to focus. Indeed, the results demonstrate a positive impact from the application to the autistic kids.

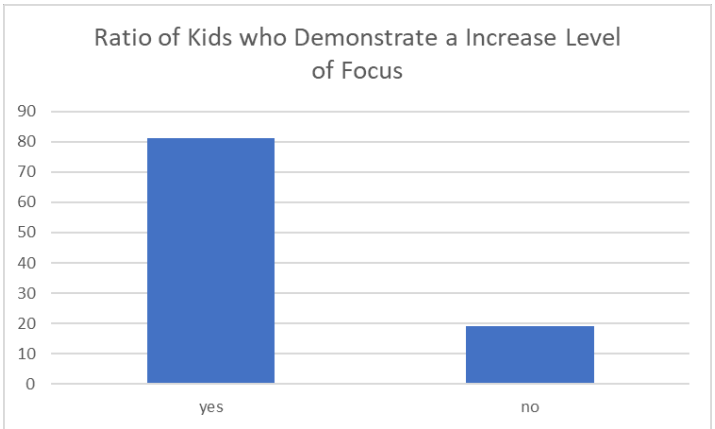

Figure 8. Result of Experiment 3

5. RELATED WORK

This paper demonstrates a reevaluation of the previous inventions that were used as assistive technologies for autistic children [1]. After having the result of understanding the application of previous technologies, this paper gives deep insights into limitations of these tools and future directions for these technologies. By interpreting the content while holding the end to develop a better tool for autistic children, this research follows the direction that was summarized from previous tools by combining the advantages and taking the limitations into consideration [15].

This research developed a word processing and organization as an assistance tool to assist students with disabilities to type or increase their performance on computational literature writing [2]. This application provides a similar goal as this paper that is to provide a more equal opportunity for people who were born with disadvantages. The difference between this application and the goal of this paper is that this paper holds the belief that the children could improve their skill under the proper circumstances, but this application could be used more as part of their input process instead of maximizing their own potential. To achieve a better tool for the autistic kids, it might be more helpful for them if this application's design could be incorporated with this paper's design in order to maximize children's potential more efficiently.

This paper analyzes the impact of literature writing on computers to paper and pen writing [3]. Through a demonstration of a positive effect, this analysis inspires the idea of this paper that to design a writing assistive tool for autistic children by strengthening the idea that this application might not only evoke their motivation to write using a reward system, but it may also improve their performance on paper and pen writing, which may eventually help them in social circumstances in reality [12].

6. CONCLUSIONS

The purpose of this research is to develop an assistive technology as a tool for autistic kids who lack the ability to focus and socialize to perform better in literature writing and ability to focus [11]. To the end of creating a more equal competing environment for everyone, this paper holds the belief similar to John Rawls that people with advantages have the responsibility to share them with people who don't. Indeed, through a platform utilizing a general variable that connects the basic reward theory and literary writing input surface, this application was formed. After doing experiments and taking other researchers' results into consideration, this research paper did reach its basic goal of assisting the autistic kids.

However, there are several improvements that could be made to increase the performance of this application regarding the limitations. First, there is a limit on the users for the present design since it only allows users to focus on inputting words, and playing tetris as a reward. However, taking other situations into consideration, this application will not be useful if the autistic kids were given homework to read. Also, tetris may not inspire all autistic kids' internal motivations. Moreover, this application could only help kids with a slight autism who could focus more easily than those with more severe symptoms.

As this paper had mentioned in the beginning, these limitations were already taken into consideration when this research was initialized. The idealized way to approach these limitations was to use a modularized design to reach the goal of individualizing this application for each user. This goal could be reached by developing more tools like linking to specific web pages or as a google extension that assist the autistic kids to focus using other ways. In addition, connection to more games is imperative as well to satisfy different autistic children's demands on games to evoke each of their internal motivation to maximum. For more severe autistic children, further investigation is required to find the way to help them.

REFERENCES

[1] Knight, V., McKissick, B.R. & Saunders, A. A Review of Technology-Based Interventions to Teach Academic Skills to Students with Autism Spectrum Disorder. J Autism Dev Disord 43, 2628 – 2648 (2013). https://doi.org/10.1007/s10803-013-1814-y

[2] Montgomery, D. J., & Marks, L. J. (2006). Using Technology to Build Independence in Writing for Students With Disabilities. Preventing School Failure: Alternative Education for Children and Youth, 50(3), 33–38. doi:10.3200/PSFL.50.3.33-38

[3] Langone, J., Willis, C., Malone, M., Clees, T., & Koorland, M. (1994). Effects of Computer-Based Word Processing Versus Paper/Pencil Activities on the Paragraph Construction of Elementary Students With Learning Disabilities. Journal of Research on Computing in Education, 27(2), 171–183. doi:10.1080/08886504.1994.10782125

[4] Bleuler, Eugen. "20. Autistic Thinking." Organization and pathology of thought. Columbia University Press, 1951. 399-437.

[5] Speaks, Autism. "What is autism." Retrieved on November 17 (2011): 2011.

[6] National Research Council. Educating children with autism. National Academies Press, 2001.

[7] Cook, Albert M., and Susan M. Hussey. "Assistive technologies." Principles and practice (2002).

[8] Scherer, Marcia J., and Rob Glueckauf. "Assessing the benefits of assistive technologies for activities and participation." Rehabilitation Psychology 50.2 (2005): 132.

[9] Weber, Keith. "The relationship of interest to internal and external motivation." Communication Research Reports 20.4 (2003): 376-383.

[10] Castree, Noel. "Socializing nature: Theory, practice, and politics." Social nature: Theory, practice, and politics (2001): 1-21.

[11] Edyburn, Dave L. "Rethinking assistive technology." Special Education Technology Practice 5.4 (2004): 16-23.

[12] Gómez, Luis Felipe. "Time to socialize: Organizational socialization structures and temporality." The Journal of Business Communication (1973) 46.2 (2009): 179-207.

[13] Weiss, Peter. The investigation. Marion Boyars, 1966.

[14] Howlin, Patricia, et al. Treatment of autistic children. John Wiley & Sons, 1987.

[15] Rutter, Michael. "The treatment of autistic children." Journal of Child Psychology and Psychiatry 26.2 (1985): 193-214.

THE CHALLENGES AND VIABILITY OF USING BLOCKCHAIN FOR WSN SECURITY

Muhammad R. Ahmed, Thirein Myo and Badar Al Baroomi

Military Technological College, Muscat, Oman

ABSTRACT

Wireless Sensor Network (WSN) comprises of cheap and multifunctional resources constrain nodes that communicate at a fair distances through wireless connections. It is open media and underpinned by an application scenario for data collecting and processing. It can be used for many exclusive applications range from military implementation inside the battlefield, environmental tracking, fitness quarter as well as emergency response of surveillance. With its nature and application scenario, protection of WSN had drawn an attention. It is understood that the sensor nodes are valuable to the attacks because of the construction nature of the sensor nodes and distributed network infrastructure. In order to ensure its capability especially in malicious environments, security mechanisms are essential. In this paper, we have discussed the challenges and the viability of the blockchain to implement in the WSN in order to protect WSN from the attacks.

KEYWORDS

Wireless Sensor Network, Security, challenges, blockchain.

1. INTRODUCTION

Wireless sensor networks (WSNs) and their applications are becoming part of our daily life, they have a great advantage for various applications in our real life [1], such as habitat monitoring, battlefield surveillance, intelligent agriculture, home automation, etc. However, the properties of WSN inevitably have the natures that are extremely restricted by their resources, including energy, memory, computing complexity, bandwidth, and communication capacity. Normally the base station is a more powerful node, which can be linked to a central station via satellite or internet communication to form a network. There are many deployments for wireless sensor networks depending on various applications, such as, environmental monitoring, volcano detection [1-3], distributed control systems [4], agricultural and farm management [5], detection of radioactive sources [6], and computing platform for tomorrows' internet [7]. However, the open nature of the wireless medium therefore offers chances for an adversary to easily eavesdrop information from the sensors, or actively do something such as replay or inject fabricated messages. A Typical WSN is shown in Figure 1.

In all communication network including WSN, Security provisioning is a critical requirement. Security in the wireless sensor network is challenging and important task because of its characteristics that includes, open nature of wireless medium, unattended operation, limited energy, memory, computing power, communication bandwidth, and communication range. So, it is more susceptible to the security attack compared to the traditional wired network. It is well known that for the protection from the some WSNs attacks, various cryptographic methods are widely used but sometimes are not very efficient and effective [8-9].

David C. Wyld et al. (Eds): CCNET, NLTM, CICS - 2022
pp. 11-18, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120302

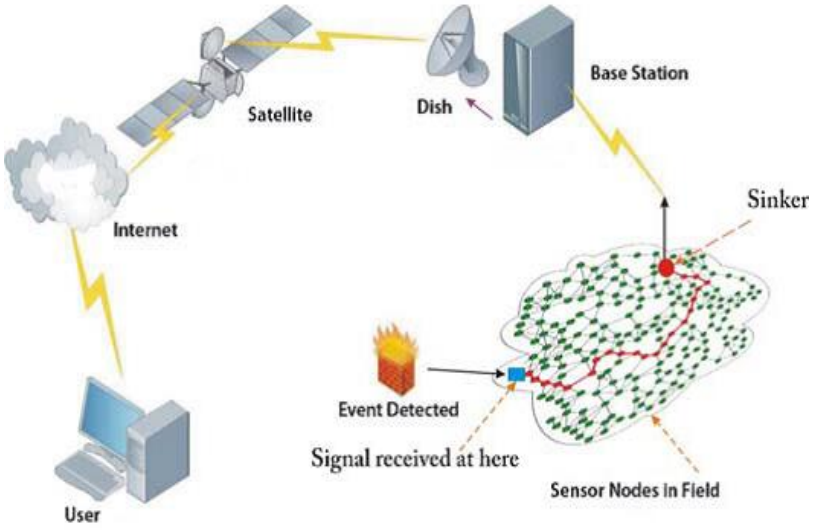

Figure 1. A typical WSN [1]

Blockchain is a method that lets the transmission of data securely based on an incredibly complex encryption mechanism. In this mechanism, each block includes records about its advent time and is related to the previous block by using hash code and transaction information. When the data is recorded in the blockchain, there's almost impossible to change it [10]. Blockchain is designed to resist fraud and alteration of facts. Implement blockchain into WSNs will deliver few advantages:[11]

a. The distributed nature of blockchain will allow a large number of sensor node connection easily
b. Computing and storage needs are distributed to all devices in the network, this will reduce the cost of the large central data
c. Centralized Server and Client Model will be eliminated when peer-to-peer messaging, file distribution, and automatic coordination between devices in the network are used

In this paper, we are focusing on investigating the possible viability and major challenges to implement the blockchain to secure WSN.

This paper consists of five sections. A brief discussion of the characteristics of WSN is in Section II. Section III describes the architecture of WSN. Section IV describes the blockchain structure. A brief discussions of WSN security with blockchain is given in Section V. conclusion is in Section VI

2. CHARACTERISTICS OF WSN

WSN is currently used for real-world unattended physical environment to measure numerous parameters. So, the characteristics of WSN must be considered for efficient deployment of the network. The significant characteristics of WSN are described as follows [12]:

Low cost: in the WSN normally hundreds or thousands of sensor nodes are deployed to measure any physical environment. To reduce the overall cost of the whole network, the cost of the sensor node must be kept as low as possible.

Energy efficient: energy in WSN is used for different purpose such as computation, communication, and storage. Sensor node consumes more energy compared to any other for communication. If they run out of power, they often become invalid as we do not have any option to recharge. So, the protocols and algorithm development should consider the power consumption in the design phase.

Computational power: normally node has limited computational capabilities as the cost and energy need to be considered.

Communication Capabilities: WSN typically communicate using radiowaves over a wireless channel. It has the property of communicating in short range, with narrow and dynamic bandwidth. The communication channel can be either bidirectional or unidirectional. With the unattended and hostile operational environment, it is difficult to run WSN smoothly. So, the hardware and software for communication must have to consider the robustness, security, and resiliency.

Security and Privacy: Each sensor node should have sufficient security mechanisms to prevent unauthorized access, attacks, and unintentional damage of the information inside of the sensor node. Furthermore, additional privacy mechanisms must also be included.

Distributed sensing and processing: the large number of sensor node is distributed uniformly or randomly. WSNs each node can collect, sort, process, aggregate and send the data to the sink. Therefore, the distributed sensing provides the robustness of the system.

Dynamic network topology: in general, WSN a dynamic network. The sensor node can fail for battery exhaustion or other circumstances, communication channel can be disrupted as well as the additional sensor node may be added to the network that result the frequent changes in the network topology. Thus, the WSN nodes must be embedded with the function of reconfiguration, self-adjustment.

Self-organization: the sensor nodes in the network must have the capability of organizing themselves as the sensor nodes are deployed in an unknown fashion in an unattended and hostile environment. The sensor nodes have work in collaboration to adjust themselves to the distributed algorithm and form the network automatically.

Multi-hop communication: a large number of sensor nodes are deployed in WSN. So, the feasible way to communicate with the sinker or base station is to take the help of an intermediate node through routing path. If one need to communicate with the other node or base station which is beyond its radio frequency it must be through the multi-hop route by intermediate node.

Application oriented: WSN is different from the conventional network due to its nature. It is highly dependent on the application ranges from military, environmental as well as health sector. The nodes are deployed randomly and spanned depending on the type of use.

Robust Operations: Since the sensors are going to be deployed over a large and sometimes hostile environment. So, the sensor nodes must be fault and error tolerant. Therefore, sensor nodes need the ability to self-test, self-calibrate, and self-repair.

Small physical size: sensor nodes are generally small with the restricted range. Due to its size its energy is limited which makes the communication capability low.

