

David C. Wyld,
Jan Zizka (Eds)

Computer Science & Information Technology

6th International Conference on Computer Science, Engineering And
Applications (CSEA 2020),
December 18 - 19, 2020, Sydney, Australia



AIRCC Publishing Corporation

Volume Editors

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

Jan Zizka,
Mendel University in Brno, Czech Republic
E-mail: zizka.jan@gmail.com

ISSN: 2231 - 5403

ISBN: 978-1-925953-32-9

DOI: 10.5121/csit.2020.101801- 10.5121/csit.2020.101812

This work is subject to copyright. All rights are reserved, whether whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the International Copyright Law and permission for use must always be obtained from Academy & Industry Research Collaboration Center. Violations are liable to prosecution under the International Copyright Law.

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

Preface

The 6th International Conference on Computer Science, Engineering And Applications (CSEA 2020), December 18 - 19, 2020, Sydney, Australia, 6th International Conference on Data Mining and Database Management Systems (DMDBS 2020), 4th International Conference on Networks and Security (NSEC 2020), 4th International Conference on Networks & Communications (NETWORKS 2020), 6th International Conference on Fuzzy Logic Systems (Fuzzy 2020), 6th International Conference on Natural Language Computing (NATL 2020) and 6th International Conference on Signal, Image Processing and Embedded Systems (SIGEM 2020) was collocated with 6th International Conference on Computer Science, Engineering and Applications (CSEA 2020). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

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

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

In closing, CSEA 2020, DMDBS 2020, NSEC 2020, NETWORKS 2020, Fuzzy 2020, NATL 2020 and SIGEM 2020 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the CSEA 2020, DMDBS 2020, NSEC 2020, NETWORKS 2020, Fuzzy 2020, NATL 2020 and SIGEM 2020.

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

David C. Wyld,
Jan Zizka (Eds)

General Chair

David C. Wyld,
Jan Zizka (Eds)

Organization

Jackson State University, USA
Mendel University in Brno, Czech Republic

Program Committee Members

AbdalhosseinRezai,
Abdel-Badeeh M. Salem,
AbdelhadiAssir,
Addisson Salazar,
Afaq Ahmad,
Ahmed Abdulqader Hussein,
Ahmed Farouk AbdelGawad,
AjithNongmaithem,
Akhil Gupta,
Alfredo Ferreira Costa,
Amalazeroual,
Ana LuísaVarani Leal,
AnandNayyar,
Ankur Singh Bist,
AnouarAbtoy,
AtikaQazi,
Azeddine WAHBI,
AzidaZainol,
BalanSethuramalingam ,
Benyamin Ahmadnia,
BijanDavvaz,
Bipan Hazarika,
boukarinassim,
Chengliang Huang,
ChidubemOkwudiriEmedo,
Chikh Mohammed Amine,
DadmehrRahbari,
Daniel Hunyad,
Daniel Rosa Canedo,
Deepak Mishra,
Domenico Ciuonzo,
Donatella Giuliani,
DorraDriss,
EmekaOgbuju,
Eng Islam Atef,
FaizaTabbana,
FarshidSahba,
Francesco Zirilli,
G.Malema,
Gabriella Casalino,

ACECR, Iran
Ain Shams University, Cairo, Egypt
Hassan 1st University, Morocco
Universitat Politècnica de València, Spain
Sultan Qaboos University, Oman
University of Technology, Iraq
Zagazig University, Egypt
South East Manipur College, India
Lovely Professional University, India
Federal University, UFRN, Brazil
Mohammed V University, Morocco
University of Macau, China
Duy Tan University, Vietnam
Signy Advanced Technology, INDIA.
AbdelmalekEssaadi University, Morocco
University Brunei Darussalam, Brunei
Hassan II University, Morocco
University of Jeddah, Saudi Arabia
Bharathiar University, India
University of California, USA
Yazd University, Iran
Gauhati University, India
skikda university, algeria
Ryerson University, Canada
University of Nigeria, Nigeria
Tlemcen University, Algeria
University of Qom, Iran
"Lucian Blaga" University of Sibiu, Romania
Federal Institute of Goias, Brazil
ISRO, India
University of Naples Federico, Italy
University of Bologna, Italy
University of Sfax, Tunisia
Federal University Lokoja, Nigeria
Alexandria University, Egypt
Military Academy, Tunisia
Jamshid Kashani University, Iran
Sapienza Università Roma , Italy
University of Botswana, Botswana
University of Bari, Italy

Gelu-OvidiuTirian,	Politehnica University Timisoara, Romania
Govindraj B. Chittapur,	Basaveshwar Engineering College, India
Grigorios N. Beligiannis,	University of Patras, Greece
Gulden Kokturk,	DokuzEylül University, Turkey
HamzehKhalili,	Polytechnic University of Catalonia, Spain.
Hanene Ben-Abdallah,	Higher Colleges of Technology, UAE
Hang Su,	Politecnico di Milano, Italy
Hari Mohan Srivastava,	University of Victoria, Canada
Himani Mittal,	GGDSD College, India
Hiroshi Ban,	Nagaoka University of Technology, Japan
Hyunsung Kim,	Kyungil University, Korea
ilangovelchamy,	CMR Institute of Technology, India
Isa Maleki,	Islamic Azad University, Iran
Israa Shaker Tawfic,	Ministry of Science and Technology, Iraq
Ivan Izonin,	Lviv Polytechnic National University, Ukraine
J.Naren,	Research Scholar, India
Jagadeesh HS,	APS COLLEGE OF ENGINEERING, India
JanuszKacprzyk,	Systems Research Institute, Poland
Javier Montero,	Universidad Complutense de Madrid, Spain
Jawad K. Ali,	University of Technology, Iraq
Jean-Charles LAMIREL,	Université de Dalian, France
Jeckleou s. mapa,Saint	Joseph Institute of technology, Philippines
JesukKo,	Universidad Mayor de San Andres (UMSA), Bolivia
Jianyi Lin,	Khalifa University, United Arab Emirates
Joao Calado,	Instituto Superior de Engenharia de Lisboa, Portugal
Joshua Abah,	Federal University of Lafia, Nigeria.
Kamel Hussein Rahouma,	Minia University, Egypt
Karim El Moutaouakil,	FPT/USMBA, Morocco
Kazuyuki Matusmoto,	Tokushima University, Japan
Ke-Lin Du,	Concordia University, Canada
Khalid M.O Nahar,	Yarmouk University, Jordan
Khalid Nazim Abdul Sattar,	Majmaah University, Saudi Arabia
KirilAlexiev,	Bulgarian Academy of Sciences, Bulgaria
Kirtikumar Patel,	I&E, India
Klenilmar Lopes Dias,	Federal Institute of Amapa, Brazil
Koh You Beng,	University of Malaya, Malaysia
Krishna Prasad,	Sri Venkateswara College of Engineering, India
László T. Kóczy,	Budapest University of Technology, Hungary
Iham,	Istanbul Aydin University, Turkey
LjiljanaTrajkovic,	Simon Fraser University, Canada
Luisa Maria ArvideCambra,	University of Almeria, Spain
M.Keyvanpour,	Alzahra University, Tehran, Iran
Mahdi Sabri,	Islamic Azad University, Iran
Mahmood Hashemi,	Beijing University of Technology, China
Malka Halgamuge,	University of Melbourne, Australia
Manish Mishra,	University of Gondar, Ethiopia
Manoj Sahni,	Pandit Deendayal Petroleum University, India
Mansour Bader,	Al-Balqa Applied University, Jordan
Manyok Chol David,	University of Juba, South Sudan
Mario Versaci,	DICEAM - University Mediterranea, Italy
Masoud Rashidinejad,	Queens University, Canada
Md. Farukuzzaman Khan,	Islamic University, Bangladesh

Mehamed Ahmed Abdurahman,	Aksum University, Ethiopia
Mekonnen Wagaw,	Bahir Dar University, Ethiopia
Melissa P. Picones-Juanillo,	Salalah College of Technology, Oman
Metais Elisabeth,	Le Cnam, France
Michail Kalogiannakis,	University of Crete, Greece
Mohamed Saad AZIZI,	Moulay-Ismail University, Morocco
Mohammad Abu Omar,	Al-Quds Open University, Palestine
Mohammad Khaled Al Maghasbeh,	Ministry of education, Jordan
Mohammad Mahdi Feraidooni,	ShahidBeheshti University, Iran
MortezaAlinia Ahandani,	University of Tabriz, Iran
Mu-Chun Su,	National Central University, Taiwan
Muhammad Arif,	Guangzhou University, China
Muhammad Sarfraz,	Kuwait University, Kuwait
Mu-Song Chen,	Da-Yeh University, Taiwan
Nadia Abd-alsabour,	Cairo University, Egypt
Naoyuki Ishimura,	Chuo University, Japan
Narinder Singh,	Punjabi University, India
Nikola Ivkovic,	University of Zagreb, Croatia
Omar Al-harbi,	Jazan University University, Saudi Arabia
Ouided SEKHRI,	Frères Mentouri University, Algeria
PariaAssari,	Islamic Azad University, Iran
Pavel Loskot,	Swansea University, United Kingdom
Picky Butani,	Savannah River National Laboratory, USA
Poonam Kumar Sharma,	D.A.V. College, India
Pranita Mahajan,	Mumbai University, India
Prasan Kumar Sahoo,	Chang Gung University, Taiwan
Punnoose A K,	Flare Speech Systems, India
Rahul Johari,	GGSSIP University, India
Rajkumar,	N.M.S.S.Vellaichamy Nadar College, India.
Ramadan Elaieess,	University of Benghazi, Libya
Rekha Jain,	Poornima Institute of Engineering & Technology, India
Robert SsaliBalagadde,	Kampala International University, Uganda
Rodrigo,	Instituto Federal de Educação, Brazil
RoumianaKountcheva,	T&K Engineering Mladost Sofia, Bulgaria
RuaaAlsabah,	University of Kerbala, Iraq
RuhaidahSamsudin,	UniversitiTeknologi Malaysia, Malaysia
RustemPopa,	University of Galati, Romania
Saeed Farzi,	AI K. N. Toosiuniversity of technology, Iran
Sahar Saoud,	Ibn Zohr University, Morocco
Sahil Verma,	Lovely Professional University, Phagwara India
Said AGOUJIL,	University of Moulay Ismail Meknes, Morocco
Saif Aldeen Saad Obayes,	Imam ALKADHIM college Babylon, Iraq
Saleh Al-Daajeh,	Abu Dhabi polytechnic, UAE
sanjeevkumar, Dr.	BhimraoAmbedkar University, India
Seppo Sirkemaa,	University of Turku, Finland
Shadi Abudalfa,	University College of Applied Sciences, Palestine
Shamneesh Sharma,	Poornima University, India
Shanta Phani,	Bengal Institute of Technology, India
Shereena Vb,	Mahatma Gandhi University, India
Shing-Tai Pan,	National University of Kaohsiung, Taiwan
Siddhartha Bhattacharyya,	Christ University, India
Simanta Sarmah,	Alpha Clinical Systems Inc, USA

Sin ThiYarMyint,
Siva Kumar PV,
Smain Femmam,
Solomiia Fedushko,
soubhikchakraborty,
Stefania Tomasiello,
Stefka Stoyanova Fidanova,
Thenmalar S,
Tom Chen,
Tome Dimovski,
VasileGoldis,
Veena MN,
Venkata Raghavendra,
Veslavaosińska,
Vilem Novak,
Xin Liu,
Xinxin Fan,
Yilun Shang,
Yogendra Kumar Jain,
Yousef FARHAOUI,
Youssef TAHER,
Youssefi My Abdelkader,
Yu-Chen Hu,
Yu-Lin (Eugene) Song,
YuriySyerov,
Zahera Mekkioui,
Zoran Bojkovic,
Zurab Kiguradze

Myanmar Institute of Information Technology
VNR VJIET, India
UHA University, France
Lviv Polytechnic National University, Ukraine
Birla Institute of Technology, India
University of Tartu, Estonia
Bulgarian Academy of Sciences, Bulgaria
SRM Institute of Science and Technology, India
University of London, United Kingdom
St. Kliment Ohridski University, Bitola
Western University of Arad, Romania
P. E. S. College of Engineering, India
Annie Besent women's College, India
Nicolaus Copernicus University, Poland
University of Ostrava, Czech Republic
University of Oulu, Finland
State Grid Tongling Electric Power Supply Company
Northumbria University, UK
Samrat Ashok Technological Institute, India
Moulay Ismail University, Morocco
Mohammed V University, Maroc
Hassan First University, Morocco
Providence University, Taiwan
Asia University, Taiwan
Lviv Polytechnic National University, Ukraine
University of Tlemcen, Algérie.
University of Belgrade, Serbia
Missouri University of Science and Technology, USA

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)



Artificial Intelligence Community (AIC)



Soft Computing Community (SCC)



Digital Signal & Image Processing Community (DSIPC)



Organized By



Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

6th International Conference on Computer Science, Engineering and Applications (CSEA 2020)

New Algorithms for Computing Field Of Vision Over 2d Grids.....01 - 18
Evan R.M. Debenham and Roberto Solis-Oba

A Multilevel Technology Acceptance Management Model.....19 - 35
Gilbert Busolo, Lawrence Nderu and Kennedy Ogada

Minimum Viable Model Estimates for Machine Learning Projects.....37 - 46
John Hawkins

6th International Conference on Data Mining and Database Management Systems (DMDBS 2020)

An Intellectual Approach to Design Personal Study Plan via Machine Learning47 - 60
Shiyuan Zhang, Evan Gunnell, Marisabel Chang and Yu Sun

4th International Conference on Networks and Security (NSEC 2020)

PNFUZZ: A Stateful Network Protocol Fuzzing Approach Based on Packet Clustering61 - 69
HuiHui He and YongJun Wang

Towards Assessing Critical Infrastructures' Cyber-Security Culture During COVID-19 Crisis: A Tailor-made Survey.....71 - 80
Anna Georgiadou, Spiros Mouzakitis and Dimitrios Askounis

4th International Conference on Networks & Communications (NETWORKS 2020)

The Statistical Correlation between Distortion and Adding Signal for PAPR Reduction in OFDM based Communication Systems.....81 - 96
Désiré Guel, Boureima Zerbo, Jacques Palicot and Oumarou Sié

6th International Conference on Fuzzy Logic Systems (Fuzzy 2020)

A Fuzzy BI-Objective Model for SVM with an Interactive Procedure to Identify the Best Compromise Solution.....97 - 105
Hager Ali Yahia, Mohammed Zakaria Moustafa, Mohammed Rizk Mohammed and Hatem Awad Khater

**6th International Conference on Natural Language
Computing (NATL 2020)**

Importance of the Single-Span Task Formulation to Extractive Question-answering.....107 - 116
Marie-Anne Xu and Rahul Khanna

Automatic Generation of Text for Match Recaps using Esport Caster Commentaries117 - 131
Oluseyi Olarewaju, Athanasios V. Kokkinakis, Simon Demediuk, Justus Roberstson, Isabelle Nölle, Sagarika Patra, Daniel Slawson, Alan P. Chitayat, Alistair Coates, Ben Kirman, Anders Drachen, Marian Ursu, Florian Block and Jonathan Hook

SMARTAJWEED Automatic Recognition of Arabic Quranic Recitation Rules.....145 - 152
Ali M. Alagrami and Maged M. Eljazza

**6th International Conference on Signal, Image Processing and
Embedded Systems (SIGEM 2020)**

Profiling NVIDIA Jetson Embedded GPU Devices for Autonomous Machines.....133 - 144
Yazhou Li and Yahong Rosa Zheng

NEW ALGORITHMS FOR COMPUTING FIELD OF VISION OVER 2D GRIDS

Evan R.M. Debenham and Roberto Solis-Oba

Department of Computer Science, The University of Western Ontario, Canada

ABSTRACT

The aim of this paper is to propose new algorithms for Field of Vision (FOV) computation which improve on existing work at high resolutions. FOV refers to the set of locations that are visible from a specific position in a scene of a computer game.

We summarize existing algorithms for FOV computation, describe their limitations, and present new algorithms which aim to address these limitations. We first present an algorithm which makes use of spatial data structures in a way which is new for FOV calculation. We then present a novel technique which updates a previously calculated FOV, rather than recalculating an FOV from scratch.

We compare our algorithms to existing FOV algorithms and show they provide substantial improvements to running time. Our algorithms provide the largest improvement over existing FOV algorithms at large grid sizes, thus allowing the possibility of the design of high resolution FOV-based video games.

KEYWORDS

Field of Vision (FOV), Computer Games, Visibility Determination, Algorithms.

1. INTRODUCTION

1.1. Background

A Field of Vision (FOV) is the set of locations that are visible from a specific position in a scene of a computer game. FOV is calculated over a two-dimensional finite grid which is referred to as the FOV grid. An FOV grid partitions a game's environment into rectangular cells. If a cell within this grid contains an object which is vision-blocking (such as a wall, a tree, etc.) then that entire cell is vision-blocking. Games use an FOV grid as it allows them to consider the visibility of a whole region at once, which is significantly less computationally expensive than considering the visibility for every point in the environment. Because of this, the FOV grid may be constructed at a different resolution than the resolution at which the game will be displayed. One grid cell is specified as the source of vision and is referred to as the FOV source cell. An FOV algorithm must determine which cells are visible from the source and which cells are not visible based on the cells that are vision-blocking; the resulting grid with cells set to visible or non-visible is called the FOV.

Figure 1 shows an FOV for a scene of Crypt of the Necrodancer by Brace Yourself Games. The game environment is shown on the left with the FOV grid superimposed in purple. The yellow brick walls block vision from the character located near the bottom left corner. In this example each FOV grid cell is a 48*48 pixel region of the game's display. A data representation of this

scene is shown on the right. The FOV source cell is marked with an S, vision-blocking cells are yellow, visible cells are in light brown, and non-visible cells are darkened.

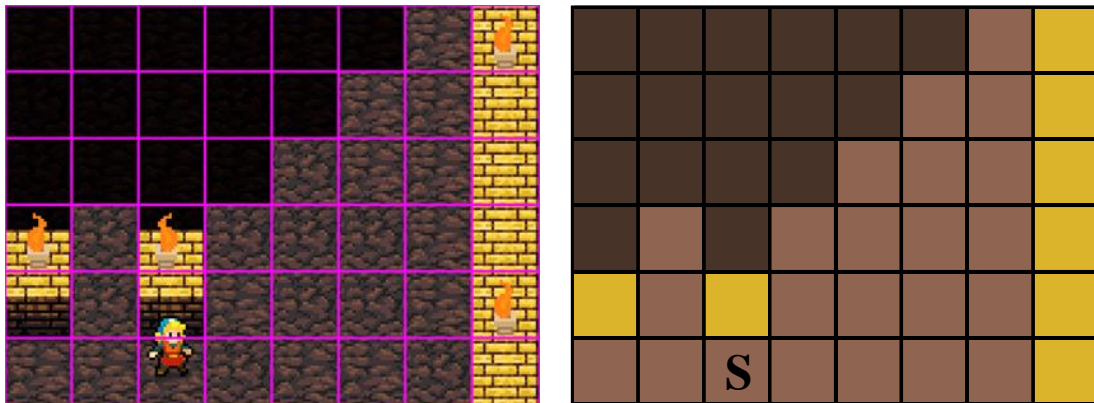


Figure 1. FOV in a game with simple 2D graphics. Left image taken from Crypt of the Necrodancer, by Brace Yourself Games. Right image shows a data representation of the FOV.

Calculating an FOV is useful for computer games with a top-down perspective. In these games the player views the game world from above and thus sees much more of the game world than an individual character inside the game. Top-down games may make use of FOV to provide accurate visibility information to computer-controlled game actors, so that they may then make decisions based on what they can see. Top-down games may also use FOV to convey to a player which areas of the world their character cannot see by visually darkening them. This visual effect is referred to as a fog of war.

FOV is used by a number of popular computer games, such as League of Legends [1, 2] and Defense of the Ancients 2 [3]. The time needed to compute an FOV must be considered when designing games with complex environments, and both of the above games calculate FOV at reduced resolutions in order to improve performance.

Games are expected to render their scenes many times per second in order to give the impression of smooth motion. These rendered scenes are then displayed as frames on a computer screen. Most computer displays support a maximum of 60 frames per second (one new frame roughly every 17ms), but in newer displays this number has increased to up to 240 frames per second (one new frame roughly every 4ms). The faster and more consistently a game is able to render its scenes the smoother the display of the game will be for the user. A game might not need to calculate the FOV every time a scene is rendered, but it must be able to calculate it fast enough such that rendering is not delayed. Games also have many different processes that must share system resources. Because of these conditions, an FOV algorithm must complete extremely quickly so as not to delay rendering or starve other processes of system resources. We show that existing FOV algorithms do not scale well with grid size and have inadequate performance as a result.

1.2. Existing FOV Algorithms and Related Work

All existing FOV algorithms make use of rays cast from the FOV source. One such algorithm is Mass Ray FOV, which casts a ray from the center of the FOV source cell to the center of each cell in the FOV grid, and if that ray does not intersect any vision-blocking cells then that ray's destination cell is marked as visible. Mass ray FOV has very poor performance because it casts as many rays as there are cells in the grid. Another algorithm based on direct ray casting is Perimeter Ray FOV. This algorithm casts a ray to the center of every cell on the perimeter of the

grid and for each ray it sets to visible all cells that the ray touches before the ray finds a vision-blocking cell. While Perimeter Ray FOV does cast fewer rays than Mass Ray FOV, all algorithms which directly cast rays to specific grid cells may cast more rays than necessary because they cast a fixed number of rays regardless of the terrain.

More intelligent FOV algorithms selectively cast visibility rays to the corners of vision blocking cells and use these rays to determine the boundaries of the visible space. Recursive Shadowcasting by Björn Bergström [4] is the most popular publicly available algorithm which is based on this selective ray casting approach. Recursive Shadowcasting starts by initializing all grid cells to not visible. It then splits the FOV grid into eight octants centered on the FOV source and traverses the cells within each octant as shown in Figure 2. This traversal occurs within each octant by rows or columns in ascending order of distance from the FOV source. As a cell is traversed, its visibility status is set to visible. However, when a vision-blocking cell is encountered, an octant is split into two smaller sub-octants which are bounded by rays cast from the FOV source to the corners of the vision-blocking cell. The cell traversals are then continued within each sub-octant.

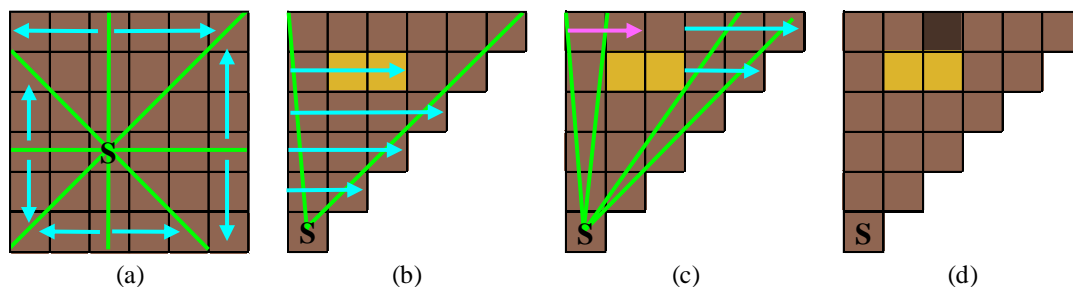


Figure 2. (a) An FOV grid split into octants with row/column traversal shown. (b) Traversal of an octant until vision-blocking cells are encountered. (c) Traversal of sub-octants after the traversal is split by the vision-blocking cells. (d) The resulting FOV.

Permissive FOV by Jonathon Duerig [5] is another popular FOV algorithm which selectively casts rays. Recursive Shadowcasting and Permissive FOV are efficient at low grid sizes but become slow as grid size increases because they perform a relatively large number of operations per-cell. It is important to note that FOV grids can be made of tens of thousands of cells. For an algorithm to improve on the performance of selective ray casting, it must determine cell visibility and set cell visibility statuses in a more efficient manner.

There are some problems in computer games that involve determining visibility information, like determining how light and shadow interact with the objects in a scene, and how to ensure that parts of objects which are not visible are not rendered. Techniques that have been designed to address these problems, such as shadow mapping [6], shadow volumes [7], real-time ray tracing [8], z-buffering [9], the painter's algorithm [10], frustum culling [11], and portal-based occlusion culling [12] cannot be used to calculate an FOV.

The rest of the paper is organized in the following manner. In Section 2 we propose a new FOV algorithm named Rectangle-Based FOV, which represents vision-blocking cells in a very compact and efficient way by using rectangles. In Section 3 we propose a second new algorithm named FOV Update, which adjusts a previously calculated FOV instead of calculating an FOV from scratch. In Section 4 we compare both of these algorithms to Recursive Shadowcasting (as it is known to have the best performance among existing FOV algorithms [13, 14]) and determined that our algorithms offer superior performance to Recursive Shadowcasting when the

grid size becomes large. In Section 5 we summarize our results and make recommendations to implementors who may wish to use our algorithms.

2. RECTANGLE-BASED FIELD OF VISION

2.1. Representing Vision-Blocking Cells with Rectangles

In computer games, the FOV usually needs to be calculated every time the FOV source moves, but the game environment changes rather infrequently. Because of this, it is possible to pre-process the environment and represent the vision blocking cells in a compact and efficient manner using rectangles. This efficient representation can then be used for many FOV calculations, and only needs to be updated if the game environment changes. The performance of an FOV algorithm which uses such a compact representation of vision-blocking cells is less dependent on the size of the FOV grid than selective ray casting algorithms.

We use rectangles to represent groups of vision-blocking cells, this allows us to compactly represent the vision-blocking terrain in a game environment. The rectangle-based representation of vision-blocking cells is created with the following process. Adjacent vision blocking cells are first grouped together into rectilinear regions. These rectilinear regions are then split into a minimal number of rectangles using the algorithm in [15]. The rectangles which represent vision-blocking cells on a grid are stored in a quadtree [16] which allows us to rapidly access them, instead of having to search for them within the grid. Quadtrees are simple to build and update, which is important if the game environment does change.

A quadtree stores rectangles by hierarchically dividing the FOV grid into successively smaller quadrants. The root node of the quadtree represents the entire FOV grid. Each internal node (including the root) has exactly four children, each representing one quadrant of the space represented by their parent. When the region that a node represents contains fewer than a predefined number N of rectangles, that node is made into a leaf and it stores all rectangles that intersect the region that it represents (see Figure 3). Note that a rectangle may be stored in multiple leaf nodes.

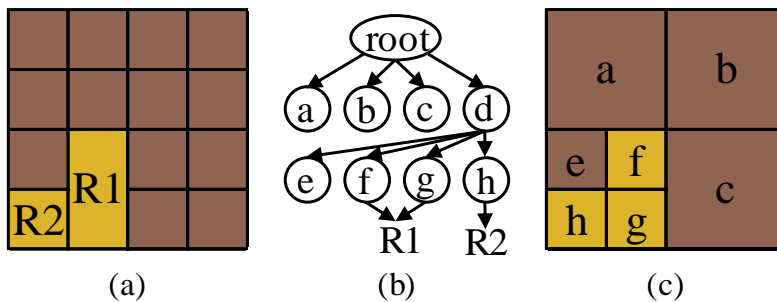


Figure 3. (a) A grid containing two rectangles R1 and R2. (b) The quadtree that represents this grid, with $N = 1$. (c) The region represented by each leaf node of the quadtree

We consider a cell to be visible if any point within that cell can be seen from the FOV source. We use this definition of visibility as it is the most commonly used in computer games, and it matches the definition of visibility used by Recursive Shadowcasting. However, our algorithm can be adapted to use other definitions of visibility as well, such as a cell being visible only if its center can be seen by the FOV source.

When determining what areas of the FOV grid are occluded behind a rectangle, we use two of the rectangle's four vertices, and refer to these as a rectangle's relevant points for determining visibility. The relevant points are always the two vertices of the rectangle which are farthest apart from each other among the vertices which are visible to the FOV source when considering that rectangle in isolation (see Figure 4).

We cast a ray from each relevant point in the opposite direction of the FOV source. The area between the rectangle and these two rays contains the area occluded behind that rectangle. Note that this area does not include the rays themselves, as they are considered to be visible from the FOV source. A cell is occluded by a rectangle if it is entirely within the area (see Figure 4).

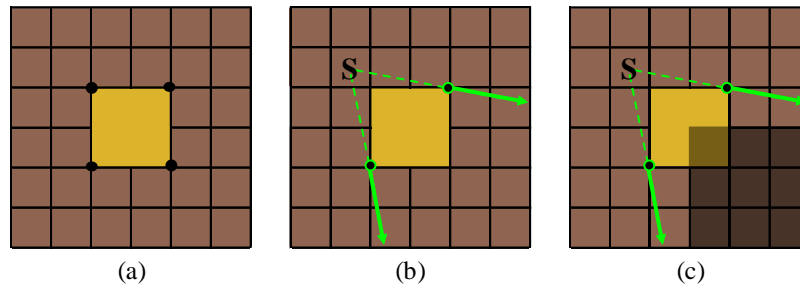


Figure 4. (a) A rectangle with its four vertices highlighted. (b) The two relevant points and casted rays for this rectangle. (c) The cells occluded by the rectangle are darkened.

2.2. Calculating an FOV using Rectangles

We process each rectangle as specified below in order to assign visibility statuses to the cells of the FOV grid. However, first we note that we cannot process each rectangle in isolation. If two rectangles R1 and R2 have (part of) a common side, there might be some cells which are not visible from the FOV source but which are not fully contained in the occluded region of either rectangle. If R1 and R2 are processed independently then these cells might be incorrectly labelled as visible. Figure 5 gives an example of this: Cells i and ii are not fully occluded by either rectangle R1 or R2 but they are not visible to the FOV source. Note that i and ii are not considered to be occluded by R1 because one of the rays cast from R1 touches the corners of i and ii. Figure 5(a) and (b) show the individually occluded regions for R1 and R2 respectively, and Figure 5(c) shows how these combined occluded areas do not result in a correct FOV.

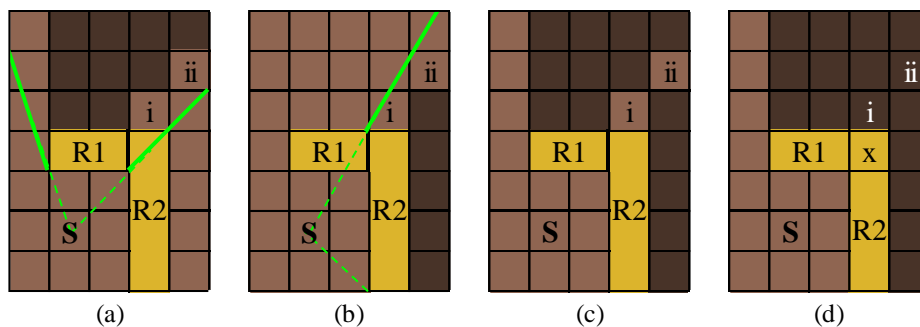


Figure 5. An environment where rectangles R1 and R2 are adjacent. (a) The region occluded by R1. (b) The region occluded by R2. (c) cells i and ii are incorrectly labelled as visible. (d) Correct labelling of the cells.

We address this issue by slightly increasing the size of one of the two adjacent rectangles, as this will ensure that all cells are correctly set to not visible. For each relevant point P of R1, we check

if P also belongs to another rectangle. This check can be efficiently performed using the quadtree: We recursively traverse the nodes of the quadtree which represent regions containing P until we reach a leaf node. Upon reaching a leaf node we check all rectangles stored in it (except R_1) to see if any of them contains P . If some rectangle R_2 contains P , we check if R_2 occludes P . If P is not occluded, we extend the size of R_1 by one row or column so that it overlaps with R_2 . In the example in Figure 5(d), R_1 is extended by one column to the right into R_2 , so the rectangles now overlap at cell x . This results in cells i and ii being correctly set to not visible as the extended R_1 occludes them.

After rectangles have been extended (if needed), the visibility statuses of the grid cells are assigned: Initially each cell is assigned a status of visible; then we process the rectangles one by one. For each row of the FOV grid that is within the occluded region of a rectangle, the algorithm calculates the leftmost and rightmost cells which are fully within the occluded region; the algorithm then traverses the row from this leftmost cell to the rightmost cell, setting every cell traversed to not visible. Once this process is repeated for every rectangle, every grid cell that is occluded by a rectangle will have its visibility status set to not visible, while all other cells will remain as visible.

2.3. Optimizations to Rectangle-Based FOV

The algorithm described in Section 2.2 will correctly calculate an FOV but it may perform redundant work. In this section we describe two performance optimizations that reduce the amount of work performed by the algorithm. Firstly, when the rectangle being processed is partially or totally occluded behind other rectangles, the regions that these rectangles occlude overlap. As we set the cell visibility statuses within these occluded regions, some cells might be set to not visible multiple times. We address this by reducing the size of the rectangle we are processing such that it is no longer occluded by other rectangles.

If we process rectangles in an order such that a rectangle is processed after any rectangles which occlude it, then when processing a rectangle we could use any partially computed visibility information to determine whether a rectangle is (partially) occluded. We can traverse the quadtree to efficiently arrange the rectangles so that many of them are processed after any rectangles that (partially) occlude them, as we explain below.

We define the distance between the FOV source and a rectangular region of the FOV grid as the distance between the FOV source and the closest point within that region. Starting from the root, we recursively process the children of each internal node of the quadtree in order from closest to furthest distance from the regions represented by the child nodes to the FOV source. For a leaf node, we process its rectangles in order from closest to furthest from the FOV source. Since a rectangle can be stored in more than one node, we only process a rectangle the first time it is encountered. Note that our ordering does not guarantee that a rectangle will always be processed after rectangles which occlude it, but for many rectangles this will be the case.

Because we determine the entire area occluded by a rectangle at once, we can take advantage of spatial locality in order to set visibility statuses efficiently. Spatial locality refers to a property of computer memory where accessing data sequentially is substantially faster than accessing it in a random order. Memory has this property due to the nature of the CPU cache, which stores recently accessed data and the data which is adjacent to it. By accessing data in a sequential manner, we take advantage of the cache storing adjacent data, and so are able to access data from the cache instead of main memory and we can set cell visibility statuses quickly.

2.4. The Rectangle-Based FOV Algorithm

First the vision-blocking terrain must be converted to a quadtree of rectangles. The steps for this process are summarized in pseudocode below:

Algorithm: Vision-Blocking Cells to Rectangles(G, N)

Input: FOV grid G , integer N specifying the maximum number of rectangles in a leaf node.

Result: The root of a quadtree which contains the vision-blocking rectangles of G .

Let L be an empty list of rectangles.

for each rectilinear region E of vision-blocking cells **in** G :

Dissect E into rectangles (as described in Section 2.1) and add those rectangles to L .

Let Q be a quadtree node representing the entire grid G . // Q is the root of the quadtree

for each rectangle R **in** L :

Add R to every leaf node of the quadtree representing a region that intersects R .

while there is a leaf node P which has more than N rectangles:

Convert P into an internal node with 4 children, each representing one quadrant of P .

Move each rectangle r in P into every one of P 's children which intersect r .

return Q

Once the quadtree has been constructed and all cells of the grid have been initialized as visible, the rectangle-based FOV algorithm is used to compute the FOV from a given source cell. The steps of the Rectangle-Based FOV algorithm are summarized in pseudocode below:

Algorithm: Rectangle-Based FOV(N, S, G)

Input: Quadtree node n , FOV source cell S , FOV grid G .

When first called: n will be the root node of the quadtree and all cells in G are set to visible.

Result: Cells in G which are not visible from S are set to not visible.

if n is a leaf node **then:**

for each rectangle R **in** n , from closest to farthest from S :

if R has not already been processed **then**

Extend R if needed, as described in Section 2.2.

Shrink R if needed, as described in Section 2.3.

Let E be the region occluded behind R .

for each row X of the grid G that intersects E :

Set to not visible the cells in X contained in E , from left to right.

Mark R as processed

else:

for each child node C of n , from closest to farthest from S :

Rectangle-Based FOV(C, S, G)

3. UPDATING AN EXISTING FIELD OF VISION

All FOV algorithms we have discussed so far calculate the FOV from scratch and require the cells of the FOV grid to be initialized as either all non-visible or all visible. This initialization significantly affects performance at large grid sizes. Additionally, these FOV algorithms completely discard the previously calculated FOV when they reset the visibility information stored in the grid cells.

Our second new FOV algorithm uses the previously calculated FOV rather than discarding it. This improves performance as this algorithm does not need to clear the grid and so it will need to assign fewer cell visibility statuses. Updating an FOV is possible because an FOV often needs to be re-calculated when the FOV source moves to an adjacent grid cell. Hence, it is likely that most FOV grid cells will have the same visibility status in both FOV calculations. Therefore, we may be able to compute the new FOV more efficiently if we update a previously calculated FOV instead of calculating it from scratch.

3.1. Cones of Changing Visibility

Updating a previously calculated FOV is conceptually more complex than calculating an FOV from scratch. When updating an FOV two FOV sources must be considered: S1, the source for which the FOV was previously calculated, and S2, the new source for which the FOV is being updated. Each vision-blocking rectangle has two relevant points when considering S1, and two relevant points when considering S2. A ray is cast from each relevant point in the opposite direction of each FOV source. A ray is considered to be visible from the FOV source from which it is cast. The four rays cast from a rectangle are grouped to form two ray pairs such that:

- Each pair is made of one ray directed away from S1 and one ray directed away from S2.
- Two of the rays in a pair share a relevant point and the other two rays either share a relevant point or lie on a common side of the rectangle.

The area between two rays in a pair, and possibly the common side of the rectangle containing their relevant points, is called a cone (see Figure 6(c)). When considered in isolation, a cone contains space which is occluded behind its rectangle from either S1 or S2, but not both. The two cones cast from a rectangle represent space where the visibility status may change when updating the FOV. The point within a cone that is closest to both S1 and S2 is called the origin of the cone. Figure 6 gives an example of how cones are created. From the preceding discussion we obtain the following observations.

Proposition 1: The set of cones cast from all rectangles represent the only space where the visibility status may change when the FOV sources moves from S1 to S2.