3. ARCHITECTURE OF WSN

WSN is dynamic which can consist of various types of sensor nodes. The environment is heterogeneous in terms of both hardware as well as software. The sensor node construction focuses to reduce cost, increase flexibility, provide fault tolerance. Improve development process and conserve energy. The structure of sensor node consists of sensing unit (sensor and analog to digital converter), processing unit (processor and storage), communication unit (transceiver), and power supply unit. [13] The major blocks shown in Figure 2. A concise description of different unit is as follows [13-17]:

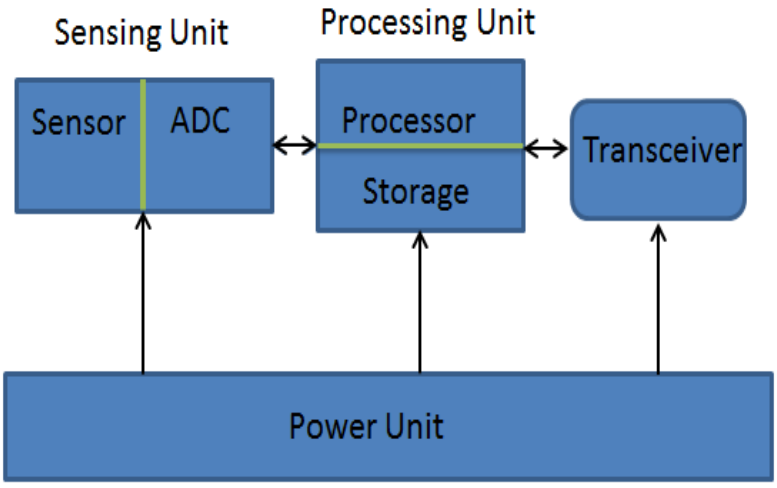

Figure 2. Structure of a sensor node

Sensing unit: it is composed of collection of different types of sensors which is needed for measurement of different phenomenon of the physical environment. Sensors are selected based on its application. Sensor out is electric signal which is analog. So, analog-to-digital converter (ADC) is used to transform the signal to digital in order to communicate with the microcontroller.

Processing unit: it consists of a processor (microcontroller) and storage (RAM). In addition, it has operating systems as well as timer. The responsibility of the processing unit includes collecting data from various sources than processing and storing. Timer is used to do the sequencing for the sequence.

Communication unit: it uses a transceiver which consists of a transmitter as well as a receiver. The communication is performed through the communication channels by using the network protocol. Based on the application requirements and relevance to communicate it normally uses suitable method such as radio, infrared or optical communication.

Power unit: the task of the power unit is to provide the energy to the sensor node for monitoring the environment at a low cost and less time. The life of the sensor depends on the battery or power generator which is connected to the power unit. Power unit is required for the efficient use of the battery.

WSN communication architecture is a bit different from the conventional computer communication and computer network. The communication architecture can be classified in different layers. To get the maximum efficiency with limited resources and low overhead WSN does not adhere as closely to the layered architecture of OSI model of conventional network.

Nevertheless, the layered model is useful in WSN for categorizing protocols, attacks and defence. So, in contrast to the traditional seven layers it is reduced to the five layers [13] that include physical layer, Data link layer, network layer, transport layer and application later. The advantage of the layered model is conceptually similar functions are combined at one layer. Figure 3 shows the communication protocol model of wireless sensor network.

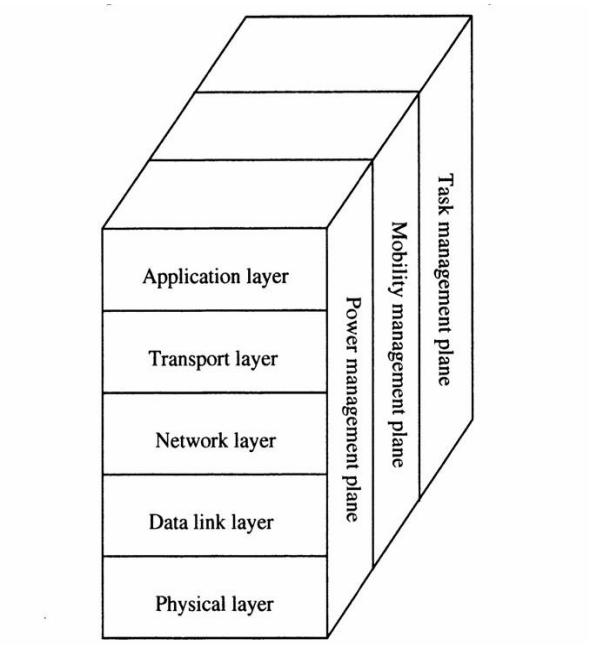

Figure 3. Protocol Stack of WSN [13]

The physical layer addresses the hardware detail of wireless communication mechanism. This layer is responsible for frequency selection, carrier frequency generation, signal detection, modulation, and data encryption. The data link layer is concerned with the media access control (MAC) protocol. Since the wireless channel is susceptible to the noise and sensor nodes may be changing the location MAC protocol at the data link layer has to be power-aware and should have the capability of minimizing the collisions.[4] The network layer manages the routing the data supplied by the transport layer or between the nodes. Whereas the transport layer can maintain the data flow if the WSN application requires that. Various type of application can be implemented in the application depending on the physical environmental sensing.

Orthogonal to the five-layer Akyildiz et al. [13] defined three management plan named power, mobility and task management. These plans are responsible for monitoring the power, movement and task distribution among the sensor nodes. These management plans help the sensor nodes to coordinate sensor tasks and minimize the overall power consumption font.

4. BLOCKCHAIN STRUCTURE

Blockchain technology allow untrusting parties with common interests to exchange the information without relying on the external entities. Core characteristics of blockchain is decentralization, accountability, and security. This technique can improve operational efficiency and save costs of the network significantly. In order to achieve the decentralization blockchain, it uses the peer-to-peer (P2P) network architecture [18]. Early blockchain technology based on P2P networking has improved the decentralized network architecture. Figure 4 shows decentralized blockchain. it consists of three important concepts [19]:

1. Blocks,
2. Nodes and
3. Miners.

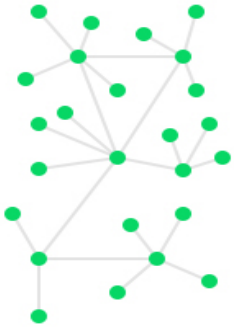

Figure 4. Decentralized Blockchain

Blocks: Each chain consists of several blocks and each block has information data, a nonce and the hash. Whenever the primary block of a chain is formed, a nonce then creates the cryptographic hash. The information data within the block is considered signed and for all time tied to the nonce and hash except it's far mined.

Nodes: Every node has its specific copy of the blockchain and the network have to algorithmically approve any newly excavated block for the chain to be up to date, trusted and verified. Since blockchains are transparent, so it is easy to check and verify each action and record within the ledger. Each node is given a unique number that shows their transactions.

Miners: Miners normally use specific program to solve the extremely complex mathematical problem of finding a nonce that generates hash which can be accepted.

5. DISCUSSIONS OF WSN SECURITY WITH BLOCKCHAIN

The security provisioning in WSN is a critical task. Several works has been done by researchers to secure the WSN. Recently, the focus was given to blockchain for the data security, management and storage of WSN. Moinet et al. proposed a blockchain based multi sensor technique [20]. This technique collects and verify the information data gathered by the sensor. Casado vara et al. proposed stochastic model of blockchain [21]. This can do the early prediction of the degeneration of the sensor accuracy from the current state of the sensor. Cui et al. proposed a hybrid blockchain model [22]. This method, realize the authentication from different communication scenarios. The previous studies mentioned have several constraints. Moreover, they did not consider the decentralized model of blockchain. A Comparison is drawn in the Table 1. To compare the WSN with the blockchain implementation and without the blockchain.

Table 1. WSN with and without Blockchain

Attributes	Without Blockchain	With Blockchain
Architecture	It is centralised with client-server architecture	Based on the distributed ledger architecture with decentralization
Power consumption	Low	High
Security	Low	High
Device Requirements	It requires limited processing capability with low storage capacity	It requires high processing capability with high storage capacity
Implementation	Simple	Difficult
Maintenance	Easy	Difficult

Considering the current architecture of WSN, sensor nodes are normally with the low capacity, low possessing power, limited storage, and limited battery capacity. Blockchain technology will bring a new arena of the WSN security as it has decentralised capacity and it can give better security. To implement Blockchain in WSN it requires high configuration sensors with high processing capability and high storage and high battery support. Moreover, WSN with the blockchain is difficult to implement and maintain. As a result, there are few challenges to implement blockchain for WSN security. Blockchain distributed character will be lost if the WSN need to be extended. So WSN will loss the scalability. Current sensor nodes do not have the high power and processing time. So, implementing Blockchain with current sensors will face difficulties with the processing speed and battery power. Blockchain requires to store transection and device ID. So, it requires high storage. Moreover, implement and maintain blockchain need highly skilled professionals.

It is difficult to implement blockchain in current resource constrain WSN, the technology requires huge resources. In order to incorporate blockchain in WSN and make it reality, we need to look forward for advancement of microelectronics to make the sensors resourceful. As the microelectronics is evolving very fast, so the blockchain soon will become the reality for WSN security . With the current infrastructure of WSN, blockchain is possible to implement but it will not be able to support the voice and video data because of storage and battery life of the sensor node.

6. CONCLUSION

Currently, the implementation of blockchain in WSN is challenging task as blockchain is still in early stage in implementing in engineering applications. The main challenge is the requirement of higher memory in each node in the network with blockchain as it does not use central server. The distributed the nature of blockchain require high memory and hardware in each node. However, it is believed that the current trend of advancement in memory technology will provide the high memory capacity with very small size in very near future. Some of the research is showing promising result which can fulfil this requirement. However, the distributed and secure architecture of blockchain will fulfil the challenges currently facing in WSN network. In future, we would like to implement blockchain for WSN security utilizing the text data transmission.

REFERENCES

[1] X. Huang, M. Ahmed, and D. Sharma, “Timing control for protecting from internal attacks in wireless sensor networks,” in 2012 International Conference on Information Networking (ICOIN), 2012, pp. 7 –12.

[2] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, “Detection, classification, and tracking of targets,” IEEE Signal Processing Magazine, vol. 19, no. 2, pp. 17 –29, Mar. 2002.

[3] C. Meesookho, S. Narayanan, and C. S. Raghavendra, “Collaborative classification applications in sensor networks,” in Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002, 2002, pp. 370 – 374.
[4] T. He, S. Krishnamurthy, J. A. Stankovic, T. Abdelzaher, L. Luo, R. Stoleru, T. Yan, L. Gu, J. Hui, and B. Krogh, “Energy-efficient surveillance system using wireless sensor networks,” in Proceedings of the 2nd international conference on Mobile systems, applications, and services, New York, NY, USA, 2004, pp. 270–283.
[5] B. Sinopoli, C. Sharp, L. Schenato, S. Schaffert, and S. S. Sastry, “Distributed control applications within sensor networks,” Proceedings of the IEEE, vol. 91, no. 8, pp. 1235 – 1246, Aug. 2003.
[6] P. Sikka, P. Corke, P. Valencia, C. Crossman, D. Swain, and G. Bishop-Hurley, “Wireless ad hoc sensor and actuator networks on the farm,” in The Fifth International Conference on Information Processing in Sensor Networks, 2006. IPSN 2006, 2006, pp. 492 –499.
[7] F. Zhao, “Wireless sensor networks: a new computing platform for tomorrow’s Internet,” in Proceedings of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication, 2004, May-2 June, vol. 1, pp. I–27 Vol.1.
[8] M. Ahmed, X. Huang, and D. Sharma, “A Taxonomy of Internal Attacks in Wireless Sensor Network,” in World Academy of Science, Engineering and Technology, Kuala Lumpur, Malaysia, 2012, pp. 427–430.
[9] M. Ahmed, X. Huang, and D. Sharma, “A Novel Framework for Abnormal Behaviour Identification and Detection for Wireless Sensor Networks,” International Journal of Computer and Communication Engineering, vol. 6, no. 2, pp. 148–151, 2012.
[10] Fran Casino, Thomas K. Dasaklis, Constantinos Patsakis, “A systematic literature review of blockchain-based applications: Current status, classification and open issues”, Telematics and Informatics, Volume 36, 2019, Pages 55-81, ISSN 0736-5853
[11] David Berdik, Safa Otoum, Nikolas Schmidt, Dylan Porter, Yaser Jararweh, “A Survey on Blockchain for Information Systems Management and Security, “Information Processing & Management, Volume 58, Issue 1, 2021, 102397, ISSN 0306-4573
[12] Ahmed, M. , Huang, X. , Sharma, D. , Cui, H. (2012), 'Wireless Sensor Network: Characteristics and Architectures', World Academy of Science, Engineering and Technology, Open Science Index 72, International Journal of Information and Communication Engineering, 6(12), 1398 - 1401.
[13] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E, Cayirci, “A Survey on Sensor Networks”, IEEE Communications Magazine , August 2002.
[14] H. Karl, A, Willing, “Protocols and Architectures for Wireless Sensor Networks”. New York: Wiley, 2005. 314−340, 2005.
[15] C. Buratti , A. Conti , D. Dardari and R, Verdone, “An Overview on Wireless Sensor Networks Technology and Evolution” , Sensors 2009, ISSN 1424-8220, pp 6869-6896, 2009.
[16] K. v. madhav , C ,rajendra and R. L. selvaraj, “A study of security challenges in wireless sensor networks”, Journal of Theoretical and Applied Information Technology, 2010.
[17] J. Feng, F. Koushanfar, and M. Potkonjak, ” Sensor Network Architecture”, supported by the national science foundation under Grant No. NI-0085773 and NSF CENS Grant, 2005.
[18] Z. Zheng, S. Xie, H. Dai, X. Chen and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," 2017 IEEE International Congress on Big Data (BigData Congress), 2017, pp. 557-564, doi: 10.1109/BigDataCongress.2017.85.
[19] Wu, J.; Tran, N.K. Application of Blockchain Technology in Sustainable Energy Systems: An Overview. Sustainability 2018, 10, 3067. https://doi.org/10.3390/su10093067.
[20] Moinet, A., Darties, B., & Baril, J. L. (2017). Blockchain based trust & authentication for decentralized sensor networks. arXiv preprint arXiv:1706.01730.
[21] Casado-Vara, R. (2018, June). Stochastic approach for prediction of WSN accuracy degradation with blockchain technology. In International Symposium on Distributed Computing and Artificial Intelligence (pp. 422-425). Springer, Cham.
[22] Cui, Z., Fei, X. U. E., Zhang, S., Cai, X., Cao, Y., Zhang, W., & Chen, J. (2020). A hybridBlockChain-based identity authentication scheme for multi-WSN. IEEE Transactions on Services Computing, 13(2), 241-251.

AN APPLICATION TO PROVIDE TRANSLATED SUBTITLES AND PICTURES FOR YOUTH ENGLISH LEARNERS USING SPEECH-TO-TEXT AND NLP TECHNIQUES

Harry Cao[1], Yu Sun[2] and Ariel Jiang[3]

[1]SuZou Industry Park Foreign Language School, Jiangsu, China
[2]California State Polytechnic University, CA, 91768, USA
[3]Cornell Tech, Pomona, New York, NY, USA

ABSTRACT

Currently, thousands of free K-12 educational videos exist online with the aim of trying to help young students learn outside of the typical scholastic environment. However, most of these videos are in English, so without subtitles it may be difficult for non-native English-speaking students to fully understand them. These students may need to spend time searching for translations and understanding content, which can distract them from grasping the important concepts within the videos. The state-ofthe- art of speech-to-text and NLP techniques might help this group digest the content of instructional videos more effectively. This paper proposes an application that uses speech-to-text, machine translation, and NLP techniques to generate translated subtitles and visual learning aids for viewers of instructional videos. This video application supports more than 20 languages. We applied our application to some popular online educational videos and conducted a qualitative evaluation of its approach and effectiveness. The results demonstrated that the application could successfully translate the English of the videos into the viewers' native language(s), detect keywords, and display relevant images to further facilitate contextual understanding.

KEYWORDS

Educational videos, mobile applications, language translation applications.

1. INTRODUCTION

Many instructional videos for language learning do not provide an ideal interface for second language learners, and this is particularly true when it comes to younger students. While at school, teachers often have students watch videos. Since these videos are often shown within a short amount of time, second language learners may not be able to learn new vocabulary words or phrases or understand content, placing them at a disadvantage. Now, students who do not understand all the information they are presented in class can watch online learning videos while at home, usually for free, through sources such as Khan Academy [1]. By using sources such as this, students can review instructional content they missed in class. However, since there are often still no translated subtitles available within these videos, second language learners may still have difficulty digesting the material. With our application, second language learner students can better understand these (video) course materials and their content, and even provide feedback to teachers.

David C. Wyld et al. (Eds): CCNET, NLTM, CICS - 2022
pp. 19-27, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120303

With the aid of artificial intelligence (AI), the technologies that turn the spoken word into written text (speech-to-text) are developing at a pace once thought unimaginable. Recently, Google's speech recognition model broke the 95% threshold for human accuracy, meaning that it has officially become better at recognizing spoken words than most humans [2]. Dating back to the 1950s, the first speech recognition systems were used for translating numbers into text [3]. In 1952, Bell Laboratories designed a system called "Audrey," which could recognize some spoken digits (the names of numbers) when using a single voice. Ten years later, "Shorebox" was able to understand 16 words in English. In the 1980s, the number of words that could be understood using AI went from hundreds to thousands. While the accuracy has been slowly improving, little progress was made until Google launched Google Voice Search. Voice Search was an important milestone for Google's ML-driven voice recognition model, which has so far collected over 230 billion actual voice recordings from users. For now, many of these technology components are shared through Google's publicly accessible library, "Speech-to- Text," e.g., Global vocabulary, Noise robustness, and Speech adaptation, all of which help make transcribing the spoken word possible [4]. These technologies are also applied to Google's popular video uploading site, YouTube, where viewers can turn on "CC subtitles" for most videos to get transcriptions displayed at the bottom. This library of technologies, alongside Google's Cloud Translation [5], serve as the foundation for our application's development.

YouTube's CC subtitles depend on Google's latest speech-to-text technologies to auto-generate English subtitles. However, this approach assumes viewers understand English, since it only provides English subtitles or sometimes, though not often, no subtitles at all. YouTube's implementation is also limited by the fact that videos only auto-generate subtitles within the same language used in the videos, with few other languages being supported. Other video transcription sites such as "Rev" [6] and "Scribie" [7] provide a sophisticated level of translation, but require user registration and membership fees that potentially make them less accessible to younger students. One could argue that it is reasonable to charge transcription fees since longer videos would require lots of computational resources [8]. However, most educational videos are short, so it may be beneficial to have a service that offers the transcription of shorter videos (of less than ten minutes) via upload for students only. In addition, other sites do not provide images to help with context. When students need to learn a new word, visual assistance has proven to be an efficient learning tool to understanding and memorization [9]. Science videos frequently use unfamiliar words or technical jargon that students find difficult to visualize. Even if students use a dictionary, this requires additional time, and definitions may still be still be too short or obscure to grasp the required context. If the transcription includes images, however, students don't have to go out of their way to search for meanings and can immediately get a mental picture of the concepts being explained. Image insertion technology is in high demand, and as a result is also very accessible. Currently, there are many video editing sites (e.g., "Kapwing" [10] and "Pixiko" [11]) as well as applications that allow users to insert images at a specified time during viewing. Unlike these sites, our application automatically searches for meaningful words and inserts them into videos to provide visual aids for students, quickly and conveniently.

Our application's main function is to take a short video as input and generate a streamlined final product consisting of three parts: a video transcription, a video translation, and key images for the video. First, we used speech-recognition technology to transcribe existing videos in English. Secondly, we translated the videos into users' native languages and embedded subtitles relevant to these languages. Then the videos were successfully translated into various languages of choice using speech-to-text and machine translation technologies. As previously mentioned, there are several existing services that may also produce video translations, but very few are free and focus exclusively on short, educational videos for students. Finally, we used NLP techniques to analyze the English transcripts and extract keywords and phrases based on a certain set of algorithms that are empirically proven to be the most accurate for this purpose.

After these keywords were found, we inserted related images back into the original videos. This will allow these images to appear briefly whenever the relevant word(s) are spoken to help students clarify meanings and generate mental images of whatever is being discussed. In order to evaluate the performance of our application, we believed it best to conduct experiments arising from two angles: (1) the accuracy of speech-to-text and translation done by our system and (2) user experience. First, we tested the accuracy of speech-to-text for our application. We expected the best metric for this to be numeric accuracy and the gold standard method to be human (speech). For a selection of educational short videos, we let native speakers watch and record what they heard, since humans are very accurate in terms of identifying the words they hear in their own language. We then fed the video into our application and compared our transcription with the human's, word for word, and recorded the accuracy of each. We also conducted similar experiments to test the accuracy of our translation and the combination of both. Since our application is made to serve young English learners, their progress in learning was an important metric of success for us. We invited primary school students who do not speak English well to watch a selection of educational videos in English that they would normally encounter in either a home or school environment. Afterwards, we put the same videos through our translation system and let students re-watch them. We then surveyed the students about their comparative experiences, specifically asking how our embedded images might have helped their understanding of unfamiliar words or contexts.

To outline the rest of this paper, Section 2 provides details on the challenges we met during the experiment and while designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges mentioned in Section 2; Section 4 presents details relevant to our experiment, followed by related work in Section 5. Finally, Section 6 provides concluding remarks, as well as possible directions for this project in the future.

2. CHALLENGES

In order to devise an application that uses speech-to-text, machine translation, and NLP techniques to generate translated subtitles and visual learning aids for viewers of instructional videos, a few challenges were identified as follows.

2.1. Challenge 1: Inferring Sentence Breaks within Transcribed Text

The first challenge we faced was how to yield an accurate translation of our English transcripts into our user's native language. Specifically, there is the difficulty of inferring the meaning of a given sentence, since text created from speech-to-text is oftentimes just a continuation of words instead of naturally segmented sentences. Inferring the breakpoint of sentences from a sequence of words without punctuation is a known difficulty surrounding NLP. First, one has to infer the correct number of breakpoints given a sequence of words. Second, one has to find the probability of a breakpoint being inserted between word w_i and word w_j for all i and j. There have been few studies done on this topic [12], mainly because it is rare for people to have to break up a sequence of words without punctuation or capitalization.

2.2. Challenge 2: Extraction of Keywords

The second challenge we faced was the extraction of keywords from the transcripts. This task had two goals: (1) extracting keywords relevant to the videos, and (2) extracting keywords related to the key concepts of the videos that students are likely encountering for the first time. The first goal can be achieved through established NLP techniques [13], but when combined with the second goal, the problem becomes more difficult to solve. For example, for a video providing a

biological explanation of how a human body works, although the word “human” might appear multiple times and would be a sensible guess as to the topic of the video, key scientific words like “heterozygous” or “oligosaccharide” that only appear once or twice would likely be neglected by traditional NLP methods of choosing keywords or key phrases.

2.3. Challenge 3: Selection of Images

The third challenge we faced was the selection of images. Because we intend for our application to insert images into videos as a fully automated process, the selection of images must also be automated. This means that we could only use keywords extracted from the videos themselves to conduct the selection and searching process. This process can be tricky as a simple word like “market” can have different meanings. Even if we know the implied meaning within the context is “economy,” it would still be impossible to identify this meaning through images and pick the correct one. To solve this challenge, more pre-processing needs to be done before the searching phase, e.g., adding the word “grocery” before “market” to emphasize one meaning over another.

3. SOLUTION

3.1. Overview of the Solution

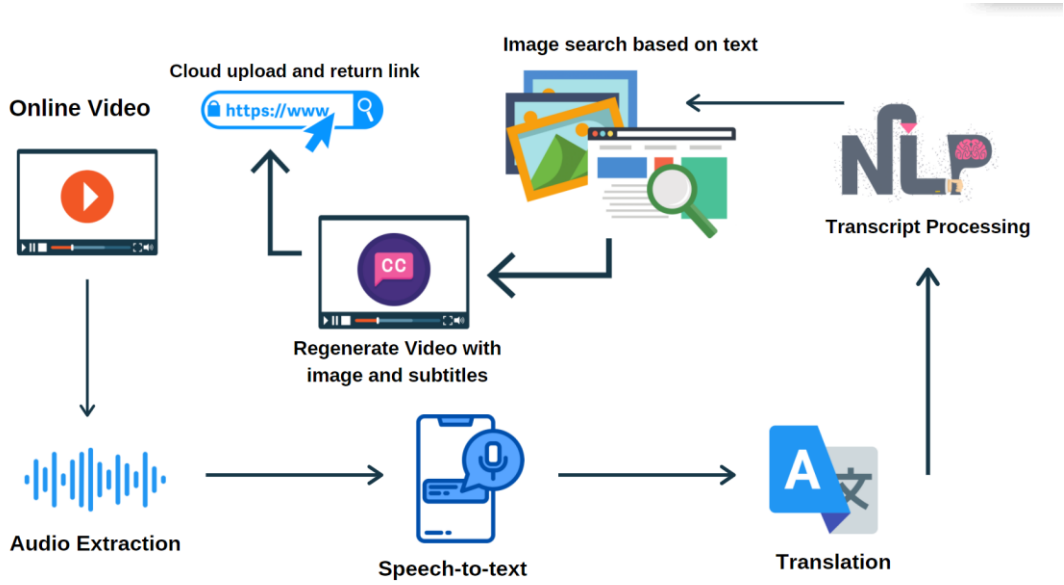

Figure 1. Overview

The application initializes itself and internally creates a Cloud Manager that controls the connection with several third-party APIs (firebase, speech-to-text, translation, image search). After the video is uploaded, it is first processed with audio extraction through efficient video editing tools such as ffmpeg. Then, the audio is uploaded to cloud storage through Cloud Manager to do speech-to-text transcription using Google’s library. During the transcription phase, three types of data are stored: (1) the transcript of the video in English, (2) the “srt” format of the transcript where “srt” is a human-readable file format that stores text sequentially, along with the timing information [14]. In our case, the text we store is the original transcript after getting translated into the user's native language using machine translation provided by a third-party tool. (3) We also store the timeline of the video during the processing phase. The timeline of the video is stored as a dictionary object in our Python implementation, where keys are words spoken

during the video and values are a list of timestamps during which such words appear. This dictionary becomes useful later on, when we need to insert images back into the video at the time(s) they appear. At this point, the transcription of the video is already complete. Next, the system analyzes the English transcripts using NLP technologies and extracts the keywords and phrases that are both representational of the video and add values to the user's knowledge base. We then search the web to find related images for these keywords and insert them back at the time they appear, using the timeline dictionary we generated earlier. After the video's subtitles and images are both embedded back into the original video, we upload the video to the cloud using Cloud Manager and return the cloud link to the user.

3.2. Component Breakdown

3.2.1. Speech-to-Text Transcription and Translation

These two components are the foundation of our application and are done sequentially in the same iteration using Google's Speech-to-Text and Translate APIs. After the API starts running and listening to the audio, it generates a streamline of results. For every ten words spoken in the stream, we first transcribe the original English words, then translate and record the translation in "srt" format. Meanwhile, the exact timeline of the video is also recorded. The code snippet shows the nested "for" loop where we process each word and store its timestamp to our dictionary for future processing, as mentioned in the previous section. Figure 1 shows where in the sequence the video is regenerated with subtitles and images using tools like ffmpeg to insert "srt."