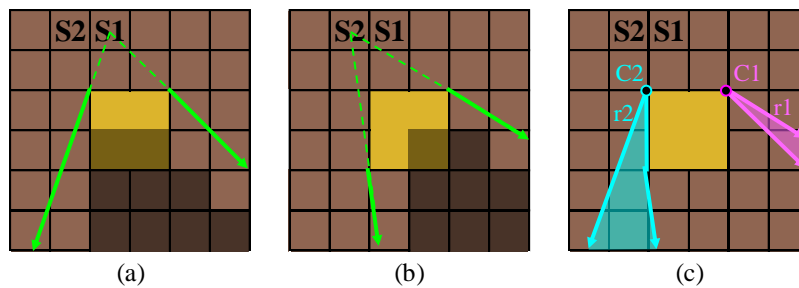


Figure 6. (a) Rays and cell visibility for S1. (b) Rays and cell visibility for S2. (c) The rays pairs forming two cones C1 and C2, with their origin points marked with a circle. The rays forming C1 are cast from the same point, while the rays forming C2 are cast from different points.

When considering a cone in isolation and the rectangle from which that cone is cast, one ray will always be adjacent to space which is visible to both FOV sources. This ray is called an outer ray, while the other is referred to as a cone's inner ray. For example, in Figure 6(c), the two rays r1 and r2 which are further away from the rectangle are the outer rays, and the two other rays are the

inner rays. The outer ray of a cone is the only part of it which is visible to both FOV sources, this property will be used in Section 3.2.

We classify cones into three categories based on the visibility of their origin. If the origin of a cone is visible to both S_1 and S_2 , that cone is fully visible. If the origin of a cone is visible from S_1 or S_2 , but not both, that cone is said to be transitioning visible. If the origin of a cone is neither visible from S_1 nor from S_2 , that cone is said to be not visible. See Figure 7.

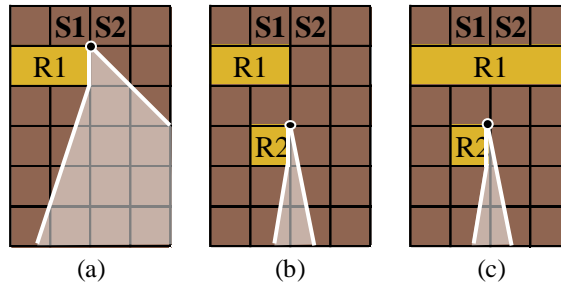


Figure 7. (a) A fully visible cone. (b) A transitioning visible cone. (c) A not visible cone.

3.2. Inverting Cones to Update Cell Visibility

Note that from here onward we assume that the FOV source always moves from its current location to an adjacent cell of the grid, i.e. S_1 and S_2 are assumed to be at the centers of grid cells which share a side. This restriction means that the FOV source cannot move through or around a vision blocking rectangle in a single FOV update. The less common computer game scenario of the FOV source moving to a non-adjacent grid cell can be addressed by performing several FOV updates in which the FOV moves to adjacent cells, or by re-calculating the FOV from scratch if the FOV source moves a large distance.

Inverting a cone C means to invert the visibility status of all grid cells which intersect C , but do not intersect C 's outer ray. Recall that we define a cell as visible to an FOV source if any part of it is visible, even the very edges. This definition is consistent with the definition of visibility used by Recursive Shadowcasting and Rectangle FOV. Figure 8 shows an example of cells whose visibility status change when a cone is inverted.

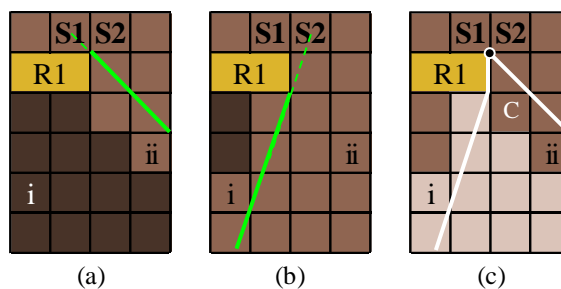


Figure 8. (a) Cells not visible from S_1 are darkened. (b) Cells not visible from S_2 are darkened. (c) The visibility status of the cells highlighted in white is changed when cone C is inverted. Note that cell i just touches the inner ray of C , so its visibility status changes as it is visible only from S_2 . Cell ii just touches the outer ray of C , and so it can be seen by both sources.

We present a series of lemmas below, which will allow us to show that to update the FOV we only need to invert all fully visible cones. Note that the proofs for the lemmas, theorems, and corollaries have been omitted due to space limitations.

Lemma 1: If the origin of a cone C_2 is within another cone C_1 , then C_2 is entirely within C_1 .

Lemma 2: Any part of a non-visible cone C that is visible to S_1 or S_2 must be contained within another cone.

Theorem 1: The FOV can be correctly updated when the FOV source moves from S_1 to S_2 by only addressing cells which intersect the fully visible cones.

3.3. Handling Intersections of Rectangles and Fully Visible Cones

Consider a fully visible cone C and the rectangle R from which it is cast. When considering C and R in isolation, the outer ray of C is visible to S_1 and S_2 , and the rest of C is only visible to one of the FOV sources. However, in an environment with many rectangles, some rectangles may intersect C , causing some parts of C to be visible to neither FOV source. If any rectangle R_i intersects C , then C is shrunk as described below, such that we remove from it any regions occluded from both FOV sources by R_i .

Before it is shrunk, a fully visible cone C is a region C' which is bounded by an outer ray, an inner ray, and possibly a side of R . We call this region C' the expanded cone of C . Note that if C is not shrunk then $C = C'$, otherwise C is contained in C' .

Shrinking C is accomplished by shortening the existing rays of C' and/or by adding new rays. There are three cases of intersection to consider between C and a single rectangle R_i :

- If R_i intersects both rays of C' , then both rays of C' are shortened so that they end as they intersect R_i . C is bounded by these new shorter rays, R_i , and possibly a side of R .
- If R_i intersects only one ray r_1 of C' , then r_1 is shortened so that it ends where it intersects R_i . A new ray r_2 is then added to C , which is directed away from the same FOV source as the inner ray of C' ; r_2 is cast from the relevant point of R_i which is inside of C' . C is then bounded by the shortened ray r_1 , the other ray of C' , R_i , r_2 , and possibly a side of R .
- If R_i does not intersect either ray of C' , then R_i is completely within C' and it effectively splits C' in a manner which is similar to Recursive Shadowcasting. This requires adding two rays to C ; one ray is cast from each relevant point of R_i with respect to the FOV source from which the inner ray of C' is cast, and both rays are directed away from that FOV source. In this case C is bounded by the rays of C' , R_i , the two rays cast from R_i , and possibly a side of R .

Note that for the purposes of cone shrinking, a ray of C is considered to intersect R_i even if it is colinear to a side of R_i , or it contains one of R_i 's vertices.

If multiple rectangles intersect C' , the cone is shrunk using the same process described above, considering the intersecting rectangles in increasing order of distance from the FOV sources. Figure 9 gives an example of a cone intersecting multiple rectangles. Rectangle R_5 intersects two rays as in the first case, rectangles R_3 and R_4 intersect one ray as in the second case, and rectangle R_2 intersects neither rays as in the third case.

An important note to keep in mind is that if any rectangle intersects the outer ray of C' , then the outer ray is shortened as explained above. Even after being shrunk, the outer ray remains the only part of C which is potentially visible to both FOV sources.

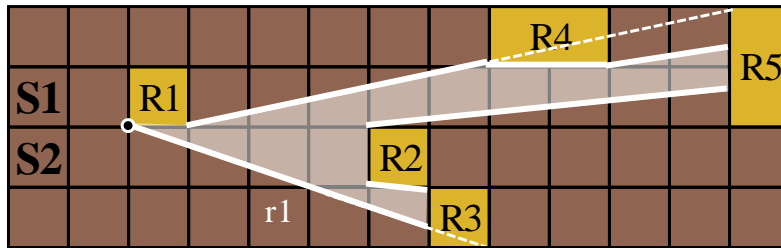


Figure 9. An example of a cone C intersecting multiple rectangles. The solid line and colored regions represent C , while the dotted lines represent the rays of the expanded cone C' . Note that the outer ray $r1$ is shortened due to intersection with $R3$. All rays except for $r1$ are cast from $S2$.

When a cone C , the rectangle R from which C is cast, and all rectangles intersecting C are considered in isolation, each cell intersecting C but not intersecting its outer ray is visible to only one FOV source, while a cell intersecting the outer ray of C is visible to $S1$ and $S2$. However, additional rectangles not intersecting C may occlude parts of C from $S1$ or $S2$; these occluded parts then must be within the intersection of C and other cones. Hence, from the above discussion we derive the following observation.

Proposition 2: Each cell g which intersects a fully visible cone C but not its outer ray is either:

- Visible to only one FOV source.
- Not visible to either source. In this case g intersects the intersection of C with another fully visible cone $C1$, but g does not intersect the outer ray of $C1$.

Each cell g intersecting the outer ray of a fully visible cone C is either:

- Visible to both FOV sources.
- Visible to only one source. In this case g intersects the intersection of C with another fully visible cone $C1$, but g does not intersect the outer ray of $C1$.

Lemma 3: If the outer ray of a cone C is cast from an FOV source S , no part of C , except possibly its outer ray, is visible from S .

We define the inner intersection of two intersecting cones as the intersection of the two cones excluding their outer rays.

Lemma 4: The grid cells completely contained within the inner intersection of two fully visible cones $C1$ and $C2$ do not change visibility status when the FOV source moves from $S1$ to $S2$.

Corollary 1: Three or more fully visible cones cannot have a non-empty inner intersection.

Theorem 2: Inverting all fully visible cones in any order will correctly update the FOV when the FOV source moves from $S1$ to $S2$.

In order to use Theorem 1 to efficiently update the FOV when the FOV source moves, we must be able to quickly determine the visibility of the origin of each cone. This can be accomplished using a line of sight check: If a straight line traced from a cone's origin to an FOV source intersects any rectangles, then that cone's origin is not visible from that FOV source.

We can efficiently identify the visibility status of the origins of cones by ordering the cones such that a fully visible cone always precedes all transitioning visible cones which are inside of it. The quadtree which stores the vision-blocking rectangles can be used to efficiently create the above ordering using a process similar to how rectangles were ordered in Section 2.2. By inverting the cones in this order, we ensure that any fully visible cone will be inverted before considering any transitioning visible cones within it. This enables us to identify and then ignore transitioning visible cones without using lines of sight.

3.4. Inverting Fully Visible Cones

After ordering the cones as described above, we invert the first fully visible cone in the order.

Consider a cone C and the line b that bisects it. If the slope of b is between $-\pi/4$ and $\pi/4$, or between $3\pi/4$ and $5\pi/4$, then C is said to be primarily horizontal. Otherwise C is said to be primarily vertical. To invert a fully visible cone C , we invert the visibility status of the cells which intersect it, except for those cells which intersect its outer ray. If C is primarily horizontal we invert the visibility of each column of grid cells intersecting the cone, one at a time from closest to furthest to the cone's origin. If C is primarily vertical, then we process the cells by rows. We process cells in this order to ensure that if a given cell is part of a vision blocking rectangle R , it will be processed before any of the cells in further rows or columns which may be occluded by R . Below we explain how to process columns of cells of a primarily horizontal cone; a similar process can be used to process rows of cells of a primarily vertical cone.

When processing a column of cells within a cone C , we invert all cells of that column which intersect C , except for any cells which intersect C 's outer ray. After the visibility status for all cells in a column have been inverted, we check if any rectangle R intersects C at that column, as then R would occlude cells in further columns. If an intersection is found, the cone C is shrunk as described in Section 3.3.

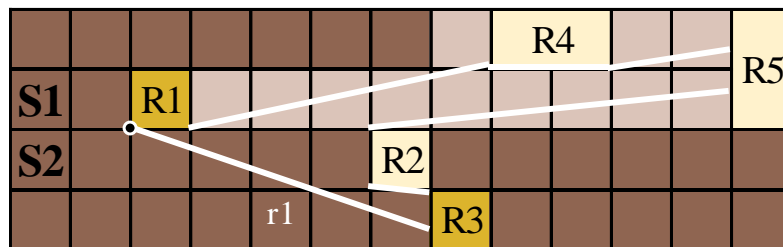


Figure 10: An example of cell visibility inversion for the primarily horizontal cone shown in Figure 9. Cells which have their visibility status inverted (highlighted in white), are all cells which intersect the cone but do not intersect the outer ray $r1$. Note that the cells within $R2$, $R4$, and $R5$ have their visibility status inverted.

Note that while checking for intersection between a rectangle R and a cone C , we can also quickly identify any cones which are inside of C by storing the vertices of R which are inside of C in a hash table H . If any of those vertices are later found to be the origin of a cone, we know that such a cone must be inside of C and therefore it must be transitioning visible.

After completing the first cone inversion as described above, we iterate over the remaining cones in the aforementioned order. If the origin of a cone is in the hash table H , we know that such a cone is transitioning visible and hence we discard it. If the origin of a cone is not in H , we check if that cone is fully visible or not visible. If the cone is fully visible it is inverted as described above, otherwise it is skipped. After iterating through every cone in this manner, we will have

inverted only the cells which are visible from one FOV source, and therefore we will have correctly updated the FOV.

3.5. The FOV Update Algorithm

The steps of the FOV Update algorithm are summarized in pseudocode below. We assume that an FOV has already been calculated and a quadtree of rectangles has been created:

Algorithm: FOV Update (S1, S2, G, Q)

Input: Grid cells S1 & S2, grid G containing FOV from S1, quadtree Q of rectangles

Result: The grid G will contain the FOV from S2

Let T be the set of all cones cast from the rectangles in Q.

Let H be an empty hash table.

for each cone C **in** T, sorted as described at the end of Section 3.3:

if C's origin is not in H **then:**

if the line traced from C's origin to S1 does not intersect any rectangles in Q **then:**

 Invert the cells of G within C as described in Section 3.4.

 Store in H the relevant points of the rectangles in L that are inside of C.

4. EXPERIMENTAL EVALUATION

We present an experimental evaluation of Recursive Shadowcasting, Rectangle-Based FOV, and FOV Update. All tests were run on a computer with an Intel Xeon E5-2683 processor and 24 gigabytes of system memory. Our algorithm implementations were written in C++, compiled using GCC 8.3.1, and run under Linux Kernel 5.6.8.

We show test results for four environments which are meant to emulate terrain which may appear in a computer game:

Environment 1: A fixed indoors environment made of 160 rectangles, with 36 square rooms connected by 74 corridors. This environment is constructed such that there is never an alignment of corridors which would allow the FOV source to see across many rooms. This is an enclosed environment where many cells and rectangles/cones will be occluded.

Environment 2: A randomized environment where 200 rectangles of random sizes are placed at random positions on the FOV grid. This simulates a less "structured" environment, such as a forest. Each rectangle has a uniformly distributed random width and height between one and six cells. The position of each rectangle is chosen uniformly at random from all locations that do not intersect another rectangle.

Environment 3: A randomized environment where 200 rectangles of random sizes are densely grouped around the center of the FOV grid and fewer rectangles appear further from the center. This simulates a more organized environment, such as a town. Each rectangle has a uniformly distributed random width and height between one and six cells. The position of each rectangle is chosen using a random distribution which results in more positions near the center of the grid.

Environment 4: A fixed environment which uses 300 rectangles to emulate the visibility grid used in League of Legends [1]. This tests the FOV algorithms using an environment taken from an existing game that includes a mixture of enclosed spaces and large open pathways.

For the four above environments (shown in Figure 11) we tested the algorithms with 25 randomly generated paths of 100 cells each. Each path was constructed by randomly selecting a non-vision

blocking starting cell and a random direction. The starting cell and selected direction define a ray. Cells which intersect this ray were added to the path in ascending order of distance from the starting cell. If the next cell to add to the path belongs to a vision-blocking rectangle, a new random direction was generated and further cells were added to the path using that new direction. This continued until the path contained 100 cells.

The FOV was calculated for each cell along the path in the order in which the cells appear in the path. We used paths in these test environments to mimic scenarios arising in computer games, where the FOV source moves through the grid following a game character. In the case of FOV Update, for each path we first computed an initial FOV using the Rectangle-Based FOV algorithm, and then measured the running time of updating the FOV for every position of the FOV source along the path. Each test environment uses a fixed number of rectangles; as the grid size increases the sizes of the rectangles is increased by the same proportion.

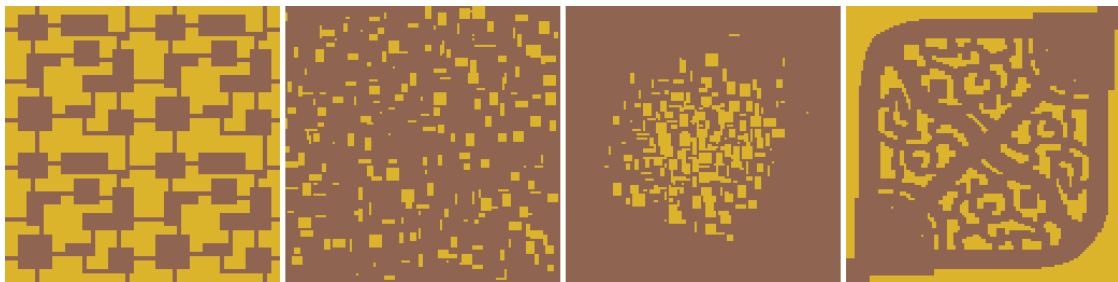


Figure 11. Environments 1, 2, 3, and 4 on a grid of size 128*128.

Table 1. Running times for Environment 1.

Smallest mean running times and smallest standard deviations are highlighted in blue.

Grid Size	Shadow		Rectangle		Update	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
128*128	6.5 μ s	1 μ s	205 μ s	20 μ s	170 μ s	24 μ s
256*256	21 μ s	3 μ s	259 μ s	26 μ s	174 μ s	25 μ s
512*512	80 μ s	14 μ s	401 μ s	39 μ s	188 μ s	27 μ s
1024*1024	290 μ s	43 μ s	774 μ s	68 μ s	204 μ s	46 μ s
2048*2048	1,342 μ s	278 μ s	2,001 μ s	163 μ s	249 μ s	77 μ s
4096*4096	6,665 μ s	1473 μ s	10,269 μ s	765 μ s	356 μ s	140 μ s

Environment 1 is designed to have a relatively low number of visible cells from any FOV source position. The low number of visible cells is an advantage for Recursive Shadowcasting and a disadvantage for Rectangle FOV. Rectangle FOV is only about 50% slower than Recursive Shadowcasting at large grid sizes, despite having to assign many more cell visibility statuses, as Rectangle FOV starts with all cells set to visible. This shows that when compared to Recursive Shadowcasting, Rectangle FOV's more efficient method of cell assignment makes a very significant difference to running time, even in environments where Rectangle FOV has to assign many more cell visibility statuses than Recursive Shadowcasting.

The low number of visible cells in this environment is also an advantage to FOV Update, as it results in a high number of non-visible cones and a low number of cells which change visibility status as the FOV source moves. Because of this, FOV Update is faster than Rectangle FOV and it is up to 20 times faster than Recursive Shadowcasting at large grid sizes. FOV update's running

time is not strongly affected by grid size in this case due to the low number of cell visibility assignments that it needs to make.

The running times for Environment 1 have the lowest standard deviations of Environments 1-4. This is expected as Environment 1 has a low variance in the number of cells visible from any point, as the shape of the environment ensures that only one room can be fully visible regardless of where the FOV source is positioned. Because of this, the running times of the algorithms are not as strongly affected by the position of the FOV source.

Table 2. Running times for Environment 2.

Grid Size	Shadow		Rectangle		Update	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
128*128	17 μ s	6.5 μ s	300 μ s	49 μ s	468 μ s	137 μ s
256*256	54 μ s	16 μ s	358 μ s	52 μ s	504 μ s	135 μ s
512*512	201 μ s	53 μ s	494 μ s	77 μ s	595 μ s	152 μ s
1024*1024	777 μ s	289 μ s	943 μ s	172 μ s	763 μ s	243 μ s
2048*2048	3,898 μ s	1,747 μ s	2,176 μ s	277 μ s	1,073 μ s	366 μ s
4096*4096	19,345 μ s	8,426 μ s	7,347 μ s	1,059 μ s	1,863 μ s	821 μ s

Recursive Shadowcasting has larger running times in Environment 2 when compared to Environment 1, as Environment 2 has many more visible cells. Due to this increased running time, it is slower than Rectangle FOV when grid size is large. When comparing Rectangle FOV's running times between Environments 1 and 2 the running times for Environment 2 are slower at most grid sizes, but are faster at grid size 4096*4096. At lower grid sizes Rectangle FOV's running time is primarily determined by the number of rectangles, and so its running time is slower than in Environment 1 as more rectangles are present in Environment 2. At 4096*4096 Rectangle FOV's running time is primarily determined by its efficiency of assigning cell visibility statuses, and so it is faster than in Environment 1 as fewer cell visibility statuses need to be assigned in Environment 2.

Many more cells may change visibility status in Environment 2 than in Environment 1 when the FOV source moves, and so FOV Update's running time is more significantly affected by grid size than in the previous environment. FOV Update is much faster than the other two algorithms as grid size becomes large, as it changes the visibility status of fewer cells.

Table 3. Running times for Environment 3.

Grid Size	Shadow		Rectangle		Update	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
128*128	25 μ s	9.7 μ s	272 μ s	35 μ s	471 μ s	138 μ s
256*256	83 μ s	35 μ s	314 μ s	43 μ s	466 μ s	142 μ s
512*512	343 μ s	169 μ s	431 μ s	64 μ s	489 μ s	146 μ s
1024*1024	2,132 μ s	809 μ s	832 μ s	117 μ s	676 μ s	173 μ s
2048*2048	11,529 μ s	5,592 μ s	2,072 μ s	226 μ s	969 μ s	269 μ s
4096*4096	46,203 μ s	25,962 μ s	6,710 μ s	1,007 μ s	1,331 μ s	539 μ s

The running times of Recursive Shadowcasting are much higher for Environment 3 than for Environment 2 because the clustering of the vision blocking rectangles results in a high number of visible cells when the FOV source is not in the center of the grid, which causes Recursive Shadowcasting to calculate the visibility status of many cells. This also explains the high standard deviation of Recursive Shadowcasting, as visibility is low if the FOV is near the center of the grid, and high otherwise. FOV Update's faster running times here than in Environment 2 are due to the clustering increasing the number of rectangles that occlude each other. This reduces the number of cones that the algorithm needs to process.

Table 4. Running times for Environment 4.

Grid Size	Shadow		Rectangle		Update	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
128*128	13 μ s	6.5 μ s	403 μ s	57 μ s	558 μ s	220 μ s
256*256	46 μ s	24 μ s	482 μ s	78 μ s	566 μ s	223 μ s
512*512	163 μ s	75 μ s	656 μ s	100 μ s	590 μ s	219 μ s
1024*1024	844 μ s	468 μ s	1,173 μ s	210 μ s	687 μ s	328 μ s
2048*2048	4,157 μ s	2,780 μ s	2,643 μ s	472 μ s	802 μ s	432 μ s
4096*4096	22,007 μ s	13,698 μ s	8,692 μ s	1,724 μ s	1,247 μ s	765 μ s

Despite Environment 4 being more enclosed than Environments 2 and 3, Recursive Shadowcasting still performs poorly here. Because the large open pathways in this environment can result in a high number of visible cells. Recursive Shadowcasting's running time is primarily determined by the number of visible cells. This also explains the high standard deviation of Recursive Shadowcasting, as if the FOV source is not in a pathway, but in one of the four clusters of rectangles, then there will be relatively few visible cells. Both Rectangle FOV and FOV update perform similarly to Environment 3 here. This makes sense as Environment 4 also involves many clustered rectangles which may occlude each other.

5. CONCLUSION

In this paper we presented two new algorithms for calculating Field of Vision over 2D grids, with the goal of making FOV calculation feasible at high grid sizes. Rectangle FOV accomplishes this by representing vision-blocking terrain in a compact and efficient manner. FOV Update uses this compact representation and a previously calculated FOV to calculate a new FOV with a minimal number of cell visibility status assignments. We then compared these algorithms to Recursive Shadowcasting, the previously fastest FOV algorithm. From this experimental evaluation, we made the following observations:

Our algorithms address the deficiencies of Recursive Shadowcasting when many cells are visible; however, our algorithms have limitations of their own when few cells are visible or when an FOV has not yet been calculated. When calculating an FOV from scratch, Recursive Shadowcasting performs best at low grid sizes and enclosed environments, while Rectangle FOV performs best at high grid sizes. Our FOV update algorithm is superior to the other two algorithms at medium and high grid sizes, but an FOV must be calculated first. Because of this, there is no universal best FOV algorithm.

Based on our experimental results, we recommend the use of a combination of algorithms if fastest computation of the FOV is desired:

- An FOV should be updated using **FOV Update** when grid size is above 512*512. At lower grid sizes, or when FOV needs to be calculated from scratch, one of the below algorithms should be used.
- An FOV should be calculated with **Recursive Shadowcasting** when grid size is 512*512 or lower, or when the environment is very “enclosed” (e.g. Environment 1).
- An FOV should be calculated with **Rectangle FOV** when grid size is above 512*512 and the current environment is not very “enclosed” (e.g. Environments 2-4).

However, not all implementors will care about minimizing average running time and would perhaps prefer to minimize the chance that FOV calculation takes long enough to be problematic. Recall that, as discussed at the end of Section 1.1, computer games generally compute a new display frame roughly every 17ms, though this time can be as low as 4ms on a modern display. FOV does not need to be calculated for every frame, but its computation must be fast enough to not delay frame rendering or starve other game processes of system resources.

Our algorithms have much more consistent performance than Recursive Shadowcasting due to their running time not being as strongly affected by the number of visible cells. Recursive Shadowcasting’s highest mean running time was 46ms (with a very high standard deviation), Rectangle FOV’s was 10ms, and FOV Update’s was 1.8ms. Therefore, if an implementor would prefer to ensure that the running time of the FOV algorithm is always adequate, our algorithms can be used in all cases. The initial FOV calculation can be performed with Rectangle FOV, and subsequent FOV calculation can be performed with FOV Update in. Additionally, the initial FOV calculation can occur while the game is loading an environment and displaying a loading screen, which would eliminate the possibility for Rectangle FOV’s running time to cause issues.

When evaluating these algorithms we tried to not make assumptions about how a particular game may use the FOV, so there is room for further improvements when certain assumptions are made. For example, the running times of Rectangle FOV and FOV Update might be improved in games with environments similar to Environment 1 or Environment 4 by using portal-based occlusion culling. Portal-based culling would allow these algorithms to determine which portions of a more “structured” environment are not visible, which would allow the algorithms to skip many visibility checks and avoid resetting the entire FOV grid. Portal based culling is likely to not be effective in less “structured” environments such as Environment 2 or Environment 3, where the added overhead might result in increased running time.

ACKNOWLEDGEMENTS

Authors partially supported by the Natural the Natural Sciences and engineering Research Council of Canada grants 04667-2015 RGPIN and 06423-2020 RGPIN

REFERENCES

- [1] League of Legends. [Online]. Riot Games, 2020. Available: <https://na.leagueoflegends.com/en/>
- [2] Jung, J. (2016), “A story of fog and war”, [Online]. Available: engineering.riotgames.com/news/story-fog-and-war [Accessed Sept. 12, 2020]
- [3] Defense of the Ancients 2. [Online]. Valve Software, 2020. Available: <http://blog.dota2.com>
- [4] Bergström, B. (2001), “FOV using recursive shadowcasting”, [Online]. Available: www.roguebasin.com/index.php?title=FOV_using_recursive_shadowcasting [Accessed Sept. 12, 2020]

- [5] Duerig, J. (2007), "Precise permissive field of view", [Online]. Available: http://www.roguebasin.com/index.php?title=Precise_Permissive_Field_of_View [Accessed Sept. 12, 2020]
- [6] Scherzer, D., Wimmer, M., & Purgathofer, W. (2011) "A survey of real-time hard shadow mapping methods", *Computer graphics forum*, Vol. 30, No. 1, 169-186.
- [7] Liu, N., & Pang, M. Y. (2009) "A survey of shadow rendering algorithms: Projection shadows and shadow volumes", 2009 Second International Workshop on Computer Science and Engineering, Vol. 1, 488-492.
- [8] Moreau, P., Pharr, M., & Clarberg, P. (2019) "Dynamic many-light sampling for real-time ray tracing", *High-Performance Graphics - Short Papers*, 21-26.
- [9] Theoharis, T., Papaioannou, G., & Karabassi, E. A. (2001) "The magic of the z-buffer: A survey", *Ninth International Conference in Central Europe on Computer Graphics Visualization and Computer Vision*, 379-386.
- [10] de Berg, M., Cheong, O., van Kreveld, M., & Overmars, M. (2008) "Binary space partitions: The painter's algorithm", *Computational Geometry: Algorithms and Applications*, 259-281.
- [11] Assarsson, U., & Moller, T. (2000) "Optimized view frustum culling algorithms for bounding boxes", *Journal of Graphics Tools*, 5(1), 9-22.
- [12] Luebke, D., & Georges, C. (1995) "Portals and mirrors: Simple, fast evaluation of potentially visible sets", *Proceedings of the 1995 Symposium on Interactive 3D graphics*, 105-106.
- [13] Jice. (2009), "Comparative study of field of view algorithms for 2D grid based worlds", [Online]. Available: http://www.roguebasin.com/index.php?title=Comparative_study_of_field_of_view_algorithms_for_2D_grid_based_worlds [Accessed Sept. 12, 2020]
- [14] Debenham, E., Solis-Oba, R. (2019) "New Algorithms for Computing Field of Vision over 2D Grids" *Electronic Thesis and Dissertation Repository*. 6552. The University of Western Ontario. Available: <https://ir.lib.uwo.ca/etd/6552> [Accessed Sept. 12, 2020]
- [15] Ohtsuki, T. (1982, May) "Minimum dissection of rectilinear regions", *IEEE International Symposium on Circuits and Systems*, 1210-1213.
- [16] Samet, H. (1984). "The quadtree and related hierarchical data structures", *ACM Computing Surveys*, 16(2), 187-260.

A MULTILEVEL TECHNOLOGY ACCEPTANCE MANAGEMENT MODEL

Gilbert Busolo, Lawrence Nderu and Kennedy Ogada

School of Computing and Information Technology, Department of Computing,
Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

ABSTRACT

Knowledge is a strategic resource for successful data driven decision making in any organization. To harness this knowledge, successful adoption of a technological intervention is key. Institutions leverage on technology to drive knowledge management (KM) initiatives for quality service delivery and prudent data management. These initiatives provide the overall strategy for managing data resources. They make available knowledge organization tools and techniques while enabling regular updates. Derived benefits of positive deployment of a technological intervention are competency enhancement through gained knowledge, raised quality of service and promotion of healthy development of e-commerce. Successful and timely adoption of technological interventions through which knowledge management initiatives are deployed remains a key challenge to many organizations. This paper proposes a wholesome multilevel technology acceptance management model. The proposed model takes into account human, technological and organizational variables, which exist in a deployment environment. This model will be vital in driving early technology acceptance prediction and timely deployment of mitigation measures to deploy technological interventions successfully.

KEYWORDS

Model, Technology Acceptance, Knowledge, Management, Multilevel

1. INTRODUCTION

There is timely realization of knowledge as a strategic resource for success in any organization. Its subsequent adoption through a knowledge management system (KMS) is key. Quite palpable, there is literature that manifests in-depth research on science and technological products. This is coupled by the desire of organizations to increase in quality of service, efficiency and optimized operations. It has led to rapid development and integration of KMS as they are now a core part of our livelihood. As technological generations change, new products whose aim is to enrich our living experience through automation, knowledge sharing, simplified and quality decision-making with feedback systems are developed. These have a rich digital repository for reference of the vast amount of information that bombard us from every angle. Uriarte (2008) defines knowledge management (KM) is the process through which individuals, organizations and even Governments generate value from their intellectual and knowledge-based assets. The process of knowledge creation is defined by stepwise and continuous evolution, not one that can be compactly planned and controlled. Emergent, motivation in which inspiration and pure change play is an important role in many situations. In addition, it has been widely accepted among scholars that social processes heavily influence organizational knowledge creation.

Kwanya, Kiplang'at & Wamukoya (2017) and supported by Fernandez & Sabherwal (2010) state that the most vital resource of today's enterprise is the collective knowledge residing in the minds

of an organization's employees, customers, and vendors. Learning how to manage organizational knowledge has many benefits, some of which are readily apparent, others are not. These benefits are not limited to leveraging on core business competencies, accelerating innovation and time-to-market while improving cycle times and decision-making. They also strengthen organizational commitment by helping to build sustainable competitive advantage in service delivery.

From a global perspective, Talisayon (2013) examines the role and success of KMS. He notes that organizations worldwide face technological challenges as they continue to evolve into an electronic work environment. This is pushed by paperwork and cost reduction mandates. All this is mirrored to requirements pointing towards increased workloads with fewer personnel and the rapid addition of electronic communication channels for use. Some of the common challenges that affect the service delivery sector include enhancing efficiencies across all agencies, improving accountability of resources and making data driven decisions. In addition, enhancing collaboration and strategic partnerships with stakeholders, capturing knowledge of an aging workforce as well as improving operational excellence cannot be ignored. It is noteworthy to keep in mind that computerized knowledge management systems plays an imperative role in providing strategies and techniques to manage institutional content, making knowledge more accessible and usable. They provide the overall strategy to manage organizational data resources by making available knowledge organization tools and techniques. This aids in monitoring knowledge contents while making sure they are accessible and updated accordingly.

Among the derived benefits of knowledge management are enhanced competence, raising quality of service, and promotion of healthy development of e-commerce. Institutions and governments leverage on technology to provide long-term development blueprint for attaining global competitiveness. ICT as an enabler or foundation for socio economic transformation recognizes the role of science, technology and innovation in modern economy. Through this, new knowledge plays an essential role in boosting wealth creation, social welfare and international competitiveness. However, this benefit is eroded through complexities experienced when users have challenges embracing new technologies. These efforts to embrace ICT tools in KM, encapsulating the desire to gain a competitive edge through data driven decision making for growth and development are at risk. Odinioha & Chukwuma (2013) underscore the pivotal importance of quality information management in any organization. This is encouraged through focused acquisition, storage, dissemination and its utilization. It is evident that acquisition and application of a KMS does not guarantee its acceptance and effective use. According to Maier (2007), he states that even though ICT can be the enabling factor in a knowledge management initiative, there are limits to its use and the installation of a good platform does not always guarantee success. The sure assurance that any technological implementation is positively received and adopted is a hard question in the hearts and minds of many implementers. Great care is given to the product development life cycle through adoption of existing system development life cycle models. However, the greatest challenge and hindrance is acceptance of these technologies by the end users. Feedback in the public domain shows several initiatives that leverage on technology have fallen short notwithstanding the resources invested in their implementation.

This paper proposes a multilevel technology acceptance model. The proposed model takes into account human, technological and organizational variables which exist in a deployment environment. This model is key in driving early technology acceptance prediction and timely deployment of mitigation measures for successful deployment of a technological intervention. This aims at shedding more light and demystifying the technology implementation and acceptance process. Secondly, the researchers proposes new constructs that can be considered in a technology implementation domain. These constructs mirror emerging trends and realities such as mandatory systems and e-government in a technology acceptance model.

2. RESEARCH METHODOLOGY

This process involves research and collection of data from several academic and scientific literature. Where possible, priority was given to sources that had published their data in the last five years. This is in cognizance to the fact that Information Communication Technology world rapidly changes and the foundations of this research are anchored on current pieces of information and practice. Constructs derived from the literature survey provided in-depth understanding of the technology acceptance process.

2.1. Technology Adoption

A decade old challenge in the technological world is gaining understanding as to why people accept or reject computer systems, as these are the basic infrastructure of any system. According to Davis, Bagozzi & Warshaw (1989), this has proven to be one of the most challenging issues in information systems (IS) research. Investments in ICT-based tools that support planning, decision-making, and communication processes are inherently risky. Unlike clerical paperwork processing systems, these newer computing tools and systems often require managers and professionals to interact directly with hardware and software. However, end-users are sometimes unwilling to use available computer systems that, if used, would generate significant performance gains. The acceptance and use of information systems and information technology innovations has been a major concern for research and practice. Over the last several decades, a plethora of theoretical models have been proposed and used to examine technology acceptance and usage, Dwivendi, Rana, Jeyaraj, Clement & Williams (2019). Forward-looking institutions and governments have identified ICT as an enabler and foundation for socio economic transformation and a tool to leverage on for success. In the quest of service delivery, key and focused ICT initiatives that leverage on acceptance of technological interventions are taking shape. In Kenya, adoption of technology as a service enabler is experienced at Huduma Centres. This is a e-government initiative whose aim is to provide a one-stop shop service for government products to its citizens. Additionally other technological interventions like Kenya Revenue Authority (KRA) Integrated Tax Management System (ITMS) and E-Citizen platform enable the public access and pay for Government services online. All these technologies aim to enhance service delivery and provide a citizen centric one-stop non-stop model of service.

Hwang, Al-Arabiati & Shin (2016) note that as new technologies, processes, procedures and systems infiltrate the world, research on potential adopters and acceptance of innovations receives attention from professionals and academic researchers. Developers of new technologies, senior management, and those responsible for managing the changes associated with the implementation of innovations are increasingly realizing that the lack of user acceptance can and most probably will lead to loss of money and resources. This greatly affects an organization's financial standing. It is evident that change creates a sense of uncertainty and lost control. Employees' resistance and lack of support are some of the most cited causes for failures associated with organizational change. This resistance represents a major barrier to organizational benefits derived from adoption of technological innovations. The most important concern for any manager is to ensure user buy-in through adoption of various techniques that ensure a positive attitude and outlook towards technology. This greatly affects its usage and success. Identification and adoption of correction measures that dissuade low automation levels and provide efficiency has led to various research initiatives on models that speak to the technology acceptance process. These aim to demystify complexities experienced when users have challenges embracing new technologies. This trend is observed as institutional systems around the world are incorporating technology as a core enabler towards growth, productivity and global competitiveness, Scherer, Siddiq & Tondeur (2018).

Great strides in technological investments and interventions are evident. Lessons learnt from previous failures in acceptance and adoption of these technological interventions have turned the focus to technological models that deal with prediction of user behavior. This gives out data that informs the decision whether to adopt the technology, apply modifications to it or reject it in its totality, Aizstrauta, Ginters & Eroles (2015). Discourse about benefits derived through adoption and acceptance of technology for any institution has been positively framed with a view that technology predominantly empowers societies. These are through savings on time, reduced costs and enhanced possibilities. Lowe, Dwivedi & D'Alessandro (2019). Technology acceptance models strive towards deconstructing user attitude and behavior that consequently affect acceptance, use, performance and ultimately success or failure of any new technology.