3.2.2. Transcripts Analysis

For transcripts analysis, our goal is to extract the most relevant and valuable words that will help users both learn new concepts and expand their vocabulary or knowledge base. We perform basic text processing techniques such as stopwords removal, tokenization, and tagging sequences with part-of-speech tags all using Python's nltk library. We discovered that words with tags such as "NNP" and "NNPS" (proper nouns) tend to be more valuable and represent the video better, since they are less likely to be generic. Therefore, we applied more significance to these words. Also, we used a sublinear term frequency approach to normalize frequent words, meaning that instead of using the raw count of the terms, we instead used a logarithm of the count plus one. This helped us prevent words that are not stopwords but common within the video, such as "human," from taking too much significance. The number of keywords also depends on the video's length, since we output a number of keywords based on a 1:2 ratio, meaning one minute of a video equals 2 keyword outputs.

Figure 2. Video frame showing an embedded image

3.2.3. Image Selection and Insertion

After retrieving the important vocabularies from performing text processing, we use Google's search API to search for the most relevant images and download them. We then use our timeline dictionary to construct an inverted dictionary where each key is the time an image should appear on the screen and the values are the corresponding images. We experimented with various durations of time for images to remain on the screen and settled on three seconds as our final implementation. We also took into consideration the possibility of image overlap, and instituted a minimum one second gap between image displays. After these steps are completed, we insert the images back into the original video using openCV. Depending on the original video frame rate and dimensions, these images will appear with modified heights and widths such that they always only occupy at most . of the screen. After the images and subtitles are both embedded into the original video, we upload the video back to firebase storage and return the publicly accessible link to the user.

4. EXPERIMENT

4.1. Experiment 1

To evaluate the accuracy of speech-to-text, we used three scientific videos which do not have English subtitles provided and let our application do the transcriptions. Then we invited five teachers to write down the text spoken in the videos. We summarized and unified the data of several teachers, and regarded this as the correct transcript. We then compared this data with the computer speech-to-text transcriptions. We counted all the words that had appeared in the videos, and extracted the words that were different from the teachers' transcriptions and our application had transformed. We then counted the words that were different between the application and the teachers and divided this number by the total number of words to calculate the rate of error. One hundred minus the rate of error number is the accuracy rate of our application. With this margin of error, the application is able to effectively interpret the verbal context of the videos.

Table 1. Deviation rates for accuracy (application versus human)

Video name	Number of participants who viewed	Rate of deviation between application and participants
Crash Course	5	3.1%
Organic Chemistry Preview	5	1.7%
Psychological Research: Crash Course Psychology #2	5	5.2%
Biology: Cell Structure | Nucleus Medical Media	5	3.9%

4.2. Experiment 2

We showed three children who don't speak English a scientific video containing a lot of jargon and no subtitles. We chose videos that were less than three minutes long and only explained the basic concepts. We allowed the children to take notes while watching the video so that they might memorize the content. After this, we asked the children to give us a brief summary of the video,

which they were unable to do. We then played the same video using our application. After viewing, all three children were able to provide a brief summary of the video. When evaluating these summaries, we used three standards of ability: clarity, understanding of key concepts, and ability to reiterate. The range of marks given were between 0-5. For clarity, 0 meant that the child had no idea of the main topic, while 5 meant that they could explain the main topic well. For understanding of key concepts, 0 meant that the child could not explain the key concepts, while 5 meant that they could explain them well. For ability to reiterate, 0 meant that the child could not explain the video without notes, while 5 meant that they could explain the video well to others.

Table 2. Respective scoring for clarity, understanding of key concepts, and ability to reiterate

Name of video	Average scores before use application	Summary after use application
Photosynthesis: Crash Course Biology #8	● clarity: 1.2 ● understanding of key concept: 1.53 ● ability to reiterate: 0.24	● clarity: 4.76 ● understanding of key concept: 4.87 ● ability to reiterate: 3.98
Simple Harmonic Motion: Crash Course Physics #16	● clarity: 2.87 ● understanding of key concept: 0.77 ● ability to reiterate: 0.2	● clarity: 4.85 ● understanding of key concept: 4.67 ● ability to reiterate: 4.01

4.3. Analysis

The result of the first experiment shows that there is a high degree of overlap between what people hear and what our application processed, which shows that the speech-to-text feature of our application has an acceptable rate of error. In the second experiment, by contrasting the input from children using the application versus the control group, we can see that our application greatly improved the understanding of the video for non-native English speaking children.

5. RELATED WORK

The paper, “Review of Speech-to-Text Recognition Technology for enhancing learning” [15] introduced how speech-to-text technology applies to enhanced learning, especially for nonnative English speakers. Shadiev [16] has shown that non-native English speakers take advantage of nineteen strategies in using STR-generated transcripts. As a result, participants are able to use them to study, compose summaries, and generate their own ideas, which helps them receive higher scores.

The paper, “Review of studies on recognition technologies and their applications used to assist learning in instruction,” [17] reviews and summarizes studies on recognition technologies within the last ten years. Technologies using haptic input were employed mostly within science based education [18]. Students use this technology to take notes during lectures and enhance understanding when information is missed or misunderstood, or for completing homework assignments.

The paper, “Investigating the effectiveness of speech-to-text recognition applications on learning performance, attention, and meditation” [19] examines the effectiveness of speech-totext recognition technology as a learning and concentration aid to encourage a calm state of mind. In this experiment, students who do not have STR (speech-to-text recognition), when unable to

understand their lectures, feel stressed and lose concentration. Meanwhile, students who used STR felt less pressure while trying to learn, since they could properly read and understand texts.

Our goal is to make an application that helps children who don't understand English well to be able to understand and summarize videos presented in English. Speech-to-text technology can help children understand the content of instructional videos in English with less pressure so they can enhance their learning skills.

6. CONCLUSION AND FUTURE WORK

In this paper, we proposed an application that automatically transcribes online educational videos and provides contextual visual aids to help K-12 non-native English speakers understand the content better and faster. With the help of speech recognition technology, we built this application's foundation using advanced speech-to-text and machine translation technologies provided by Google. For visual aids, such as images, we effectively applied text processing techniques to first extract key information from the transcript(s), then embedded these images back into the videos to help students obtain a visualization of the video's important concepts. During the implementation phase, we met with a few challenges, including translation inaccuracy and the unpredictability of the image contents as mentioned in Section 2. We were able to eliminate some of the effects these issues had on our application by modifying the system to avoid them. In the end, our evaluation results indicate that the application is effective. In our first experiment, we observed that the error rate of our speech-to-text transcription is negligible and did not seem to affect people's ability to understand the content of the video(s). In our second experiment, we observed that non-English speaking children were better able to summarize videos when using our application versus a control group that viewed the videos without it. This is a good indication that our application does in fact help non-English speaking viewers better understand instructional videos presented in English.

Our application cannot guarantee complete accuracy, however, when generating subtitles for medium to longer-length videos. In addition, many English words have more than one meaning, so there will almost always be errors in translation. While our application also adds images to help understand the most important keywords of the videos, we couldn't predict what these images would be, since the application chooses the first image found by Google and inserts it into the video. While testing the application, some images would appear with strange words or pictures. Although this doesn't affect the aid provided by the images, overall, it doesn't provide a perfect watching experience.

We hope we can further adapt and evolve our application to reduce such errors. We would like our application to be able to differentiate which images are valid and which are not, e.g., which images best illustrate the context and subject matter of the videos. The accuracy of the machine translation could also be improved so that non-native English speakers can understand the videos even more clearly.

REFERENCES

[1] Free online courses, lessons & practice. (n.d.). Retrieved May 01, 2021, from https://www.khanacademy.org/

[2] A brief history of speech recognition. (n.d.). Retrieved from https://sonix.ai/history-of-speechrecognition

[3] The machines that learned to listen. (n.d.). Retrieved from https://www.bbc.com/future/article/20170214-the-machines-that-learned-to-listen

[4] Speech-to-Text: Automatic Speech Recognition. Google Cloud. (n.d.). Retrieved from https://cloud.google.com/speech-to-text
[5] Cloud Translation. Google Cloud. (n.d.). Retrieved from https://cloud.google.com/translate
[6] Rev Speech-to-Text Services: Convert Audio & Video to Text. (n.d.). Retrieved from http://www.rev.com/
[7] Corporation, C. (n.d.). Transcribe your audio/video files rates starting at $0.10/min. Retrieved from https://scribie.com/
[8] Goddard, W. (2021, March 17). Speech Recognition Algorithm - Brought to you by ITChronicles. Retrieved from https://itchronicles.com/speech-to-text/speech-recognition algorithm/
[9] Vavra, Karen L., et al. "Visualization in science education." *Alberta Science Education Journal* 41.1 (2011): 22-30.
[10] Make Something Awesome. (n.d.). Retrieved from https://www.kapwing.com/
[11] Online video editor: Improve your videos in a few clicks. (n.d.). Retrieved from https://pixiko.com/
[12] Barrows Jr, Randolph C., M. Busuioc, and Carol Friedman. "Limited parsing of notational text visit notes: ad-hoc vs. NLP approaches." *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2000.
[13] Aone, Chinatsu, et al. "A scalable summarization system using robust NLP." *Intelligent Scalable Text Summarization*. 1997.
[14] What is an SRT File? How to Create & Use SRT Files. (2021, March 09). Retrieved from https://www.rev.com/blog/resources/what-is-an-srt-file-format-create-use-srt-files
[15] Rustam Shadiev, et al. “Review of Speech-to-Text Recognition Technology for Enhancing Learning.” *Journal of Educational Technology & Society*, vol. 17, no. 4, 2014, pp. 65–84. *JSTOR*, www.jstor.org/stable/jeductechsoci.17.4.65. Accessed 1 May 2021.
[16] Shadiev, R., Hwang, W. Y., & Huang, Y. M. (in press). Investigating applications of speech to text recognition for face to face seminar to assist learning of non-native English participants. Technology, Pedagogy and Education.
[17] Shadiev, Rustam, et al. “Review of Studies on Recognition Technologies and Their Applications Used to Assist Learning and Instruction.” *Educational Technology & Society*, vol. 23, no. 4, 2020, pp. 59–74. *JSTOR*, www.jstor.org/stable/26981744. Accessed 1 May 2021.
[18] Neri, L., Noguez, J., Robledo-Rella, V., Escobar-Castillejos, D., & Gonzalez-Nucamendi, A. (2018). Teaching of classical mechanics concepts using visuo-haptic simulators. Educational Technology & Society, 21(2), 85–97.
[19] Shadiev, Rustam, et al. “Investigating the Effectiveness of Speech-to-Text Recognition Applications on Learning Performance, Attention, and Meditation.” *Educational Technology Research and Development*, vol. 65, no. 5, 2017, pp. 1239–1261. *JSTOR*, www.jstor.org/stable/45018724. Accessed 1 May 2021.

W&G-BERT: A CONCEPT FOR A PRE-TRAINED AUTOMOTIVE WARRANTY AND GOODWILL LANGUAGE REPRESENTATION MODEL FOR WARRANTY AND GOODWILL TEXT MINING

Lukas Jonathan Weber[1], Alice Kirchheim[2], Axel Zimmermann[3]

[1]Mercedes-Benz AG, Stuttgart, Germany
[2]Department Mechanical Engineering,
Helmut-Schmidt-University, Hamburg, Germany
[3]esz-partner Eber, Schwarzer, Zimmermann GbR, Kirchheim,
Baden-Württemberg, Germany

ABSTRACT

The request for precise text mining applications to extract information of company based automotive warranty and goodwill (W&G) data is steadily increasing. The progress of the analytical competence of text mining methods for information extraction is among others based on the developments and insights of deep learning techniques applied in natural language processing (NLP). Directly applying NLP based architectures to automotive W&G text mining would wage to a significant performance loss due to different word distributions of general domain and W&G specific corpora. Therefore, labelled W&G training datasets are necessary to transform a general-domain language model in a specific-domain one to increase the performance in W&G text mining tasks.

The article describes a concept for adapting the generally pre-trained language model BERT with the popular two-stage language model training approach in the automotive W&G context. We plan to use the common metrics recall, precision and F1-score for performance evaluation.

KEYWORDS

Natural language processing, Domain-specific language models, BERT, Labelled domain-specific datasets, Automotive warranty and goodwill.

1. INTRODUCTION

In the automotive W&G sector, an unprecedented number of written feedback texts are generated by global workshops and customer studies every day, making the manual analysis of these texts and the extraction of actionable insights from them an extremely difficult task for individuals. Various of such customer feedbacks are used to analyze and predict current and future product weaknesses. Both for the resulting failure elimination and for internal reporting, the mining of customer feedback plays a crucial role in product warranty. A large part of the written feedbacks are “not in a [directly] machine-processable [...] data [form]" available, which can be made usable for decision-making by text mining approaches [2]. Due to this, the demand for accurate text mining tools to extract information of company based W&G data is steadily increasing. The

David C. Wyld et al. (Eds): CCNET, NLTM, CICS - 2022
pp. 29-35, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120304

analytical competence of text mining methods for information extraction is among others based on the developments and insights of deep learning techniques applied in *natural language processing* (NLP) [3, 4]. Current deep-learning language models are trained on general text corpora (e.g. Wikipedia). Especially in domains such as medicine [5], biology [4, 6], finance [7], common science [8] and W&G, word distributions of general and specific corpora are different, which can be a significant performance problem for NLP-based text mining models [4]. Therefore, the application of generally trained language models to W&G tasks is only suitable to a limited extent [9]. Labelled W&G training datasets are necessary to transform a general-domain language model in a specific-domain one to increase the performance in W&G text mining tasks. To the best of our knowledge, there are not any existing labelled training datasets in the W&G domain public available.

In this paper, we describe the concept of creating a pre- and fine-tuned W&G-specific language model *BERT* [1]. Therefore, in Section 2 we briefly describe the already published articles in the area of language models with a deep dive in their domain specific adaptations. In section 3, we describe the creation of suitable training datasets for the following pre-training and fine-tuning activities of a W&G-specific context-dependent language model. Finally, in section 4 we give an outlook on the overall text mining architecture in which a W&G specific language model can be integrated.

2. RELATED WORK

Silvestri *et al.* [10] describe in their publication the significant performance improvement of NLP tasks through the use of the word embedding language models *word2vec* and *GloVe*. The basis for the *word2vec* language model is a continuous skip-grams (SG) or a continuous Bag-of-Words (CBOW) architecture. The SG architecture predicts the context of an entity based on the entity, whereas the CBOW approach predicts a missing entity based on the immediate context [11, 12]. In contrast to Mikolov *et al.* [12], Jeffrey Pennington *et al.* [13] developed *GloVe* as a global log-bilinear regression model for unsupervised word representation learning that outperforms SG- and CBOW- models on word analogy, word similarity and named-entity recognition tasks. Onan [14] increases the potential of the *word2vec* language model in the topic extraction domain by developing a two-stage architecture using an improved word embedding model and a cluster ensemble framework. The improved word embedding model consists of the fusion of conventional word embedding architectures (*word2vec*, *pos2vec*, *word-position2vec* and *LDA2vec*). In addition, the well-known clustering algorithms *K-means* [15], *k-modes* [16], *k++* [17], *self-organising maps* [18] and *DIANA* [19] are combined specifically unweighted into a cluster ensemble framework. In the *word2vec* and *GloVe* language models, emergent entities in a different context are characterised with the same word vector (context independent language models). Since the same entities take on a different meaning in a different context, the results of the already discussed language models can be improved. The context-based consideration of entities is taken into account in the model *Bidirectional Encoder Representations from Transformers* (BERT) [1]. The pre-training of the language model *BERT* is realised with the help of the static *Masked Language Model* (MLM) [20] and the *Next-Sentence-Prediction* (NSP). In contrast to Devlin *et al.* [1], Liu *et al.* [21] chose a dynamic masking approach in their robustness-optimised language model *RoBERTa* without implementing the NSP task. *BERT* by Devlin *et al.* [1] provides the basis for multi-layered applications in the fields of medicine, biology and finance. The simple adaptation of the *word2vec*, *ELMo* [22] or *BERT* language models would yield weak results in domain specific *Named Entity Recognition* (NER) and *Relation Extraction* (RE) metrics in domain-specific application case, as they were pre-trained on general input data. The authors Lee *et al.* [4] developed the *BioBERT* language model which was pre-trained on bio-medical data. The superiority of a specifically pre-trained *xBERT* language model over conventional language models was also confirmed by Peng *et al.* [23]. Furthermore,

Beltagy *et al.* [8] present the potential of their pre-trained *SciBERT* language model in their publication. By implementing a sentence-piece library based on unsupervised tokenisation of the scientific corpus using WordPiece, this generates a performance advantage in the F1 score compared to the *BioBERT*. In addition, Zhuang Liu *et al.* [7] designed the *FinBERT* language model. *FinBERT* is characterised by a parallelised pre-training approach modelled on the Horovod architecture [24] using a mixed-precision training approach [25] on general and finance-specific corpora. The language model developed by Zhuang Liu *et al.* shows not only an improved performance compared to *BERT*, but also considerable success in the common quality validation metrics for a small pre-training corpus (20% of the original corpus). In addition to the context-dependent architecture *BERT* and the specific language models derived from it by Lee *et al.*, Peng *et al.*, Beltagy *et al.* and Zhuang Liu *et al.*, the language model *ERNIE 1.0* developed by Sun *et al.* [26] is based on a masked language models approach. However, Sun *et al.* differentiates itself from the general *BERT* approach by extending the MLM approach in terms of basic-level, phrase-level and entity-level masking. Furthermore, *ERNIE 1.0* differs from *BERT* by a five-stage pre-training phase and the application of the Dialogue Language Model (DLM) to improve the learning ability of semantic representations. In contrast to *ERNIE 1.0*, its extension *ERNIE 2.0* [27] is based on a continuous multi-task learning pre-training architecture. Here, lexical, syntactic and semantic information are learned by using several word-aware, structure-aware and semantic-aware tasks. This allows Sun *et al.* [27] to ensure that the learned parameters encode the previously learned knowledge. The evaluation of the continuous multi-task training method using defined training tasks which demonstrates the potential of this approach. Furthermore, the *ERNIE 2.0* language model represents a considerable increase in performance in the common metrics of the GLUE benchmark compared to *XLNet* [28] and *ERNIE 1.0*.

3. APPROACH

As shown in Figure 1, we develop our W&G-BERT model based on the architecture of *BERT*, which was elaborated as one of the state-of-the-art language representation models in section 2. Furthermore, we use the popular two-stage language model training approach, which consists of a pre-training and fine-tuning phase to increase the performance of its application on domain specific tasks. Subsequently in this section, we describe the proposed *BERT* model and the planned pre-training and fine-tuning activities of W&G-BERT.

3.1. BERT: Bidirectional Encoder Representations from Transformers

The process of learning word representations in an unsupervised way from a great amount of unannotated semi and unstructured data is a well-known and long-established method. Developed models in the past such as *word2vec* and *GloVe* are focused on learning context independent word representations [4]. To improve the performance of language models, more recent ones like *ELMo*, *XLM* [30], *XLNet*, *ERNIE1.0* and *ERNIE 2.0* are focused on learning context dependent word representations.

BERT is a context dependent word representation language model which is based on a masked language model and a next sentence prediction pre-training architecture by using bidirectional transformer encoder [31]. *BERT* uses the MLM approach by predicting randomly masked tokens in given sentences, while the NSP approach guarantees the sentence-relationship understanding between different sentences in the same corpus. This allows Devlin *et al.* to ensure that the inputs are represented efficiently as the sum of token-, segmentation- and position embeddings.

3.2. Pre-Training W&G-BERT

Authors like Alsentzer *et al.*, Beltagy *et al.*, Gu *et al.*[6], Lee *et al.*, Peng *et al.* and Zhuang Liu *et al.* provided in their results that general pre-trained models usually achieve poor results in common performance metrics on domain specific text mining tasks. Similar to biomedical or scientific texts, automotive W&G texts contain of a huge amount of domain specific terms and expressions, which requires expert knowledge of the corresponding researcher (e.g. *condensation water drain hose*, *center fill* or *open pore wood trim*). In the planned work, we pre-train W&G-BERT on English automotive specific *customer surveys* and *workshop complaints* in an unsupervised way with the MLM-approach in addition to the already pre-trained general corpora and initially weighted *BERT* model (see Table 1).

Table 1. Pre-training corpora

Domain	**Corpus type**	**English (approx. tokens)**
General[1]	Wikipedia	2500M
	BookCorpus	800M
Auto. warranty and goodwill	Customer survey	~50M
	Workshop complaints	~200M

3.3. Fine-Tuning W&G-BERT

With manageable architectural modification, W&G-BERT can be successfully applied to several text mining tasks. Therefore, we plan to fine-tune W&G-BERT on two self-constructed supervised automotive W&G specific datasets and their related text mining tasks: *NER* and *RE*.

W&G Named entity recognition is one of the most important W&G text mining tasks, which deals with the identification of domain-specific automotive W&G expressions and terms. *BERT* and its specifications [4–8] are build up on a multi-layer transformer encoder. Therefore, the simple specification process will be adapted to generate a *BERT* based W&G-BERT-language model. To the best of our knowledge, no specific fine-tuning datasets exist in the NER area for the English automotive W&G sector. Therefore, it is up to us to create a suitable annotated dataset, which is in the range of already annotated domain specific datasets [4, 7]. For performance evaluation, we will use the common metrics *recall*, *precision* and *F1-score*.

W&G Relation Extraction is a task to predict and classify the relationship between named entities in a corpus. Similar to the named entity recognition task, there are no suitable fine-tuning datasets available. Therefore, we will annotate on ourselves English automotive W&G datasets to fine-tune our language model in the specific context. Furthermore, we are planning to anonymize target entities in a sentence by using pre-defined tags like "@failure location#" or "@failure type#" [4]. For example, an annotated sentence with two target entities is represented by *"The customer claims that the @failure location# has @failure type#."* Similar to the utilisation of the planned performance metrics for the NER task, we will use the same performance metrics for the RE task.

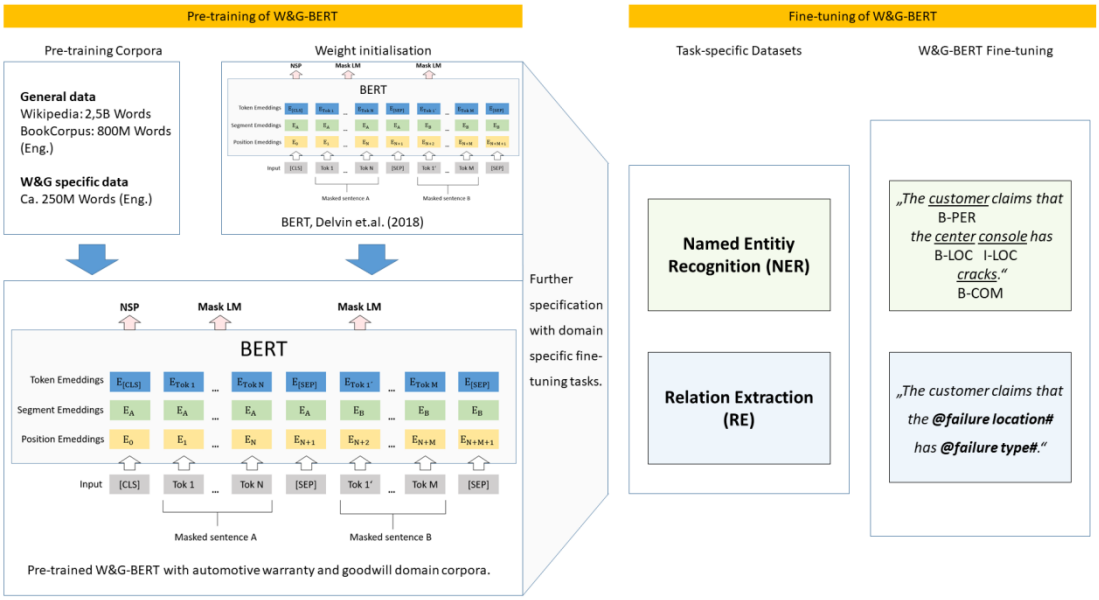

Figure 1. Pre-training and fine-tuning procedure of W&G-BERT

4. OUTLOOK

Based on the planned scientific contributions in the field of language models (W&G-BERT) and in agnostic sensitivity-based explanatory models (xLIME survey), we design a holistic approach to identify similar unstructured and semi-structured datasets in large domain specific corpora (see Figure 2). Therefore, we plan to map given data sets into vector representations using the context-dependent domain-specific language model W&G-BERT. After vectorization, the data will be presented in a high-dimensional numeric format. Due to this, the implementation of a dimension reduction algorithm may be necessary (this will be evaluated during the overall implementation). With the help of the context-based vector representations, semantic similarity values are to be calculated with the help of the cosine distance. Based on these, a state-of-the-art community detection algorithm can be selected and used to identify homogeneous network patterns in a heterogeneous data environment. After detecting communities in the network, we use the automatically labelled data set in means of the detected communities and the manually annotated clusters. Finally, a classifier will be trained with the help of the (un-) supervised clusters to recognise the identified clusters in future datasets. To ensure the acceptability with regulators and the general need for human understandable Blackbox predictions in sensitive areas, we plan to apply the best performing LIME-based explainability method to the domain-specific language model based classification model.

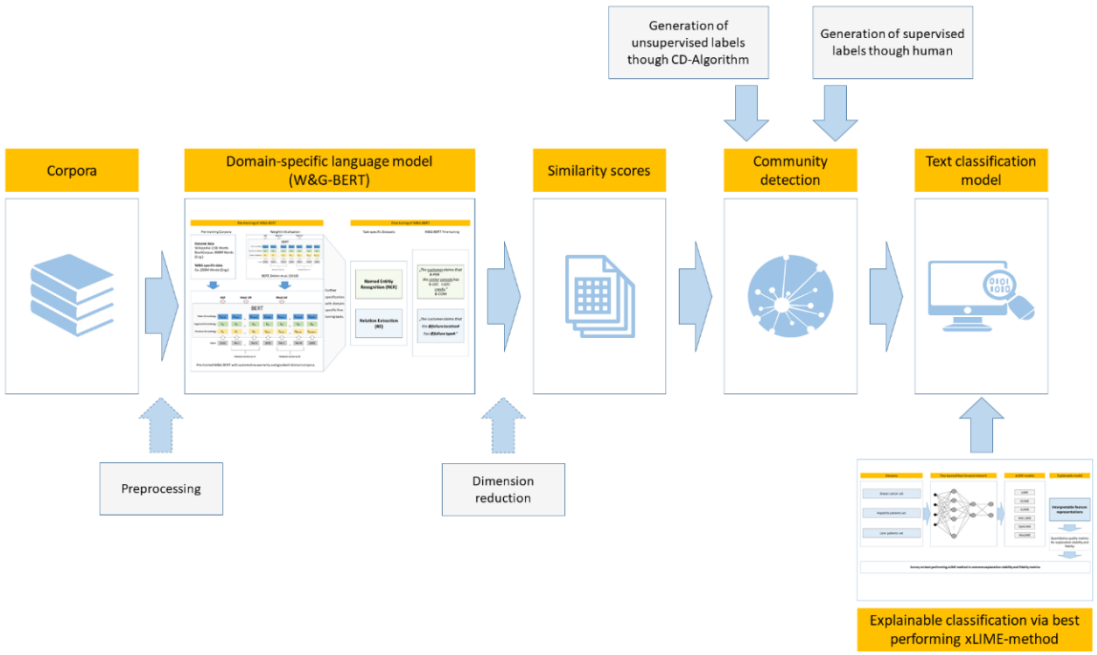

Figure 2. Overall approach with contributions in domain specific language model and performance based xLIME survey

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Oct. 2018. [Online]. Available: http://arxiv.org/pdf/1810.04805v2

[2] C. Felden, "Extraktion, Qualitätssicherung und Klassifikation unstrukturierter Daten," HMD Praxis der Wirtschaftsinformatik, no. 247, pp. 54–62, 2006.