2.2. Technology Adoption Models

The acceptance and increased utilization of technological innovations are crucially beneficial for both the service provider and user during the whole process of engagement, Gucin & Berk (2015). Maier (2007) notes that success of any information system cannot be measured directly, but has to be assessed using a number of measures which are relevant for success. He notes that since the 70s, many authors have developed approaches to assess the success of an IS through several proposed variables, indicators and measures. As the researchers ponder on the time long challenge of technology rejection, this concern led to the study of several models that aim to explain computer acceptance and usage behavior. A technology adoption model explains the acceptance and consequent use of a technological intervention.

2.2.1. Theory of Reasoned Action

Theory of Reasoned Action (TRA) is an extensively studied model from the field of social psychology. It greatly concerns itself with the determinants of consciously intended behaviors. It deduces that a person's performance of a specified behavior is determined by his or her behavioral intent to perform the behavior and this is jointly determined by the person's attitude and subjective norm concerning the behavior as first studied by Davis, Bagozzi and Warshaw and cited in Lai (2017). In 1967, Martin Fishbein first introduced TRA, Ajzen & Fishbein (1980) with an aim of bringing clarity to behavior of an individual in a voluntary environment. It picks beliefs, attitudes, intentions and behaviors as the main variables, articulating their relationship amongst one another in determining the end motivation that leads to an individual's action. Being a well-researched but general intention model, Lai (2017) that has been quite successful in the prediction and enlightenment of human behavior in a variety of domains, it should therefore be an appropriate choice model for reviewing computer usage determinants. Quite interestingly, TRA being a general model as shown in Figure 1 does not specify the beliefs that are operative for a particular behavior. To adopt it, researchers first identify beliefs that are salient for subjects regarding the behavior under investigation. Davis, Bagozzi & Warshaw (1989) continue to clarify that the model proposes five to nine salient beliefs, gathered through free response interviews with representative members of the subject population. These beliefs are most frequently elicited from a representation of the sample population considered for use.

Greene (2009) expounds that the development of TRA was to explain influences of several variables on behaviors that involve non-mandatory decision-making. Behaviors that are impulsive, habitual, or scripted are specifically excluded. TRA is not advised to be deployed where expounding of behaviors that are mandatory in nature like a frequent traveler getting through airport security but the theory is aptly applied to behaviors that the individual has some level of choice on the type of decision to be made. This aspect has been noted as a limitation to the theory as its focus and scope of application is on voluntary behavior.

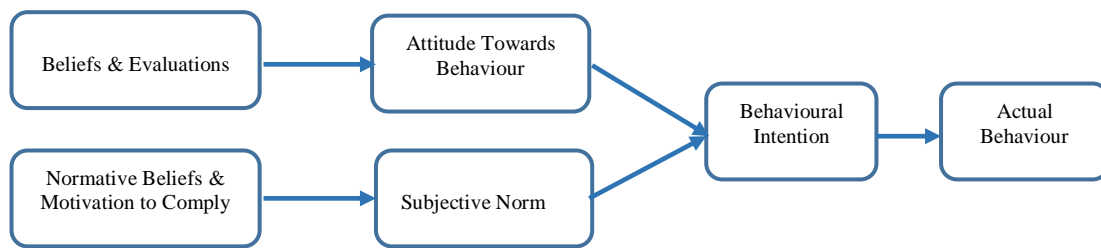


Figure 1: Theory of Reasoned Action (Lai, 2017)

$$\text{Person's Performance} = \text{Behavioral Intent (BI)}$$

According to Greene (2009), Behavioral Intent (BI) that revolves around a person's plans, motivations or desires are the immediate forecaster of how one will behave. An individual's plan or activity revolves around an intention to accomplish something. A precursor of the proposed result is planning and accomplishing activities that make way for smooth execution of the main plan. A good analogy is the intent to further one's education is precipitated by the desire to increase in skill and consequently have a better life. When broken down, it entails looking for a good institution of higher learning, availing resources for the venture, prioritizing and rescheduling one's activities for the duration of the course, applying for the course among other foundational undertakings. The intent component greatly affects person's performance and choice of course as the theory can be applied in a KM environment to gauge the level of acceptance of a proposed technological intervention. As intentions can change depending on other conditions, it is a good component to tap on for positive behavior change, influencing an individual to undertake a positive action.

$$\text{Behavioral Intent (BI)} = \text{Attitude (A). Subjective Norm (SN)}$$

Greene concurs that intentions result from underlying attitudes and subjective norms therefore they are not independent. An attitude being a general orientation toward a behavior based on a variety of beliefs and evaluations is determined by identification of a set of relevant beliefs, measuring their strength or certainty of these beliefs, and measuring their evaluation. Then the measures are summed together resulting in an attitude measurement. Attitudes are specific to performing a particular behavior. To determine an attitude, belief strength and evaluation are empirically weighted for a specific behavior and group before they are added together. These relative weights are discovered through surveys of individuals in the target audience about a behavior. The importance of attitudes, as determined through surveys, is helpful in designing intervention drives to address the component that best predicts behavioral intentions. Subjective norms being the social component of behavioral intentions are composed of normative beliefs that are influenced by the view of other people concerning that behavior and the drive to comply with the directive or task at hand. Depending on the issue at hand, influence may be heavier from one quarter than from another quarter hence influencing the direction of action of the person. The pressure to be in good standing with a group of people or a certain individual has great impact in a decision.

$$\text{Attitude towards Behavior (A)} = \sum \text{Salient Beliefs (b1). Evaluations (e1)}$$

Salient behavioral beliefs held to determine the attitude toward the behavior and the envisioned consequences of performing that behavior. Salient normative beliefs (beliefs about the views of significant others) are held to determine subjective norm. Salient control beliefs (beliefs about

factors that may facilitate or impede performance of the behavior) are assumed to determine perceived behavioral control. Salient beliefs are those that first come to an individual's mind as repercussions of performing a certain act or engaging in a specific activity, Sutton et al. (2003). The behavioral beliefs can be either positive or negative. An individual can believe that technology will ease the process of doing work but may also believe that computer use will affect his eyesight or even increase chances of getting cancer. Evaluations are appraisals taken by the individual attached to a behavioral outcome or attribute. This includes negative or positive views about the behavior and greatly contributes to attitude along with behavioral beliefs.

$$\text{Subjective Norm (SN)} = \sum \text{Normative Beliefs (nb1). Motivation to Comply (mc1)}$$

Davis, Bagozzi and Warshaw (1989) and Lai (2017) agree that an individual's subjective norm is determined by a multiplicative function of his or her normative beliefs that are guided by perceived expectations of specific referent individuals or groups. This is greatly determined by social pressure and affects his or her motivation to comply with these expectations. This is influenced by the presumed perception of what others think of the issue at hand. Fishbein and Ajzen (1975) rank TRA as a general model as it does not specify the belief parameters that are operative for a particular behavior. In its application, researchers have first to undertake the task of identifying the beliefs that are outstanding for subjects regarding the behavior under investigation. They do recommend researchers to identify and engage with a minimum of five and a maximum of nine salient beliefs using free response interviews to elicit data with representative members of the subject population. It is recommended that the beliefs most commonly elucidated from the sample population be used. During research and application of TRA, its main drawbacks arise from lack of addressing the role of habit, cognitive deliberation and usage voluntariness, Taherdoost (2018).

2.2.2. Technology Acceptance Model

Technology Acceptance Model (TAM) is a less general adaptation of TRA introduced in 1986 by Davis and heavily researched in this day and age, Hwang, Al-Arabi & Shin (2016). It has a specific application area for modelling user acceptance in an information systems environment as an early stage user acceptance tool. This model is envisaged to provide an understanding of the parameters that influence computer acceptance that are general, capable of explaining user behavior across a broad range of end-user computing technologies and user populations, while at the same time being both parsimonious and theoretically justified. According to Davis, Bagozzi & Warshaw (1989) and Hwang et al. (2016), TAM posits that two particular beliefs, perceived usefulness and perceived ease of use are of primary relevance for computer acceptance behaviors as shown in Figure 2 below.

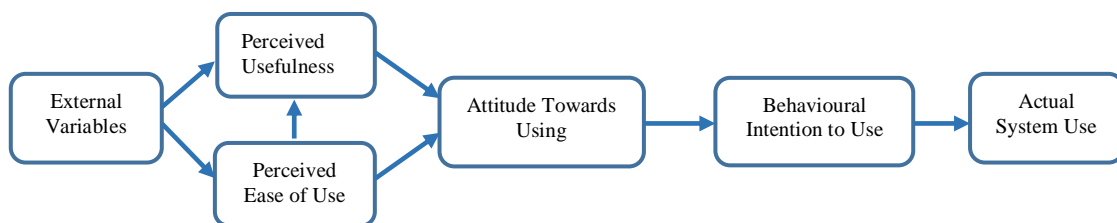


Figure 2: Technology Acceptance Model (Hwang, Al-Arabi & Shin, 2016)

TAM model revolves around these two parameters to determine attitude towards using a technological intervention. This is underscored by Scherer et al. (2018) pointing out that perceived usefulness and perceived ease of use are considered as the key variables of TAM as

they generally describe the outcomes when applied in a technological environment. These are perceived usefulness and perceived ease of use as the main determinants towards technological acceptance behavior as illustrated below.

$$\text{Attitude (A)} = \text{Perceived Usefulness (U)} + \text{Perceived Ease of Use (E)}$$

Davis, Bagozzi & Warshaw (1989) defines the two major TAM identifiers, Perceived Usefulness (U) as the prospective user's subjective probability that using a specific application system will increase his or her job performance within an organizational context. Perceived Ease of Use (EOU) refers to the degree to which the prospective user expects the target system to be free of effort. The Perceived Usefulness and Ease of Use are intertwined and affect each other because when a user positively identifies and feels that a system is easy to use to achieve the set organizational goals, then the system becomes useful for them, hence having a positive attitude towards technology and influencing behavior.

$$\text{Behavioral Intention to Use (BI)} = \text{Attitude towards Using (A)} + \text{Perceived Usefulness (U)}$$

The relationship between Attitude (A) and Behavioral Intention (BI) commonly referred to as The A-BI relationship implies that when all other factors are constant, system users will form an intention to use a technological intervention, which they perceive to have a positive effect on them. Attitude is recognized as a major factor that influences technology acceptance or rejection during deployment and subsequent use. The U-BI relationship between Perceived Usefulness (U) and Behavioral Intention (BI) is grounded on the idea that in an organization setting, users form an intention on behavior that they believe when undertaken, will generally increase their work performance over any other positive or negative perception evoked. The base reason being that achieving good results in such an environment like a work environment is instrumental to acquiring other rewards that are extrinsic to the content of the work itself, such as better pay or a higher grade through promotions.

$$\text{Perceived Usefulness (U)} = \text{Perceived Ease of Use (E)} + \text{External Variables}$$

Any technological intervention deployed that offers an easy to use interface, defined and logical workflows and systems that enable the user to accomplish more work with less effort is perceived as easier to use. One that outputs high quality graphs with detailed precision scores better. This enables improved performance when leveraging on technology and directly affects perceived usefulness. Additional features that ease the interaction between the system functionalities and the user while achieving the set system objectives enhance usability. When the external variables in the operating environment are conducive, optimally supporting technological interventions, the expected output will be quite positive. External factors in the operating environment that impinge on the optimum working atmosphere and environment hinder the perceived usefulness of that technology and by large the user may form a negative attitude towards technology.

$$\text{Perceived Ease of Use (E)} = \text{External Variables}$$

There is a direct correlation between Perceived Ease of Use (E) and External Variables that are present in the deployment environment. External Parameters like experience working on computer systems, years spent on automated systems, exposure to technology among others greatly influence the user's perspective on how easy or difficult it is to use the system. A system which is positively evaluated is more acceptable and useable, Hwang et al. (2016). Davis (1993) encourages more research on the role of additional TAM variables for the model to be more responsive to technological environments that have more parameters that greatly influence the acceptance and deployment of technology. Parameters like choice in a mandatory environment is

a key variable that influences attitude as the user has to interface with the system for daily work irrespective of his attitude towards it. One of the greatest drawback of TAM is its inability to inform the implementation process, Hwang et al. (2016). It leans heavily towards user acceptance, leaving out system designers and project implementers. For TAM to continue being relevant to modelling user interactions with technology, researchers are called to consider it as a basic model that offers the benefit and flexibility of integrating extended and contextualized motivational influences and user behaviors based on emerging realities in contemporary technological environments, Lim (2018). For clarity in understanding and adoption of TAM, the model should be augmented with new variables that are context specific to the application environment and mirrored with the times, Lowe et al. (2019) to provide more consistent prediction of system use, Taherdoost (2018).

2.2.3. Task Technology Fit Model

Goodhue and Thompson as cited by Lai (2017) ponders on the linkage between information technology adoption and its corresponding effect on individual performance on a given a task. The Task Technology Fit Model (TTF) as shown in Figure 3 below underscores the degree to which a technology deployment assists an individual in performing his portfolio of tasks. An important factor to note with TTF is that there is little explicit recognition that more utilization of a system will not necessarily lead to higher performance. Utilization of a poor information system will not improve an individual's performance. TTF draws its constructs from two assertions; utilization of a technology and the leveraged technology should be a good fit to the task deployed as predictors of performance.

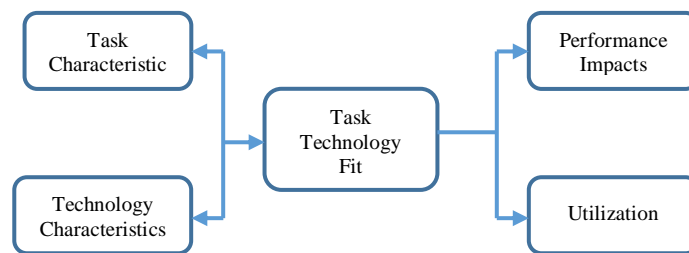


Figure 3: Task Technology Fit Model (Lai, 2017)

Utilization of a technological intervention is based on human perspective and other situational factors like attitude, behavior, perceived satisfaction and social norm. These factors affect the initial technological perspective when a system is introduced and continue having weight on its repeated utilization. From this theory, it is implied that increased utilization of a system having positive human outlook impacts positively on its performance rating. When a technological intervention offers features that are just right for the purposed task, the user will feel that this is the technological tool of choice, hence the fit. This perceived sense of a tool that eases the task load through features that break down the task into doable subtasks while simplifying the whole process optimally leads to better utilization and performance. The level of technological fit is a strong predictor of better performance. Further research on the TTF model by Goodhue & Thompson (1995) led to Technology to Performance Chain theory (TPC) as shown in Figure 4. This is a more individualized model that looks at both the technology and its utilization in relation to its relevance to the perceived task. Its underlying mantra is that technologies must be utilized and fit the task they support to have a performance impact. It provides clear relationship on the technological, user tasks and utilization variables as they relate towards progressive or retrogressive performance. It is important to note that in as much individuals embrace

technological tools to accomplish given tasks, the individual's characteristics could greatly influence how well the technological utilization will be. Some of these individual characteristics are motivation, prior training and experience with technological interventions.

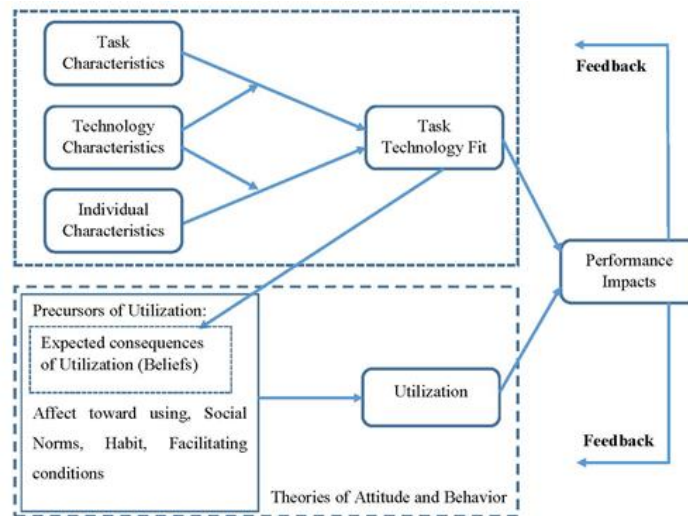


Figure 4. Technology to Performance Chain Model (Goodhue and Thompson, 1995)

2.2.4. Unified Theory of Acceptance and Use of Technology

Unified Theory of Acceptance and Use of Technology (UTAUT) technology acceptance model was formulated out of the gaps experienced with the use of several technology acceptance models. Venkatesh, Morris, Davis & Davis (2003) comprehensively reviewed eight technology acceptance models and formulated a unified model as shown in Figure 5 after data synthesis, whose aim was to integrate elements across the eight models while addressing their shortcomings. The eight models reviewed were the technology acceptance model by Davis (1989), the theory of reasoned action by Fishben & Ajzen (1975), the motivational model by Davis, Bagozzi & Warshaw (1992). In addition, the theory of planned behavior by Taylor & Todd (1995), the model of PC utilization by Thompson, Higgins & Howell (1991), the innovation diffusion theory by Moore & Benbasat (1991), the social cognitive theory by Compeau & Higgins (1995) and a combined version of technology acceptance model (TAM) and theory of planned behavior were reviewed. They used data drawn from four different organizations over a period of six months and conducted their research based on three points of measurement. This research was borne out of an observation that several underlying concepts drawn from technology acceptance theories were inherently comparable, Zhou, Marfo, Antwi, Antwi, Kachie & Wireko (2019).

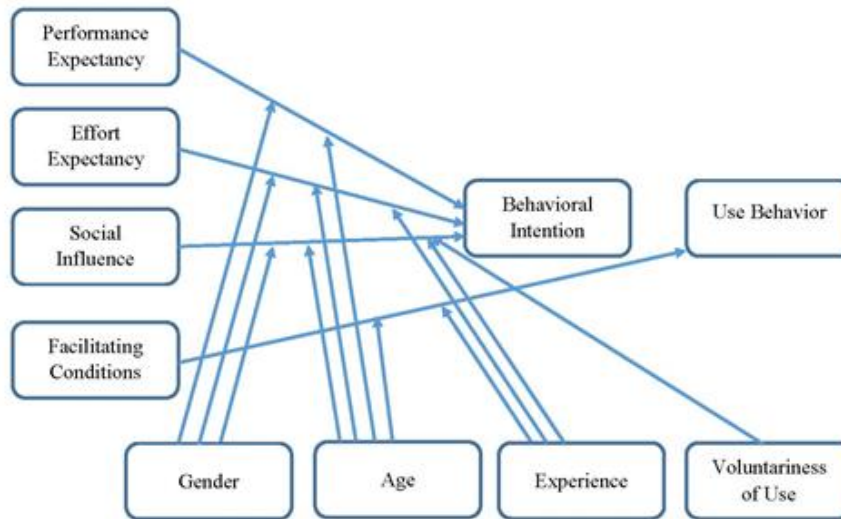


Figure 5: UTAUT Research Model (Venkatesh, Morris, Davis & Davis 2003)

The eight choice models were adaptable to only between 17 percent and 53 percent variance in the results concerning user intentions on their use of information technology, Venkatesh et. al. (2003). From these findings, the proposed UTAUT technology acceptance model was applied to the original data and fared better, outperforming the eight models with adaptability of 69 percent. Four constructs stood out and were earmarked as indicators that were direct determinants towards technology acceptance. These were performance expectancy, effort expectancy, social influence and facilitating conditions. Venkatesh et. al. (2003) in their research concluded that attitude towards using technology; self-efficacy and anxiety did not meet the threshold to be considered as main determinants towards intention of using a technological intervention. The UTAUT research model was as shown in figure 5 above.

Dwivedi et al. (2019) and Venkatesh et al. (2003) expound performance expectancy as the degree or weight to which an individual believes that using a particular technological intervention will enable the user attain gains in their work performance and subsequently, their output. The term performance expectancy is drawn from the amalgamation of five factors that were weighed in the formation of perceived ease of use during research on technology acceptance model. These five constructs are external motivation, job fit, relative advantage as well as outcome expectancy, Venkatesh & Davis (2000). Noted as a positive factor towards technology acceptance and use, performance expectancy construct has received a lot of attention and research, Hamzat & Mabawonku (2018).

Effort Expectancy is the level of ease experienced when a technology is in use, Alam, Hu & Gani (2019). A technology whose effort expectancy is good exhibits ease of use tendencies. Users feel drawn to such system and the observed percentage of acceptance to such systems is usually high making adaptation to these technologies easier. This construct can be applied in environments that exhibit voluntary and also mandatory tendencies. Some of the techniques that effort expectancy can be achieved is through easy to use interfaces and user involvement during technology development and deployment, to the level where the user feels familiar with the technology and believes that the system is easy to use to deliver the set objectives. Effort expectancy is derived from three constructs: perceived ease of use, complexity and ease of use, Alam, Hu & Gani (2019) and Venkatesh et. al. (2003).

Social influence is explained as the degree at which a user believes that the technology adapted is easy to use based on the views of others concerning that technology. Venkatesh et. al. (2003) expounds it as the degree to which a user perceives that others who hold an important role believe that he or she should use the new system. This construct is directly fed from subjective norm, social factors and perceived image derived when one uses the new technology. This construct generally is adaptable when a technology is deployed in mandatory conditions and is experienced due to the fact that the user has not yet had any or minimal interaction with the system. Consequently, his views are based on other people's views, who may have had a higher level of interaction with it. Normally, this position changes when the user gets to use the system and from the personal interaction, a solid view of the system is derived.

The degree to which a user believes that there exist enough organizational and technical infrastructure to support the use of the new technology is referred to as facilitating condition, Alam, Hu & Gani (2019) and Venkatesh et. al. (2003). During development and deployment of any system, it is prudent to consider the environmental and technological system use needs for removing any obstacles that may cause technology rejection. Deployment of a good technological intervention in an environment that is not ready or responsive to its adoption and optimum use is already a recipe for failure. As the important system development factors are dealt with, a keen look should be given to the environmental and technological optimal system use needs for successful technological deployment.

Alam, Hu & Gani (2019) underscore the pivotal need for high reliability and maintainability of any technological intervention during its lifetime. This is an important factor of any system and in the extended UTAUT model, it comes out clearly as perceived reliability which is critical to user acceptance, satisfaction and usage. The costs associated with acquisition, use and maintenance of technological interventions is a pivotal factor to consider when considering any system. Growing needs and changes in the operational environment call for changes in the structure of any system. This may be new laws governing the use of the technology, changed business model or just upgrades for increased dependability, safety and quality of service call for periodic maintenance and updates. Users therefore are quite aware of these needs and their corresponding financial effect on the overall budget for maintenance. In this regard, the extended UTAUT model includes price value and perceived reliability as key constructs in addition to performance expectancy, effort expectancy, social influence and facilitating condition in its matrix.

2.3. Findings of The Study

The researchers summarizes the choice models as shown in Table 1 below:

Table 1: Choice Models Comparison

Summary Findings			
Construct	Application Area	Strengths	Limitations
Theory of Reasoned Action (TRA)			
a) Beliefs b) Attitudes c) Intentions d) Behaviors e) Subjective Norm	Social Psychology. Used to determine consciously intended behaviors	Extensively studied model on consciously intended behaviors	a) Difficult to implement in environments where choice is limited. b) Sufficiency of attitudes and subjective norm to clearly predict behaviors. c) Does not address role of habit. d) Does not address role of

Summary Findings			
Construct	Application Area	Strengths	Limitations
			cognitive deliberation
Technology Acceptance Model (TAM)			
a) Perceived usefulness b) Perceived ease of use	Modelling user acceptance in and information systems environment	Widely applied model in information technology environments with several adopted variations to it	a) The two variables do not aptly represent and model current technological environments. b) Focuses more on user, not operating environment c) No variable represents design process or informs implementation phases
Task Technology Fit (TTF)			
a) Utilization of a technology b) Technology Fit	Technological environment	Adaptable to new products already in the market	a) Difficult to implement in environments where choice is limited. b) Utilization of a system with low TTF will not increase performance.
Unified Theory of Acceptance and Use of Technology (UTAUT)			
a) Performance Expectancy b) Effort Expectancy c) Social Influence d) Facilitating Conditions	Technological environment	Uses moderators of gender, age, experience and voluntariness of use that can be adjusted according to environment.	a) Scales used to measure the constructs are not conclusive.

The objective of this research was to propose a multilevel technology acceptance management model that took into account human, technological and organizational variables. Review of existing models and their limitations aim to provide deeper reason for research on a model that mitigates the identified limitations. Identification of new constructs that exist in a deployment environment is pivotal to this research. Successful deployment of technologies through early technology acceptance prediction and timely deployment of mitigation measures is the desire of many organizations. The proposed model that incorporates new constructs aim at explaining technology acceptance of emerging technologies such as e-government, internet-of-things, e-board and Enterprise Resource Planning Systems. This is grounded in the need for a comprehensive understanding of IT adoption and factors that determine its success or failure.

2.4. Proposed Technology Acceptance Management Model

Derived from analyzing the choice models while mirroring this knowledge to the current deployment environment, the researchers propose a multilevel technology acceptance management model as illustrated in Figure 6. Every system project manager would love to be able to predict whether the technological deployment will be acceptable to the envisioned users. Davis, Bagozzi & Warshaw (1989) laid a foundation to this discussion by recommending that a model should also be in a position to diagnose the base reasons why a proposed system may not be fully acceptable to the users. It should also provide an opportunity for corrective action to be taken thus increasing acceptability of a technological intervention. This will increase business

impact, saving time and resources that would otherwise go to waste deploying and supporting a system that is unproductive and not supportive to the business goals.

Taherdoost (2018) explains that although a number of models and frameworks have been developed, explaining user adoption or rejection of new technological interventions needs more than one theoretical approach for a wholesome understanding. This creates a cumbersome process and consequently, need for a model that when applied, encompasses most of the possible variables that exists in a deployment environment is key. The apparent need for further research into models that provide an ever-richer understanding of technology adoption is key. This is proposed through consideration of constructs that relate to each other with representation from the human, technological and organizational perspective. The proposed model aims to bring more clarity to technological deployment process through assessing key parameters derived from the deployment environment. It groups the key variables into three main categories:

- i. Human variables
- ii. Technological variables
- iii. Organizational variables

It is envisioned that this will bring understanding and inform the technology deployment process on the parameters and key variables that need to be focused on for technology acceptance. The proposed variables with their corresponding factors are synthesized into a multilevel knowledge management acceptance model that forms a foundation for the study. The illustrated model below depicts the three multilevel dimensions towards technology acceptance.

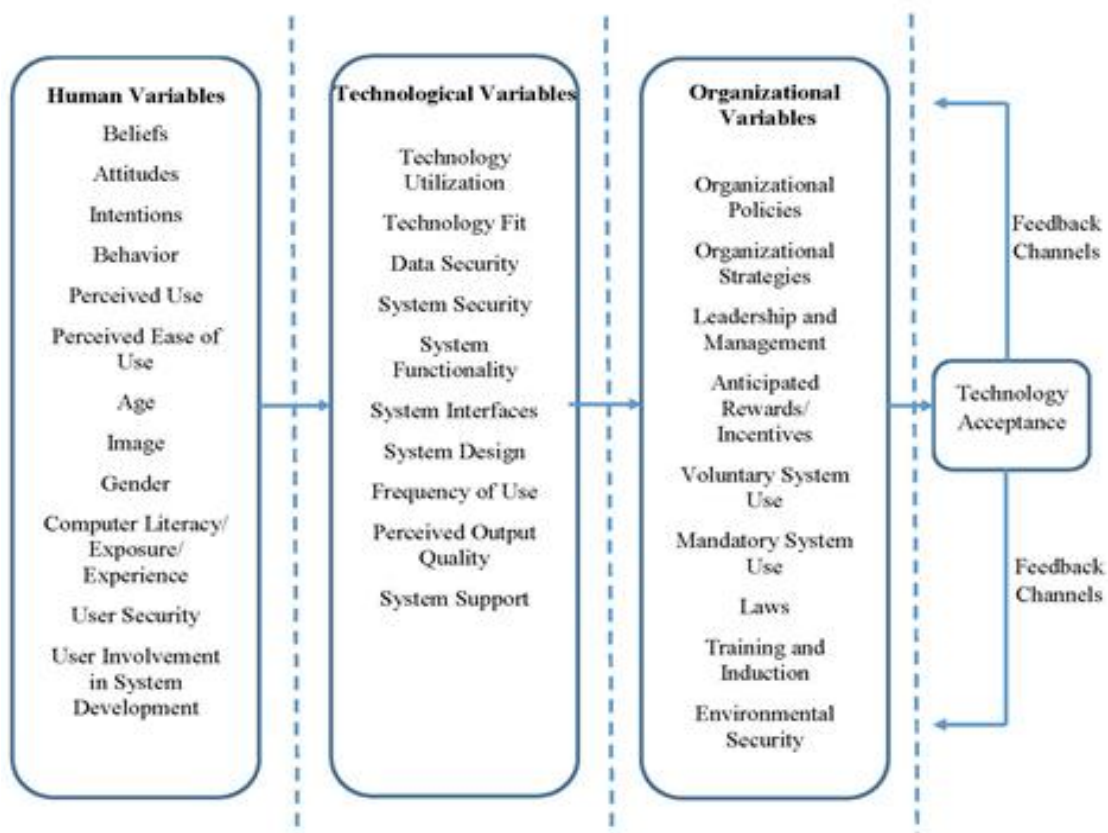


Figure 6. Proposed Multilevel Technology Acceptance Management Model

It has been observed that the application of variables to TAM that are dependent to the application environment address its limitations and potentially confirm its robustness and generalizability, Fayad & Paper (2015).

2.4.1. Technological Variables

Research shows that the level of end-user satisfaction with an information technology intervention has widely been accepted as an indicator of IT success. This is according to Mahmood, Gemoets & Jacquez (1999). These variables are human related characteristics such as beliefs, attitudes, intentions, perceived ease of use and perceived use that directly have emotional impact the user's perceived value of the system. Variables such as age of the user, their gender and image have a direct bearing on where the individual places himself in terms of personal relevance to technology. Computer literacy and exposure to technology, perceived personal security from any technology related harm, user involvement in the system development process brings about a sense of ownership to the proposed technological deployment, the user feels they have a part to play towards technological success and are more inclined towards technology use. User acceptance and satisfaction, a key ingredient in technology adoption starts at the point where the individual's perception concerning technology is positive and its perceived value is high.

2.4.2. Human Variables

Technology related variables encompass the technological need on the ground and the solutions availed through a technological intervention. Deployments of technological solutions increase day by day as the world finds an easy partner to a myriad of challenges in technology. Being enlightened on the antecedents of technology use early enough during the development and deployment process informs key project steps and is a recipe for success. In as much as the world recognizes technological inventions as a sure bet on easing many processes, that are characterized as loads, the level of acceptance of these interventions is quite wanting and a major concern to product developers and solution providers. When a user perceives technology as a positive tool in their sphere of operation, he will evaluate the technological capabilities and risks associated with its adoption. Variables such as technology utilization in accomplishing the laid objectives, technology fit to the projected tasks, system functionality to accomplish the desired procedures endear a user and underscore the usefulness of technology. System interfaces that are easier to navigate to achieve the set goals increase system utilization. Perceived and actual period that the system will be in use plays an important factor in its acceptance.

A balance should be sought on time and frequency of use for maximum gains. As all systems have an input and output, the quality of system output should reflect the effort infused in the input process. A clear, concise, comprehensible and quality output should be the target for any technological intervention. The data in use should be secure from instances that may compromise its quality. The constant worry of any organization when deploying any technological intervention is the availability of technical support for the product over a period of time. This includes but is not limited to availability of periodic updates, upgrades and support when the need arises. Any technological system that does not have an established support base is destined to fail and its level of acceptance may be below par.

2.4.3. Organizational Variables

Every technology operates in its required environment. The process of deploying technology may be successfully controlled with key acceptance variables dealt with early enough to ensure all-out acceptance and business gains derived from it. During deployment and operations, other external

factors that may be within or beyond the scope of management may change influencing the technology adopted either positively or may negate the gains derived through its adoption. Organizational policies should be friendly to technology adoption through decisive policy statements that support adoption and deployment of relevant technologies that aid in fulfilling the goals of any institution. These should be enablers that give flesh to fulfilling organizational strategies and goals. Laws governing the use of any technology should be well balanced to accord users an equilibrium, gaining the benefits of technological adoption while shielding them from instances of abuse, harm and insecurity. Top management attitude towards technology dictates how far technology adoption and use in any organization goes. Formulation of strategies and policies that encourage technology adoption and harness its use in organizational procedures is a recipe for success. Set incentives and rewards to users who leverage on technology to accomplish tasks increases its adoption and frequent use.

Laws set and reviewed by the legislature should be void of collisions with adoption of technology, but rather encourage its use. Regulations that clearly spell out procedures like conveyance should be amended to encompass use of technology as an enabler to quality service delivery. Yearly, the government requires all of its citizenry with personal identification numbers to file tax returns. Great efforts through capacity building campaigns and awareness is vigorously conducted in the social space but still the response to this exercise is below par with unmet targets. The mandatory or voluntary nature to the adoption and use of any system directly influences the user perception towards it. Induction and training on use of technology helps bridge technology related fear and enables the user appreciate its use in achieving the set goals. All these deployments should be in a controlled environment and should not cause any harm to its users or environment. Any perceived threats should be addressed in a sober, informative and inclusive manner to the benefit of its adoption. The way an organization handles its internal and external (environmental) variables reflect on its readiness on the adaption of technology. How a country or organization handles these external variables directly affect the perception towards technology adoption and subsequent use. Every system has users, the system itself has some technical capabilities and all these exist in an environment that the user cannot exclusively control. It is pivotal for any successful technological adoption, these three variables be keenly appreciated as they directly affect human perception towards technology and its successful use.

The model inculcates feedback as a key element in adoption, successful acceptance, usage, and further upgrades that may be undertaken for making the technological intervention more responsive to user needs. Feedback mechanism may also influence the variables in the three groupings as each deployment environment has unique users, deployed technology and external variables and therefore is not exhaustive. This model is more suited for a technological deployment environment.

2.5. Further Research

As the world continues to embrace technology, different technological application areas exist. These deployments sometimes are not successful due to inconclusive and comprehensive data driven understanding of the pivotal variables that exist in that environment. This calls for further research and validation of the proposed model with its constructs for a robust and more inclusive technological acceptance model. This may include specific application environments like health and education as this may bring rise to different constructs that are technological environment-specific while maintaining a wholesome outlook to the technology acceptance process.

3. CONCLUSIONS

Institutions and individuals desire to get it right the first time when implementing a technological deployment. This desire led the researchers to propose a model that provides new variables that span beyond the human focus of the reviewed models. The proposed model when applied to a technological intervention environment hopes to provide a better understanding of the deployment environment and candidate system. Data derived from the proposed model aims to illuminate the adoption path through data driven decision-making efforts. The proposed model adopts independent variables for a more inclusive and comprehensive evaluation when testing the dependent variable. In addition, this stimulates research and a drive for deeper understanding of user relationships with newer technologies in their technological environments.

REFERENCES

- [1] Uriarte F. A (2008). Introduction to Knowledge Management, Jakarta, Asean Foundation.
- [2] Kwanya T. Kiplang'at J. & Wamukoya J (2017). Emerging Trends in Information and Knowledge Management, Eldoret, Moi University Press.
- [3] Fernandez, I.B & Sabherwal, R (2010). Knowledge Management systems and Processes, New York, Armonk Publishers.
- [4] Talisayon S. (2013). Knowledge Management for the Public Sector, Tokyo, Asian Productivity Organization.
- [5] Odinioha, G.J and Chukwuma, O.L (2013). The International Journal of Management, Volume 2, Issue 3, Port Harcourt, Wiley-Blackwell Publishers.
- [6] Maier R. (2007). Knowledge Management Systems, Third Edition, Innsbruck, Springer. DOI 10.1007/978-3-540-71408-8.
- [7] Davis F., Bagozzi R. & Warshaw P. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models, Michigan, Ann Arbor, Institute of Management Science.
- [8] Dwivedi Y. K, Rana P.N. Jeyaraj A., Clement M. & Williams M. D. (2019). Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT): Towards a Revised Theoretical Model. 21, 719–734. DOI 10.1007/s10796-017-9774-y.
- [9] Hwang, Y. Al-Arabi M. & Shin, D.H (2016). Understanding Technology Acceptance in a Mandatory Environment, UK, Sage Publishers.
- [10] Scherer R., Siddiq F. & Tondeur J. (2018). The Technology Acceptance Model (TAM): A Meta-analytic Structural Equation Modelling Approach to Explaining Teachers' Adoption of Digital Technology in Education: Norway, Elsevier.
- [11] Aizstrauta D., Ginters E. & Eroles M. P (2015). Applying Theory of Diffussion of Innovations to Evaluate Technology Acceptance and Sustainability. 43. 69-77. DOI 10.1016/j.procs.2014.12.010.
- [12] Lowe B., Dwivedi Y. & D'Alessandro S. (2019). Consumers and Technology in a Changing World. Retrieved 28 December, 2019 from <https://kar.kent.ac.uk/73789/>.
- [13] Gucin N. O.& Berk O. S. (2015). Technology Acceptance in Health Care: An Integrative Review of Predictive Factors and Intervention Programs. 195, 1698-1704. DOI 10.1016/j.sbspro.2015.06.263.
- [14] Lai P. C (2017). The Literature Review of Technology Adoption Models and Theories for the Novelty Technology, Electronic Journal of Information Systems and Technology Management, 14(1), 21-38, Retrieved 20 April, 2019, from <http://www.jistem.fea.usp.br>
- [15] Ajzen, I., & Fishbein, M. (1980). Understanding Attitudes and Predicting Social Behavior. Engelwood Cliffs, NJ, Prentice Hall.
- [16] Greene K., (2009). Reasoned Action Theory Scope Components, New Jersey, Rutgers University Press.
- [17] Sutton, S., French, D.P., Hennings, S.J. et al. Curr Psychol (2003). 22: 234. DOI.org/10.1007/s12144-003-1019-1.
- [18] Fishbein, M., & Ajzen, I. (1975). Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley.
- [19] Taherdoost H. (2018). A review of Technology Acceptance and Adoption Models and Theories. 22. 960-967. DOI 10.1016/j.promfg.2018.08.137.

- [20] Davis, F. D. (1993). User Acceptance of Information Technology: System Characteristics, User Perceptions And Behavioral Impacts. *International Journal of Man-Machine Studies*, 38, 475-487. DOI:10.1006/imms.1993.1022.
- [21] Lim W. M (2018). Dialectic Antidotes to Critics of the Technology Acceptance Model: Conceptual, Methodological and Replication Treatments for Behavioral Modelling in Technology-Mediated Environments, Swinburne, *Australasian Journal of Information Systems*
- [22] Goodhue D. L & Thompson R. L (1995). Task-Technology Fit and Individual Performance, Management Information Systems Research Center, University of Minnesota.
- [23] Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of The Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*. 46 (2) 186-204.
- [24] Venkatesh V., Morris M. G., Davis G. B. & Davis F. D. (2003). User Acceptance of Information Technology: Toward A Unified View, *MIS Quarterly*, Volume 27 (3),425-478.
- [25] Taylor S & Todd P. A. (1995). Understanding Information Technology Usage: A Test of Competing Models. The Institute for Operations Research and the Management Sciences, DOI:10.1287/isre.6.2.144.
- [26] Zhou L. L., Marfo J. O., Antwi H. A., Antwi M. O., Kachie A.D. & Wireko S. A. (2019). Assessment of the social influence and dactilating conditions that support nurses adoption of hospital electronic information management systems (HEIMS) in Ghana using the unified theory of acceptance and use of technology (UTAUT) model. Retrieved 20 January, 2020. DOI.org/10.1186/s12911-019-0956-z.
- [27] Hamzat S. & Mabawonku I. (2018). Influence of Performance Expectancy and Facilitating Conditions on use of Digital Library by Engineering Lecturers in universities in South-west, Nigeria. *Library Philosophy and Practice* (e-journal).1620. from <https://digitalcommons.unl.edu/libphilprac/1620>
- [28] Alam M. Z., Hu W., & Gani M. O. (2019). An Empirical Analysis of the Influencing Factors of Adoption of Mobile Health Services in Bangladesh Based on Extended UTAUT Model. WHICEB 2019 Proceedings. 53 from <https://aisel.aisel.org/whiceb2019/53>
- [29] Fayad R. & Paper D. (2015). The Technology Acceptance Model E-Commerce Extension: A Conceptual Framework, Lebabon, Elsevier.
- [30] Mahmood M. A., Gemoets L. A., & Jacquez C. (1999). Variables Affecting Information Technology End-User Satisfaction. DOI:10.1006/ijhc.1999.0353.

AUTHORS

Gilbert Busolo Lusweti Results-driven professional offering a progressive result driven fifteen -year career in Information Technology. A process improvement champion with a history of implementing procedures and technologies to enhance operational efficiency, value addition and quality based decision making through systematic automation and digitization efforts. Currently a MSc. Student at Jomo Kenyatta University of Agriculture and Technology, focused on Technology Adoption and Knowledge Management.



Dr. Lawrence Nderu Experienced University Lecturer with a demonstrated history of working in the higher education industry. Skilled in Software Development, (Python, Java, C#, C++, R, JavaScript, Android, PHP, C++, and Databases. Strong education professional with a Doctor of Philosophy - PhD focused in Computer Science from University of Vincennes in Saint-Denis, Paris, France. Research Interest on Artificial Intelligence, Data Science and Software Development. Goal: Technology for building the resilience of communities.



Dr. Kennedy Ogada Champion for effective and efficient Data Mining classification techniques. Experienced University Lecturer on Information Technology, Information Policy, Telecommunications Policy, and Media and Foreign Policy. PhD in Information Technology.



MINIMUM VIABLE MODEL ESTIMATES FOR MACHINE LEARNING PROJECTS

John Hawkins

Transitional AI Research Group, Sydney, Australia

ABSTRACT

Prioritization of machine learning projects requires estimates of both the potential ROI of the business case and the technical difficulty of building a model with the required characteristics. In this work we present a technique for estimating the minimum required performance characteristics of a predictive model given a set of information about how it will be used. This technique will result in robust, objective comparisons between potential projects. The resulting estimates will allow data scientists and managers to evaluate whether a proposed machine learning project is likely to succeed before any modelling needs to be done. The technique has been implemented into the open source application MinViME (Minimum Viable Model Estimator) which can be installed via the [PyPI python package management system](https://pypi.python.org/pypi/MinViME), or downloaded directly from the GitHub repository. Available at <https://github.com/john-hawkins/MinViME>

KEYWORDS

Machine Learning, ROI Estimation, Machine Learning Metrics, Cost Sensitive Learning

1. INTRODUCTION

In an ideal world we would prioritise all our machine learning projects according to their expected payoff. This would, in turn, require a reasonable estimate of both the probability of success and the return given success. Creating such an estimate is difficult and instead we tend to be guided by heuristics, implicit and explicit, about the size of the problem and its difficulty.

Difficulty estimation is often limited to a discussion about what it would take to feasibly productionise a model. Consideration of the difficulty of building the model, to the extent that it is done at all, is usually an informal process relying on the experience of data scientist working on the project. In this work we present a structured approach to estimating the difficulty of a machine learning task by estimating the minimum required performance of a model that would meet the business objectives.

The most common approach to which the details of the business case are considered in the building of a machine learning solution is through development of domain specific methods. For example, work on designing loss functions for specific applications (See [1], [2]). Alternatively, there is a practice of using cost sensitive learning [3, 4, 5, 6] to permit general techniques learn a solution that is tuned to the specifics of business problem (for example [7]). In general, you can achieve the same outcome by learning a well calibrated probability estimator and adjusting the decision threshold on the basis of the cost matrix [8]. The cost matrix approach to analysing a binary classifier can be also be used to estimate its expected ROI [9].

These approaches allow you to improve and customise a model for a given application, but they do not help you decide if the project is economically viable in the first place. There have been extensive efforts to characterise project complexity in terms of data set requirements for given performance criteria [10], however the task of determining these criteria from business requirements has not, to our knowledge, been addressed. Fortunately, as we will demonstrate, some of the machinery for the cost matrix analysis can be used in reverse to estimate the baseline performance a model would require to reach minimal ROI expectations.

We use the framework of a cost matrix to develop a method for estimating the minimum viable binary classification model that satisfies a quantitative definition of the business requirements. We demonstrate that a lower bound on *precision* can be calculated *a priori*, but that most other metrics require knowledge of either the expected number of false positives or false negatives. We devise an additional *simplicity* metric which allows for direct comparison between machine learning projects purely on the basis of the business criteria. Finally, we demonstrate that lower bounds on the AUC and other metrics can be estimated through numerical simulation of the ROC plot.

We use the minvime ROC simulator to conduct a range of experiments to explore the manner in which the minimum viable model changes with the dimensions of the problem landscape. The results of these simulations are presented in a manner to help develop intuitions about the feasibility of machine learning projects.

2. GLOBAL CRITERIA

In order to estimate the required performance characteristics of any machine learning model there are several global criteria that need to be defined. Firstly, we need to determine the overall ROI required to justify the work to build and deploy the model. This can take the form of a project cost plus margin for expected return. Given an overall estimate we then need to amortise it to the level of analysis we are working at, for example, expected yearly return.

In addition, we require the expected number of events that will be processed by the machine learning system for the level of analysis mentioned above. For example, this may be the number of customers that we need to analyse for likelihood of churn every year. Included in this we require to know the *base rate* of the event we are predicting. In other words how many people tend to churn in a year.

Given the above criteria, we next need to evaluate how a model would need to perform in order to process the events at a sufficient level of competence. To do this analysis we need to look at the performance of the model as it operates on the events we are predicting. We start this process by looking at the impact a binary classification model using a cost matrix analysis.

3. COST MATRIX ANALYSIS

The idea of a cost matrix in machine learning emerged from the literature on training models on imbalanced datasets. Specifically there are a set of papers in which the learning algorithm itself is designed such that the prediction returned is the one that minimises the expected cost [5, 4]. Typically this means not necessarily returning the class that is most probable, but the one with the lowest expected cost. In essence, the model is learning a threshold in the probability space for making class determinations. The same idea can be applied after training an arbitrary classifier that returns a real-value result in order to determine a threshold that minimises the error when the model is applied.

It has been observed that the later approach is not dissimilar from a game theoretical analysis [11]. In game theory the cost matrix would delineate the outcomes for all players of a game in which each player's fortunes are affected by their own decision and that of the other player. Each axis of the matrix describes the decisions of a player and by locating the cell at the nexus of all decisions we can determine the outcome of the game.

In binary classification we replace the players of the game with two sources of information: the prediction of the model and the ground truth offered up by reality. For binary classification problems this is a 2x2 matrix in which each of the cells have a direct corollary in the confusion matrix. In the confusion matrix we are presented with the number of events that fall into each of the four potential outcomes. In a cost matrix we capture the economic (or other) impact of each specific type of event occurring. From here on in we will follow Elkan [5] and allow the cells of the matrix to be either depicted as costs (negative quantities) or benefits (positive quantities).

In table 1 we demonstrate the structure of a cost matrix in which we also depict the standard polarity of the costs and benefits in most applications.

Table 1: Cost/Benefit Matrix for a Binary Classification System.

	Actual Negative	Actual Positive
Predict Negative	True Negative (TN)	False Negative (FN)
Predict Positive	False Positive (FP) Cost	True Positive (TP) Benefit

If we assume, without loss of generality, that the positive class corresponds to the event against which we plan to take action, then as observed by Elkan [5] the entries of the cost matrix for the predicted negative row should generally be identical. The impact of taking no action is usually the same. It is also usually zero because the model is typically being applied to a *status quo* in which no action is taken. In most cases you will only need to estimate the cost/benefits of your true positives and false positives. This is because the assumed ground state (the status quo) is the one in which the actual costs of false negatives are already being incurred. Similarly the true negatives are the cases in which you have taken no action on an event that does not even represent a loss of opportunity.

The overall outcome of using a model can be calculated by multiplying the cells of the confusion matrix by the corresponding values in the cost matrix and summing the results. This tells us what the overall impact would have been had we used the model on the sample of data used to determine the confusion matrix. The advantage of using a cost/benefit structure in the matrix is that we can then read off whether the result of using the model was net-positive or net-negative depending on the polarity of the total impact. This approach can then be used to optimise a decision threshold by varying the threshold and recalculating the expected impact.

Before continuing with the analysis of the cost matrix we make some observations about the process of determining the content of the cost matrix.

3.1 Deterministic Outcomes

In some situations we can say in advance what the impact of a specific conjunction of prediction and action will be. We might know, for example, that giving a loan to a customer who will default will result in a loss of money. If we take the action of rejecting that loan then the money is no longer lost.

This determinism in outcomes is typically true of situations where once we have taken our action there is no dependence on additional decisions by other people to impact whether our action is effective. This is not always true. There are certain circumstances in which knowing what will happen does not guarantee a given outcome. In these cases there is stochastic relationship between the prediction and the outcome.

3.2 Stochastic Outcomes

In the stochastic situation we understand that even in the event of possessing a perfect oracle for the future events, we would still be left with an element of randomness in the outcomes of our actions. This is typically true where the action we take will involve third parties or processes beyond our control to achieve their result.

The canonical example of a predictive process with a stochastic outcome is a customer churn model. A churn model is used to intervene and attempt to prevent a customer from churning. However, our ability to influence a customer to not churn is not guaranteed by accurately predicting it.

In these situations we need to make additional estimates in order to get the contents of the cost matrix. For example, if we expect that our interventions will succeed one in every five attempts, and the value of a successful intervention is \$1,000, then the benefit of a true positive is \$200. Defining a cost/benefit matrix for a situation with stochastic outcomes will require additional assumptions.

3.3 Metrics from the Matrix

If we have an existing model then we can use the cost matrix to estimate the economic impact by combining it with a confusion matrix. How then can we estimate the required properties of a model such that it renders a project viable?

Once the model is deployed it will make predictions for the N cases that occur in the period of analysis. We can demarcate the content of the unknown confusion matrix with the symbol tp true positives and fp for false positives. The benefit of each true positive is captured by the value \mathcal{B} from our cost matrix, similarly the cost of each false positive is demarcated \mathcal{C} . In order to satisfy that the model meets our ROI minimum \mathcal{M} then the following will need to hold:

$$tp \cdot \mathcal{B} - fp \cdot \mathcal{C} \geq \mathcal{M} \quad (1)$$

From Equation 1 we can derive explicit definitions for both the number of true positive and false positive predictions that would satisfy our minimal criteria.

$$tp \geq \frac{fp \cdot \mathcal{C} + \mathcal{M}}{\mathcal{B}} \quad (2)$$

$$fp \leq \frac{tp \cdot \mathcal{B} - \mathcal{M}}{\mathcal{C}} \quad (3)$$

Note, that these expressions involve a reciprocal relationship between the true positives and false positives. Moderated by the ratio between the costs and benefits of taking action in relation to the overall required return. We would like to be able to estimate the True Positive Rate (TPR) or *recall*, and the False Positive Rate (FPR) or *fall-out*. To define these we need to introduce one of the other

global terms required to estimate model requirements: the base rate r at which the event we are predicting occurs. Which allows us to define TPR and FPR as follows:

$$TPR = \frac{tp}{N \cdot r} \quad (4)$$

$$FPR = \frac{fp}{N \cdot (1 - r)} \quad (5)$$

One of the most important metrics for evaluating a binary classification model is precision (or Positive Predictive Value). This determines the proportion of all predicted positive events that were correct. We can now generate an expression for what this would be at the exact point of meeting of ROI requirements.

$$precision = \frac{1}{1 + \frac{B}{c} - \frac{M}{tp \cdot c}} \quad (6)$$

Unfortunately, the presence of the number of true positives in this expression resists additional simplification. However, if we focus our attention on the *break even* point we can eliminate this component and derive the following lower bound on the precision of the model.

$$precision > \frac{1}{1 + \frac{B}{c}} \quad (7)$$

This lower bound corresponds to the line on which the benefits of the true positives are counteracted by the cost of the false positives. Equation 6 corresponds to the line that lies $\frac{M}{B}$ units above this lower bound. We depict both of these lines schematically in Figure 1.

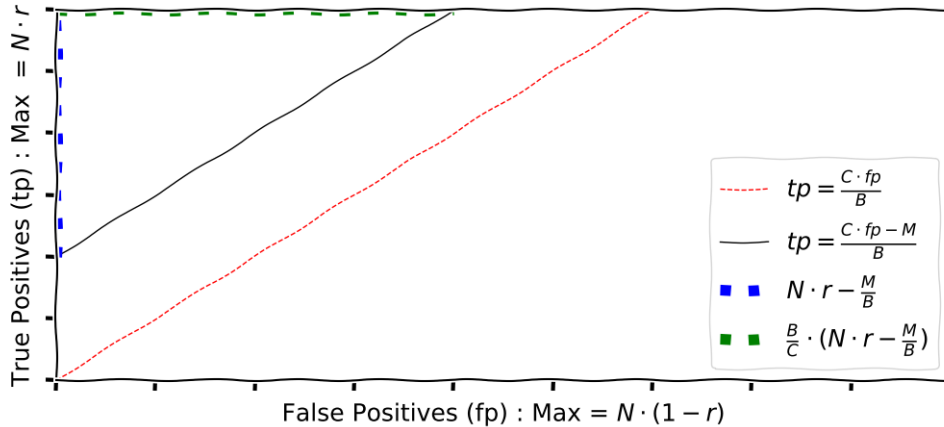


Figure 1: Schematic representation of potential model outputs depicting the boundary between viable and inviable models.

This geometric representation suggests an additional metric for the feasibility of any given project. Note, that the triangle in the upper left corner represents all possible model outputs that would satisfy the stated criteria. The dimensions of this triangle are shown in the legend by the lines of

blue and green squares. We can calculate the area of this triangle and express it as a proportion of the total area of the rectangle. This will result in a metric which is the proportion of all potential model outputs that are economically viable. We can define an expression for this ratio as shown in Equation 8.

$$simplicity = \frac{\mathcal{B}}{2 \cdot \mathcal{C} \cdot N^2 \cdot r(1-r)} \cdot \left(N \cdot r - \frac{\mathcal{M}}{\mathcal{B}} \right) \quad (8)$$

This simplicity measure allows us to rank the difficulty of problems based purely on their business criteria. As this metric is to be calculated in a discrete space, and it must deal with edge cases in which the triangle is truncated, we refer the reader to the source code for its complete calculation.

Our efforts to estimate additional requirements in terms of standard machine learning metrics are hindered by the requirement to know either the number of true positives or false positives in the minimal viable model. To move beyond this limitation we will employ numerical techniques to estimate the bounds on the model performance.

3.4 ROC Simulation

The ROC plot provides us with a method of examining the performance of a binary classification model in a manner that does not depend on a specific threshold. As shown in the example plot 2 it depicts a range of TPR and FPR values across all the different potential thresholds of the model. Additionally, the ROC plot permits the definition of a non-threshold specific metric to evaluate the overall discriminative power of the model (regardless of chosen threshold), the Area Under the Curve (AUC) [12].

We can generate synthetic ROC plots by exploiting their key properties.

1. They are necessarily monotonically increasing
2. They will necessarily pass through the points (0,0) and (1,1)
3. They remain above or on the diagonal between these points (unless the model is inversely calibrated)

We can simulate such curves by designing a parameterised function that satisfies these criteria. We have designed the function shown in equation 9 for this purpose. It is composed of two parts, both of which independently satisfy the criteria.

$$y = \alpha \cdot (-(x-1)^{2\beta} + 1) + (1-\alpha) \cdot x \quad (9)$$

In this function α is a weighting between zero and one that determines the mix between the two component functions. The first component function is an inverted even-powered polynomial function of x , offset so that its origin lies at coordinates (1,1). The second of function is just x itself, which is the diagonal line. Both functions necessarily pass through the points (0,0) and (1,1), and their mixture determines the shape of the curve between these points.

Figure 2 shows some examples of simulated ROC plots with different values of the parameters α and β . Note, it is critical to restrict the space of these parameters such that the search space does

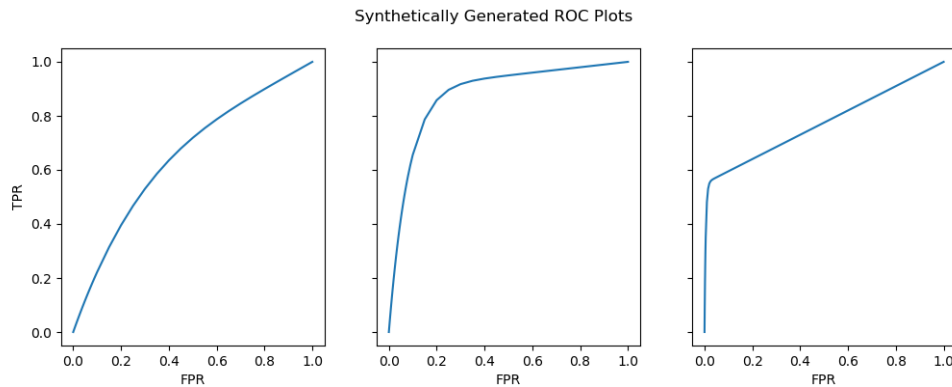


Figure 2: Examples of ROC plots generated through simulation.

not result in unrealistic ROC plots. The third plot in Figure 2 demonstrates that at extreme values the curve starts to represent a spline of two linear functions.

By running simulations across a variety of values of α and β we can generate multiple synthetic ROC plots. Each plot provides a set of hypothetical values for the true positive rate and false positive rate across a range of thresholds. If we combine these with the pre-defined characteristics of the problem we are solving (the number of cases, baseline rate, costs and benefits). Then we can identify which of these synthetic ROC plots would represent acceptable models. We can then rank them such that we search for the ROC curve with the minimum AUC that satisfies our ROI constraints.

Given a minimum viable AUC, we can then calculate estimates for the other metrics discussed above by taking the TPR and FPR values at the minimum viable threshold.

We have built a Python tool for running such simulations and released it as a Python package to the PyPI library (Available at <https://github.com/john-hawkins/MinViME>).

4. SIMULATIONS

The MinViME tool allows us to explore the viable model landscape across a broad range of business criteria in order to understand what drives feasibility. In order to do so we need to represent the space of business problems in an informative and comprehensible way. We do so using the following three ratios.

1. **Benefit to Total ROI.** In other words how significant to the total required ROI is each additional true positive. We explore a range of values between zero and one.
2. **Cost to Benefit.** In other words how many false positives can we tolerate for every true positive case we identify. In principle this could take values greater than one, however this is almost never seen in practice so we explore a range of values between zero and one
3. **Base Rate.** The rate at which the event being predicted occurs in the population.

In addition, we use a static value of one million for the number of cases processed per unit of time This value is realistic for many problems and our experimentation with values an order of magnitude either side demonstrates little variation on the results presented below.

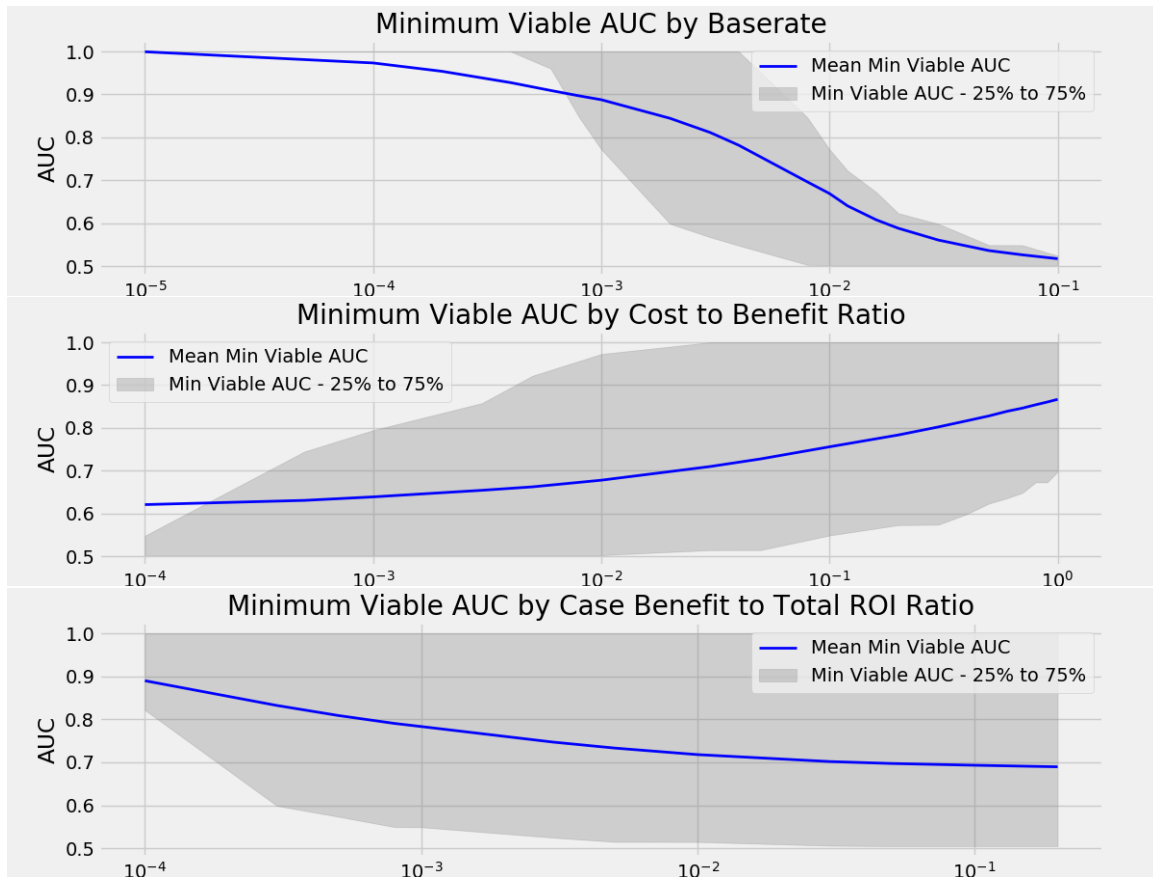


Figure 3: Relationship between the AUC of the minimum viable model and the criteria that determine the business problem. Note: We use a log scale on the X axis to demonstrate how the difficulty varies with each order of magnitude change in the nature of the problem.

4.1 Results

In Figure 3 we see the mean minimum required AUC across each of the three dimensions. The gray bounds indicate the lower and upper quartiles. We see that problems get easier as either the base rate or the benefit to total ROI increases. The opposite occurs for the cost to benefit ratio as we would expect. The base rate plot demonstrates the most variability across the spectrum, with base rates between 10^{-3} and 10^{-2} demonstrating the largest variability. Above this range problems generally have less strenuous requirements and below it most problems require demanding model performance.

Elkan [5] observes that if we scale the entries in a cost matrix by a positive constant or if we add a constant to all entries, it will not affect the optimal decision. However, this is not true for determining the performance characteristics of the minimum viable model. The interaction between the number of cases, the required ROI and the values in the cost matrix will mean that these changes can affect the characteristics of the minimum viable model.

When the benefit to total ROI ratio is very small we see that the model requirements are predominantly for an AUC above 0.8. To further understand the relationship between the base rate and the cost to benefit ratio we plot the surface of required AUC for these values when the benefit to total

ROI is 10^{-4} .

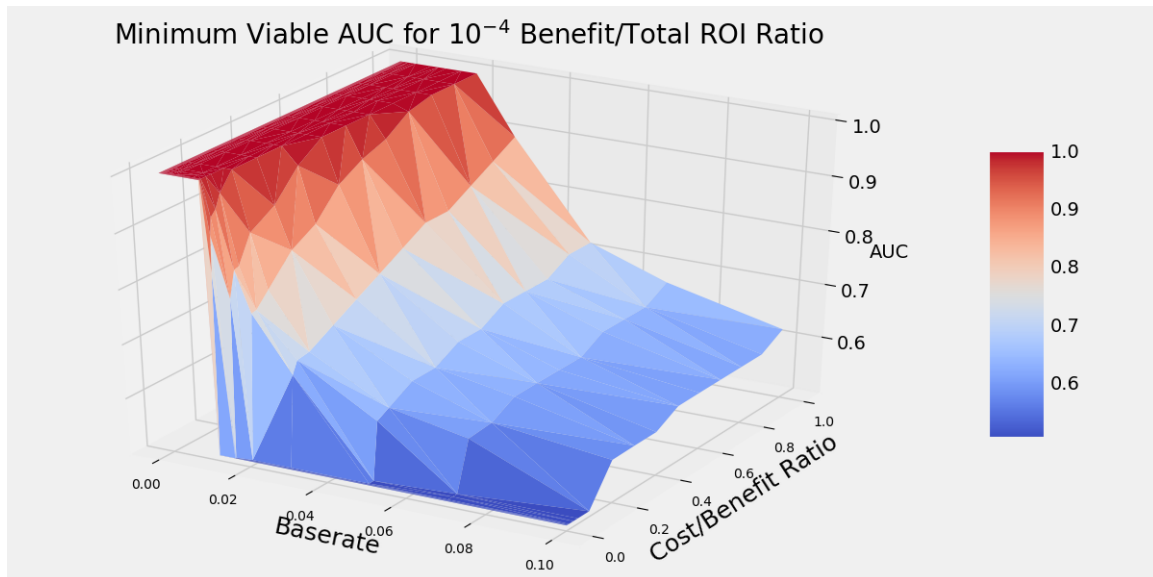


Figure 4: Surface of the AUC landscape for business problems in which each TP prediction returns 10^{-4} of the total required ROI

In Plot 4 we can see clearly that very low base rates can generate intractable problems that are likely to be extremely difficult to solve. As the base rate increases we see a general weakening of the effect of the cost/benefit ratio. However, at very low cost/benefit ratios the impact of the base rate is almost entirely mitigated.

5. CONCLUSION

We have discussed the problem of estimating the minimum viable performance characteristics of a machine learning model. The goal is to be able to do this before projects are undertaken so that we can make informed decisions about where our resources are most effectively deployed.

We investigated whether analytical lower bounds of standard machine learning performance metrics could be calculated from business criteria alone. We derived this for the model *precision*, and derived a novel *simplicity* metric which permits *a priori* comparisons of project complexity for binary classification systems.

We then demonstrated a numerical approach to estimating the minimum viable model performance through simulation of ROC plots. This in turn allows lower bound estimation of all other standard binary classification metrics. Using this method we explored the space of minimum viable models for a range of business problem characteristics. What we observe are some critical non-linear relationships between the base rate and the cost/benefit ratio that will determine whether a project is feasible or likely to be intractable.

REFERENCES

- [1] J. Johnson and T. Khoshgoftaar, "Medicare fraud detection using neural networks," *Journal of Big Data*, vol. 63, no. 6, 2019. [Online]. Available: <https://doi.org/10.1186/>



10.40537-019-0225-0 Cited on page(s): 33

- [2] C. Hennig and M. Kutlukaya, "Some thoughts about the design of loss functions," *REVSTAT - Statistical Journal Volume*, vol. 5, pp. 19–39, 04 2007. Cited on page(s): 33
- [3] P. Domingos, "Metacost: A general method for making classifiers cost-sensitive," *Proceedings of the Fifth ACM SIGKDD Int'l. Conf. on Knowledge Discovery & Data Mining*, pp. 155–164, 8 1999. Cited on page(s): 33
- [4] D. D. Margineantu, "On class-probability estimates and cost-sensitive evaluation of classifiers," in *In Workshop on Cost-Sensitive Learning at the Seventeenth International Conference on Machine Learning (WCSL at ICML2000, 2000*. Cited on page(s): 33, 34
- [5] C. Elkan, "The foundations of cost-sensitive learning," *Proceedings of the Seventeenth International Conference on Artificial Intelligence: 4-10 August 2001; Seattle*, vol. 1, 05 2001. Cited on page(s): 33, 34, 35, 40
- [6] Y. Tian and W. Zhang, "Thors: An efficient approach for making classifiers cost-sensitive," *IEEE Access*, vol. 7, pp. 97 704–97 718, 2019. Cited on page(s): 33
- [7] H. Fatlawi, "Enhanced classification model for cervical cancer dataset based on cost sensitive classifier," 08 2017. Cited on page(s): 33
- [8] N. Nikolaou, N. Edakunni, M. Kull, P. Flach, and G. Brown, "Cost-sensitive boosting algorithms: Do we really need them?" *Machine Learning*, vol. 104, 08 2016. Cited on page(s): 33
- [9] O. Ylijoki, "Guidelines for assessing the value of a predictive algorithm: a case study," *Journal of Marketing Analytics*, vol. 6, 01 2018. Cited on page(s): 33
- [10] S. Raudys and A. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, pp. 252–264, 04 1991. Cited on page(s): 34
- [11] I. E. Sanchez, "Optimal threshold estimation for binary classifiers using game theory," *ISCB Comm J*, p. 2762, 2017. [Online]. Available: <https://doi.org/10.12688/f1000research.10114.3> Cited on page(s): 35
- [12] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30(7), pp. 1145–1159, 1997. Cited on page(s): 38

AN INTELLECTUAL APPROACH TO DESIGN PERSONAL STUDY PLAN VIA MACHINE LEARNING

Shiyuan Zhang¹, Evan Gunnell², Marisabel Chang², Yu Sun²

¹Arnold O. Beckman High School, Irvine, CA 92620, USA

²Department of Computer Science,
California State Polytechnic University, Pomona, USA

ABSTRACT

As more students are required to have standardized test scores to enter higher education, developing vocabulary becomes essential for achieving ideal scores. Each individual has his or her own study style that maximizes the efficiency, and there are various approaches to memorize. However, it is difficult to find a specific learning method that fits the best to a person. This paper designs a tool to customize personal study plans based on clients' different habits including difficulty distribution, difficulty order of learning words, and the types of vocabulary. We applied our application to educational software and conducted a quantitative evaluation of the approach via three types of machine learning models. By calculating cross-validation scores, we evaluated the accuracy of each model and discovered the best model that returns the most accurate predictions. The results reveal that linear regression has the highest cross validation score, and it can provide the most efficient personal study plans.

KEYWORDS

Machine learning, study plan, vocabulary.

1. INTRODUCTION

Vocabulary is essential in reading, writing and essay passages of standardized tests such as the Scholastic Assessment Test (SAT) and American College Test (ACT). Higher-level vocabulary may appear in reading section since diverse variety of authors from historical to modern period wrote these reading passages. Similarly, questions of sophisticated vocabulary may appear in the writing passage; though it would not ask for the synonym of a word instead it would ask for the proper diction of the word. Building vocabulary will be efficient in improving scores, and learning new terms requires memorization skills.

Knowing the definition of complex words help students to answer the section of sentence completion and word definition correctly. However, reading, writing and essay sections cannot be completed without using memorization.[13] Since memorization is a common and vital process of developing a sophisticated level of vocabulary, learners spend many hours study to retain and recall vocabulary for their standardized test; but most of the time memorizing the definition of vocabulary can be very challenge. Although learners may know the definition of a new complex words, they may forget the definition a month later. Thus, memorizing the definition of a vocabulary may not be a satisfactory choice.

Open Problem: Although there are many strategies to learn English vocabulary, these memorization strategies cannot bring quality result. [9][11] Learners may use books that provide the vocabulary that they must know before the standardized test. [10] However, they need to spend a lot of time memorizing the definition. Also, these books may be a old edition and contain a few words that may appear in the standardized test. Other technique is using word-search-puzzle games for improving vocabulary, but this game may not word for everybody since different learners think in different way. [12] Other strategies is using flashcards to recall the definition of words. This technique cannot be too sophisticated since learner need to memorize the definition. Furthermore, they would spend a lot of time when they need to search a word in multiple stacks of flashcard.

Solution: A pictorial method and automated personal study plan using machine learning and mobile computing. In this paper, we presented a new approach for recalling and retaining vocabulary. We have developed a mobile application that enable learners to customize their vocabulary by adding picture that assist them to remember the meaning of a vocabulary. Furthermore, this application can enable learner to input their desirable ACT/SAT score for the standardized test to obtain a personal study plan. In order to get the personal study plan learners must take an assessment. Once a learner finishes the assessment, the system would predict the group of words that is appropriate it for him/her.

In three experiment scenarios, we demonstrated the accuracy of different machine learning algorithms to predict our vocabulary. First, we showed the accuracy of Support Vector Machine (SVM), Linear regression and polynomial regression algorithms to predict difficulty order of learning words. Second, we demonstrated the accuracy of machine learning models to predict Distribution in Terms of Problem Difficulty. Finally, the accuracy of the models to predict Optimal Typing Distribution for Studying. In all case, Linear regression model has a better performance than other two training model. We obtained that the linear regression algorithm performs better in Distribution in Terms of Problem Difficulty model with an accuracy of 41%.

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

Nowadays, machine learning has been used to develop mobile application to make people life easier.[17][18] Although machine learning can analyze and interpret pattern of structure of a data to learn, reason and solve an issue, we would encounter some challenges during the training process. We will use machine learning to provide a learning plan which can be used during the SAT/ACT study process. In our project, there are a few limitations at the time to train different machine learning models.

2.1. Challenge 1 - Deciding student study methods which can fit for most of learners.

Using an appropriated study design may impact learners' memorization. Many learners memorize definition of words by reading word definition, using flashcards, and playing games. Others may group their vocabulary by its type, level of difficulty, alphabetic order etc. However, if learners

use the inappropriate study method, they may forget the meaning of the words after a period. Thus, choosing the correct study plan methods is important to improve learners' vocabulary. In this project, we plan to use existing survey to analyze and obtain the better learning vocabulary methods that suit most of the learners.

2.2. Challenge 2 – Effectively design a study plan

Designing a study plan may be difficult if a learner does not know what to expect and what study methods are more efficient than others. Learners may ask others to share their study methods and try them. Nevertheless, trying the learning methods may be costly and inefficient since learners would spend a lot of time by trying them out and not getting satisfactory result. We proposed an intelligence approach using different machine learning algorithm to elaborate an appropriate learning plan. In this approach, learners would get a study plan based on their performance of an assessment and their desirable SAT/ACT score.

2.3. Challenge 3 – Training a machine learning model with small dataset

The volume of the dataset is one of the most important portions of training a model. When a dataset is small the training model might not be very accurate to predict a specific feature. Thus, the accuracy of a training model might be a little lower than we expected. We plan to use a dataset that contains over 500 data to investigate the accuracy of our training model. In our experiment, we will use three different machine learning algorithms which are SVM, linear regression and polynomial regression

3. SOLUTION

3.1. Machine Learning Model and Prediction

We used open source machine learning library calls scikit-learn to predict the difficulty distribution, difficulty order of learning words, and the types of vocabulary in our model [1]. Scikit-learn provides state-of-the-art machine learning algorithms for supervised and unsupervised problems. We utilized 3 different machine learning algorithms and compare them to obtain the most accurate approach to our problem. [2]In our experiment, we used SVM, linear regression and polynomial regression algorithm. Linear regression is a machine learning algorithm in which two variables is linearly correlated [3]. Polynomial regression is also considering a linear regression with higher order term [4]. SVM is a supervised machine learning algorithm that is utilized for classification, outlier detection and regression.

3.2. Mobile Application

As shown in Figure 1, our app, Memelish, contains a home screen that provides 4 different options which users can choose during their learning process. The button in top left is a search bar, in which the user can type words and search for definitions and memes of the words. The button on top right can direct the user to the personal settings page after being clicked. It contains a banner on the top area which can tap on. After tapping it, the client will be directed to the latest unfinished study set in history. If there is no unfinished set in history, the banner will be titled with "Start a new set?" and automatically suggest a new set based according to the client's personal study plan. There are four buttons in the middle area which connect to the Learn, DIY, and Me pages. The Random button will randomly generate a list to test the user's skill level.