[3] M. Allahyari et al., "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques," Jul. 2017. [Online]. Available: http://arxiv.org/pdf/1707.02919v2

[4] J. Lee et al., "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," Bioinformatics (Oxford, England), vol. 36, no. 4, pp. 1234–1240, 2020, doi: 10.1093/bioinformatics/btz682.

[5] E. Alsentzer et al., "Publicly Available Clinical BERT Embeddings," 2019. [Online]. Available: https://arxiv.org/pdf/1904.03323.pdf

[6] Y. Gu et al., "Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing," 2021. [Online]. Available: https://arxiv.org/pdf/2007.15779.pdf

[7] Zhuang Liu, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao, "FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining," 2020. [Online]. Available: https://www.ijcai.org/proceedings/2020/0622.pdf

[8] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A Pretrained Language Model for Scientific Text," Mar. 2019. [Online]. Available: http://arxiv.org/pdf/1903.10676v3

[9] R. Zhu, X. Tu, and J. X. Huang, "Utilizing BERT for biomedical and clinical text mining," in Data Analytics in Biomedical Engineering and Healthcare: Elsevier, 2021, pp. 73–103. [Online]. Available: https://doi.org/10.1016/B978-0-12-819314-3.00005-7

[10] S. Silvestri, F. Gargiulo, and M. Ciampi, "Improving Biomedical Information Extraction with Word Embeddings Trained on Closed-Domain Corpora," in 2019 IEEE Symposium on Computers and Communications (ISCC), Barcelona, Spain, Jun. 2019 - Jul. 2019, pp. 1129–1134.

[11] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Oct. 2013. [Online]. Available: http://arxiv.org/pdf/1310.4546v1

[12] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” Jan. 2013. [Online]. Available: http://arxiv.org/pdf/1301.3781v3
[13] Jeffrey Pennington, Richard Socher, and Christopher Manning, “Glove: Global Vectors for Word Representation,” Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543, 2014. [Online]. Available: https://nlp.stanford.edu/pubs/glove.pdf
[14] A. Onan, “Two-Stage Topic Extraction Model for Bibliometric Data Analysis Based on Word Embeddings and Clustering,” IEEE Access, vol. 7, pp. 145614–145633, 2019, doi: 10.1109/ACCESS.2019.2945911.
[15] A. K. Jain, “Data clustering: 50 years beyond K-means,” Pattern Recognition Letters, vol. 31, no. 8, pp. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.
[16] Z. Huang, Ed., A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining, 1997. [Online]. Available: http://citeseer.ist.psu.edu/viewdoc/download;jsessionid=08889727BE3ABD82AEC0A6888F74EED4?doi=10.1.1.6.4718&rep=rep1&type=pdf
[17] D. Arthur and S. Vassilvitskii, “k-means++: The Advantages of Careful Seeding: Arthur, David & Vassilvitskii, Sergei. (2007). K-Means++: The Advantages of Careful Seeding. Proc. of the Annu. ACM-SIAM Symp. on Discrete Algorithms. 8. 1027-1035. 10.1145/1283383.1283494.,” no. 8, pp. 1027–1035, 2007. [Online]. Available: https://dl.acm.org/doi/10.5555/1283383.1283494
[18] T. Kohonen, “The self-organizing map,” no. 21, pp. 1–6, 1990. [Online]. Available: https://doi.org/10.1016/S0925-2312(98)00030-7
[19] H. Chipman and R. Tibshirani, “Hybrid hierarchical clustering with applications to microarray data,” Biostatistics (Oxford, England), vol. 7, no. 2, pp. 286–301, 2006, doi: 10.1093/biostatistics/kxj007.
[20] W. L. Taylor, ““Cloze Procedure”: A New Tool for Measuring Readability,” Journalism & Mass Communication Quarterly, no. 30, pp. 415–433, 1953, doi: 10.1177/107769905303000401.
[21] Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” Jul. 2019. [Online]. Available: http://arxiv.org/pdf/1907.11692v1
[22] M. E. Peters et al., “Deep contextualized word representations,” Feb. 2018. [Online]. Available: http://arxiv.org/pdf/1802.05365v2
[23] Y. Peng, S. Yan, and Z. Lu, “Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets,” Jun. 2019. [Online]. Available: http://arxiv.org/pdf/1906.05474v2
[24] A. Sergeev and M. Del Balso, “Horovod: fast and easy distributed deep learning in TensorFlow,” Feb. 2018. [Online]. Available: http://arxiv.org/pdf/1802.05799v3
[25] P. Micikevicius et al., “Mixed Precision Training,” Oct. 2017. [Online]. Available: http://arxiv.org/pdf/1710.03740v3
[26] Y. Sun et al., “ERNIE: Enhanced Representation through Knowledge Integration,” 2019a. [Online]. Available: http://arxiv.org/pdf/1904.09223v1
[27] Y. Sun et al., “ERNIE 2.0: A Continual Pre-training Framework for Language Understanding,” 2019b. [Online]. Available: https://doi.org/10.1609/aaai.v34i05.6428
[28] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. Le V, “XLNet: Generalized Autoregressive Pretraining for Language Understanding,” Jun. 2019. [Online]. Available: http://arxiv.org/pdf/1906.08237v2
[29] Y. Wang, Y. Sun, Z. Ma, L. Gao, Y. Xu, and T. Sun, “Application of Pre-training Models in Named Entity Recognition,” Feb. 2020. [Online]. Available: http://arxiv.org/pdf/2002.08902v1
[30] G. Lample and A. Conneau, “Cross-lingual Language Model Pretraining,” 2019. [Online]. Available: https://arxiv.org/abs/1901.07291
[31] A. Vaswani et al., “Attention is All you Need,” 2017. [Online]. Available: https://arxiv.org/abs/1706.

INVESTIGATION OF OPTIMIZATION TECHNIQUES ON THE ELEVATOR DISPATCHING PROBLEM

Shaher Ahmed, Mohamed Shekha,
Suhaila Skran and Abdelrahman Bassyouny

Department of Mechatronics Engineering, Faculty of Engineering and Materials Science, The German University in Cairo, Egypt

ABSTRACT

In the elevator industry, reducing passenger journey time in an elevator system is a major aim. The key obstacle to optimising elevator dispatching is the unpredictable traffic flow of passengers. To address this difficulty, two main features must be optimised: waiting time and journey time. To address the problem in real time, several strategies are employed, including Simulated Annealing (SA), Genetic Algorithm (GA), Particle Swarm Optimization Algorithm (PSO), and Whale Optimization Algorithm (WOA). This research article compares the algorithms discussed above. To investigate the functioning of the algorithms for visualisation and insight, a case study was created. In order to discover the optimum algorithm for the elevator dispatching problem, performance indices such as average and ideal fitness value are generated in 5 runs to compare the outcomes of the methods. The goal of this study is to compute a dispatching scheme, which is the result of the algorithms, in order to lower the average trip time for all passengers. This study builds on previous studies that recommended ways to reduce waiting time. The proposed technique reduces average wait time, improves lift efficiency, and improves customer experience.

KEYWORDS

Stochastic Optimization, Elevator Dispatching Systems, Meta-Heuristics Optimization Techniques.

1. INTRODUCTION

Elevators are designed to transport passengers safely, comfortably, and efficiently from their initial floors to their destinations. The use of elevators has been increasing throughout the years, as more construction is happening. An important concern for elevator producers is the importance of elevator efficiency, the reduction of passenger waiting time, and reducing the time to the destination. People prefer to use elevators rather than stairs to reach the upper levels of a building. People experience various challenges with the traditional elevator management system with just one operating lift, such as long wait times for the lift and long trip times since several passengers share a single lift and installing another lift might be highly expensive.

The Elevator Dispatching Problem (EDP) has become a popular research topic in both academia and industry as it is an excellent example of a stochastic optimal control problem of economic importance that is too large to solve using traditional techniques such as dynamic programming. Several studies use optimization techniques to improve the performance of the controller or elevator. The research gap of this research is the need for a comparative study to assess the

David C. Wyld et al. (Eds): CCNET, NLTM, CICS - 2022
pp. 37-49, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120305

optimality and efficiency of the elevator dispatching system. Several studies were conducted on the EDP offering multiple solutions to this problem, such as the use of algorithms and dispatching scheduling, the use of multiple elevators, and finally, the assistance of optimization techniques. These solutions were introduced for the purpose of optimizing the journey time, to achieve higher efficiency of lift transportation.

This paper gives a comprehensive literature review of several strategies used to improve the efficiency of the Elevator Dispatching Problem. Optimization Techniques algorithms are some of the strategies used to optimise elevator operation.

The rest of this paper is organized as follows: in Section 2, we review appropriate literature on stochastic and dynamic vehicle routing as well as on elevator dispatching. We formulate the EDP in Section 3. Section 4 describes and presents the algorithms that were utilised to solve the EDP. Section 5 shows the output of the algorithms which is the elevator routes to be taken. Finally, in Section 6, we finalise this study and outline future research options.

2. LITERATURE REVIEW

M. Latif, M. Kheshaim, and S. Kundu proposed EDP algorithms such as Estimated Time to Dispatch (ETD), Destination Control System (DCS), and Compass Plus Destination System (CDS) in 2016. The ETD algorithm uses destination-based dispatching system, which asks users to register their destination floor, thus assigning them to an elevator car. The DCS algorithm groups users according to their registered destinations and integrates artificial intelligence, travel forecasting, fuzzy logic and genetic optimization and continuously monitors the traffic behaviour at all times to avoid malfunctions. The CDS algorithm introduces the conventional smart grouping concept. Passengers with similar destination is assigned the same elevator car, and elevators are assigned to specific floors, leading to fewer stops being made thus reducing time to reach the destination. [1]

M. Sale and V. Prakash mentioned in 2019 that the optimization strategies of Elevator Dispatching Systems may be classified as fuzzy based algorithms, genetic algorithms, neural network algorithms, artificial intelligence algorithms, and zoning algorithms. The Fuzzy Algorithm reduces waiting time and improves the service quality and performance however it is a robust method. The memory may face faults which affects the performance of fuzzy controllers. The Artificial intelligence, Neural Network have a very optimal dispatching policy but needs expert experiences and training efforts to achieve good results and are difficult to understand and implement. Zoning algorithms have reduced energy consumption and have an optimal dispatching policy but it becomes robust and less flexible in heavy traffic. [2]

In a paper conducted by Sorsa, J., et al. that studied the EDP by solving a snapshot optimization problem. Passengers register their destination floors in elevator lobbies, after which the elevator group control system (EGCS) completes the assignment immediately and cannot be changed afterwards. A Poisson and a geometric Poisson process are used to describe stochastic requests and consumers. This leads to improved forecasting accuracy than the Poisson process and several scenarios that closely match the variable realisations. As a result, the scenarios may be utilised as the foundation for a robust EDP that minimises both a passenger service quality criterion and its volatility owing to variable demand. [3]

A. So and L. Al-Sharif, considered the Idealized Optimal Benchmark concept used in various engineering projects like in thermodynamics to produce the most efficient heat engine possible, so this concept can be used in the traffic elevator design by calculating the value of the round-trip time which is the time taken by the elevator to complete the task of transporting and picking up

the passenger to the desired destination to produce the maximum performance of an elevator system. Number of stops and highest reversal floor are important parameters that should be at the minimum possible values and these values can be done by specific formulae. The number of the elevators are sequenced to the virtual sectors in the building to equalize the handling capacities of the sectors. [4]

An EGCS is one of the important matters in vertical transportation systems in buildings, it is required to tackle some optimization problems. It dispatches the elevators under its control to perform each passenger's call. While the elevators are fulfilling their present calls, more passengers may emerge. This makes the control problem dynamic. Its solution defines complete elevator route by minimizing passengers' waiting or journey times. Janne Sorsa [5] considered the EDP as a bi-level, dynamical, stochastic, multi-objective optimization problem. Since the EDP is a difficult optimization problem, its solving time using an exact algorithm increases exponentially with respect to the problem's size. The main effective solution approach shall be based on a genetic algorithm suitable for real-time optimization. Tartan et al. [6] used a Genetic Algorithm that has attracted researchers to study the suggested optimization of waiting time. The suggested solution minimises average passenger waiting time and employs a simpler encoding methodology, resulting in computational cost efficiency.

In 2018, A. Vodopija, J. Stork, T. Bartz-Beielstein, and B. Filipic discussed the S-Ring based elevator and its operation on the principle of a normal ring road. It is usually used in the elevator group control. The system works on the principle of neural network, at each state it is checked whether a new customer arrived. The controller then decides whether the elevator car should stop or continue to the next state if the current state is an active elevator state. Finally, the active state indicates whether the customer has been served or not. [7]

In addition, in 2019, Bapin, Y., and Zarikas, V. created an algorithm that employs an image collecting and processing system to make use of information regarding passenger group sizes and waiting times. The study aimed for the construction of a decision engine capable to control the elevators actions in order that improves user's satisfaction. The data is utilised by the probabilistic decision-making model to perform Bayesian inference and update variable parameters. The results demonstrated that the suggested algorithm demonstrated the intended behaviour in 94% of the circumstances examined. [8]

Z. Yang and W. Yue investigated Elevator Traffic Pattern Recognition Using Fuzzy BP Neural Network with Self-Organizing Map (SOM) Algorithm in 2017. SOM is a form of unsupervised learning network method that operates on the premise of grouping comparable inputs on the same output in order to find a better clustering centre after a sufficient number of repetitions. Alongside with the fuzzy logic theory that provides a working algorithm to solve the control system. With the help of the neural network technique that features a stronger self-learning ability which is helpful when dealing with a nonlinear and uncertain dynamics and changes of the elevator traffic. When placing both theories together, each one covers for the drawbacks of the other to obtain the best results. [9]

M. Sale and V. Prakash discussed the design of an elevator system has many factors that should be took in consideration like managing time and money. EGCS is responsible for making several efficient elevators in carrying passengers. EGCS is a solution by reducing the waiting time during the peak time where all passengers need to use the elevator, this system uses fuzzy algorithm. Fuzzy system is a problem-solving rule based. It mainly consists of four blocks: fuzzy rule for applying on data, fuzzifier to accept input values, DeFuzzifier for generating output and inference engine for operations. There are two modules in the system: Hall call module which is

responsible for sending the calling floor id to the controller and car call module which is responsible for sending the destination floor id to the controller. [10]

3. PROBLEM STATEMENT AND FORMULATION

The main objective of this problem is to provide passengers with the shortest waiting and destination time. Thus, this can be considered as multi-objective optimization problem. Both parameters can be combined into one parameter; journey time, thus, the objective is to minimise the average journey time for all passengers. The model should have some parameters that should be specified. These parameters are then used to evaluate the two main equations for the waiting and destination time.

Table 1. Table of Fixed parameters of the elevator

Fixed Parameters	Time
Opening Time (OT)	2 Seconds
Closing Time (CT)	2 Seconds
Passengers Load Time (PLT)	5 Seconds
Between Floors (BFT)	5 Seconds

$$Load\ Time\ (L_T) = O_T + C_T + PL_T \quad (1)$$
$$N_1 = |F_N - C_N| \quad (2)$$
$$N_2 = |F_N - D_F| \quad (3)$$

N_1 is defined to be the number of floors between the passenger and the elevator, F_N is the floor the elevator is at and C_N is the call floor number. N_2 is defined to be the number of floors between the passenger and the elevator, and D_F is the destination floor number.

C_B is defined to be the number of calls made before the current passenger, C_A is defined to be the number of calls made after the current passenger and D_B is defined to be the number of drops offs before the current passenger.

Thus, the first objective function equation for calculating the waiting time of a passenger is defined to be:

$$Waiting\ Time\ (W_T) = BF_T * N_1 + L_T * C_B + L_T * D_B + O_T \quad (4)$$

Thus, the second objective function equation for calculating the destination time of a passenger is defined to be:

$$Destination\ Time\ (DT) = C_T + L_T * C_A + BF_T * N_2 + L_T * D_B + O_T \quad (5)$$

Finally, the objective of the EDP is the summation of equations (4) and (5):

$$J_T = D_T + W_T \quad (6)$$

Where J_T, D_T, and W_T denote journey, destination and waiting times respectively.

The average journey time is the output of the algorithms, also known as the fitness value, which is the parameter desired to be minimized and it presents the average journey time for all passengers. Where n represents the number of passengers.

$$\textit{Average Journey Time} = \sum_{1}^{n} JT(n)/n \quad (7)$$

In order to compare the algorithms, a case study was created to analyse the functioning of the algorithms and to produce a plausible answer. The decision variables of the EDP are the floor numbers, which is represented in an array, presenting the route that the elevator will follow in order to reach the average journey time or for all passengers.

A building was selected for a case study serviced by a single elevator. The floors of the building start from the ground floor, and end in the 20th floor (20). Thus, the building has 21 floors in total. The case is studied on 10 passengers in total. The initial floor of the elevator is the 4th floor.

Table 2. Case study

	Case Study Scenario
Passenger 1	5 → 9
Passenger 2	6 → 7
Passenger 3	3 → 15
Passenger 4	11 → 0
Passenger 5	20 → 8
Passenger 6	10 → 17
Passenger 7	13 → 19
Passenger 8	1 → 14
Passenger 9	16 → 2
Passenger 10	18 → 12

4. METHODOLOGY

In this paper, four optimization techniques are discussed and implemented on the EDP. This section presents the algorithms; Simulated Annealing (SA), Genetic Algorithm (GA), Particle Swarm Optimization Algorithm (PSO), and Whale Optimization Algorithm (WOA).

The EDP is a permutation problem; thus, the output of the algorithms will be the route that the elevator must follow to reach the desired objective function.

4.1. Trajectory based Stochastic Technique

4.1.1. Simulated Annealing (SA)

One of the most used heuristic approaches for tackling optimization problems is the SA algorithm. The annealing technique describes the ideal molecular configurations of metal particles in which the potential energy of the mass is reduced, and it refers to progressively cooling the metals after they have been exposed to high heat. In general, the SA algorithm employs an iterative movement based on a changeable temperature parameter, simulating the annealing transaction of metals. A basic optimization technique compares the outputs of the objective functions running with the current and neighbouring points in the domain repeatedly, so

that if the neighbouring point produces a better result than the current one, it is saved as the base solution for the next iteration. Otherwise, the algorithm quits the operation without attempting to search the larger area for better results.

Parameters T_i (initial temperature), T_f (final temperature) and i_m (maximum number of iterations) are first initialized. The change in energy ΔE is then computed for the feasible solution. If this value is less than 0, the solution is then accepted and the best solution is also updated. If the solution is greater than 0, this solution is first evaluated by 2 parameters, *r* and *P*. *r* is a random number initialized between 0 and 1.

If the value of *r* is less than *P*, the solution will be accepted, updating the current value, however, the best solution so far is not updated and the temperature value is then updated. This loop is repeated while the stopping criteria has not been met; the temperature value is greater than the T_f value, and as long as the iterations is less than the i_m value. [11]

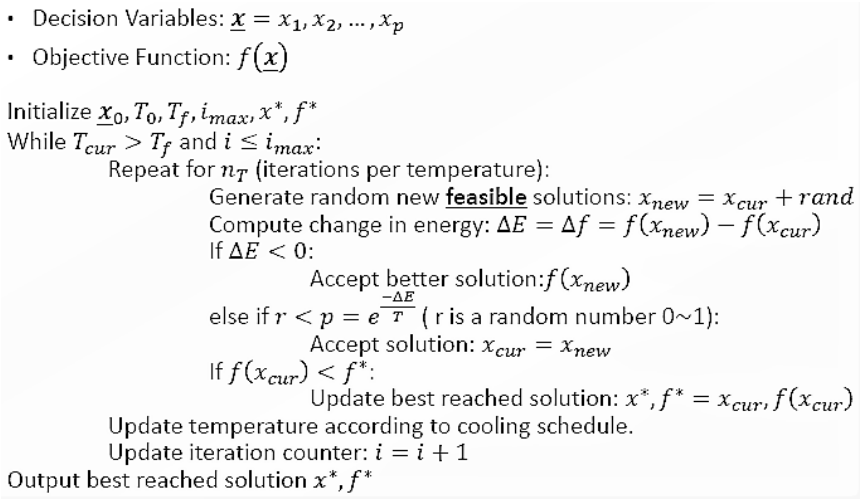

Fig. 1. Pseudo-code of the SA algorithm

Table 3. Used Parameters on the SA algorithm

Parameter	Value
Population size	1
Initial Temperature	200
Final Temperature	0.01
No. of iterations	100
No. of iteration per temperature	1

4.2. Population based Stochastic Technique – 1/2

In the population based stochastic techniques, the loop is operating while the stopping criteria has not been reached, which are: 1) current number of iterations is less than the i_m value, 2) when the objective function value has reached a certain pre-defined value 3) when there has been no improvement in the population over a number of iterations.

4.2.1. Genetic Algorithm (GA)

Genetic algorithms are randomised search algorithms designed to mimic the mechanics of natural selection and natural genetics. Genetic algorithms work on string structures, which, like biological structures, evolve over time according to the rule of survival of the fittest through a randomised yet organised information exchange. Thus, in every generation, a new set of strings is created, using parts of the fittest members of the old set. At first, the coding to be used must be defined. Then using a random process, an initial population of strings is created. Following that, a series of operators is employed to take this initial population and construct consecutive populations that should improve with time. The genetic algorithms' core operators are reproduction, crossover, and mutation. Reproduction is a process that is based on each string's goal function (fitness function). This objective function identifies how "good" a string is. Thus, strings with higher fitness value have bigger probability of contributing offspring to the next generation. Crossover is a process in which members of the last population are mated at random in the mating pool. So, a pair of offspring is generated, combining elements from two parents (members), which hopefully have improved fitness values. Mutation is the random (but infrequent) modification of the value of a string position. Mutation is, in reality, a random walk across the coded parameter space. Its goal is to prevent essential information contained inside strings from being lost prematurely.

The Genetic Algorithm (GA) is a population based stochastic technique. X_{prev} is the previous values of X from the previous loop.

Parameters m (population size) of size 4, pe (Elite members survived from X_{prev}) 1 Elite member, pc (Cross-over best members survived from X_{prev}) 2 Cross-over member, and pm (mutant worst members survived from X_{prev}) 1 Mutant member are first initialized. A new set of feasible solutions (chromosomes) are generated, and depending on the fitness value of each chromosome, it is categorized to either elite, cross-over or mutant. For the cross-over and mutant members, operations must occur on these members in order to generate a new child from its corresponding parents. Davis-order crossover is used for the cross over members and swap mutation is used for the mutant members. Post the operation, the fitness values are then computed once again for all the chromosomes and the iteration is repeated until reaching the stopping criteria. [12]

- Decision Variables: $\underline{\boldsymbol{x}} = x_1, x_2, \dots, x_p$
- Objective Function: $f(\underline{\boldsymbol{x}})$
- Initialize $\underline{\boldsymbol{x}}, i_{max}, m, p_e, p_c, p_m, x^*, f^*$

While ***Stopping Criteria*** Ex. $(\boldsymbol{i \leq i_{max}})$:
    Generate random new population with (m) chromosomes x_{new}:
        - p_e Elite members are survived from x_{prev}
        - p_c Cross-over best members from x_{prev}
        – p_m mutate worst members from x_{prev}
    Compute fitness value: $f(x_{new})$
    Update iteration counter: $i = i + 1$
Output best reached solution x^*, f^*

Fig. 2. Pseudo-code of the GA algorithm

Table 4. Used Parameters on the GA algorithm

Parameter	Value
Population size	4
No. of iterations	100
Elite member	1
Crossover member	2
Mutant member	1

4.3. Population based Stochastic Technique – 2/2

The PSO and WOA deals with arithmetic problems and the EDP is a permutation problem, thus, a space transformation must be done to use this algorithm in order to use the algorithms on the EDP.

4.3.1. Particle Swarm Optimization Algorithm (PSO)

The particle swarm optimization (PSO) is a randomized, population-based optimization method that was inspired by the flocking behaviour of birds or fish schooling. In PSO, each single solution is a "bird" in the search space. We call it a "particle". A swarm of these particles moves through the search space to find an optimal position. PSO is initialized with a group of random particles (solutions) and then searches for optima by updating generations. During every iteration, each particle is updated by following two "best" values. The first one is the position vector of the best solution (fitness) this particle has achieved so far. The fitness value is also stored. This position is called *pbest*. Another "best" position that is tracked by the particle swarm optimizer is the best position, obtained so far, by any particle in the population. This best position is the current global best and is called *gbest*. After finding the two best values, the position and velocity of the particles are updated.

Parameters *m* (Population size), *w* (Inertia weight), *c1* and *c2* (cognitive factor and social factor) and i_m (maximum number of iterations) are first initialized. Initial Velocity is an array, initialized at 0 for all elements. New population of random feasible solutions are generated, and for every particle, the velocity is first updated using its equation, then the position is then updated. The fitness values of all populations are then computed followed by updating the personal and the neighbourhood best. A previous study [13] is followed to convert the solution representation of the PSO to a permutation one. This modified algorithm is implemented using the smallest place value rule, which finds the permutations between continuous positions *Xkj*. A new sequence vector is generated and represented by *Skj* The values inside the continuous position's (*X*) vector are arranged in their ascending order, then the reflecting indices of this order are added in the *S* vector by the order of presence the corresponding arithmetic value in the original (unchanged) *X* vector. [14]

- Decision Variables: $\underline{x} = x_1, x_2, \dots, x_p$
- Objective Function: $f(\underline{x})$
- Initialize $\boldsymbol{X_0}, \boldsymbol{V_0}, i_{max}, m, w, c_1, c_2, x^*, f^*$

While ***Stopping Criteria*** Ex. $(\boldsymbol{i \leq i_{max}})$:
Generate random new population with (m) particles $\boldsymbol{X_{i+1}}$.
For each particle $\boldsymbol{x_i}$ in current population $\boldsymbol{X_i}$:
- Velocity Update:
$\boldsymbol{r_1, r_2} = random\ number\ vectors\ (same\ size\ as\ \boldsymbol{x_i})$
$\boldsymbol{v_{i+1} = w\, v_i + c_1 r_1 (Pbest_i - x_i) + c_2 r_2 (NBest_i - x_i)}$
- Position Update:
$\boldsymbol{x_{i+1} = x_i + v_{i+1}}$
Compute fitness value: $f(X_{new})$
Update personal best: $Pbest_{i+1} = x_{i+1}\ if\ f(x_{i+1}) < f(Pbest_i)$
Update neighborhood best: $Nbest_{i+1} = \underset{N}{\text{argmin}}(f(x_{i+1}))$
Update iteration counter: $i = i + 1$
Output best reached solution x^*, f^*

Fig. 3. Pseudo-code of the PSO algorithm

Table 5. Used Parameters on the PSO algorithm

Parameter	**Value**
Population size (*m*)	4
No. of iterations	100
inertia weight (*w*)	0.792
Cognitive and Social factor (*c1*, *c2*)	1.4994

4.3.2. Whale Optimization Algorithm (WOA)

Whale Optimization Algorithm (WOA) is a recently proposed optimization algorithm mimicking the hunting mechanism of humpback whales in nature. The WOA algorithm starts with a set of random solutions. At each iteration, search agents update their positions with respect to either a randomly chosen search agent or the best solution obtained so far. The '*a'* parameter is decreased from 2 to 0 in order to provide exploration and exploitation, respectively. *A* random search agent is chosen when $A>1$, while the best solution is selected when $A<1$ for updating the position of the search agents. Finally, the WOA algorithm is terminated by the satisfaction of a termination criterion.