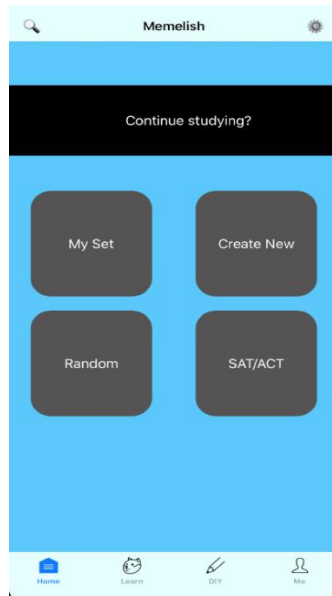


Figure 1: Home page screen

In order to create and set up the home screen buttons, we used `UIBarButtonItem` and `UIStackView` (see Figure 2). We used `UIBarButtonItem` to create the search and setting buttons. In this `UIBarButtonItem`, we passed icon image as a title and set the corresponded action that the button needs to execute when the button is pressed. For “My Set”, “Create New”, “Random” and “SAT/ACT”, we used `UIStackView`. These buttons are created and set up in the `stackview()` function. In order to locate the button in the desirable position, we used 2 different lists that contains 2 buttons and passed it to the `UIStackView`.

```

let leftButton = UIBarButtonItem(title: "🔍", style: .done, target: self,
action: #selector(handleLeftButton))
navigationItem.leftBarButtonItem = leftButton

let rightButton = UIBarButtonItem(title: "⚙️", style: .done, target: self,
action: #selector(handleRightButton))
navigationItem.rightBarButtonItem = rightButton

func stackview(){
let list = [MySet, CreateNew]

MySet.heightAnchor.constraint(equalToConstant: 150).isActive = true
MySet.widthAnchor.constraint(equalToConstant: 150).isActive = true

CreateNew.heightAnchor.constraint(equalToConstant: 150).isActive = true
CreateNew.widthAnchor.constraint(equalToConstant: 150).isActive = true

let stackView1 = UIStackView(arrangedSubviews: list)
stackView1.axis = .horizontal
stackView1.translatesAutoresizingMaskIntoConstraints = false
stackView1.distribution = .equalSpacing
stackView1.spacing = 10

view.addSubview(stackView1)
stackView1.topAnchor.constraint(equalTo: view.topAnchor, constant:
300).isActive = true
stackView1.leadingAnchor.constraint(equalTo: view.leadingAnchor, constant:
30).isActive = true

```

```

stackView1.trailingAnchor.constraint(equalTo: view.trailingAnchor,
                                     constant: -30).isActive = true
stackView1.heightAnchor.constraint(equalToConstant: 150).isActive = true

let array = [Random, ACT]

Random.heightAnchor.constraint(equalToConstant: 150).isActive = true
Random.widthAnchor.constraint(equalToConstant: 150).isActive = true

ACT.heightAnchor.constraint(equalToConstant: 150).isActive = true
ACT.widthAnchor.constraint(equalToConstant: 150).isActive = true

let stackView2 = UIStackView(arrangedSubviews: array)
stackView2.axis = .horizontal
stackView2.translatesAutoresizingMaskIntoConstraints = false
stackView2.distribution = .equalSpacing
stackView2.spacing = 10

view.addSubview(stackView2)
stackView2.topAnchor.constraint(equalTo: stackView1.bottomAnchor, constant:
20).isActive = true
stackView2.leadingAnchor.constraint(equalTo: view.leadingAnchor, constant:
30).isActive = true
stackView2.trailingAnchor.constraint(equalTo: view.trailingAnchor,
constant: -30).isActive = true
stackView2.heightAnchor.constraint(equalToConstant: 150).isActive = true
}

```

Figure 2: Code Segment to create and set up button in homescreen

Figure 3 shows the search engine mentioned in the home screen. The user needs to click the search bar and enter a word. The algorithm can find the words in the database based on user inputs. Also, it will return a list of previewed outcomes with icons and words. The user can tap on one cell of the list for details. Details include the original images, words, and definitions. There is a button on the bottom left corner for users to add the word into personal study sets.

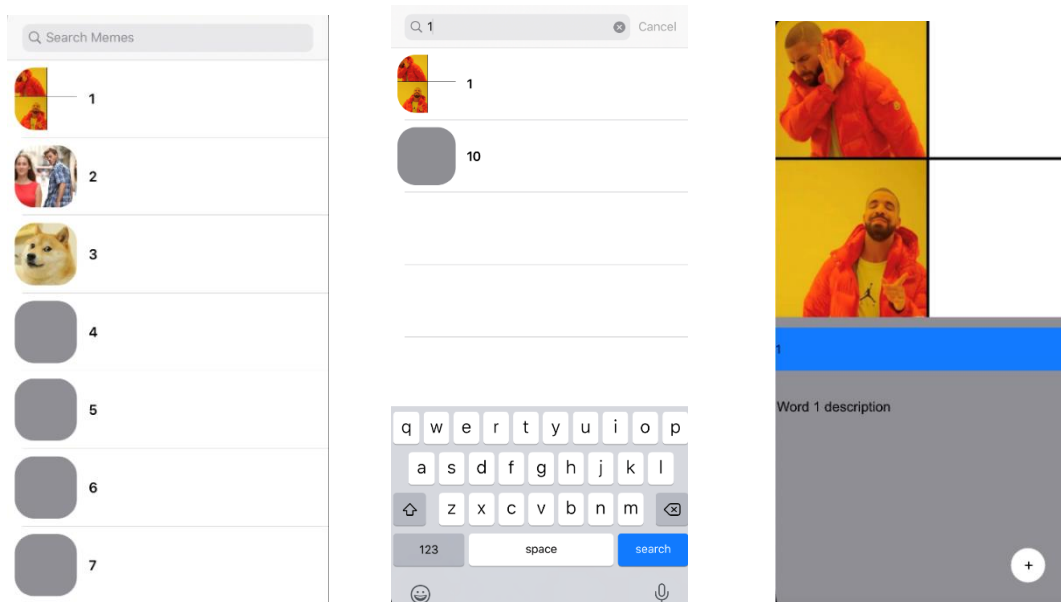


Figure 3: Searching page contains all vocabulary in alphabetic order. (left) Searching page that contains words based on user

Input. (middle) Vocabulary page that contains word, definition and image. (right)

```

Extension SearchWord{
  Override func tableView(_ tableView: UITableView, didSelectRowAt indexPath: IndexPath) {
    let explanation = ExplanationVC()

    if inSearchMode{
      explanation.memeString = searchData[indexPath.row].listNameData
      explanation.imageString=UIImage(named:
        searchData[indexPath.row].listNameImage) ?? UIImage(named: "1")!
      //send image + if no image then use imagel
      self.present(explanation, animated: true, completion: nil)
    }else{
      explanation.memeString = data[indexPath.row].listNameData
      explanation.imageString = UIImage(named:
        data[indexPath.row].listNameImage) ?? UIImage(named: "1")!
      self.present(explanation, animated: true, completion: nil)
    }
  }
}

```

Figure 4: Algorithm uses for searching word in the search engine.

Figure 5 is the editing page for users to make their own memes to help memorize the words. Users can click on the plus button to import their own images from their phones or to select templates from online sources. The text button allows users to add text on the image. Draw button can let users paint with various colors by using fingers. If users make a mistake while drawing, they can choose to undo the last step by clicking the undo button or to erase the drawing by tapping the eraser button. Below the tool buttons, there is a rectangular area that will demonstrate the word and its definition. On top right, there is a save button that allows users to save DIY memes to their study sets.

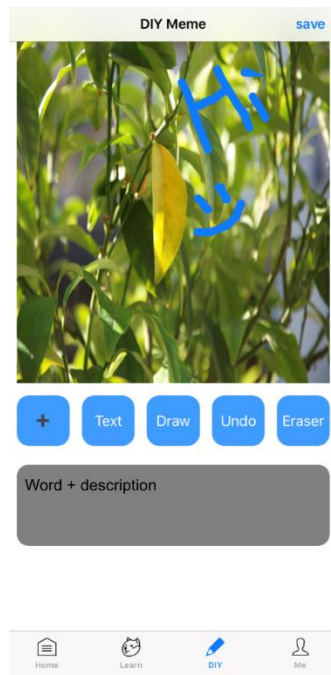


Figure 5: Editing page for adding memes to a vocabulary.

```

Override func draw(_rect: CGRect){
    super.draw(rect)
    guardlet context = UIGraphicsGetCurrentContext() else {return}
    //exception

    context.setStrokeColor(UIColor.systemBlue.cgColor) //control the color of
the pen
    context.setLineWidth(10) //control the width of the pen
    context.setLineCap(.round) // pen shape

    lines.forEach { (line) in
        for (i,j) in line.enumerated(){
            if i==0{
                context.move(to: j)
            }else{
                context.addLine(to: j)
            }
        }
    }

    context.strokePath()
}

//2D coordinate set
Fileprivate var lines = [[CGPoint]]()

Override func touchesBegan(_ touches: Set<UITouch>, with event: UIEvent?) {
    lines.append([CGPoint]()) //所有点连起来成线
}

Override func touchesMoved(_ touches: Set<UITouch>, with event: UIEvent?) {
    guardlet point = touches.first?.location(in: nil) else {return}
    guardvar endPoint = lines.popLast() else {return}
    endPoint.append(point) //pen up, append the end point when finished drawing
    lines.append(endPoint) //collect the first&last points to form a line
    setNeedsDisplay()
}

```

Figure 6: Code segment for creating the editing and setting up page

Figure 7 is a list of view of study sets. Users can scroll up and down to view the entire list. This page lists out all words in the set with words and meme previews. The preview images might be adjusted to a low resolution, but users are able to view the original image by clicking on the words. After tapping, users will enter a detailed view page that contains word, image, definition, and an explanation of how the word and the image relate to each other.

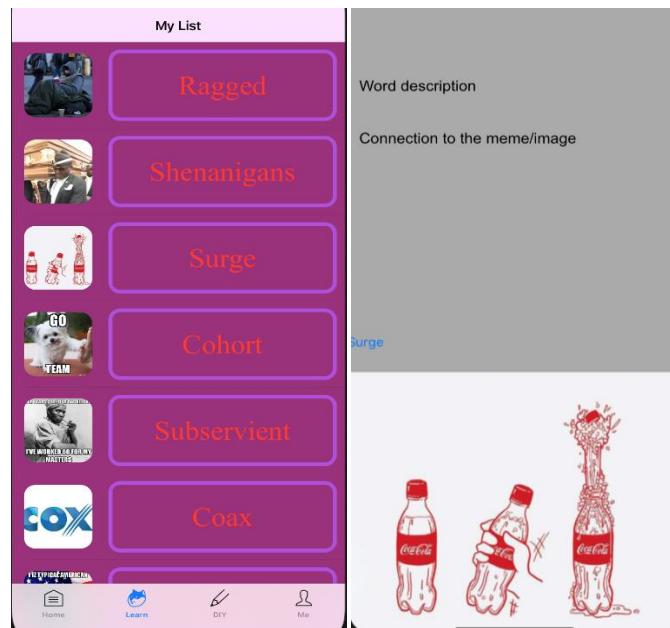


Figure 7: Study slist page contains words and memes. (left). Vocabulary page contains word, definition and image. (right)

```

Override func tableView(_ tableView: UITableView, cellForRowAt indexPath:
IndexPath) -> UITableViewCell {

    let cell = tableView.dequeueReusableCell(withIdentifier: "cell", for:
indexPath) as! ListCell

    cell.NewLabel.text = vocab[indexPath.row]
    cell.backgroundColor = CCFavColor
    cell.textLabel?.font = UIFont.boldSystemFont(ofSize: 28)
    cell.textLabel?.textColor = UIColor.systemGreen
    cell.coolerImageView.image = UIImage(named: imageList[indexPath.row])
    return cell
}

Override func tableView(_ tableView: UITableView, numberOfRowsInSectionInSection
section: Int) -> Int {

    return vocab.count
}

Override func tableView(_ tableView: UITableView, didSelectRowAt indexPath:
IndexPath) {
    let dtvc = DetailedVC()
    dtvc.imageString = UIImage(named: imageList[indexPath.row])!
    dtvc.dataString = vocab[indexPath.row]
    self.present(dtvc, animated: true, completion: nil)
}

```

Figure 8: Code segment for displaying study list page

As show Figure 9, there is a personal settings page where users can edit their profiles. The gray block on top illustrates the user's profile picture. Users can change their profile pictures by tapping on the gray block and selecting an image from a local photo library. The username will be demonstrated below the profile picture. If users click on My Account button, their account

information including name, email address and total number of memorized words will appear. My collection button will direct users to a list of the study sets they collected from others. The View History button can show the recent records of study sets users viewed. Lastly, My Plan button generates a personal study plan by utilizing machine learning.[15] [16] The user only needs to take an evaluation test and enter the goal. The program will recommend study sets corresponding to the user's skill level and record performance.

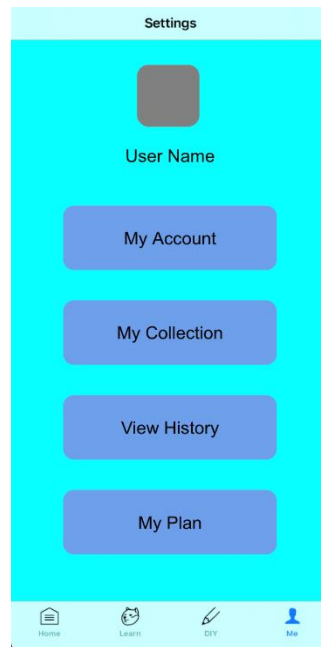


Figure 9: Personal setting page

```

Func TapToPhotoLib() {
    let tap = UITapGestureRecognizer(target: self, action:
    #selector(handleTap))
    profileImageView.addGestureRecognizer(tap)
}

@objc func handleTap() {
    //create image picker to select image from lib or camera
    let picker = UIImagePickerController()
    picker.delegate = self
    picker.sourceType = .photoLibrary
    picker.allowsEditing = true
    self.present(picker, animated: true, completion: nil)
}

Func imagePickerController(_ picker: UIImagePickerController,
    didFinishPickingMediaWithInfo info: [UIImagePickerController.InfoKey :
Any]) {
    let selectedImage = info[.editedImage] as! UIImage
    profileImageView.image = selectedImage
    self.dismiss(animated: true, completion: nil)
}

```

Figure 10: Code segment for adding image to the profile

4. EXPERIMENT

Memory retention for students through standardized studying is always a potential misstep for students as they attempt to figure out what system works best for them. By highlighting common areas of change, we can better test to see what changes will have a lasting impact. Using a study method dataset, we check for changes in vocabulary ordering, difficulty distribution, and problem typing.

4.1. Comparison of Vocabulary Ordering through Multiple Machine Learning Models

As each sub area has a largely different impact, we separate our experiments individually. When trying to retain new words such as those needed for the SAT, word groupings can often impact the level at which students remember. Testing for these involved gathering over 500 samples of various student study methods and using Linear Regression, support vector machine, and linear regression to predict the optimal study route.

Figure 11 illustrates the accuracy of each model's prediction. From the results, Linear Regression contains the highest predictive accuracy. The "line of best fit" regressive algorithm generates an algorithm that more accurately matches the general data of students that was created. Simply swapping out the ordering of vocab words allows for different groupings to emerge that mimic a linear relationship. By comparison, Polynomial Regression and SVM did not fare as well in this experiment. The sample data does not follow the polynomial algorithm nor can it be organized via support-vector machines.

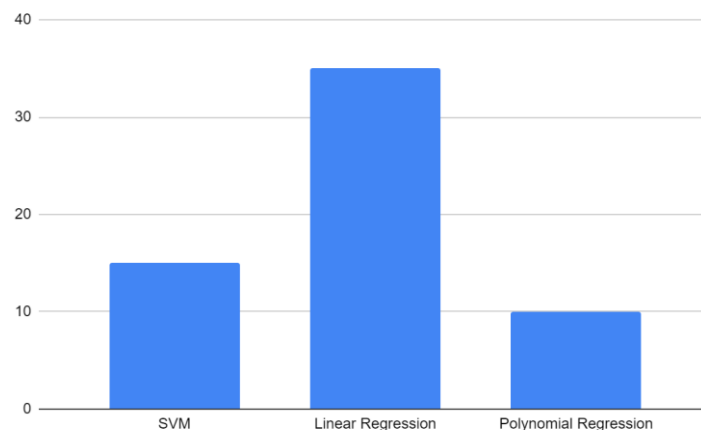


Figure 11: Accuracy of prediction of vocabulary ordering through SVM, Linear regression and Polynomial regression

4.2. Predicting the best Distribution in Terms of Problem Difficulty

Each of the following experiments follows the same formatting of over 500 generated samples for the given area, with the accuracy test for each individual algorithm: SVM, Linear Regression, Polynomial Regression. For this specific experiment we aimed to see which algorithm could predict the best distribution in terms of problem difficulty: all very hard, all very easy, variation, etc.

As seen in Figure 12, the various models all have a marginal impact on difficulty distribution, with Linear Regression and Polynomial regression both being the standout methods. Difficulty distribution is a much more noticeable area of change as even small shifts, such as from mostly easy words to mostly difficult words, can have a big impact on future memory retention. Both linear and polynomial algorithms work quite well as, while a singular line of straight fit covers a majority of the changes, there is a large swing in difference as simply studying different words has a large impact for students.

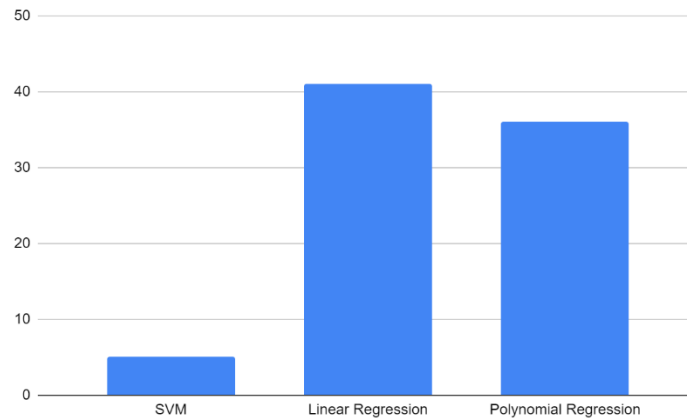


Figure 12: Accuracy of prediction of distribution in term of problem difficulty through SVM, Linear regression and Polynomial regression

4.3. Predicting the Optimal Typing Distribution for Studying

With our final experiment we once again used a sample size of 500 generated data points for testing whether the selected machine learning models can predict the optimal typing distribution for studying. This compared the memory retention of each of our study sets with the main variance being which type of vocabulary words are grouped together: (all from the same subject, even split between related topics, random, etc.). Oftentimes, learning multiple related topics can not only increase understanding, but drastically increase how long afterwards a student can remember.

Figure 13 showcases the results of the three tests and aligns well with the previously completed experiments. Linear Regression continues to be the model with the highest accuracy. This is primarily due to our common study data and its lack of polynomial factors. The primary pieces of data in this experiment as well showcase that our gathered information has a linear relationship that is impactful during the training process. SVM continues in this experiment to be the model with the lowest resulting accuracy, going so far down as to be almost completely inconsequential.

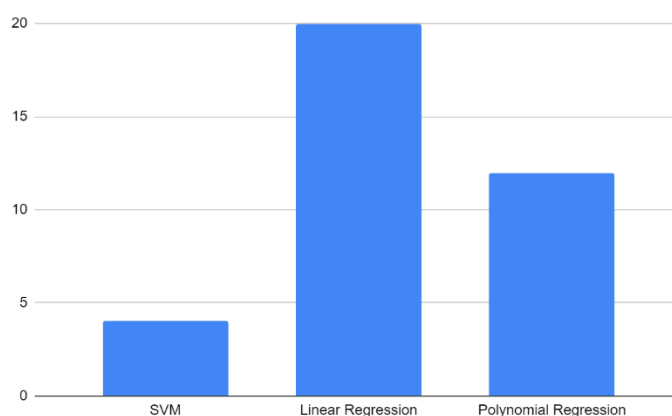


Figure 13: Accuracy of prediction of vocabulary ordering through SVM, Linear regression and Polynomial regression

4.4. Analysis

Overall, the experiment results support our goal of using machine learning algorithms to test study plan changes that impact student memory retention. Across all three experiments, Linear Regression consistently stands out with a much higher level of accuracy compared to other methods, peaking at 41%. Given the topic and organization of data this result makes sense as a major portion of the data follows a linear algorithmic relationship. For problem ordering and difficulty distribution in particular, linear regression was able to have a much stronger impact.

5. RELATED WORK

Nguyen, T and Khuat, T. proposed a method to learn English vocabulary in a classroom. [6] Even though the research was conducted for 2 weeks, they showed that using fun games can help Vietnamese students improve their English vocabulary. However, they suggested that vocabulary games need to be designed according to the number of students, proficiency level, cultural context, timing and learning topic. Thus, we decided to use machine learning model to personalize the study plan. As different of Nguyen, T and Khuat, T. method, our learning method does not have time and place restriction since it can be used anytime and anywhere.

Merriam, A. presented a comparison between translation and pictorial methods to recall and retain English vocabulary. [7] In his research, he concluded that there is not difference between translation and pictorial methods for recalling and retaining English vocabulary. He stated that pictorial method can produce misunderstanding since learners can interpret a picture in different way. In our mobile application, we used similar method; but as different of this method, learners can add their own picture that help them to remember vocabulary.

Azadeh, R., Mai, N., and Pesaranghader, A. conducted a research about the effect of using mobile application on learning English vocabulary [8]. Their study showed that using mobile application can bring positive impact on learners' performance since learners improved their English vocabulary. We took similar approach, but the difference is that we use customize vocabulary that can be showed in SAT or ACT test.

6. CONCLUSION AND FUTURE WORK

Vocabulary is vital to communicate our thought and perform satisfactory in standardized test such as ACT and SAT. In order to be successful in standardized test, learners use appropriately study plan methods. However, different learning plan can be suitable for some learners but not for others. Some of them may use flashcards, vocabulary games, pictorial method etc. Using flashcards and memorizing the definition may seem efficient, but learners may forget the definition after a while, so all the work would be wasted.

In this project, we proposed an intelligent study method to solve memorization issue using mobile computing and machine learning; and a memorization method to recall and retain vocabulary by using picture and memes. A mobile application has been developed to design an efficient study plan. To obtain a study plan, learners must enter the desirable ACT/SAT score and take an assessment. Once a learner finished the assessment, the system predicts a study plan that is suitable for the learner and return a study plan based on his/her performance.

Our experiments show that the linear regression model stand out in our three experiments. The highest accuracy of the linear regression model was 41% when the vocabulary is grouping based on level of difficulty of vocabulary.

As a future work, we plan to gather a group of high school students to use our mobile application during the SAT study period and observe if our mobile application improved learners' memorization.

One of the limitations that we encounter during the training process was that the size of the study method dataset used in our machine learning model was not enough to get a satisfactory accuracy. We plan to handle a survey to high school students who already took the SAT and input to our dataset. In this survey, we will ask students to pick the best study method and their SAT score.

REFERENCES

- [1] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research* 12 (2011): 2825-2830.
- [2] Goodfellow, Ian, Y. Bengio, and A. Courville. "Machine learning basics." *Deep learning*. Vol. 1. MIT press, 2016. 98-164.
- [3] John, Ralph C. St. "Applied linear regression models." (1983): 201-202.
- [4] Ostertagová, Eva. "Modelling using polynomial regression." *Procedia Engineering* 48 (2012): 500-506.
- [5] Soman, K. P., R. Loganathan, and V. Ajay. *Machine learning with SVM and other kernel methods*. PHI Learning Pvt. Ltd., 2009.
- [6] Huyen, Nguyen Thi Thanh, and Khuat Thi Thu Nga. "Learning vocabulary through games." *Asian EFL Journal* 5.4 (2003): 90-105.
- [7] Al Nassir, Merriam. "Meaning recall and retention: comparison between translation method and pictorial method in learning vocabulary in Saudis' school." *2000-2019-CSU Theses and Dissertations* (2012).

- [8] Rezaei, Azadeh, Neo Mai, and Ahmad Pesaranhader. "The effect of mobile applications on English vocabulary acquisition." *Jurnal Teknologi* 68.2 (2014).
- [9] Mountain, Lee Harrison. *Ten essential vocabulary strategies: Practice for success on standardized tests*. Educators Publishing Service, 2004.
- [10] Nadler, Burton Jay, Jordan Nadler, and Justin Nadler. *Words You Should Know In High School: 1000 Essential Words To Build Vocabulary, Improve Standardized Test Scores, And Write Successful Papers*. Simon and Schuster, 2004. [11]
- [11] Lessard-Clouston, Michael. "Strategies and success in technical vocabulary learning: Students' approaches in one academic context." *Indian Journal of Applied Linguistics* 34.1/2 (2008): 31.
- [12] Vossoughi, Hossein, and Marzieh Zargar. "Using word-search-puzzle games for improving vocabulary knowledge of Iranian EFL learners." (2009): 79-85.
- [13] Webb, Stuart A. "The effects of pre-learning vocabulary on reading comprehension and writing." *Canadian Modern Language Review* 65.3 (2009): 441-470.
- [14] Kornell, Nate. "Optimising learning using flashcards: Spacing is more effective than cramming." *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 23.9 (2009): 1297-1317.
- [15] Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2020.
- [16] Michie, Donald, David J. Spiegelhalter, and C. C. Taylor. "Machine learning." *Neural and Statistical Classification* 13.1994 (1994): 1-298.
- [17] Moran, Kevin Patrick, et al. "Machine learning-based prototyping of graphical user interfaces for mobile apps." *IEEE Transactions on Software Engineering* (2018).
- [18] Aggarwal, Vaneet, et al. "Prometheus: Toward quality-of-experience estimation for mobile apps from passive network measurements." *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*. 2014.

PNFUZZ: A STATEFUL NETWORK PROTOCOL FUZZING APPROACH BASED ON PACKET CLUSTERING

HuiHui He and YongJun Wang

College of Computer Science, National University of
Defense Technology ChangSha, China

ABSTRACT

Due to the interactivity of stateful network protocol, network protocol fuzzing has higher blindness and lower testcase validity. The existing blackbox-based fuzzing has the disadvantages of high randomness and blindness. The manual description of protocol specification which requires more expert knowledge, is tedious and does not support the protocol without public document, which limits the effect of current network protocol fuzzer. In this paper, we present PNFUZZ, a fuzzer that adopts the state inference based on packet clustering algorithm and coverage oriented mutation strategy. We train a clustering model through the target protocol packet, and use the model to identify the server's protocol state, thereby optimizing the process of testcase generation. The experimental results show that the proposed approach has a certain improvement in fuzzing effect.

KEYWORDS

Fuzzing, Software Vulnerabilities, Network Protocol, Network Packet Clustering.

1. INTRODUCTION

With the rapid development of the Internet and the Internet of Things, various computer technologies have brought great convenience and improvement to people's lives. At the same time, people have more and more important data stored on the Internet, and their dependence on networks and computers is increasing. However, network attacks, data theft and virus spread are becoming more and more serious. For instance, the famous WannaCry [1] ransomware attack was a cyberattack, which targeted computers on Windows operating system. It made use of "EternalBlue" exploit, which found a security flaw in Windows' Server Message Block(SMB) protocol, an implementation of the application layer protocol which was mainly used as the communication protocol of Microsoft network. The prevalence of software security vulnerabilities is one of the main reasons for the frequent occurrence of such network security incidents. Network server is more likely to be targeted and attacked by criminals because of its exposed communication ports in the public environment. With the popularization of smart devices and the development of the Internet of Things, smart devices with networking functions, its software and network protocols have sprung up one after another, which has exacerbated the severity and challenges of the network security situation [2]. Therefore, digging out potential undiscovered security vulnerabilities and repairing them play a huge role in ensuring the security of cyberspace.

In order to alleviate the losses caused by software vulnerabilities, vulnerability mining techniques are currently effective technical methods. Vulnerability mining technique based on fuzz testing

[3][4] has the characteristics of high degree of automation, less expert knowledge, and lower false positive rate. It constructs a large number of deformed and mutated data to input to the software. The software detection technique determines whether the software has caused errors, abnormalities or even crashes after receiving and processing these malformed data. If these conditions occur, analyzing them to determine whether the vulnerability has been triggered and whether it can be exploited.

This paper is organized as follows: Section 2 introduces the previous work and the advantages and disadvantages of the existing methods. Section 3 introduces the principle and components of PNFUZZ. Section 4 analyses the experimental results. Section 5 gives a conclusion.

2. PREVIOUS WORK

There are two challenges in fuzz testing of network protocol. Firstly, network protocol is a kind of agreement rule of network data communication between client and server. It specifies the type, structure and size of data sent by both sides of communication. In the process of fuzz testing, testcases need to meet the protocol specifications to a certain extent, otherwise they are easy to be discarded. Secondly, many network protocols are stateful, and the client and server need to keep communicating in a session. This kind of session needs to be maintained in fuzz testing, otherwise the two sides of communication will not be able to communicate normally, which will lead to the failure of fuzz testing. Faced with these two challenges, there are several "dumb" and "smart" protocol fuzzer. "Dumb" fuzzer sends randomly generated test cases to server under test, without knowing the network protocol adopted by server under test. Therefore, "dumb" fuzzer is suitable for any protocol and is easy to develop, because it only needs to generate testcases randomly for different servers. However, due to the randomness and blindness of "dumb" fuzzer, the effectiveness of testcases is much lower than that of "smart" fuzzer. ProxyFuzzer [5] is a typical "dumb" fuzzer, which randomly modifies the communication data between the client and the server by acting as a proxy between the client and the server. Although this fuzzer can be used in the client and server that maintain the session state, but randomly generated testcases make the crash probability of server very low.

Due to the limitation of "dumb" fuzzer in network protocol fuzz testing, "smart" fuzzer is more popular at present. It tries to understand the internal implementation and format of network protocol to generate testcases. Generally speaking, "smart" fuzzer can understand the data structure of network packets by constructing protocol Finite State Machine(FSM), referencing protocol specification or data syntax. Therefore, "smart" fuzzer can produce testcases with lower blindness and randomness and higher efficiency. For example, Peach [6], BooFuzz [7] and AFLNET [8] are "smart" fuzzers. Taking BooFuzz to test the SMTP protocol for example, BooFuzz is a kind of "smart" fuzz testing framework. Users can decide which parts of the data need to be modified and which parts cannot be modified according to the specification and characteristics of the target protocol. Before starting fuzzing, we need to provide the protocol specification of SMTP. An interaction process between the client and server of the SMTP protocol is shown in Table 1. The italic text can be changed, while other data cannot be changed. Otherwise, the interaction will be interrupted. So the fixed part of the requested data is not modified by the configuration of BooFuzz, but the part and its length of the changeable data can be modified randomly. Then BooFuzz determines the status of the server according to the server response status code, and run the next test. Therefore, the test efficiency of "smart" fuzzer depends heavily on the understanding of network protocol specification and communication data format syntax. Before fuzz testing, we need to learn and understand the target network protocol specification based on some relevant documents, and then manually build the data structure of communication data. Moreover, we also need to build a protocol FSM or state transition diagram to describe the state transition between the client and the server. Therefore, these methods has

complex preparation, low level of automation, tedious process of manually defining protocol format and does not support undocument protocol.

Table 1. A mailing interaction of SMTP.

State	Cilent Request	Server Response
220	Try to connecting...	220 xxx.com Anti-spam GT for Coremail System
250	HELO SMTP.xxx.com	250 OK
334	AUTH LOGIN	334 username
334	test@xx.com	334 password
235	123456	235 auth successfully
250	MAIL FROM: test@xx.com	250 OK
250	RCPT TO: test2@xx.com	250 OK
354	DATA	354 Enter mail, end with "." on a line by itself
250	hello mail!.	250 Message sent
221	QUIT	221 Bye

In this work, we propose PNFUZZ to deal with the challenges and limitations of the above methods. PNFUZZ adopts state inference based on packet clustering algorithm and coverage-oriented mutation strategy. Firstly, PNFUZZ infers the session state by clustering the response packets instead of manually understanding the protocol specification, and selects the corresponding test case according to the session state. This method is used to solve the maintenance of session state in stateful network protocol fuzz testing. Secondly, PNFUZZ uses the coverage oriented mutation strategy to generate the final testcases. This method mutates the request data, and does not need to manually build the data structure of the target protocol data. This method is used to solve the randomness and blindness of testcase generation.

3. TOOL DESIGN AND IMPLEMENTATION

3.1. Description and definition of problems

Before introducing PNFUZZ in detail, we need to introduce some definitions and terms.

Definition 1 (Fuzz Testing). A process of repeatedly inputting testcase into server under test.

Definition 2 (Seed). Test data before mutation.

Definition 3 (Testcase). Mutated seed, which is finally sent to server under test.

Definition 4 (Monitor). In the process of fuzz testing, checking whether server under test generates an exception or crash.

Network protocol is an agreement for communication between programs. According to International Organization for Standardization (ISO), computer network protocols can be divided into five layers: physical layer, data link layer, network layer, transport layer and application layer. Depending on the application of network programs, there are a large number of different application layer protocols, such as File Transfer Protocol (FTP), Domain Name Resolution Protocol (DNS), and Hypertext Transfer Protocol (HTTP). In this paper, the fuzz testing for network protocol mainly refers to fuzz application layer protocol. The network protocol can be classified to stateful protocol and stateless protocol depending on whether the request and

response data are independent. For the stateless protocol, there is no dependence between the request data sent by the client one after the other, and the server only responds to the received client data, and does not depend on the previous request and response data, such as HTTP. For the stateful protocol, the request data sent by the client is determined by the last response data received by the server, such as FTP. Generally, the fuzz testing for file processing software only needs to input each testcase to program under test, and monitor the running and execution of it. However, it is more difficult to fuzz network server. Basing on understanding the data format of network protocol, the testcases can be sent to the server after establishing a connection with the server through the network and fuzzer needs to maintain a session with the server.

3.2. Packet clustering algorithm

Clustering [9] is an unsupervised machine learning algorithm. It divides unlabeled datasets into several categories by similarity metric, and each data is classified into a certain cluster. The data of the same cluster have higher similarity, while the data of different clusters have lower similarity. The popular clustering algorithms are K-means, DBSCAN and Hierarchical Clustering. In this work, we use Aggregate Hierarchical Clustering [10] to divide the response datasets into multiple categories. The advantage of this algorithm is that it does not need to manually specify the number of final clusters. The similarity between datasets can be calculated by distance metric, such as Euclidean distance, Cosine distance, and Longest Common Subsequence Distance. Among them, Longest Common Subsequence Distance [11] is more suitable for the calculation of the similarity of binary data.

Given data a and b , we define the Longest Common Subsequence Distance D as follows:

$$D(a, b) = 1 - \frac{\text{length}(LCSS(a,b))}{\max(\text{length}(a), \text{length}(b))} \quad (1)$$

where $LCSS$ is the longest common subsequence algorithm.

3.3. Model and Algorithm

The overall architecture and implementation of PNFUZZ is shown in Figure 1. PNFUZZ consists of data preprocessing, instrumentation, seed selector, state recognition, mutator and two communication channels. We elaborate on each below.

(1) Two Communication Channels

In order to implement the fuzz testing of network protocol, the basic ability of PNFUZZ is to communicate with server under test. PNFUZZ realizes two communication channels. One realizes network communication between PNFUZZ and server under test through socket APIs. The other is to feed back the code coverage information of server under test to PNFUZZ through anonymous pipes and shared memory. It is worth mentioning that instrumentation can insert anonymous pipes and shared memory related code into the server under test.

(2) Data Collection and Processing.

Data collection is the first preparation after selecting the target protocol. We can capture the request and response data between the client and the server through the network sniffing tool such as Wireshark [12] and Tcpdump [13]. In this work, we select Wireshark to capture protocol data. The advantage of Wireshark is that it supports multiple platforms and can extract protocol

session easily with GUI interface. The protocol data captured by Wireshark can be saved to PCAP file for subsequent analysis and processing.

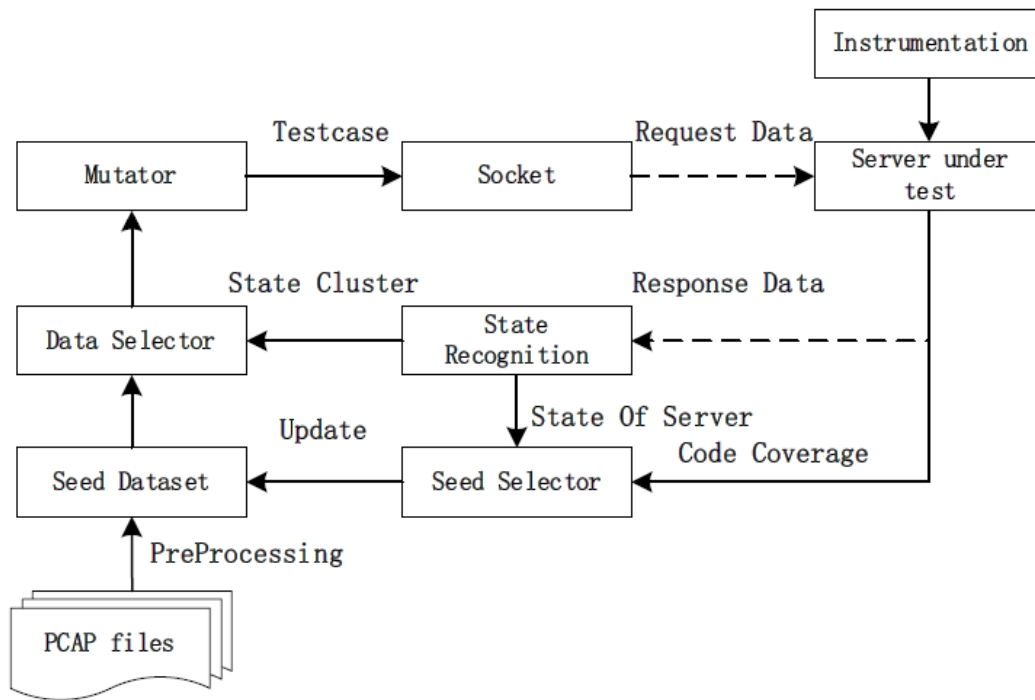


Figure 1. Architecture and Implementation of PNFUZZ

After collecting enough data, it is necessary to preprocess datasets. On the one hand, there is noise in the captured data, on the other hand, PCAP file cannot be directly applied in the fuzz testing. Data preprocessing includes the following steps:

- Removing network packets from non-target protocols.
- Removing invalid and redundant packets due to retransmission mechanism.
- The packets of the session between the client and the server are extracted completely.
- Filling all response packets with "\x00" until the length is the same as the length of the longest packet.
- Establishing relationships between request data and response data for each session.