5 Parameters are being updated at the beginning of each iteration, which are *a*, *A*, *C*, *l* and *p*. *l* and *p* are random numbers generated from 0 to 1, while the remaining parameters are updated using their corresponding equations.

Depending of the values of the above parameters, the corresponding operation is executed. If P < 0.5 and $A < 1$; the Encircling Prey method is used. If P < 0.5 and $A >= 1$; the Search for Prey method is used. If $P >= 0.5$; the Spiral Update method is used. The fitness value of all agents is calculated and the agent that has the best fitness value is compared to the overall reached best agent, and is updated if it is better than it. [15]

A space transformation used to implement a permutation problem using the WOA. The great value priority rule was used to transform the arithmetic solution to a permutation solution. This is done by sorting the elements of the solution in descending order, then taking the indices of these elements and use them as the solution of the WOA. Also, in order to improve the performance

and quality of solutions, local search was applied by selecting 2 elements at random to swap if a better tour is found.

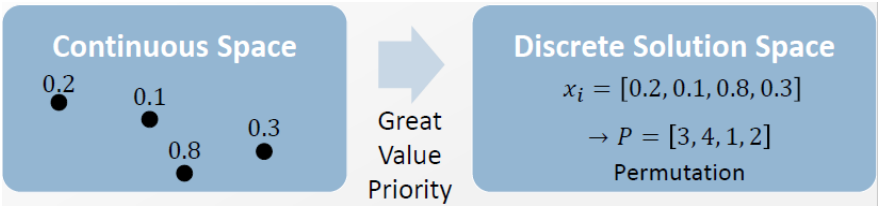

Fig. 4. Example of the great value priority transformation

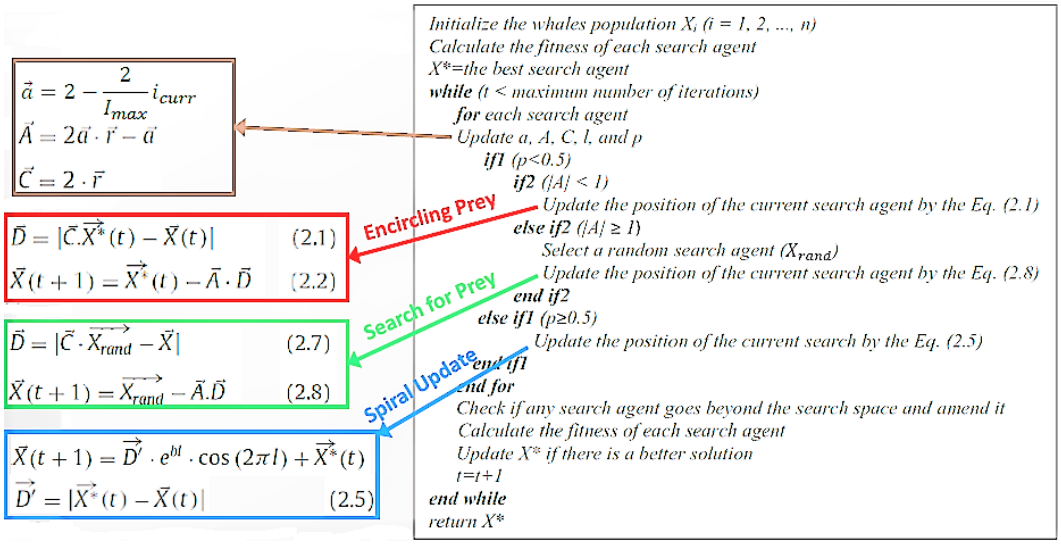

Fig. 5. Pseudo-code of the WOA algorithm

Table 6. Used Parameters on the WOA algorithm

Parameter	Value
Population size (m)	4
No. of iterations	100

5. RESULTS AND DISCUSSION

The case study is tested on all four implemented algorithms. Each case study's result is plotted on a convergence plot graph, where the X axis is the number of iterations and the Y axis presents the fitness values.

For the studied case implemented on the SA, the simulations best solution (route to be taken) is [4,1,3,5,16,10,9,14,20,13,6,18,15,11,12,19,7,8,17,2,0]. The average journey time of the optimized solution is approximately 472.3 seconds.

For the studied case implemented on the GA, the simulations best solution (route to be taken) is [4,3,1,5,6,7,9,10,11,13,14,15,16,17,18,19,20,12,8,2,0]. The average journey time of the optimized solution is approximately 222.3 seconds.

For the studied case implemented on the PSO, the simulations best solution (route to be taken) is [4,20,18,13,11,10,12,8,1,16,0,2,3,19,5,17,15,14,6,9,7]. The average journey time of the optimized solution is approximately 553.3 seconds.

For the studied case implemented on the WOA, the simulations best solution (route to be taken) is [4,5,9,10,16,18,12,1,13,2,14,19,6,20,17,3,8,15,11,7,0]. The average journey time of the optimized solution is approximately 541.3 seconds.

It is observed, that with the increase of the number of iterations, the solution converges to the best possible fitness value of the algorithm, with respect to the case study. It is also observed during trial runs, that the more the passengers in a case study, the more iterations needed to reach an optimal solution, as there are several combinations of solutions possible.

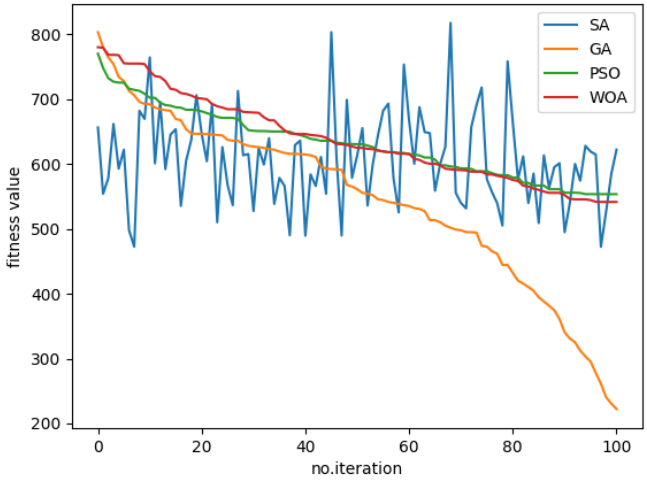

Fig. 6. Convergence Plot of SA (Blue), GA (Orange), PSO(Green) and WOA (Red) respectively

5.1. Performance Comparison

Table 7. Performance Comparison of the algorithms

Performance Index	SA	GA	PSO	WOA
Avg. Fitness value in 5 runs	531.4 s	279.1 s	600 s	620 s
Optimal Fitness Value	472.3 s	222.3 s	553.3 s	541.3 s

As per the data shown in Table 7 and Figure 6, we can see that the genetic algorithm has the best fitness value over the period of 100 iterations, while the simulated annealing algorithm is fluctuating between the fitness values. As for the PSO and WOA, the algorithms are reaching the optimal solution over the number of iterations and it is expected that if the number of iterations is increased, based on the graphs of both algorithms, it will reach an optimal solution. Both algorithms are affected by the space transformation, thus this is why both algorithms are preforming similarly and their fitness values are not so far from the GA and SA. The GA performs best over the remaining algorithms and over 5 runs as well, while the WOA preforms worst. This could be caused by the random variables that are in algorithm.

6. CONCLUSIONS AND FUTURE RECOMMENDATIONS

In conclusion, because of the growth of large buildings and the increased demand for more sophisticated elevators, elevator dispatch systems are always being enhanced and evolving. These upgraded methods are designed to reduce wait times.

This current research work evaluates the usage of four different optimization algorithms to provide a solution for the Elevator Dispatching Problem. The algorithms, Simulated Annealing (SA), Genetic Algorithm (GA), Particle Swarm Optimization Algorithm (PSO), and Whale Optimization Algorithm (WOA) are implemented using Python programming language and tested on a case to solve the Elevator Dispatching Problem in order to improve lift efficiency and provide a better user experience. To evaluate and test the performance of the four algorithms, performance indices are computed such as average and optimal fitness value in 5 runs to consider which of the algorithms is best to use in the industry. Based on the testing and findings, the Genetic Algorithm performed better than other algorithms. Our future recommendation is to use the Genetic Algorithm (GA) in the practical and dynamic elevator dispatching systems. We also highly recommend using adaptive optimization techniques. Further research is required to develop, implement and evaluate the use of multiple elevators. We advise that this study be improved further by collecting elevator traffic data throughout the course of a complete working day, running the algorithms at different periods of the day using different case studies, and taking the average of the results. A limitation that could improve the study is the inclusion of same floor calls from different passengers.

ACKNOWLEDGEMENTS

The authors thank and acknowledge the assistance and the usage of lecture notes provided by *Prof. Dr. Omar M. Shehata*, *M.Sc Eng. Shaimaa El Baklish*, and *M.Sc Eng. Catherine M. Elias* at the Department of Mechatronics Engineering of The German University in Cairo. The authors would like to thank the anonymous reviewers for their insightful remarks that helped enhance this research work.

REFERENCES

[1] M. Latif, M. Kheshaim, and S. Kundu, “A review of elevator dispatching systems,” 2016.
[2] M. Sale and V. Prakash, “Dynamic dispatching of elevators in elevator group control system: Research and survey,” International Journal of Innovative Technology and Exploring Engineering, vol. 8, pp. 98–102, 01 2019.
[3] J. Sorsa, H. Ehtamo, J.-M. Kuusinen, M. Ruokokoski, and M.-L. Siikonen, “Modeling uncertain passenger arrivals in the elevator dispatching problem with destination control,” Optimization Letters, vol. 12, 03 2017.
[4] A. So and L. Al-Sharif, “Calculation of the elevator round-trip time under destination group control using offline batch allocations and realtime allocations,” Journal of Building Engineering, vol. 22, 01 2019.
[5] J. Sorsa, “Optimization models and numerical algorithms for an elevator group control system,” 2017.
[6] E. Tartan and C. Ciftlikli, “A genetic algorithm-based elevator dispatching method for waiting time optimization,” IFAC-PapersOnLine, vol. 49, pp. 424–429, 12 2016.
[7] A. Vodopija, J. Stork, T. Bartz-Beielstein, and B. Filipi˘c, “Model-based multiobjective optimization of elevator group control,” 10 2018.
[8] Y. Bapin and V. Zarikas, “Smart building’s elevator with intelligent control algorithm based on bayesian networks,” International Journal of Advanced Computer Science and Applications, vol. 10, 01 2019.
[9] Z. Yang and W. Yue, “Elevator traffic pattern recognition based on fuzzy bp neural network with som algorithm,” Advances in Modelling and Analysis A, vol. 60, pp. 630–645, 12 2017.

[10] M. Sale and V. Prakash, “Dynamic scheduling of elevators with reduced waiting time of passengers in elevator group control system: Fuzzy system approach,” pp. 339–345, 0 2017.
[11] L. Rere, M. I. Fanany, and A. Arymurthy, “Simulated annealing algorithm for deep learning,” Procedia Computer Science, vol. 72, 11 2015.
[12] A. Thengade and R. Dondal, “Genetic algorithm – survey paper,” IJCA Proc National Conference on Recent Trends in Computing, NCRTC, vol. 5, 01 2012.
[13] I. Dubey and M. Gupta, “Uniform mutation and spv rule based optimized pso algorithm for tsp problem,” pp. 168–172, 02 2017.
[14] Eberhart and Y. Shi, “Particle swarm optimization: developments, applications and resources,” vol. 1, pp. 81–86 vol. 1, 2001.
[15] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” Adv. Eng. Softw., vol. 95, no. C, p. 51–67, May 2016. [Online]. Available: https://doi.org/10.1016/j.advengsoft.2016.01.008

AUTHORS

Shaher Ahmed is the first author and the lead researcher of this paper. He is currently pursuing a M.Sc degree focused in Industrial Automation, Machine Learning and Wireless Sensor Networks at The German University in Cairo's Department of Mechatronics Engineering. He graduated from the same university with a B.Sc in Mechatronics Engineering. His primary areas of interest include Optimization, Machine Learning, and Industrial Automation.

Mohamed Shekha graduated from The German University in Cairo with a B.Sc in Mechatronics Engineering. His major areas of interest include Automation and Intelligent Systems, particularly in the Automotive industry.

Suhaila Skran holds a bachelor's degree from the German University in Cairo, Egypt. She has a Bachelor of Science degree in Mechatronics Engineering. "System Identification of Active Quarter Car Model Using Deep Learning" was the title of her thesis. She works full-time at Valeo Egypt as an Automotive Embedded Software Engineer.

Abdelrahman Bassyouny holds a B.Sc in Mechatronics Engineering from the German University in Cairo.

AUTHOR INDEX