Finally, we formally describe the processed dataset as follows:

Define request data or seed as $Req = (cid, sid, payload)$, and define response data as $Res = (sid, cid, payload)$. Where cid denote identification of single request data in the request dataset, sid denote identification of single request data in the request dataset, and $payload$ denote binary request or response data. Define request datasets as $REQ = \{Req1, Req2, \dots, Reqm\}$ and define response datasets as $RES = \{Res1, Res2, \dots, Resn\}$

(3) Instrumentation

Instrumentation is a technique for executing some extra code during the normal execution of a program. In the aspect of fuzz testing, instrumentation has two important functions. Firstly, it can collect program runtime information, output debugging information and track program execution status. The second is to enhance the sensitivity of the program to anomalies and detect anomalies in time. For example, when some abnormal behaviours occur such as array index out of range

and corrupting local variables on the stack, the program may not crash immediately, but will affect the subsequent program behaviour, even the program will not be affected at all. After instrumentation, the program will crash immediately when it detects abnormal behaviour, and we can check the thread context and core dump in time.

(4) State Recognition

PNFUZZ infers the session state by the response message of server under test. Before fuzz testing, we need to train a clustering model based on response datasets RES . In this work, we implement Hierarchical Clustering Algorithm by the Agglomerative Clustering component of Sklearn [14] which is a popular machine learning algorithm library. We get the clustering results $\mathcal{C}=\{C_1,\dots,C_k\}$ of all session states of server under test and a clustering model M . Each session state cluster C_i contains some response data: $C_i\subseteq RES$. Clusters are disjoint: $C_i\cap C_j=\emptyset, \forall i=1k, C_i\subseteq RES (i\neq j, i\leq k; j\leq k)$. In the fuzz testing, the protocol data P of the response packet is extracted, and then the session state cluster C_r of the protocol data is predicted by the trained clustering model.

(5) Seed Selector

According to the result of the state recognition, seed selector randomly selects a seed Req_r from request datasets REQ associated with the session state C_r of server under test and pass it to the data mutator. In addition, if the code coverage cov_r of server under test is higher than the average code coverage cov through anonymous pipes and shared memory or server under test crash, the previous sent testcase Tr' should be added to the request datasets of corresponding states. Otherwise, this testcase will be discarded. Finally, the code coverage cov_r is recorded into set $COV=\{cov_1,\dots,cov_k\}$. The complete algorithm of state recognition and seed selection is shown in algorithm 1. seed set updating algorithm is shown in algorithm 2.

Algorithm 1 State recognition by packet cluster

Input: RES, REQ , clustering model M and the target protocol data P

Output: the session state cluster C_r and testcase T_r

```

1:  $Mat \leftarrow \emptyset$ 
2:  $cid \leftarrow 0$ 
3: for  $Res_a$  in  $RES$  do
4:   for  $Res_b$  in  $RES$  do
5:      $Mat \leftarrow Mat \cup D(Res_a, Res_b);$ 
6:   end for
7: end for
8:  $M \leftarrow \text{AgglomerativeClustering}(RES, Mat)$ 
9:  $C_r \leftarrow \text{Predict}(M, Mat, P)$ 
10:  $C \leftarrow \text{RandomSelect}(C_r)$ 
11:  $cid \leftarrow \text{GetMember}(C, "cid")$ 
12:  $T_r \leftarrow REQ[cid]$ 
13: return  $C_r, T_r$ 

```

Algorithm 2 seed set updating

Input: REQ , C_r , T'_r , cov_r and COV **Output:** REQ and COV

```

1:  $\overline{cov} \leftarrow \text{Average}(COV)$ 
2: if ( $cov_r > \overline{cov}$ ) | (Server under test crashed) then
3:    $REQ \leftarrow REQ \cup T'_r$ 
4:    $COV \leftarrow COV \cup cov_r$ 
5:    $\overline{cov} \leftarrow \text{Average}(COV)$ 
6: else
7:    $REQ \leftarrow REQ \setminus \{C_r\}$ 
8: end if
9: return  $REQ$ ,  $COV$ 

```

(6) Mutator

The mutator mutates the seed selected by the seed selector to produce the final testcase, which is then sent to server under test through the socket network communication channel. PNFUZZ's mutation strategy is based on AFL [15], which has very excellent extension such as LearnAFL[16].

4. EXPERIMENT AND RESULT ANALYSIS**4.1. Experiment Setting**

This experiment is based on Ubuntu 18.04 operating system and will be compared with AFLNET and BooFuzz. AFLNET is a state-oriented fuzzing tool for stateful network protocols, which can perform better fuzzing tests on common network protocols. BooFuzz is a stateful blackbox fuzzing framework with high scalability. The FTP protocol server LightFTP [17] and the RSTP protocol server Live555 [18] are the test targets. Using BooFuzz as a benchmark and comparing performance of PNFUZZ, AFLNET and BooFuzz in three indicators: fuzzing efficiency, code branch coverage and vulnerability mining capabilities.

4.2. Result Analysis

The higher code coverage means that the server under test to enter deeper code. Covering more code means more likely to trigger more bugs or vulnerabilities. The code coverage in Table 2 is based on BooFuzz and shows that PNFUZZ and AFLNET have improvement in these two protocols. The code coverage of AFLNET is higher than that of PNFUZZ on LightFTP and lower on Live555. The reason is that AFLNET will make certain adaptations to FTP, and the FTP data format is more simple, while the RTSP data format is more complex. Since BooFuzz is a blackbox fuzzer, its code coverage is inferior to that of coverage-oriented graybox fuzzing. Table 3 and Table 4 show that BooFuzz relies on manual construction of protocol specifications, is faster in the testing process, and about 10 times faster than AFLNET, but the quality of its testcases is relatively poor. So it is difficult to trigger software crash. AFLNET sends testcases at a little bit faster rate than PNFUZZ. The reason is that PNFUZZ needs to recognize server's state through a trained clustering model before sending testcase, which consumes time. Although a certain amount of time is sacrificed, the quality of the generated testcases is better, and in the final the number of crashes triggered is bigger than AFLNET. Table 4 shows that LightFTP did

not trigger any crashes. This phenomenon may be caused by the short fuzz testing time. In short, although the test speed of PNFUZZ is relatively slower, the quality of the generated testcases is relatively better and PNFUZZ is more highly automated and achieve the expected effect.

Table 2. Code coverage improvement.

	LightFTP	Live555
AFLNET	54 %	63%
PNFUZZ	51%	68%

Table 3. Sending testcase speed.

	LightFTP	Live555
BooFuzz	100~110 count/s	100~110 count/s
AFLNET	9 count/s	9 count/s
PNFUZZ	8 count/s	7 count/s

Table 4. Average number of crashes per hour.

	LightFTP	Live555
BooFuzz	0	2
AFLNET	0	6
PNFUZZ	0	7

5. CONCLUSION AND FUTURE WORK

This paper proposes PNFUZZ, a stateful network protocol fuzzer based on the packet clustering algorithm. The trained model of network packet clustering is used to identify the session state of server under test, thereby optimizing the process of selecting testcases. Combining with coverage-guided greybox fuzzing strategy, PNFUZZ improves the quality of testcases.

In the future, more effective protocol reverse techniques and machine learning algorithms can be used to improve the overall fuzzing effect.

ACKNOWLEDGEMENTS

This work was funded by the National Natural Science Foundation of China, and the project approval number was 61472439.

REFERENCES

- [1] Akbanov, M., Vassilakis, V. G., & Logothetis, M. D. (2019). WannaCry ransomware: Analysis of infection, persistence, recovery prevention and propagation mechanisms. *Journal of Telecommunications and Information Technology*.
- [2] Khan, M. A., & Salah, K. (2018). IoT security: Review, blockchain solutions, and open challenges. *Future Generation Computer Systems*, 82, 395-411.
- [3] Liu, B., Shi, L., Cai, Z., & Li, M. (2012, November). Software vulnerability discovery techniques: A survey. In *2012 fourth international conference on multimedia information networking and security* (pp. 152-156). IEEE.

- [4] Manès, V. J. M., Han, H., Han, C., Cha, S. K., Egele, M., Schwartz, E. J., & Woo, M. (2019). The art, science, and engineering of fuzzing: A survey. *IEEE Transactions on Software Engineering*.
- [5] Website, “ProxyFuzz,” <https://github.com/SECFORCE/proxyfuzz/blob/master/proxyfuzz.py> , 2020, accessed: 2020-06-17.
- [6] Wang, E., Wang, B., Xie, W., Wang, Z., Luo, Z., & Yue, T. (2020). EWVHunter: Grey-Box Fuzzing with Knowledge Guide on Embedded Web Front-Ends. *Applied Sciences*, 10(11), 4015.
- [7] Website, “BooFuzz: A fork and successor of the sulley fuzzing framework,” <https://github.com/jtpereyda/boofuzz>, 2020, accessed: 2020-06-17.
- [8] Pham, V. T., Böhme, M., & Roychoudhury, A. (2020). AFLNET: A Greybox Fuzzer for Network Protocols. In *Proc. IEEE International Conference on Software Testing, Verification and Validation (Testing Tools Track)*.
- [9] Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data* (pp. 25-71). Springer, Berlin, Heidelberg.
- [10] Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), 86-97.
- [11] Bergroth, L., Hakonen, H., & Raita, T. (2000, September). A survey of longest common subsequence algorithms. In *Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000* (pp. 39-48). IEEE.
- [12] Orebaugh, A., Ramirez, G., & Beale, J. (2006). *Wireshark & Ethereal network protocol analyzer toolkit*. Elsevier.
- [13] Fuentes, F., & Kar, D. C. (2005). Ethereal vs. Tcpdump: a comparative study on packet sniffing tools for educational purpose. *Journal of Computing Sciences in Colleges*, 20(4), 169-176.
- [14] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- [15] Website, “American fuzzy lop (afl) fuzzer,” http://lcamtuf.coredump.cx/afl/technical_details.txt, 2020, accessed: 2020-06-17.
- [16] Yue, T., Tang, Y., Yu, B., Wang, P., & Wang, E. (2019). LearnAFL: Greybox Fuzzing With Knowledge Enhancement. *IEEE Access*, 7, 117029-117043.
- [17] Website, Lightftp server, <https://github.com/hfiref0x/LightFTP>, 2020, accessed: 2020-06-17.
- [18] R. Finlayson, Live555 media server, <http://www.live555.com/mediaServer>, 2020, accessed: 2020-06-17.

TOWARDS ASSESSING CRITICAL INFRASTRUCTURES' CYBER-SECURITY CULTURE DURING COVID-19 CRISIS: A TAILOR-MADE SURVEY

Anna Georgiadou, Spiros Mouzakitis and Dimitrios Askounis

Decision Support Systems Laboratory, National Technical University of
Athens, Iroon Polytechniou 9, 15780 Zografou, Greece

ABSTRACT

This paper outlines the design and development of a survey targeting the cyber-security culture assessment of critical infrastructures during the COVID-19 crisis, when living routine was seriously disturbed and working reality fundamentally affected. Its foundations lie on a security culture framework consisted of 10 different security dimensions analysed into 52 domains examined under two different pillars: organizational and individual. In this paper, a detailed questionnaire building analysis is being presented while revealing the aims, goals and expected outcomes of each question. It concludes with the survey implementation and delivery plan following a number of pre-survey stages each serving a specific methodological purpose.

KEYWORDS

Cybersecurity Culture, Assessment Survey, COVID-19 Pandemic, Critical Infrastructures

1. INTRODUCTION

Coronavirus disease 2019, widely known as COVID-19, is an infectious dis-ease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. The disease was first detected in late 2019 in the city of Wuhan, the capital of China's Hubei province [2]. In March 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a pandemic [3]. To-day, with more than 11 million confirmed cases in 188 countries and at least half a million casualties, the virus is continuing its spread across the world. While epidemiologists argue that the crisis is not even close to being over, it soon become apparent that “the COVID-19 pandemic is far more than a health crisis: it is affecting societies and economies at their core” [4].

Terms such as “Great Shutdown” and “Great Lockdown” [5, 6, 7] have been introduced to attribute the major global recession which arose as an eco-nomic consequence of the ongoing COVID-19 pandemic. The first noticeable sign of the coronavirus recession was the 2020 stock market crash on the 20th February. International Monetary Fund (IMF) in the April World Economic Outlook projected global growth in 2020 to fall to -3 percent. This is a downgrade of 6.3 percentage points from January 2020, making the “Great Lockdown” the worst recession since the Great Depression, and far worse than the Global Financial Crisis [7]. According to the International Labour Organization (ILO) Monitor, published on 7th April 2020, full or partial lockdown measures are affecting almost 2.7 billion workers, representing around 81% of the world's workforce [8].

Organizations from different business domains and operation areas across the globe try to survive this unprecedented financial crisis by investing their hopes, efforts and working reality on information technology and digitalization. Employees are being encouraged and facilitated on teleworking while most products and services become available over the web while, in many cases, transforming and adjusting to current rather demanding reality. However, these organizations come up against another COVID-19 side-effect not that apparent: the cyber-crime increase.

The increase in the population percentage connected to the World Wide Web and the expansion of time spent online, combined with the sense of confinement and the anxiety and fear generated from the lockdown, have formulated a prosperous ground for cyber-criminals to act. Coronavirus has rapidly reshaped the dark web businesses, as buyers and sellers jump on the opportunity to capitalize on global fears, as well as dramatic shifts in supply and demand. Phishing emails, social engineering attacks, malware, ransomware and spyware, medical related scams, investment opportunities frauds, are only a few examples of the cyber-crime incidents reported during the crisis period [9, 10].

INTERPOL's Cybercrime Threat Response team has detected a significant increase in the number of attempted ransomware attacks against key organizations and infrastructure engaged in the virus response. Cybercriminals are using ransomware to hold hospitals and medical services digitally hostage; preventing them from accessing vital files and systems until a ransom is paid [11].

Cyber-security agencies, organizations and experts worldwide have issued recommendations and proposed safeguard measures to assist individuals and corporations to defend against cyber-crime. While the virus is dominating in every aspect of our daily lives and human interaction is being substituted by digital transactions, cybersecurity gains the role it was deprived from during the last years. The question that remains unanswered, given the circumstances, is: What are the COVID-19 pandemic cyber-security culture side-effects on both individual and organizational level?

This paper presents the design and delivery plan of a survey aiming to evaluate the cyber-security culture during COVID-19 pandemic in the critical infrastructure domain. Section 2 presents background information regarding the importance of public cyber-security surveys conducted over the years emphasizing on the variety and originality of their findings. Building upon their approach, a detailed methodology is presented in Sections 3 & 4, in an effort to develop a brief, targeted and comprehensible survey for the assessment of the cybersecurity readiness of organizations during the crisis with emphasis on employees' feelings, thoughts, perspective, individuality. In Section 5, we sketch the survey next steps towards its conduction and fruitful completion. Finally, Section 6 concludes by underlying the importance of our survey reasoning while focusing on the challenging scientific opportunities that arise from it.

2. BACKGROUND

Over the last decades, cybersecurity surveys have been a powerful asset for information security academics and experts seeking to explore the constantly transforming technological reality. Their aim has been to reveal the contemporary trends on cybersecurity threats, organizations investment priorities, solutions for cloud security, threat management, application security, security training and certifications, and so many other topics.

Initially, they were narrowed down and addressed to certain participants depending on the nature and specific goal of each survey. A lighthouse representative of this kind was the Computer

Crime & Security Survey conducted by the Computer Security Institute (CSI) with the participation of the San Francisco Federal Bureau of Investigation's (FBI) Computer Intrusion Squad. This annual survey, during its 15 years of life (starting from 1995 and reaching up to 2010), was probably one of the longest-running continuous surveys in the information security field [12]. This far-reaching study provided unbiased information and analysis about targeted attacks, unauthorized access, incident response, organizational economic decisions regarding computer security and risk management approaches based on the answers provided by computer security practitioners in U.S. corporations, government agencies, financial institutions, medical institutions and universities.

Following their lead, numerous organizations of the public and private sector are seeking revealing findings that shall assist them in calibrating their operations and improving their overall existence in the business world via cybersecurity surveys. Healthcare Information and Management Systems Society (HIMSS) focusing on the health sector [13]; ARC Advisory Group targeting Industrial Control Systems (ICS) in critical infrastructures such as energy and water supply, as well as in process industries, including oil, gas and chemicals [14]; SANS exploring the challenges involved with design, operation and risk management of ICS, its cyber assets and communication protocols, and supporting operations [15]; Deloitte in conjunction with Wakefield Research interviewing C-level executives who oversee cybersecurity at companies [16]; these being only some of the countless examples available nowadays.

Current trend in the cybersecurity surveys appears to be broadening their horizon by becoming available and approachable to individuals over the internet [17, 18]. Since their goal is to reach out and attract more participants, thus achieving a greater data collection and, consequently, enforcing their results, tend to be shorter, more comprehensive to the majority of common people and apparently web-based.

Recognizing the unique value of this undisputable fruitful security evaluation methodology and rushing from the special working and living circumstances due to the COVID-19 pandemic, we identified the research opportuning to evaluate how this crisis has affected the cybersecurity culture of both individuals and organizations across the suffering globe. Security threats, frauds, breaches & perils have been brought to the light, recommendations have been given and precautions have been made [19, 20, 21]. What about the cybersecurity culture and its potential scars from this virus? Addressing this concern was our aim when designing, conducting and analysing the survey presented in this paper.

3. SECURITY CULTURE FRAMEWORK

During the last months, we have been conducting a thorough scientific research related to cybersecurity tools, solutions and frameworks with a clear focus on the human factor. We have benchmarked the dominant reveals on the field, classified their possibilities and analysed their core security factors. Having identified their gaps and overlaps, common grounds and differentiations and thoroughly studied several academic principles regarding information security, including technical analyses, algorithmic frameworks, mathematical models, statistical computations, behavioural, organizational and criminological theories, we have created a foundation combining the elements that constitute the critical cyber-security culture elements [22]. The suggested cybersecurity culture framework is based on a domain agnostic security model combining the key factors affecting and formulating the cybersecurity culture of an organization. It consists of 10 different security dimensions analysed into 52 domains assessed by more than 500 controls examined under two different pillars: the organizational and the individual level. This hierarchical approach is being presented in Figure 1 while Figure 2 lists the model dimensions per level.

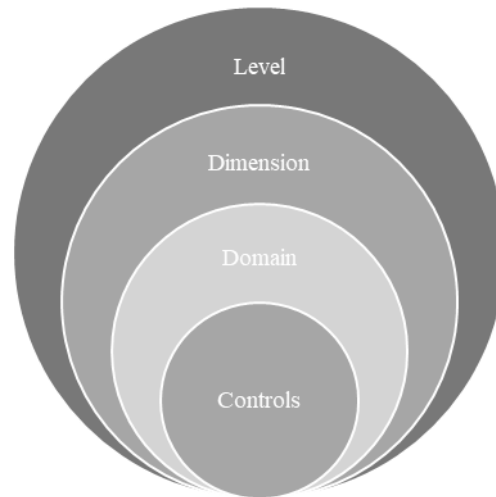


Figure 1. Cyber-Security Culture Model: Main Concepts



Figure 2. Cyber-Security Culture Model: Levels & Dimensions

An assessment methodology along with a weighting algorithm have been developed based on the previously mentioned model in order to offer a clockful cyber-security culture evaluation framework which shall be used as the basis of survey under design.

4. DESIGNING THE SURVEY

Our aim is to design a survey which shall be short and targeted getting the security pulse of current business reality in the critical infrastructure domain. One of our major goals is to keep the questionnaire small and easily addressed in a timely manner by a common employee with no special security expertise or knowledge. This way, we can facilitate participation of a broader workforce group minimizing effort and prerequisites while maximizing result variation and credibility. Towards that goal, we need to formulate questions to target specific security factors bridging various security domains while smartly extracting information depicting the existing working security routine and culture, their disruption by the COVID-19 crisis and their reaction to these special and rather demanding circumstances.

On the other hand, taking into consideration the reported cyber-crime incidents along with the fraud and attack techniques used by the criminals of the dark web during this period, we focused our evaluation on specific dimensions related to network infrastructure, asset management, business continuity, employee awareness and attitude.

In the paragraphs to follow, we outline how starting from a detailed cyber-security culture framework with more than 500 controls, we have narrowed down our objectives to a questionnaire containing no more than 23 questions, depending on the provided answers. Table 1 indexes the questions constituting the final version of our questionnaire including secondary clarification questions presented based on provided participant input whereas Table 2 correlates each of the questions to specific cyber-security levels, dimensions and domains of our model.

Table 1. Question indexing, including secondary clarification questions presented based on provided input (asterisk annotated).

Q1	Prior to the COVID-19 crisis, were you able to work from home?	Q9.2	How were you informed how to use them?	Q12.6	I am proud to work for my organization.
Q2.1	Did you receive any security guidelines from your employer regarding working from home?	Q10.1	Has your company adopted a specific collaboration solution?	Q12.7	I have access to the things I need to do my job well.
Q2.2*	Please describe the main (2-3) security guidelines provided.	Q10.2*	What abilities does it offer?	Q13	What is your age?
Q3	What kind of devices are you using to connect to your corporate working environment?	Q11.1	Did you face any of the below cyber-security related threats during the COVID-19 crisis?	Q14	What is the highest degree or level of school you have completed?
Q4	Are these devices accessed by users other than yourself?	Q11.2*	Please name any other cyber-security threats you encountered during this period, not listed above.	Q15	Please select the business domain of the organization you work for.
Q5	These devices are personal or corporate assets?	Q12.1	To what extent do you agree with the following statements: (during this specific period of the COVID-19 crisis) I prefer working from home than going to the office.	Q16	Which of the following best describes your work position?
Q6	Are these devices managed by your organization?	Q12.2	I work more productively from home.	Q17	Comments
Q7	Which of the following apply for the devices you currently use for your working from home employment? (providing security	Q12.3	I collaborate with my colleagues as effectively as when we are in office.		

	measures alternatives, e.g. antivirus, password protections)				
Q8	How do you obtain access to your corporate working environment?	Q12.4	I am satisfied by my employer's approach to the crisis.		
Q9.1	Were you asked to use applications or services that you were unfamiliar with, because of the need for remote working?	Q12.5	I have all the support I need to face any technical problems I have (e.g. corporate access issues, infrastructure failures, etc.).		

4.1. Organizational Level

Culture is defined as a set of shared attitudes, values, goals and practices that define an institution or organization. Consequently, cyber-security culture refers to the set of values, conventions, practices, knowledge, beliefs and behaviours associated with information security. Therefore, its skeleton is being outlined by the working environment along with the technological infrastructure and security countermeasures that define it.

To understand, evaluate and analyse the security cultural status of the critical infrastructure organizations participating to our survey, we have included questions Q1-Q10 that heartbeat the overall technological and security readiness and adaptability. Under the coronavirus prism, we intend to understand if teleworking was possible prior to the crisis or not and under which security policies. Thus, we have included queries polling the remote access procedures and their meeting standards as well as the types, configuration and management of the devices used to gain access to the corporate environments. In other words, we attempt to assess the way and the means of the working from home experience with a clear focus on cyber-security.

Additionally, we intend to assess the security maturity of the management, the security response team and awareness training program by introducing a number of questions clearly related to cyber-security familiarity and readiness. The most critical question of these category is the one referring to security guidelines provided during the COVID-19 crisis seeking to match their responsiveness and contemporality. Informing your workforce by issuing supportive guiding principles following the example of leading cyber-security entities and experts during rather confusing and challenging periods is a core security indicator.

Another business facet which is examined, although not directly related to information security, is the collaboration possibilities offered to employees. Communication and teamwork need to be facilitated and promoted, especially during this time period when general isolation is mandated as the only defence against the virus spread. Companies are expected to provide all means necessary to assist their employees in being productive, effective and cooperative. This notion and quality are being tested via two simplified questions included into our survey.

4.2. Individual Level

Moving down to an individual level, evaluation becomes more demanding since virus fear and emotional stress dominate every aspect of daily life directly or indirectly affecting the human related security factors. Questions Q11-Q12 attempt to probe the security behaviour, attitude and competency of the remote workers by examining their emotions, thoughts and beliefs and by asking them to report any security incidents they came up against.

Q9.2			Communication	Organizational Culture and Top Management Support					
Q10.1				Operating Procedures					
Q10.2*									
Q11.1								- Security Behaviour - Security Agent Persona	Security Skills Evaluation
Q11.2*									
Q12.1							Employee Climate		
Q12.2									
Q12.3									
Q12.4									
Q12.5							Employee Satisfaction		
Q12.6									
Q12.7									
Q13									
Q14							Employee Profiling		
Q15									
Q16									

The accruing questionnaire manages to effectively and efficiently combine the two security levels of our framework. Additionally, its contents have been tailored to rapidly yet effectually heartbeat the cyber-security reality during a disrupting chronological period, such as the COVID-19 pandemic. This agile instrument, although offering a quick and fruitful measurement method compared to similar concurrent surveys, it cannot be considered an in-depth information security assessment. Furthermore, it should not be used to label participating organisations but only to provide an overview of current status.

5. NEXT STEPS

Having developed a first questionnaire version addressing the security elements of interest based on our security culture framework, we need to carefully design the rest of the survey methodology including:

- **validity testing:** identify ambiguous questions or wording, unclear instructions, or other problems prior to widespread dissemination possibly conducted by a group of survey experts, experienced researchers and analysts, certified security and technology officers.
- **delivery method:** select the appropriate delivery method and possibly run an instrument validity testing to verify survey conduction methodology
- **sample selection:** carefully chose representatives from energy, transport, water, banking, financial market, healthcare and digital infra-structure from different European countries (e.g. Cyprus, France, Ger-many, Greece, Italy, Romania, Spain) affected by the COVID-19 crisis.
- **survey duration:** defining a specific start and end period communicated to all invited parties.

6. CONCLUSIONS AND FUTURE WORK

Our survey focuses on evaluating the security readiness and responsiveness of corporations during the Great Shutdown and more specifically it shall be addressing critical infrastructure domain representatives from different countries affected by the coronavirus.

Security cultural approach demands flexibility and concurrency. In a radically evolving and transforming environment, security and risk teams need to become part of the crisis management group, remote working employees need to remain vigilant to cyber-threats and operations life-

cycle needs to remain uninterrupted especially for the operators of essentials services. Our research aims to investigate if and in what extend is this approach embraced by the critical infrastructure organizations in different countries nowadays while revealing interesting findings related to cyber-security and inspiring further scientific researches on this field.

ACKNOWLEDGEMENT

This work has been co-funded from the European Union's Horizon 2020 research and innovation programme under the EnergyShield project "Integrated Cybersecurity Solution for the Vulnerability Assessment, Monitoring and Protection of Critical Energy Infrastructures", grant agreement No 832907 [23].

REFERENCES

- [1] T. P. Velavan and C. G. Meyer, "The COVID-19 epidemic," *Tropical Medicine and International Health*, vol. 25, no. 3, p. 278–280, 2020.
- [2] D. S. Hui, E. I. Azhar, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T. D. Mchugh, Z. A. Memish, C. Drosten, A. Zumla and E. Petersen, "The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China," *International Journal of Infectious Diseases*, vol. 91, pp. 264-266, 2020.
- [3] "WHO Director-General's opening remarks at the media briefing on COVID-19," World Health Organization (WHO), 2020.
- [4] United Nations Development Programme, "Socio-economic impact of COVID-19 | UNDP," [Online]. Available: <https://www.undp.org/content/undp/en/home/coronavirus/socio-economic-impact-of-covid-19.html>. [Accessed 07 07 2020].
- [5] M. Wolf, "The world economy is now collapsing," *Financial Times*, 14 04 2020. [Online]. Available: <https://www.ft.com/content/d5f05b5c-7db8-11ea-8fdb-7ec06edeef84>. [Accessed 01 07 2020].
- [6] E. Talamàs, "The Great Shutdown: Challenges And Opportunities," *Forbes*, 14 05 2020. [Online]. Available: <https://www.forbes.com/sites/iese/2020/05/14/the-great-shutdown-challenges-and-opportunities/#60eaf6e86f12>. [Accessed 07 07 2020].
- [7] G. Gopinath, "The Great Lockdown: Worst Economic Downturn Since the Great Depression," *IMFBlog*, 14 04 2020. [Online]. Available: <https://blogs.imf.org/2020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression/>. [Accessed 07 07 2020].
- [8] "ILO Monitor:COVID-19 and the world of work. Second edition," International Labour Organization (ILO), 2020.
- [9] A. G. Blanco, "The impact of COVID-19 on the spread of cybercrime," *BBVA*, 27 04 2020. [Online]. Available: <https://www.bbva.com/en/the-impact-of-covid-19-on-the-spread-of-cybercrime/>. [Accessed 07 07 2020].
- [10] "COVID-19 cyberthreats," INTERPOL, 2020. [Online]. Available: <https://www.interpol.int/en/Crimes/Cybercrime/COVID-19-cyberthreats>. [Accessed 07 07 2020].
- [11] "Cybercriminals targeting critical healthcare institutions with ransomware," INTERPOL, 04 04 2020. [Online]. Available: <https://www.interpol.int/en/News-and-Events/News/2020/Cybercriminals-targeting-critical-healthcare-institutions-with-ransomware>. [Accessed 07 07 2020].
- [12] C. D. Robert Richardson, "CSI Computer Crime and Security Survey 2010/2011," Computer Security Institute (CSI), 2011.
- [13] "2019 HIMSS Cybersecurity survey," Healthcare Information and Management Systems Society, 2019.
- [14] T. Menze, "The State Of Industrial Cybersecurity," ARC Advisory Group, 2019.
- [15] B. Filkins and D. Wylie, "SANS 2019 State of OT/ICS Cybersecurity Survey," SANS™ Institute, 2019.
- [16] "The future of cyber survey 2019," Deloitte, 2019.
- [17] Information Security Community, "Cybersecurity Trends Survey," [Online]. Available: <https://www.surveymonkey.com/r/Cybersecurity-DR>. [Accessed 17 04 2020].

- [18] “SMESEC - SMEs' Cybersecurity Watch,” SMESEC , [Online]. Available: https://docs.google.com/forms/d/e/1FAIpQLSdLcZ_6LKFANf6oX2XIOf_4Q55yDGR7ZZKCus7md4F1IsqvEQ/viewform. [Accessed 17 04 2020].
- [19] “COVID-19, Info Stealer & the Map of Threats – Threat Analysis Report,” Reason Labs, 09 03 2020. [Online]. Available: <https://blog.reasonsecurity.com/2020/03/09/covid-19-info-stealer-the-map-of-threats-threat-analysis-report/>. [Accessed 16 04 2020].
- [20] “Home working: preparing your organisation and staff,” National Cyber Security Centre, 17 03 2020. [Online]. Available: <https://www.ncsc.gov.uk/guidance/home-working>. [Accessed 17 04 2020].
- [21] “Working From Home - COVID19 - ENISA,” ENISA - European Union Agency for Cybersecurity, [Online]. Available: <https://www.enisa.europa.eu/topics/WFH-COVID19?tab=details>. [Accessed 17 04 2020].
- [22] K. Panetta, “7 Security Areas to Focus on During COVID-19,” Gartner, 24 04 2020. [Online]. Available: <https://www.gartner.com/smarterwithgartner/7-security-areas-to-focus-on-during-covid-19/>. [Accessed 07 07 2020].
- [23] “Energy Shield,” Energy Shield, 2019. [Online]. Available: <https://energy-shield.eu/>. [Accessed 25 03 2020].

AUTHORS

Mrs. Anna Georgiadou is a research associate in the Management and Decision Support Systems Laboratory in the School of Electrical and Computer Engineering at the National Technical University of Athens (NTUA). She has been working as a senior engineer on operation and system support services for major telecommunication providers, energy regulators, banking systems and other critical national infrastructures. She has recently become a PhD candidate on the cyber-security field inspired by her active participation in the HEDNO's (Hellenic Electricity Distribution Network Operator) information security management group. She is a certified database and network administrator, software developer and data-analyst.

Dr. Spiros Mouzakis is a senior research analyst for National Technical University of Athens (NTUA). He has 18 years of industry experience in conception, analysis and implementation of information technology systems. His research is focused on decision analysis in the field of decision support systems, risk management, Big Data Analytics, as well as optimization systems and algorithms, and enterprise interoperability. He has published in numerous journals including Computer Standards & Interfaces, International Journal of Electronic Commerce, Information Systems Management, and Lecture Notes in Computer Science, and has presented his research at international conferences.

Dr. Dimitris Askounis is a Professor in the School of Electrical and Computer Engineering at the National Technical University of Athens (NTUA). He has been involved in numerous IT research and innovation projects funded by the EU since 1988 in the thematic areas of eBusiness interoperability, eGovernance, data exploitation and management, decision support, knowledge and quality management, computer integrated manufacturing, enterprise resource planning, etc. He teaches digital innovation management, decision support and management systems, and he is a member of scientific committees on innovation and entrepreneurship competitions and incubators offered by International University networks, Financial Institutions, etc. Dr. Askounis has published over 80 papers in scientific journals and international conference proceedings.

THE STATISTICAL CORRELATION BETWEEN DISTORTION AND ADDING SIGNAL FOR PAPR REDUCTION IN OFDM BASED COMMUNICATION SYSTEMS

Désiré Guel¹, Boureima Zerbo², Jacques Palicot³ and Oumarou Sié¹

¹Department of Computer Engineering,
Joseph Ki-Zerbo University, Burkina Faso

²Thomas-Sankara University, Burkina Faso

³CentraleSupélec/IETR, Campus de Rennes, France

ABSTRACT

In recent past years, PAPR (Peak-to-Average Power Ratio) of OFDM (Orthogonal Frequency-Division Multiplexing) system has been intensively investigated. Published works mainly focus on how to reduce PAPR. Since high PAPR will lead to clipping of the signal when passed through a nonlinear amplifier. This paper proposes to extend the work related to "Gaussian Tone Reservation Clipping and Filtering for PAPR Mitigation" which has been previously published. So, in this paper, we deeply investigate the statistical correlation between PAPR reduction, and the distortion generated by three (3) adding signal techniques for PAPR reduction. Thereby, we first propose a generic function for PAPR reduction. Then, we analyse the PAPR reduction capabilities of each PAPR reduction technique versus the distortion generated. The signal-to-noise-and-distortion ratio (SNDR) metric is used to evaluate the distortion generated within each technique by assuming that OFDM baseband signals are modelled by complex Gaussian processes with Rayleigh envelope distribution for a large number of subcarriers. The results related to one of the techniques is proposed in the first time in this paper, unlike those related to the other two PAPR reduction techniques where the studies were already published. Comparisons of the proposed approximations of SNDR with those obtained by computer simulations show good agreement. An interesting result highlighted in this paper is the strong correlation existing between PAPR reduction performance and distortion signal power. Indeed, the results show that PAPR reduction gain increases as the distortion signal power increases. Through these 3 examples of PAPR reduction techniques; we could derive the following conclusion: in an adding signal context, the adding signal for PAPR reduction is closely linked to the distortion generated, and a trade-off between PAPR-reduction and distortion must be definitely found.

KEYWORDS

Orthogonal Frequency Division Multiplexing (OFDM), Peak-to-Average Power Ratio (PAPR), signal-to-noise-and-distortion ratio (SNDR), Adding Signal Techniques.

1. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM), a popular type of multicarrier transmission [5, 6, 7, 8], is an effective modulation technique for high-data-rate wireless and wireline applications, including Digital Subscriber Line (DSL) [9], Digital Audio Broadcasting (DAB) [10], Digital Video Broadcasting (DVB) [11], and Wireless Local Area Network

David C. Wyld et al. (Eds): CSEA, DMDBS, NSEC, NETWORKS, Fuzzy, NATL, SIGEM - 2020

pp. 81-96, 2020. CS & IT - CSCP 2020

DOI: 10.5121/csit.2020.101807

(WLAN) [12]. It has been adopted by the third-generation partnership program long-term evolution (3GPP-LTE) [13] and LTE-Advance. OFDM has been recently adopted for use in future 5G air interface [14]. One of the advantages of OFDM modulations is its ability to minimize the multi-path propagation effects and the impulse noise by transforming effectively the frequency selective fading channel into a flat fading channel. In addition, OFDM eliminates the need for equalizers and uses modern digital signals processing techniques, such as the fast Fourier transform (FFT) techniques [15].

Due to the large number of subcarriers, the OFDM signal has a large Peak-to-Average Power Ratio (PAPR) which makes system performance very sensitive to nonlinear distortions [16, 18]. This remains one of the major drawbacks associated to OFDM transmission technique. In practice, transmission devices such as the High-Power Amplifier (HPA) have a limited dynamic range [19]. Therefore, to ensure a distortion-less transmission, hardware with a high-power back-off is required. But it restricts the OFDM system from utilizing the transmitted power effectively. To reduce the PAPR of OFDM signals, various techniques have been proposed [20]. The most widely known techniques are based on selective mapping [21], phase shifting [22] or some form of coding. Selective mapping and phase shifting offer a considerable reduction of the PAPR [21] but, at the price of a significant increase of the overall system complexity. Coding techniques with the capabilities of both PAPR reduction, as well as error correction (Reed-Muller codes) [23], are attractive. However, these codes significantly reduce the overall throughput of the system, especially if there is a relatively large number of subcarriers [23].

The novelty of this paper, on one hand, involves proposing a generic function for PAPR reduction which embeds three methods for PAPR reduction : SC [2], HC [3] and GM [4] that have been widely studied in the past; on the other hand, it involves investigating the relation between PAPR reduction and the distortion signal power by theoretical analysis and computer simulations which follows up the work done in [1]. So, the metrics such as complementary cumulative distribution function (CCDF) is used PAPR reduction performance evaluation. Signal-to-noise-and-distortion ratio (SNDR) metric is used to evaluate the distortion generated by the different schemes for PAPR reduction. What is also new is the proposed approximation of SNDR for GM [4] which is proposed the first time in this article, while those related to SC [2], HC [3] have been the studied and been published in [1]. Through these three (3) examples of PAPR reduction techniques; it has been demonstrated that adding signal for PAPR reduction is closely linked to the distortion generated, and a trade-off between PAPR-reduction and distortion must be found. Indeed, greater will be the PAPR reduction, greater will be the distortions generated that could affect the over-all system performance in term of BLER/BER degradation and OOB (Out-Off-Band) radiation/interference.

The paper is organized as follows. Section 2 introduces the related work. Section 3 characterizes the OFDM system and defines the PAPR in OFDM signals context. The three (3) adding signal techniques for PAPR reduction and distortion characterization are described in section 4. In section 5, simulation results are shown, and conclusions are drawn in section 6. Some future directions and open problems are discussed in section 7.

2. RELATED WORK

In this section, we highlight works done in the past mainly related to "Geometric Method"(GM) which is strongly related to this paper; also, the Busgang's principle are highlighted and will undoubtedly allow to understand the conclusions of our study. Due to the simplest way for PAPR reduction, the techniques in [24, 26, 27, 28] have been proposed and seem to be promising for use in commercial systems. In [28], it is an adding signal technique whose its adding signal for PAPR reduction is directly expressed; [24, 26, 27] are clipping techniques and considered, in this paper,

as adding signal techniques whose the adding signal for PAPR reduction are the difference between original signal and its clipped version. All these techniques [24, 26, 27, 28] are nonlinear processes and may lead to significant distortion (in-band and out of band distortion) and performance loss. In [25], filtering is used to remove the out-of-band distortion whereas in [28], filtering is used to remove the part of adding signal generated over the OFDM signal bandwidth. This results in, no spectral regrowth in [24, 26] and no Bit Error Rate (BER) performance degradation in [28]. Recently, [29] presents a comparative analysis for various companding based PAPR reduction techniques and provides useful guidelines for the design and development of power and bandwidth efficient OFDM systems.

2.1. The “Geometric Method”

The “Geometric Method”(GM) has been proposed in [4] where the PAPR reduction scheme has been presented; the performance (PAPR reduction, transmission quality and execution complexity) has been presented as well in the context of a WLAN system based on the IEEE 802.11 a/g standard. The simulation results showed a loss of PAPR reduction performance when the PAPR reduction signal (“adding signal for PAPR”) was set far from the OFDM signal in frequency domain. It has been demonstrated that the loss of PAPR reduction gain was due the distortion generated within the technique; so setting the PAPR reduction signal far from the OFDM signal in frequency domain in order to mitigate the generated distortion, at the same reduces PAPR reduction gain. However [4] did not highlight the relationship which existed between the PAPR reduction signal and the transmission quality degradation in term of BLER/BER.

In order to overcome the limitations of GM [4], an Enhanced GM (EGM) which is a downward-compatible PAPR reduction technique has been proposed in [25]; the performance (reduction of PAPR, quality of transmission and complexity of execution) of the technique has been evaluated and compared to GM [4]. The results showed that, under the same simulation conditions, GM [4] is more efficient than EGM [25] in terms of PAPR reduction. However, EGM does not degrade BLER/BER unlike GM. The main reason of this outcome is that EGM filters the signal for PAPR reduction in order to preserve the BLER/BER; this results to in loss of PAPR reduction performance. EGM [25] remains a downward-compatible “adding signal” for PAPR reduction technique with none BLER/BER degradation. However, EGM does not control Out-Of-Band (OOB) spectral regrowth; this could lead to OOB (Out-Off-Band) radiation/interference. Therefore, in [25], an Enhanced “Geometric Method” based on “Tone Reservation” has been proposed. It transforms the GM [4] technique into the TR technique by filtering the PAPR reduction signal with an FFT/IFFT pair based digital filtering [37].

The main idea in [37] was to transform the adding signal techniques (like clipping techniques, GM) into TR techniques thanks to a filtering based on FTT/IFFT pair; the reason is to take benefit of the TR advantages such as none BLER/BER degradation, none Side Information (SI) transmission. As the transformation is a low-complexity process (about the FFT/IFFT complexity), the obtained technique results in a low-complexity TR technique. However, the transformation into TR techniques generates a loss of performance in PAPR reduction; it has been proposed to increase the PAPR reduction gain by iterating the process of transformation. However, the system complexity grows linearly with the number of iterations and a trade-off shall be done between the PAPR reduction performance and the PAPR reduction scheme's complexity. All these observations and PAPR reduction signal filtering seemed to indicate the link between the PAPR reduction signal and the distortion generated which affects BLER/BER performance and could lead to OOB radiation/interference.

2.2. Bussgang's principle

In a report dating from March 1952 [31], J.J. Bussgang discusses the Gaussian signals (inter and auto) correlation function estimation, when one of the two inputs of the correlator undergo to a nonlinear transformation without memory. He succeeds in expressing the link between the intercorrelation function of these two signals and that of the correlation function before distortion. Bussgang's theorem has long been used and continues to be in the adaptive equalization area as evidenced by the works proposed in [32, 33]. In [32], Z. Ding and RA Kennedy deal with the role of the AM / AM transfer characteristic in non-linear systems with memory effect and propose an extension of the theorem of Bussgang in the case of complexes.

Now, assuming a nonlinear technique characterizes by its AM/AM nonlinear function $f(\cdot)$, where the output signal $y(t)$ is expressed as:

$$y(t) = f\left[r(t)e^{j\varphi(t)}\right], 0 \leq t \leq T_s \quad (1)$$

The Bussgang's theorem [31] applied to equation (1) leads to:

$$y(t) = \alpha x(t) + d(t) \quad \text{where,} \quad \alpha = \frac{R_{yx}(0)}{R_{xx}(0)} \quad (2)$$

where the distortion term $d(t)$ is uncorrelated with the input signal $x(t)$. Using (2), the output correlation function of the signal $y(t)$ is expressed as:

$$R_{yy}(\tau) = |\alpha|^2 R_{xx}(\tau) + R_{dd}(\tau) \quad (3)$$

The attenuation coefficient α is given by;

$$\alpha = \frac{R_{yx}(0)}{R_{xx}(0)} = \frac{E[x^*(t)y(t)]}{P_x} = \frac{E[rf(r)]}{P_x} \quad (4)$$

where $R_{yx}(\tau)$ denotes the input-output cross-correlation function, $f(r)$ the AM/AM non-linear function, P_x is the OFDM signal power, and $E[\cdot]$ is the statistical expectation operator.

Different approaches may be used to compute $R_{yy}(\tau)$ including the Direct Method, the Transform Method, or the Derivative Method. In [17, 18], the Derivative Method is used, and it is possible to prove for any nonlinear function $f(\cdot)$ that:

$$R_{yy}(\tau) = \sum_{n=0}^{\infty} c_n \left[\frac{R_{xx}(\tau)}{P_x} \right]^{2n+1} = \frac{c_0}{P_x} R_{xx}(\tau) + \sum_{n=1}^{\infty} c_n \left[\frac{R_{xx}(\tau)}{P_x} \right]^{2n+1} \quad (5)$$

with C_n expressed by:

$$c_n = \frac{1}{P_x} \frac{1}{n+1} \left\| \int_{D(r)} r f(r) p(r) \cdot L_n^{(1)} \left(\frac{r^2}{P_x} \right) dr \right\|^2 \quad (6)$$

where $D(r) = \{r : 0 \leq r \leq \infty\}$ is the domain of integration $p(r)$ is the probability density function of the OFDM envelope and $L_n^{(1)}(x)$ is the Laguerre function expressed by

$$L_n^{(1)}(x) = \frac{x^{-k} e^x}{n!} \left(\frac{d}{dx} \right)^n (x^{n+k} e^{-x}) \quad (7)$$

3. OFDM SYSTEM AND PAPR

In OFDM systems, the time-domain signal $x(t)$ can be written as

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{2j\pi \frac{k}{T_s} t}, 0 \leq t \leq T_s \quad (8)$$

where X_k is mapped data for $k = 0, 1, 2, \dots, N-1$, N is the number of subcarriers and T_s is the OFDM symbol period. The PAPR of $x(t)$ can be defined as

$$\text{PAPR}[x] = \frac{\max_{t \in [0, T_s]} |x(t)|^2}{P_{\text{av}}}, \quad (9)$$

where P_{av} is the average power defined as $P_{\text{av}} = E[|x(t)|^2]$. Note that, it is more useful to consider $\text{PAPR}[x]$ as a random variable and can be evaluated by using a statistical description given by the complementary cumulative density function (CCDF), defined as the probability that PAPR exceeds ψ_0 , i.e., $\text{CCDF} = \Pr\{\text{PAPR}[x] > \psi_0\}$. To approximate PAPR in (9) in discrete time-domain, an oversampled version of (8) can be used. In this case, the oversampled time-domain signal x_n can be written as

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{2j\pi k \frac{n}{NL}}, 0 \leq n \leq NL-1, \quad (10)$$

where L is the oversampling factor. To better approximate the PAPR of continuous-time OFDM signals, $L \geq 4$ is used to capture the peaks of the continuous time-domain signals. The time-

domain samples x_n are NL -point IFFT of the data block with $(L-1)N$ zero-padding. The PAPR computed from the L -times oversampled time domain OFDM signal samples can be defined as:

$$\text{PAPR}[x] = \frac{\max_{0 \leq n < NL} |x_n|^2}{\text{E}[|x_n|^2]} \quad (11)$$

4. HC, SC AND GM ALGORITHMS AND DISTORTION CHARACTERIZATION

Let us consider the discrete-time OFDM symbol x_n , the aim of adding signal techniques is to generate an adding signal c_n such as $\text{PAPR}[y_n] < \text{PAPR}[x_n]$, where $y_n = x_n + c_n$ is the PAPR reduced signal. The adding signal c_n for PAPR reduction is a function of the OFDM symbol x_n , i.e., $c_n = f(x_n)$ where $f(\cdot)$ is the adding signal generator function.

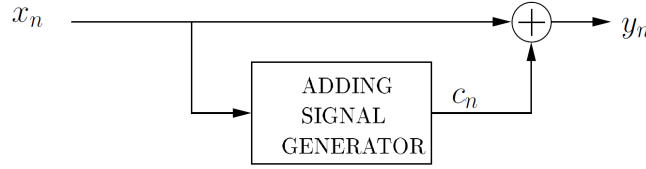


Figure 1: Adding signal scheme for PAPR reduction.

Fig.1 shows, the scheme of adding signal techniques for PAPR reduction.

4.1. HC, SC and GM Characterization and SNDR Definition

Now, let us consider the adding signal techniques, SC [2], HC [3] and GM [4] for PAPR reduction. The adding signal for PAPR reduction c_n can be formulated as

$$c_n = f_{\eta, \theta_n}(x_n) = \begin{cases} (\eta - 1)x_n e^{j\theta_n}, & |x_n| < \frac{1}{\eta} A \\ \left(\frac{A}{|x_n|} - 1\right)x_n e^{j\theta_n}, & |x_n| \geq \frac{1}{\eta} A \end{cases}, \quad (12)$$

where A is the magnitude threshold, θ_n is defined as $\theta_n = 2\pi\delta f \frac{n}{NL} T_s$, $0 \leq n \leq NL-1$, and T_s is the symbol period of the OFDM signal. The GM corresponds to (12) with $\eta = 1$ and the parameter δf is defined as $\delta f = f_r - f_c$, where f_r and f_c are the carrier frequencies of the OFDM signal and adding signal respectively; the SC is a special case of (12) with $\eta = 1$ and $\delta f = 0$; while the HC is a special case of (12) with $\eta = \infty$ and $\delta f = 0$.

These three (3) adding signal techniques for PAPR reduction, SC [2], HC [3] and GM [4] are all nonlinear processes and can be modelled as nonlinear AM/AM distorting functions $g_{\eta,\theta_n}(\cdot)$ defined below.

$$g_{\eta,\theta_n}(x) = \left[1 + (\eta - 1) e^{j\theta_n} \right] x \cdot 1_{x < \frac{1}{\eta} A} + \left[x + (1 - x) e^{j\theta_n} \right] \cdot 1_{x \geq \frac{1}{\eta} A} \quad (13)$$

It should be noted that similar threshold functions $g_{\eta,\theta_n}(x)$ have been also considered in other complex nonlinear dynamics [30]. Using (13), the PAPR reduced signal y_n can be written as

$$y_n = g_{\eta,\theta_n}(|x_n|) e^{j\theta_n} = x_n + f_{\eta,\theta_n}(x_n) \quad (14)$$

where θ_n is the x_n phase. According to Busgang's theorem [31], the nonlinear functions $g_{\eta,\theta_n}(\cdot)$ can be decomposed as

$$g_{\eta,\theta_n}(x) = \alpha_{\eta,\theta_n} x + d, \quad (15)$$

where d is the distortion created by $g_{\eta,\theta_n}(\cdot)$ and α_{η,θ_n} is chosen such that d is uncorrelated with x , i.e., $E[xd^*] = 0$. From (15), it is shown that

$$\alpha_{\eta,\theta_n} = \frac{E[x^* g_{\eta,\theta_n}(x)]}{E[|x|^2]} = \frac{E[x^* g_{\eta,\theta_n}(x)]}{P_{av}}, \quad (16)$$

The distortion power is given by

$$\varepsilon_d = E[|d|^2] = E[|g_{\eta,\theta_n}(x)|^2] - |\alpha_{\eta,\theta_n}|^2 P_{av}. \quad (17)$$

The nonlinear functions $g_{\eta,\theta_n}(\cdot)$ depend on the samples θ_n , thus we define $\text{SNDR}_{[\theta_n]}$ as the signal-to-noise-and-distortion ratio for θ_n , given.

$$\text{SNDR}_{[\theta_n]} = \frac{|\alpha_{\eta,\theta_n}|^2 P_{av}}{N_0 + \varepsilon_d} = \text{SNR} \frac{|\alpha_{\eta,\theta_n}|^2}{1 + \text{SNR} \frac{\varepsilon_d}{P_{av}}}, \quad (18)$$

where $\text{SNR} = \frac{P_{av}}{N_0}$ is the signal-to-noise ratio (SNR). The global SNDR is thus defined as

$$\text{SNDR} = \frac{1}{N} \sum_{\theta_n} \text{SNDR}_{[\theta_n]} \quad (19)$$

4.2. SNDR Close Form of the Three Adding Signal Techniques

Assuming that the baseband OFDM signal converges to a complex Gaussian process for large N , it has been shown in [34] that $|x_n|$ is an i.i.d Rayleigh random variable of which the probability density function (PDF) is :

$$p_X(x) = \frac{2x}{P_{\text{av}}} e^{-\frac{x^2}{P_{\text{av}}}}, \quad x \geq 0. \quad (20)$$

To derive a SNDR close form of the three adding signal techniques for PAPR reduction, we consider the equations (15) to (19) where the variable x is an i.i.d Rayleigh random variable whose PDF has been given in (20).

Let M_η and N_η the expressions defined as

$$M_\eta = \int_0^{\frac{1}{\eta}} x^2 p_X(x) dx \quad \text{and} \quad N_\eta = \frac{A}{\eta} \int_{\frac{1}{\eta}}^{\infty} x p_X(x) dx \quad (21)$$

We show by developing (21) and using (20) that

$$M_\eta = \left[1 - \left(1 + \frac{A^2}{\eta^2 P_{\text{av}}} \right) e^{-\frac{A^2}{\eta^2 P_{\text{av}}}} \right] P_{\text{av}}, \quad \text{and} \quad N_\eta = \left[\frac{A^2}{\eta^2 P_{\text{av}}} e^{-\frac{A^2}{\eta^2 P_{\text{av}}}} + \frac{A\sqrt{\pi}}{\eta\sqrt{P_{\text{av}}}} Q\left(\frac{A\sqrt{2}}{\eta\sqrt{P_{\text{av}}}}\right) \right] P_{\text{av}}, \quad (22)$$

where $Q(\cdot)$ is the Q-function defined as $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{\tau^2}{2}} d\tau$

Now, using the definition of M_η and N_η given in (21), we can very easily show that

$$E[x^* g_{\eta, \theta_n}(x)] = \int_0^{\infty} x^* g_{\eta, \theta_n}(x) p_X(x) dx = (1 - e^{j\theta_n}) P_{\text{av}} + \eta (M_\eta + N_\eta) e^{j\theta_n}, \quad (23)$$

and

$$\begin{aligned}
\mathbb{E}\left[|g_{\eta,\theta_n}(x)|^2\right] &= \int_x^\infty |g_{\eta,\theta_n}(x)|^2 p_X(x) dx \\
&= |1 - (\eta - 1)e^{j\theta_n}|^2 M_\eta + \frac{A^2}{P_{av}} e^{-\frac{A^2}{\eta^2 P_{av}}} P_{av} + 2(1 - \cos\theta_n) [P_{av} - (M_\eta + \eta N_\eta)]
\end{aligned} \quad (24)$$

Now, substituting (22) into (23) and (24), we show that

$$|\alpha_{\eta,\theta_n}|^2 = 1 + \left[1 - \eta \left(1 - \Gamma\left(\frac{A}{\eta\sqrt{P_{av}}}\right)\right)\right]^2 - 2 \left[1 - \eta \left(1 - \Gamma\left(\frac{A}{\eta\sqrt{P_{av}}}\right)\right)\right] \cos\theta_n, \quad (25)$$

And

$$\frac{\mathcal{E}_d}{P_{av}} = \eta^2 \left(1 - e^{-\frac{A^2}{\eta^2 P_{av}}}\right) - \eta^2 \left(1 - \Gamma\left(\frac{A}{\eta\sqrt{P_{av}}}\right)\right)^2, \quad (26)$$

Where $\Gamma(x)$ is the function defined as $\Gamma(x) = e^{-x^2} - x\sqrt{\pi}Q(x\sqrt{2})$, $x \geq 0$.

4.2.1. Case of Soft-Clipping (SC) [2]

Setting $\eta = 1$ and $\delta f = 0$ (i.e., $\theta_n = 0$) in (25) and (26); and substituting (25) and (26) into (18), we obtain the SNDR for SC expressed by (27).

$$\text{SNDR} = \text{SNR} \frac{\left(1 - \Gamma\left(\frac{A}{\sqrt{P_{av}}}\right)\right)^2}{1 + \text{SNR} \left(1 - e^{-\frac{A^2}{P_{av}}} - \left(1 - \Gamma\left(\frac{A}{\sqrt{P_{av}}}\right)\right)^2\right)}, \quad (27)$$

4.2.2. Case of Heavyside-Clipping (HC) [3]

When $\eta = \infty$ and $\delta f = 0$, we have the HC case. Let us start by setting $\delta f = 0$ (i.e., $\theta_n = 0$) in (25); we obtain

$$|\alpha_{\eta,\theta_n}|^2 = \eta^2 \left(1 - \Gamma\left(\frac{A}{\eta\sqrt{P_{av}}}\right)\right)^2$$

Due to $\eta \left(1 - \Gamma\left(\frac{A}{\eta\sqrt{P_{av}}}\right)\right) \square \frac{A}{2} \sqrt{\frac{\pi}{P_{av}}}$ and $\eta \sqrt{1 - e^{-\frac{A^2}{\eta^2 P_{av}}}} \square \frac{A}{\sqrt{P_{av}}}$ for $\eta \square \infty$, the SNDR for HC is given by

$$\text{SNDR} = \text{SNR} \frac{\frac{\pi A^2}{4 P_{\text{av}}}}{1 + \text{SNR} \left(\left(1 - \frac{\pi}{4} \right) \frac{A^2}{P_{\text{av}}} \right)}, \quad (28)$$

4.2.3. Case of Geometric Method (GM) [4]

The GM corresponds to (12) with $\eta = 1$ and the parameter δf is defined as $\delta f = f_r - f_c$. Setting only $\eta = 1$ in (25) and (26); we obtain

$$\text{SNDR}_{[\theta_n]} = \text{SNR} \frac{1 - 2\Gamma \left(\frac{A}{\sqrt{P_{\text{av}}}} \right) \cos \theta_n - \Gamma^2 \left(\frac{A}{\sqrt{P_{\text{av}}}} \right)}{1 + \text{SNR} \left(1 - e^{-\frac{A^2}{P_{\text{av}}}} - \left(1 - \Gamma \left(\frac{A}{\sqrt{P_{\text{av}}}} \right) \right)^2 \right)}, \quad (29)$$

According to the sum of Riemann theorem, for large N

$$\frac{1}{N} \sum_{n=0}^{N-1} \cos \theta_n \approx \frac{1}{2\pi\delta f T_s} \int_0^{2\pi\delta f T_s} \cos \theta d\theta = \frac{\sin(2\pi\delta f T_s)}{2\pi\delta f T_s}, \quad (30)$$

Using (30) in (19), the SNDR for GM is derived by

$$\text{SNDR} \approx \text{SNR} \frac{1 - 2\Gamma \left(\frac{A}{\sqrt{P_{\text{av}}}} \right) \frac{\sin(2\pi\delta f T_s)}{2\pi\delta f T_s} - \Gamma^2 \left(\frac{A}{\sqrt{P_{\text{av}}}} \right)}{1 + \text{SNR} \left(1 - e^{-\frac{A^2}{P_{\text{av}}}} - \left(1 - \Gamma \left(\frac{A}{\sqrt{P_{\text{av}}}} \right) \right)^2 \right)}, \quad (31)$$

5. SIMULATION RESULTS

Computer simulations were performed on an OFDM system with 64-subcarriers and 16-QAM symbol mapping; 10^4 OFDM symbols were randomly generated. GM is applied to the OFDM system with $\delta f T_s = \frac{1}{8}$. Indeed, in [4], it is shown that, when $\delta f T_s$ approaches 1/8, GM realizes a significant PAPR reduction gain.

Fig.2 compares the theoretical SNDR as a function of SNR with the simulation results for the three adding signal techniques. The proposed approximations diverge from the simulation results for a low SNR. However, for an SNR ≥ 5 dB, they show good agreement with the simulation results. This is because as for a low SNR, the approximation error in (27), (28) and (31) becomes dominant. Let us note that, the accuracy of approximation can be improved by increasing the sampling rate.

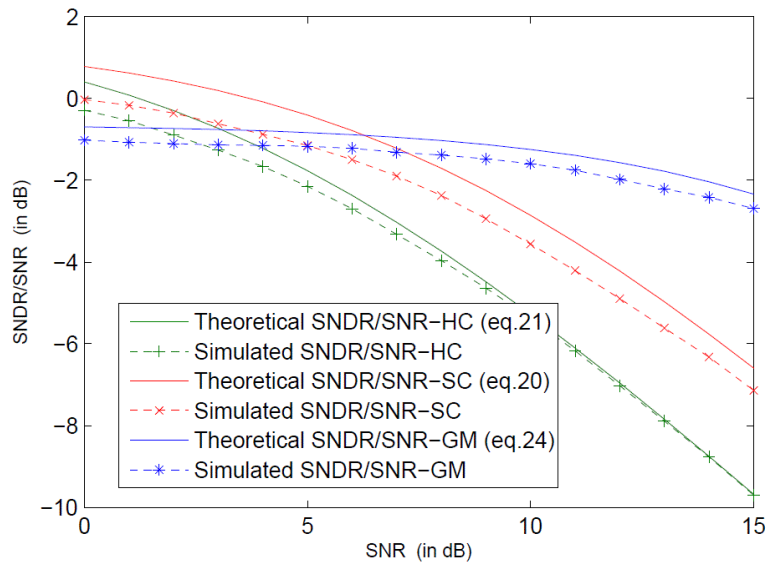


Figure 2: Comparison of the proposed SNDR as a function of SNR with the simulation results for

$$\frac{A}{\sqrt{P_{av}}} = 3\text{dB}$$

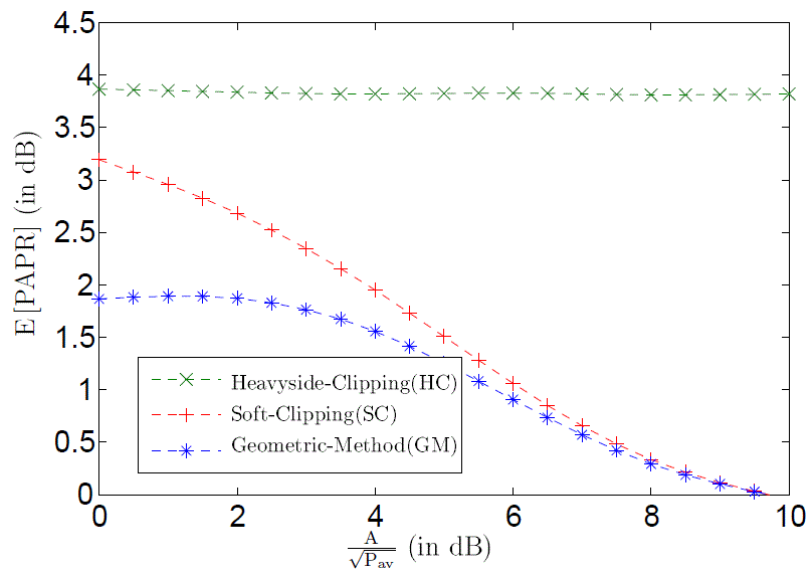


Figure 3: PAPR reduction performance as a function of $\frac{A}{\sqrt{P_{av}}}$ for SC, HC and GM

Fig.3 plots the PAPR reduction performance of the three adding signal methods for PAPR reduction. In order to evaluate the PAPR reduction performance, $E[\Delta\text{PAPR}]$ defined as the difference between the initial PAPR and the reduced PAPR is used. For SC and GM, $E[\Delta\text{PAPR}]$

decreases as $\frac{A}{\sqrt{P_{av}}}$ increases, whereas for HC, $E[\Delta\text{PAPR}]$ is almost constant.

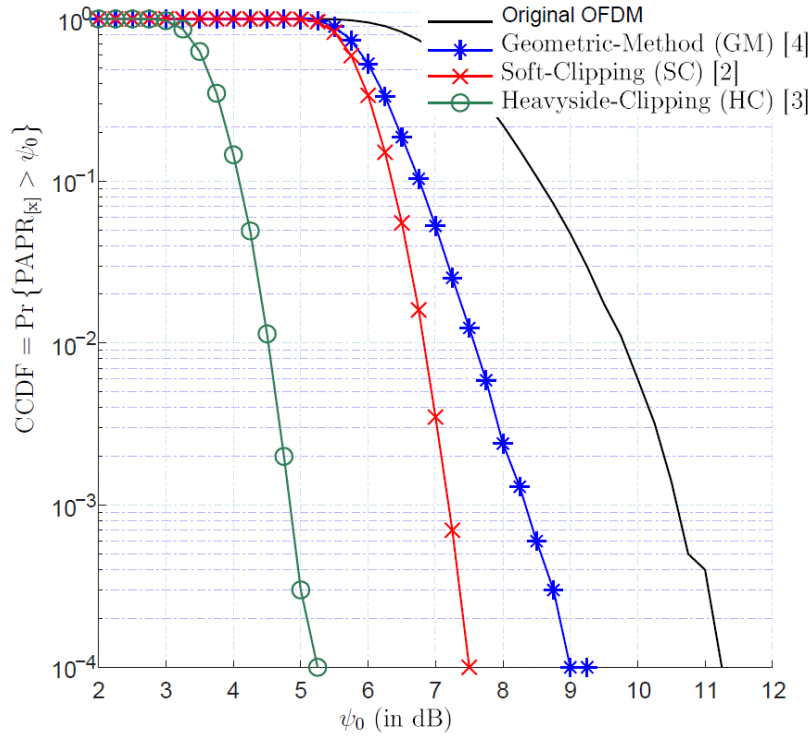


Figure 4: CCDF of SC, HC and GM for $\frac{A}{\sqrt{P_{av}}} = 5$ dB.

Note that, for the same value of $\frac{A}{\sqrt{P_{av}}}$, HC is more efficient than SC which is more efficient than GM in terms of PAPR reduction as shown in Fig.4. For example in Fig.4, at 10^{-2} of the CCDF, HC reduces about 5.25 dB the PAPR, SC reduces about 2.75 dB, whereas in GM case, the reduction is about 2.25 dB.

Now, let us look at closely the simulation results about the distortion analysis. Fig.5 plots the simulated SNDR to SNR as a function of $\frac{A}{\sqrt{P_{av}}}$ for the three different adding signal methods

with SNR = 10 dB. It shows that, for the same value of $\frac{A}{\sqrt{P_{av}}}$, HC introduces more distortion than SC which introduces more distortion than GM. For example, for $\frac{A}{\sqrt{P_{av}}} = 3$ dB, SNDR/SNR is equal to -5.5 dB, -2.75 dB and -2.25 dB for HC, SC and GM respectively.

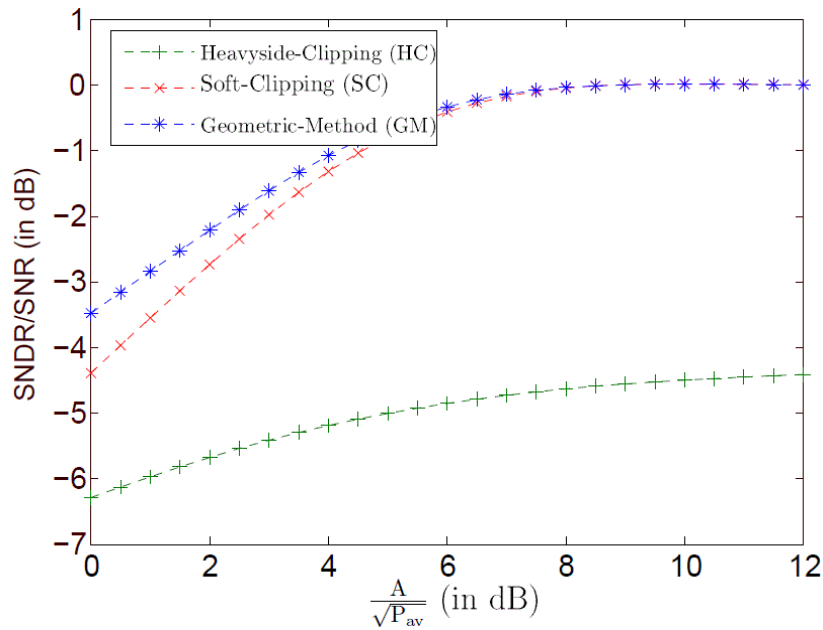


Figure 5: SNDR to SNR of SC, HC and GM for SNR = 10 dB.

According to the results of Fig.3 and Fig.5, we remark that

- for SC and GM, larger is $\frac{A}{\sqrt{P_{av}}}$, more significant will be the PAPR reduction and more significant will be the distortion generated,
- HC reduces greatly the PAPR but leads to a significant distortion generation.

With above remarks, we can conclude that, a significant distortion lead to a considerable reduction of the PAPR.

6. CONCLUSION

In this paper, we have proposed a generic function for PAPR reduction which embeds three (3) methods for PAPR reduction widely studied in the past : these three (3) methods are known also as adding signal techniques. The PAPR reduction performance of these three (3) adding signal techniques and the distortion generated by these techniques were studied. We analysed the distortion generated by deriving a close form expression of the SNDR while assuming that OFDM baseband signals are modelled by complex Gaussian processes with Rayleigh envelope distribution for a large number of subcarriers. The proposed approximation of SNDR for these adding signal techniques show good agreement with the simulation results. The results related to “Geometric Method” (GM) are proposed in the first time in this paper, unlike those related to Soft-Clipping (SC), Heavyside-Clipping (HC) where the studies were already published.

Comparisons between these techniques are made in terms of PAPR reduction performance and distortion generated. Results show that, the PAPR reduction increases as the distortion generated increases (in terms of power).

From this study, we have drawn the following conclusion namely, in adding signal context, the adding signal for PAPR reduction is closely linked to the distortion generated. So, a trade-off between PAPR-reduction and distortion must be done.

7. FUTURE WORK

As it has been shown in this article through these 3 techniques, SC [2], HC [3] and GM [4], OFDM PAPR can be reduced at the price of more or less significant distortion generated which could affect BLER/BER performance and lead to OOB (Out-Of-Band) radiation/interference.

To follow up this work, the following directions have been identified and deserve to be developed:

- An optimal solution should be investigated in such way the PAPR will be significantly reduced at the price of “weak distortion” or “none distortion” generated. A single iteration should be considered in order to keep the PAPR reduction scheme low power consumption. As the PAPR reduction techniques investigated here are characterized by a nonlinear function $f_{\eta, \theta_n}(\cdot)$ which acts on the amplitude of the signals for PAPR reduction. It would be interesting to find the optimal PAPR reduction function $f_{\eta, \theta_n}^{opt}(\cdot)$ which will lead to significant PAPR reduction with low computational complexity.
- A further direction for the work improvement is to combine the PAPR reduction techniques investigated in this paper with High Power Amplifier (HPA) linearization techniques especially with an adaptive predistortion method taking into account the memory effect of the amplifier. The goal should be to provide a big picture in terms of the overall system performance enhancement and not focus only on PAPR reduction techniques.

ACKNOWLEDGMENT

The authors would like to thank the authors and contributors whose works were included in this study.

REFERENCES

- [1] Y. Louet, J. Palicot and D. Guel, “Gaussian Tone Reservation Clipping and Filtering for PAPR Mitigation”, International Journal on Advances in Telecommunications, vol 12 no 1 & 2, year 2019.
- [2] X. Li and L.J. Cimini, “Effects of Clipping and Filtering on the Performance of OFDM,” IEEE Communication Letters, vol. 2, no. 5, pp. 131-133, May 1998.
- [3] Qian Hua, R. Raich, G.T. Zhou, “On the benefits of deliberately introduced baseband nonlinearities in communication systems”, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04), vol. 2, pp. 905-908, May 2004.
- [4] D. Guel, J. Palicot and Y. Louet, “A Geometric Method for PAPR Reduction in a Signal Adding Context for OFDM Signals”, IEEE International Conference on Digital Signal Processing, pp. 347-350, 1-4 July 2007.
- [5] J. Bingham, “Multicarrier modulation for data transmission: an idea whose time has come,”IEEE Commun. Mag., vol. 28, pp. 5-14, May 1990.
- [6] L. Litwin, “An introduction to multicarrier modulation,”IEEE Potentials, vol. 19, pp. 36-38, Apr.-May 2000.
- [7] Z.Wang and G. B. Giannakis, “Wireless multicarrier communications,”IEEE Signal Processing Mag., pp. 29-48, May 2000.

- [8] R. van Nee and R. Prasad, "OFDM Wireless Multimedia Communications." London, UK: Artech House, 2000.
- [9] "Asymmetrical Digital Subscriber Line (ADSL) Transceivers," ITU-T Recommendation G.992.1, July 1999.
- [10] ETSI, "Radio broadcasting systems: digital audio broadcasting to mobile, portable and fixed receivers", European Telecommunication Standard, ETS 300-401, Feb. 1995.
- [11] ETSI, "Digital video broadcasting: framing structure, channel coding and modulation for digital terrestrial television", European Telecommunication Standard, ETS 300-744, Aug. 1997.
- [12] IEEE 802.11, "IEEE standards for wireless LAN medium access control (MAC) and physical layer (PHY) specifications", Nov. 1997
- [13] I. Pospishny, V. Vasyuk, S. Romanchyk, O. Dovzhenko, O. Dony and V. Shvaichenko, "3GPP long term evolution (LTE)," 2010 International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), Lviv- Slavske, 2010, pp. 192-192.
- [14] A. Ghosh, A. Maeder, M. Baker and D. Chandramouli, "5G Evolution: A View on 5G Cellular Technology Beyond 3GPP Release 15,"in IEEE Access, vol. 7, pp. 127639- 127651, 2019, doi: 10.1109/ACCESS.2019.2939938.
- [15] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform,"IEEE Trans. Commun. Technol., vol. 19, pp. 628-634, Oct. 1971.
- [16] S. Merchan, A. G. Armada, and J. L. Garcia, "OFDM performance in amplifier non-linearity,"IEEE Trans. Broadcast., vol. 44, pp. 106-114, Mar. 1998.
- [17] W.B. Davenport, W.L. Root, "An Introduction to the Theory of Random Signals and Noise", McGraw-Hill, 1958
- [18] P. Banelli, G. Baruffa, and S. Cacopardi, "Effects of HPA nonlinearity on frequency multiplexed OFDM signals,"IEEE Trans. Broadcast., vol. 47, pp. 123-136, June 2001.
- [19] A. Barbieri, G. Cariolaro, and L. Vangelista, "Nonlinear models of TWT revisited for OFDM systems," in Proc. of the 38 th Midwest Symposium on Circuits and Systems, vol. 1, Aug. 1996, pp. 522-525.
- [20] Yves Louet and Jacques Palicot, "A classification of methods for efficient power amplification of signals", Annals of Telecom, vol. 63, nb. 7-8, pp. 351-368, Aug. 2008
- [21] R. W. Bumbl, R. F. H. Fischer, and J. B. Huber, "Reducing the peak-to average power ratio of multicarrier modulation by selected mapping,"Electronics Letters, vol. 32, pp. 2056-2057, Oct. 1996.
- [22] L. J. Cimini, Jr. and N. R. Sollenberger, "Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences,"IEEE Commun. Lett., vol. 4, pp. 86-88, Mar. 2000.
- [23] J. Davis and J. Jedwab, "Peak-to-mean power control in OFDM, Golay complementary sequences, and Reed-Muller codes,"IEEE Trans. Inform. Theory, vol. 45, pp. 2397- 2417, Nov. 1999.
- [24] J. Armstrong, "New OFDM peak-to-average power reduction scheme", Vehicular Technology Conference, VTC 2001 Spring. IEEE VTS 53rd, vol.1 , 2001 pp. 756 - 760, 2001.
- [25] D. Guel, J. Palicot, and Y. Louet, "Tone reservation technique based on geometric method for orthogonal frequency division multiplexing peak-to-average power ratio reduction", IET Commun., November 2010 Volume 4, Issue 17, pp. 2065-2073.
- [26] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," Electronics Letters, vol. 38, pp. 246-247, Feb. 2002.
- [27] H. Ochiai and H. Imai, "On clipping for peak power reduction of OFDM signals, in IEEE Global Telecommun. Conf., 2000. (GLOBECOM'00)., vol. 2, Nov. 27-Dec. 1, 2000, pp. 731-735.
- [29] A. Mohammed, M. Shehata, A. Nassar and H. Mostafa, "Performance Comparison of Companding-Based PAPR Suppression Techniques in OFDM Systems," 8th International Conference on Modern Circuits and Systems Technologies (MOCAS) Year: 2019
- [30] Yilun Shang, "Deffuant model of opinion formation in one-dimensional multiplex networks,"Journal of Physics A: Mathematical and Theoretical, 48(39), 395101 - October 2015, <https://doi.org/10.1088/1751-8113/48/39/395101>
- [31] J. Bussgang, "Crosscorrelation function of amplitude-distorted Gaussian signals,"Research laboratory of electronics, Massachusetts Institute of Technology, Cambridge. Technical Report 216, (1952).
- [32] J. Minkoff, "The Role of AM-to-PM Conversion in Memoryless Nonlinear Systems," vol. 33, pp. 139-144, Feb 1985.
- [33] J. B. D. Filho, G. Favier, and J. M. T. Romano, "New Bussgang methods for blind equalization," in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-97, vol. 3, pp. 2269-2272, 21-24 April 1997.

- [34] S. Wei, D. L. Goeckel, and P. A. Kelly, "The complex envelope of a bandlimited OFDM signal converges weakly to a Gaussian random process: proof and application," to appear in IEEE Transactions on Information Theory, <http://www.ece.lsu.edu/swei>.
- [35] D. Guel and J. Palicot, "Clipping Formulated As an Adding Signal Technique for OFDM Peak Power Reduction," VTC Spring 2009 - IEEE 69th Vehicular Technology Conference
- [36] T. Jiang and Y. Wu, "An overview: Peak-to-average power ratio reduction techniques for OFDM signals," IEEE Transactions on broadcasting, 54(2):257-268, 2008.
- [37] D. Guel and J. Palicot, "FFT-IFFT Pair Based Digital Filtering for the Transformation of Adding Signal PAPR Reduction Techniques in Tone Reservation Techniques," in International Conference on Wireless and Mobile Communications (ICWMC), 2009.

A FUZZY BI-OBJECTIVE MODEL FOR SVM WITH AN INTERACTIVE PROCEDURE TO IDENTIFY THE BEST COMPROMISE SOLUTION

Hager Ali Yahia¹, Mohammed Zakaria Moustafa²,
Mohammed Rizk Mohammed¹ and Hatem Awad Khater³

¹Department of Communication and Electronics Engineering,
Alexandria University, Alexandria, Egypt

²Department of Electrical Engineering (Power and Machines Section)
Alexandria University, Alexandria, Egypt

³Department of Mechatronics, Faculty of Engineering,
Horus University, Egypt

ABSTRACT

A support vector machine (SVM) learns the decision surface from two different classes of the input points. In many applications, there are misclassifications in some of the input points and each is not fully assigned to one of these two classes. In this paper a bi-objective quadratic programming model with fuzzy parameters is utilized and different feature quality measures are optimized simultaneously. An α -cut is defined to transform the fuzzy model to a family of classical bi-objective quadratic programming problems. The weighting method is used to optimize each of these problems. An important contribution will be added for the proposed fuzzy bi-objective quadratic programming model by getting different efficient support vectors due to changing the weighting values. The experimental results show the effectiveness of the α -cut with the weighting parameters on reducing the misclassification between two classes of the input points. An interactive procedure will be added to identify the best compromise solution from the generated efficient solutions.

KEYWORDS

Support vector machine (SVMs); Classification; Multi-objective problems; Weighting method; fuzzy mathematics; Quadratic programming; Interactive approach.

1. INTRODUCTION

Support Vector Machines (SVMs) are a classification technique developed by Vapnik at the end of '60s [1]. The theory of support vector machines (SVMs) is a new classification technique and has drawn much attention on this topic in recent years [6]. Since then the technique has been deeply improved, being applied in many different contexts. In many applications, SVM has been shown to provide higher performance than traditional learning machines [6]. SVMs are known as maximum margin classifiers, since they find the optimal hyperplane between two classes as shown in figure1, defined by a number of support vectors [4].

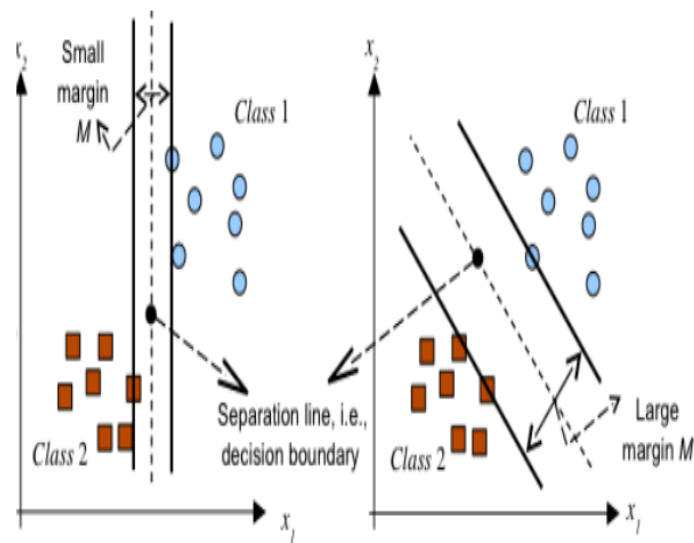


Figure 1: maximization of the margin between two classes

The well-known generalization feature of the technique is mainly due to the introduction of a penalty factor, named C that allows us to prevent the effects of outliers by permitting a certain amount of misclassification errors.

In this paper, the idea is to apply the fuzzy multi-objective programming technique for developing the set of all efficient solutions for the classification problem with minimum errors. An α -cut is taken to transform the fuzzy multi-objective problem model to a classical one (α problem). The weighting method is used to solve the α problem proposed to generate the set of efficient solutions for the proposed model. The remainder of this paper is organized as follows. A brief review for the SVM is described in section 2. The proposed fuzzy bi-objective model for the Support Vector Machine will be derived in section 3. NEXT, section 4 presents three numerical examples corresponding to three different α -cut. Section 5 provides our general conclusions.

2. SUPPORT VECTOR MACHINES

SVM is an efficient classifier to classify two different sets of observations into their relevant class as shown in figure 2 where there are more than straight line separates between the two sets. SVM mechanism is based upon finding the best hyperplane that separates the data of two different classes of the category.

The best hyperplane is the one that maximizes the margin, i.e., the distance from the nearest training points [2].

Support vector machine has been utilized in many applications such as biometrics, chemoinformatics, and agriculture. SVM has penalty parameters, and kernel parameters that have a great influence on the performance of SVM [3]. We review the basis of the theory of SVM in classification problems [7].

Let a set S of labelled training points

$$(y_1, x_1) \dots (y_l, x_l). \quad (1)$$

Where, $x_i \in \mathcal{R}^N$ belongs to either of two classes and is given a label $y_i = \{-1, 1\}$ for $i = 1, \dots, l$.

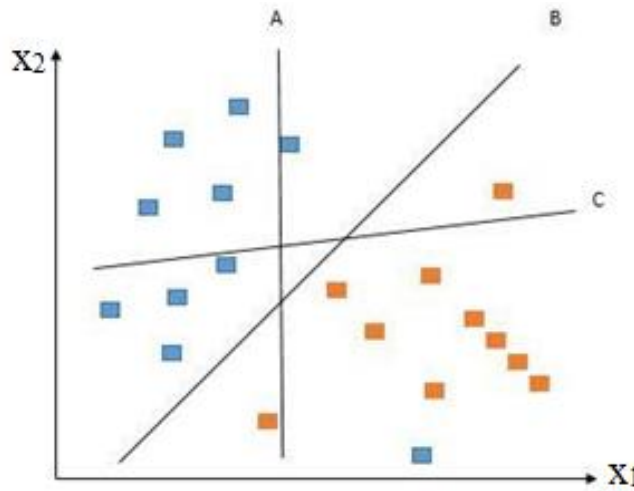


Figure 2: Data classification using support vector machine

In some cases, to get the suitable hyperplane in an input space, mapping the input space into a higher dimension feature space and searching the optimal hyperplane in this feature space.

Let $z = \varphi(x)$ denote the corresponding feature space vector with mapping φ from \mathcal{R}^N to a feature space z . We wish to find the hyperplane

$$w \cdot z + b = 0 \tag{2}$$

defined by the pair (w, b) according to the function

$$f(x_i) = \text{sign}(w \cdot z_i + b) = \begin{cases} 1, & \text{if } y_i = 1 \\ -1, & \text{if } y_i = -1 \end{cases} \tag{3}$$

where $w \in z$ and $b \in \mathcal{R}$. For more precisely the equation will be

$$\begin{cases} (w \cdot z_i + b) \geq 1, & \text{if } y_i = 1 \\ (w \cdot z_i + b) \leq -1, & \text{if } y_i = -1, \end{cases} \quad i = 1, \dots, l \tag{4}$$

For the linearly separable set S , we can find a unique optimal hyperplane for which the margin between the projections of the training points of two different classes is maximized.

For the data that are not linearly separable figure 3, the previous analysis can be generalized by introducing some nonnegative variables $\xi_i \geq 0$ then,

$$y_i(w \cdot z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l. \tag{5}$$

The term $\sum_{i=1}^l \xi_i$ can be thought of as some measure of the amount of misclassifications.

The optimal hyperplane problem is then regarded as the solution to the problem

$$\begin{aligned}
& \text{minimize } \frac{1}{2} w \cdot w + C \sum_{i=1}^l \xi_i \\
& \text{subject to } y_i(w \cdot z_i + b) \geq 1 - \xi_i, \quad (6) \\
& \quad \quad \quad i = 1, \dots, l \\
& \quad \quad \quad \xi_i \geq 0, \quad i = 1, \dots, l
\end{aligned}$$

where, C is a constant. The parameter C can be regarded as a regularization parameter [5]. SVM algorithms use a set of mathematical functions that are defined as the kernel.

The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. For example, linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.

Basically, the training part consists in finding the best separating plane (with maximal margin) based on specific vector called support vector. If the decision is not feasible in the initial description space, you can increase space dimension thanks to kernel functions and may be find a hyperplane that will be your decision separator.

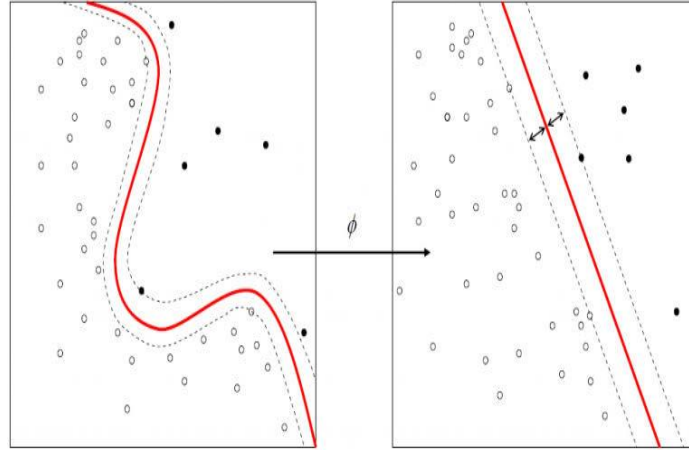


Figure 3: linearly separable and nonlinearly separable

3. FORMULATION OF THE FUZZY BI-OBJECTIVE QUADRATIC PROGRAMMING MODEL OF SVM

In this section, we make a detail description about the idea and formulation of the fuzzy biobjective programming model for the SVM. SVM is a powerful tool for solving classification problems, but due to the nonlinearity separable in some of the input data, there is an error in measuring the amount of misclassification. In the same time, in many real-world applications, each of the input points does not exactly belong to one of the two classes [11].

From this point of view, we reformulate the classical model of the SVM to the following bi-objective programming model with fuzzy parameters.

3.1. The fuzzy bi-objective support vector machine (FSVM):

Now, we add another objective function with fuzzy parameters $\tilde{v}_i, i = 1, 2, \dots, l$ for the previous model in section 2 to be in the form

$$\begin{aligned}
 & \text{Min } \|w\|^2, \\
 & \text{Min } \sum_{i=1}^l \tilde{v}_i \xi_i \\
 & \text{Subject to} \tag{7} \\
 & y_i(w \cdot x_i + b) \geq 1 + \xi_i \quad , i = 1, 2, \dots, l \\
 & \xi_i \geq 0 \quad , i = 1, 2, \dots, l
 \end{aligned}$$

By taken an α -cut for the membership functions corresponding to the fuzzy parameters $\tilde{v}_i, i = 1, 2, \dots, l$, we get the following α -problem:

$$\begin{aligned}
 & \text{Min } \|w\|^2, \\
 & \text{Min } \sum_{i=1}^l \alpha_i \xi_i \\
 & \text{Subject to} \tag{8} \\
 & y_i(w \cdot x_i + b) \geq 1 + \xi_i \quad , i = 1, 2, \dots, l \\
 & \xi_i \geq 0 \quad , i = 1, 2, \dots, l \\
 & \sigma \leq \alpha_i \leq 1 \quad , i = 1, 2, \dots, l
 \end{aligned}$$

With sufficient small $\sigma > 0$.

Where the parameter ξ_i is a measure of the error in the SVM and the term $\alpha_i \xi_i$ is a measure of the error with different degrees α_i . The (α -problem) is solved by the weighting method to get the set of all efficient solutions.

This problem is a bi-objective quadratic programming problem. The first objective is to maximize the gap between the two hyperplanes which used to classify the input points. The second objective is to minimize the error (with different degrees $\alpha_i, i = 1, 2, \dots, l$) in measuring the amount of misclassification in case of nonlinearity separable input points [11].

Problem 8 can be solved by the weighting method to get the set of all efficient solutions for the classification problem.

The right choice of weightage for each of these objectives is critical to the quality of the classifier learned, especially in case of the class imbalanced data sets. Therefore, costly parameter tuning has to be undertaken to find a set of suitable relative weights [10].

3.2. The weighting method

In this method each objective $f_i(X), i = 1, 2, \dots, k$, is multiplied by a scalar weight $w_i \geq 0$ and $\sum_{i=1}^k w_i = 1$. Then, the k weighted objectives are summed to form a weighted-sums objective function [8][12].

$$\text{Assume } W \text{ as } \left\{ \begin{array}{l} w \in R^k: w_i \geq 0, \\ i = 1, 2, \dots, k \\ \text{and } \sum_{i=1}^k w_i = 1 \end{array} \right\} \quad (9)$$

be the set of nonnegative weights. Then the weighting problem is defined as:

$$\begin{array}{l} P(W): \text{Min } \sum_{i=1}^k w_i f_i \\ \text{Subject to } M = \left\{ \begin{array}{l} X \in R^n: g_r(X) \leq 0, \\ r = 1, 2, \dots, m \end{array} \right\}. \end{array} \quad (10)$$

Then, in this paper the weighting method takes the form

$$\begin{array}{l} \text{Inf } z = w_1 \|w\|^2 + w_2 \sum_{i=1}^l \alpha_i \xi_i \\ \text{Subject to} \\ y_i(w \cdot x_i + b) \geq 1 + \xi_i, i = 1, 2, \dots, l \\ \xi_i \geq 0, i = 1, 2, \dots, l \\ w_1 > 0, w_2 \geq 0 \\ w_1 + w_2 = 1 \\ \sigma \leq \alpha_i \leq 1, i = 1, 2, \dots, l \\ \text{With sufficient small } \sigma > 0 \end{array} \quad (11)$$

Here we use “Inf” instead of “Min” since the set of constraints is unbounded, where $w_1 > 0$.

Also, we avoid the redundant solutions by adding the constraint $w_1 + w_2 = 1$.

3.3. An interactive procedure to identify the best compromise solution

For the version of our bi-objective (SVM) model which applies to determine the best compromise solution, we need the following hypothesis (after the interaction with the decision maker)[13]:

The best compromise solution for the set of the generated efficient solution is that efficient one corresponding to

$$\min_{\alpha} N^+ \leq \min_{\alpha} N^-$$

Where, N^- is the number of support vectors of the negative class, N^+ is the number of support vectors of the positive class.

We must notice that this hypothesis can be reversed according to the preference of the decision maker (see Yaochu Jin, 2006) [9].

4. NUMERICAL EXAMPLES

By using python program, we can solve the previous problem and show the effect of different values of the weighting parameters. The data set that is used in these examples consist of 51 points and each point has two features, table 1 shows part of this data.

Table 1: Description of part of datasets used in our study.

X1	X2	Y
1.9643	4.5957	1
2.2753	3.8589	1
2.9781	4.5651	1
2.932	3.5519	1
3.5772	2.856	1
0.9044	3.0198	0
0.76615	2.5899	0
0.086405	4.1045	0

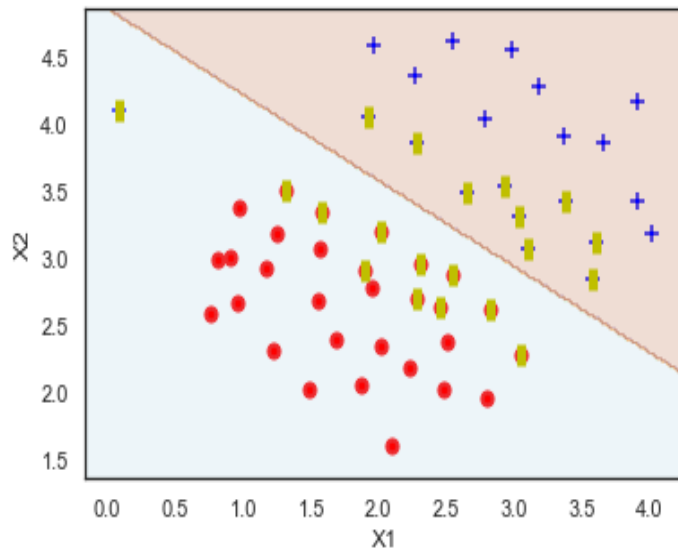


Figure 6: $w_2 = \frac{1}{2}, w_1 = \frac{1}{2}, \alpha = \frac{1}{4}$, number of support vectors = 20

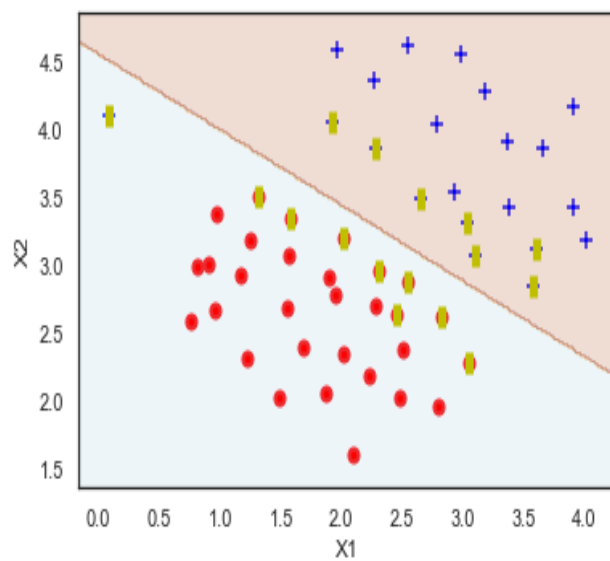


Figure 5: $w_2 = \frac{1}{2}, w_1 = \frac{1}{2}, \alpha = \frac{1}{2}$, number of support vectors = 16

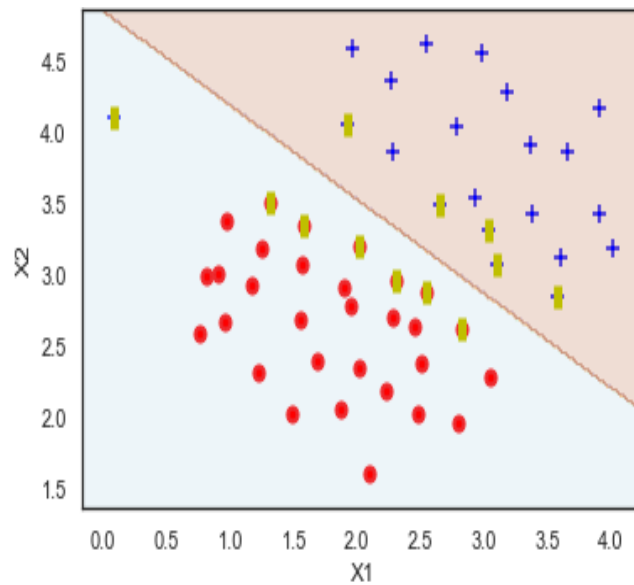


Figure 4: $w_2 = \frac{1}{2}, w_1 = \frac{1}{2}, \alpha=1$, number of support vectors = 12

So, the previous results, by using different degrees (α) at the same weights (w_1 & w_2), show how these parameters (α, w_1, w_2) effect on the performance of SVM. When the value of α is increased the number of support vectors is reduced.

There are good reasons to prefer SVMs with few support vectors (SVs). In the hard-margin case, the number of SVs (#SV) is an upper bound on the expected number of errors made by the leave-one-out procedure [9].

So, we can control the performance of SVM according to our requirements by adjusting the values of the parameters (α, w_1, w_2).

According to our hypothesis that presented in section 3.3, the best compromise solution is that corresponding to $\alpha=1$.

5. CONCLUSIONS

This paper introduced the fuzzy multi-objective programming technique for developing the set of all efficient solutions for the classification problem with minimum errors. The weighting method is used to solve our fuzzy model after defuzzification by using the α – cut technique. The experimental evaluation was carried out using 51 datasets, each one has two features. The experimental results show the effect of the parameters (α, w_1, w_2) on the misclassification between two sets.

Our future work is to try to construct a utility function to select the best compromised hyperplane from the generated set of the efficient solutions.

An interactive procedure is added to identify the best compromise hyperplane from the generated efficient set.

REFERENCES

- [1] Cortes, Corinna; Vapnik, Vladimir N (1995) "Support vector networks" (PDF). *Machine learning*. 20 (3):273297. CiteSeerX 10.1.1.15.9362. DOI:10.1007/BF00994018.
- [2] Asmaa Hamad^{1,3(B)}, Essam H. Houssein^{1,3}, Aboul Ella Hassanien^{2,3}, and Aly A. Fahmy² :Hybrid Grasshopper Optimization Algorithm and Support Vector Machines for Automatic Seizure Detection in EEG Signals. Faculty of Computers and Information, Minia University, Minya, Egypt. January 2018. DOI: 10.1007/978-3-319-74690-6_9.
- [3] Alaa Tharwat¹;, Thomas Gabel¹, Aboul Ella Hassanien²;; Parameter Optimization of Support Vector Machine using Dragon_y Algorithm. Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences, Frankfurt am Main, Germany ,Faculty of Computers and Information, Cairo University, Egypt. January 2018 DOI: 10.1007/978-3-319-64861-3_29.
- [4] Gray, D., Bowes, D., Davey, N., Sun, Y., Christianson, B.: Using the Support Vector Machine as a Classification Method for Software Defect Prediction with Static Code Metrics. In: Palmer Brown, D., Draganova, C., Pimenidis, E., Mouratidis, H. (eds.) *EANN 2009. Communications in Computer and Information Science*, vol. 43, pp. 223–234. Springer, Heidelberg (2009).
- [5] Chun-Fu Lin and Sheng-De Wang: "Fuzzy Support Vector Machines". Article in *IEEE Transaction on neural networks*, VOL.13, NO.2, March 2002. DOI:10.1109/72.991432.
- [6] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol.2, no.2,1998.
- [7] C. Cortes and V. N. Vapnik, "Support vector networks," *Machine Learning*, vol.20,pp.273-297,1995.
- [8] Chankong V. and Haimes Y.Y., *Multi-objective Decision-Making: Theory and Methodology* (North Holland Series in System Science and Engineering, 1983).
- [9] YaochuJin (Ed.), *Multi-objective Machine Learning Studies in Computational Intelligence*, Vol. 16, pp. 199-220, Springer-Verlag, 2006.
- [10] Shounak Datta and Swagatam Das, Multiobjective Support Vector Machines: Handling Class Imbalance With Pareto Optimality, *IEEE Transactions on Neural Networks and Learning Systems*, 2019.DOI:10.1109/TNNLS.2018.2869298.
- [11] Y.Wang, Shouyang Wang, K.K.Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Transactions on Fuzzy Systems*, VOL.13, NO.6, January 2006, DOI:10.1109.
- [12] Mohammed Zakaria Moustafa, Mohammed Rizk Mohammed, Hatem Awad Khater and Hager Ali Yahia, Building a Bi-objective Quadratic Programming Model for The Support Vector Machine, 8th International Conference on Artificial Intelligence, Soft Computing (AISC 2020), DOI: 10.5121/csit.2020.100208.
- [13] Mohammed Zakaria Moustafa, Mohammed Rizk Mohammed, Hatem Awad Khater and Hager Ali Yahia, A bi-objective Model for SVM with an interactive procedure to identify the best compromise solution, *International Journal of Artificial Intelligence and Applications (IJAIA)*, Vol.11, No.2, March 2020.

IMPORTANCE OF THE SINGLE-SPAN TASK FORMULATION TO EXTRACTIVE QUESTION-ANSWERING

Marie-Anne Xu¹ and Rahul Khanna²

¹Crystal Springs Uplands School, CA, USA

²University of Southern California, CA, USA

ABSTRACT

Recent progress in machine reading comprehension and question-answering has allowed machines to reach and even surpass human question-answering. However, the majority of these questions have only one answer, and more substantial testing on questions with multiple answers, or multi-span questions, has not yet been applied. Thus, we introduce a newly compiled dataset consisting of questions with multiple answers that originate from previously existing datasets. In addition, we run BERT-based models pre-trained for question-answering on our constructed dataset to evaluate their reading comprehension abilities. Among the three of BERT-based models we ran, RoBERTa exhibits the highest consistent performance, regardless of size. We find that all our models perform similarly on this new, multi-span dataset (21.492% F1) compared to the single-span source datasets (~33.36% F1). While the models tested on the source datasets were slightly fine-tuned, performance is similar enough to judge that task formulation does not drastically affect question-answering abilities. Our evaluations indicate that these models are indeed capable of adjusting to answer questions that require multiple answers. We hope that our findings will assist future development in question-answering and improve existing question-answering products and methods.

KEYWORDS

Natural Language Processing, Question Answering, Machine Reading Comprehension

1. INTRODUCTION

Machine Reading Comprehension (MRC), particularly extractive close-domain question-answering, is a prominent field in Natural Language Processing (NLP). Given a question and a passage or set of passages, a machine must be able to extract the appropriate answer or even set of answers from the passage(s). Solving this task has various real-world implications, particularly in industry areas such as customer support. Some application examples include chatbots, voice assistants, and automated customer service. Using these applications can greatly increase efficiency for both companies and customers by reducing time spent hunting for answers that a machine can find in seconds.

Many groundbreaking question-answering datasets such as the Stanford Question Answering Dataset, SQuAD [1], consist of only single-span question-answer pairs, or answers that require only one extraction. Numerous datasets have been created with several answer categories, NewsQA [2], DuReader [3], MS MARCO [4], DROP [5], however, the majority of the answers are single-span. Models that are trained on these datasets are therefore primarily focused on extracting just a single answer, possibly precluding their effectiveness when multiple answers are

required. For example, one sample of a multi-span question-answer pair along with its relevant passage from the DROP dataset is shown in Figure 1.

The first issue in 1942 consisted of denominations of 1, 5, 10 and 50 centavos and 1, 5, and 10 Pesos. The next year brought "replacement notes" of the 1, 5 and 10 Pesos while 1944 ushered in a **100 Peso note** and soon after an inflationary **500 Pesos note**. In 1945, the Japanese issued a 1,000 Pesos note. This set of new money, which was printed even before the war, became known in the Philippines as Mickey Mouse money due to its very low value caused by severe inflation...

Which new peso notes were the highest created by 1944?
100 Peso note, 500 Pesos note

Figure 1: A shortened DROP dataset passage with its multi-span question-answer pair.

Various models such as Retro-Reader [6], SA-Net, and ALBERT [7] have already surpassed human performance on MRC datasets like SQuAD 2.0 [8], which introduces questions that may not have answers given the passage(s). However, model performance on datasets such as Discrete Reasoning Over Paragraphs (DROP) have not yet matched human quality, and models trained on datasets like DuReader are even further away from reaching human performance. These contrasting results show that it is unclear whether or not models are able to execute machine reading comprehension. More recently, models such as the Multi-Type Multi-Span Network [9] are specially designed to extract either one or multiple answer spans. While a specialized model is capable of returning multi-span answers, we seek to investigate if the current state-of-the-art models can adapt without fine-tuning to produce similar results.

Thus, this research project proposes to assess the performance of the current state-of-the-art models when evaluated on only the multi-span questions of existing datasets. By exploring the MRC abilities of models trained on single-span extraction, we can determine if the model is simply concentrating on returning only one answer or if it is actually processing and comprehending the task of question-answering. Future researchers will be able to use this work in order to identify where question-answering models can be improved and recognize the limitations of using a model trained on a single-span dataset. Additionally, single-span answers could potentially be overly specific for a given question, thus exploring multi-span answer question-answering can potentially provide the end-user with more information to answer their questions. The new multi-span dataset compiled from the DROP and NewsQA datasets is also available for future research.

2. RELATED WORKS

As previously mentioned, the majority of existing question-answering pairs have a single answer. The type of dataset we are particularly interested in for our evaluation is extractive closed-domain question-answering. The appropriate answer(s) must be directly extracted from only the given passage(s). One example of this kind of dataset is the SQuAD dataset, with over 100,000 single-span questions, making it larger than most previous question-answering datasets. SQuAD has several answer types, such as Date, Person, and Location, and its passages cover an extensive number of subjects. SQuAD 2.0, the latest version of SQuAD, combines the original 100,000 questions with new unanswerable questions, forcing models to learn when to abstain from

answering. Similarly, the DROP and NewsQA datasets are also extractive, closed-domain datasets that have a small percentage of multi-span question-answer pairs (6.0% for DROP and 5.68% for NewsQA). The DROP dataset contains questions that may need numerical operations (e.g. “Who threw the longest touchdown pass?”) or entity co-referencing (tracking the multiple appearances of an entity). Compared to SQuAD, DROP has more complex passages and questions that require more complicated reasoning. Like SQuAD 2.0, NewsQA has more ambiguous questions and questions without answers, and its answer types mirror those of SQuAD (Numeric, Clause Phrase, Date/Time, Location, etc). However, the human performance of NewsQA is much lower than the previous two datasets, despite it being a similar size. While other datasets such as the MS MARCO and DuReader datasets have multi-span answers, they are either generative (answers cannot be directly extracted), or are in another language like Chinese.

Some of the most popular NLP models are BERT [10] and its variations, RoBERTa [11] and ALBERT [7]. BERT, which stands for Bidirectional Encoder Representations from Transformers, is capable of executing a variety of general NLP tasks, such as next sentence prediction. The key feature of BERT is that it can look at the context on both sides of a given text span, hence the part “bidirectional”. RoBERTa, or Robustly optimized BERT approach, introduces alternate strategies for the BERT training process in order to improve performance. ALBERT (A Lite BERT) contains fewer parameters than BERT, reducing training time and memory restrictions while simultaneously maintaining and even producing better results.

3. PURPOSE

Multi-span question-answering can be used to improve existing applications of question-answering and NLP in general. As mentioned earlier, MRC plays a key role in industry, forming the foundation for tools such as virtual assistants. However, these machines are still flawed. If a certain question posed by a user requires multiple answers, the machine needs to be able to find different, quality answers in order to be as productive and helpful as possible. Models that are trained on mainly single-span questions may not exhibit high-quality performance when faced with multi-span questions. Using our evaluations, future research will improve question-answering models and hopefully implement them to refine existing applications. If the models are capable of answering multi-span questions posed by an end-user, they will minimize the user’s time spent looking for an answer that the machine was not able to return. By contributing to the field of MRC with our evaluations, we also hope to further the application of MRC and NLP as a whole in the real world.

4. METHODS

To analyze model performance on multi-span question answering, we collect all question-passage-answer triplets that have multi-span answers from the DROP and NewsQA datasets. The newly compiled, strictly multi-span dataset consists of almost 30K questions. As shown in Table 1, most passages are quite long, while answers are only a few words.

Table 1. Multi-span dataset statistics.

Statistic	
Number of Questions	29288
Avg question len (words)	7.10
Avg passage len (words)	546.54
Avg answer len (words)	4.69
Avg answers/question	2.22

Refer to Figure 2 for some details on the top first and second words of the questions. Similar to a pie chart, the first words are represented by the inner circle, and the second words are shown as subcategories in the outer ring. 16,674 of the total 29,288 questions are represented here, roughly 57% of the entire dataset. The first words are dominated by “what,” “who,” and “where” questions; the “what” section forms approximately 59% of these 16,674 questions, “who” makes up 14%, “where” 10%, and “which” 8%. The unusual “players” section under “which” most likely originates from the large number of National Football League (NFL) questions in the DROP dataset. For example, “which players scored touchdowns longer than 20 yards?” would fall under the first word category “which” and second word category “players”.

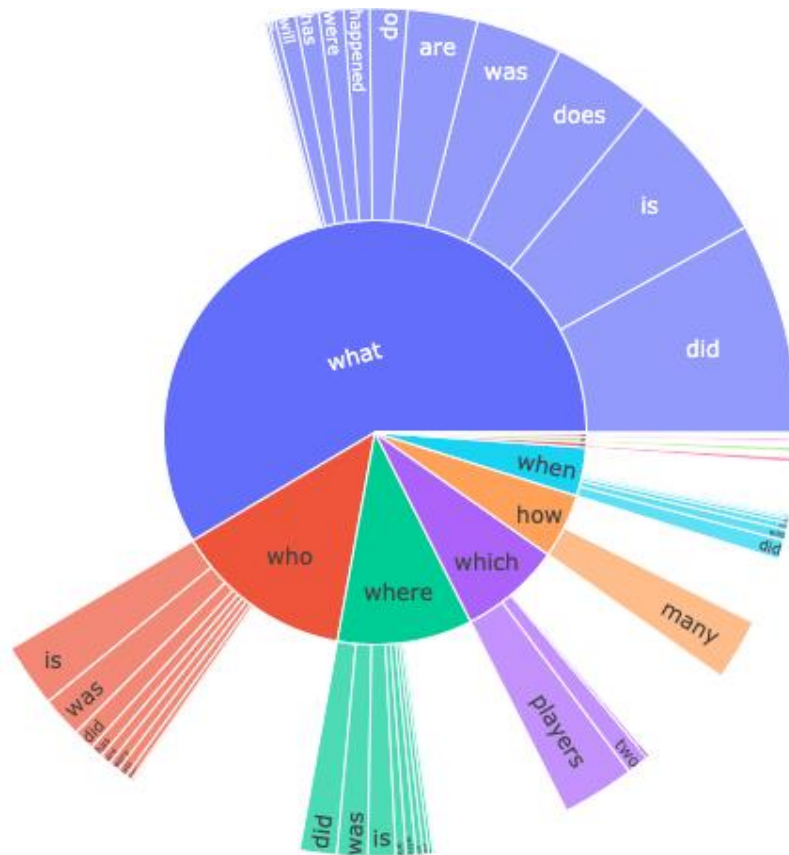


Figure 2: The top 10 first and top 14 second question words of the multi-span dataset.

The models run on this dataset are the most downloaded BERT-base [14], BERT-large [15], RoBERTa-base [16], RoBERTa-large [17], ALBERT-base [18], and ALBERT-xlarge [19] models fine-tuned on the SQuAD2.0 dataset. We choose these models because they have already been fine-tuned for single-span extractive question-answering and can be easily modified and evaluated on their multi-span question-answering abilities. We follow the standard preparation of question and passage pairs for the extractive task, [CLS] question text [SEP] passage text. As BERT based models have a token limit of 512 tokens, we follow common practice of truncating all constructed sequences to the 512 token limit, this affects 19,837 of our question, passage pairs. Due to limited runtime, we run all models on the 9,451 shorter question, passage pairs and the three base models on the entire dataset. Evaluation runtime is approximately three days using a Google Cloud Platform Compute Engine CPU. After passing the tokens into the model, the model produces the start and end scores for each token; for example, the token with the highest starting score marks the start of the best span, and the token with the highest end score is the last token in the span. We utilize a relatively naive, greedy approach by extracting the top n non-

overlapping answer spans with the best scores. The number of predicted answer spans is determined by the number of true answer spans. We construct the predicted answers by concatenating the relevant tokens; all of the tokens between the start and end token are part of the predicted answer span. If the model is unable to find an acceptable span, it returns an empty string.

To assess the models, we create two arrays that are the same length as the passage's character length. One array is for the predicted answers while the other is for the true answers. For each answer span, we find its location in the passage by string matching; the answer span is a substring of the passage. In order to accurately match the strings, we must lowercase the original passage as BERT and ALBERT answer spans are not case-sensitive. We do not lowercase for RoBERTa because the most downloaded models, and subsequently the ones we used in this experiment, are case-sensitive. Characters in the passage that are part of an answer are labeled as a one (referred to as Class 1) while the rest are zeroes (Class 0). The result is an array of ones and zeros that is the same length as the character length of the passage. We can then compare the indices of the characters in the true and predicted answers, or the ones in the arrays. Due to a time constraint, we choose the first occurrence of the answer in the passage if it appears multiple times. This does not affect the evaluations because the answer spans are neither overlapping nor repeated.

For each predicted answer set, we calculate the average exact match: how many predicted answers are exactly the same as the true answers? As stated above, we lowercase the true answers for BERT and ALBERT before matching and do not lowercase for RoBERTa. We also calculate the micro and Class 1 precision, recall, and F1 scores between the true and predicted binary arrays that we created earlier. Precision measures the number of characters in the predicted answers that are also true; micro-precision calculates the global precision while Class-1 reports the precision for each answer. Conversely, recall measures the number of characters in the true answers that are also in the predicted answers. We report the precision, recall, and F1 scores because they judge how similar our predicted answers are to the true answers when they are not identical.

Our final metric is BLEU [12], an automatic evaluation method that originated from the field of machine translation. BLEU compares the number of *n-grams*, or set of words, present between the candidate (predicted answer, in our case) and reference (true answers) sentences. BLEU also penalizes the candidate (predicted) sentence when it is shorter than the reference(s), called the brevity penalty. We use the BLEU-1 score—the majority of answers are only one word, thus we only use unigrams when computing our BLEU scores. We calculate the BLEU scores between the predicted and true answer spans for each question and then find the arithmetic mean. BLEU differs from the previous metrics, such as precision, in that it introduces the brevity penalty and also measures by word, while precision measures by character. We can then compare these all of the mentioned metrics to the original DROP and NewsQA evaluations to determine the performance of our models.

5. RESULTS AND DISCUSSION

Of the three types of models, RoBERTa stands out as consistently exhibiting strong scores in all evaluation metrics. Other models like BERT-base and ALBERT-xlarge had peaks in certain metrics, like exact match and F1 respectively, but their remaining metrics were not as prominent. RoBERTa-large had the more consistently high scores, most notably the highest Class-1 F1 and recall scores. Class-1 metrics are calculated when we only consider the true (Class-1) characters in our binary arrays. A few key metrics (EM, micro-F1, and BLEU) are displayed in Figure 3 and Table 2. The remaining evaluations (precision, recall) for each model can be seen in Table 3.

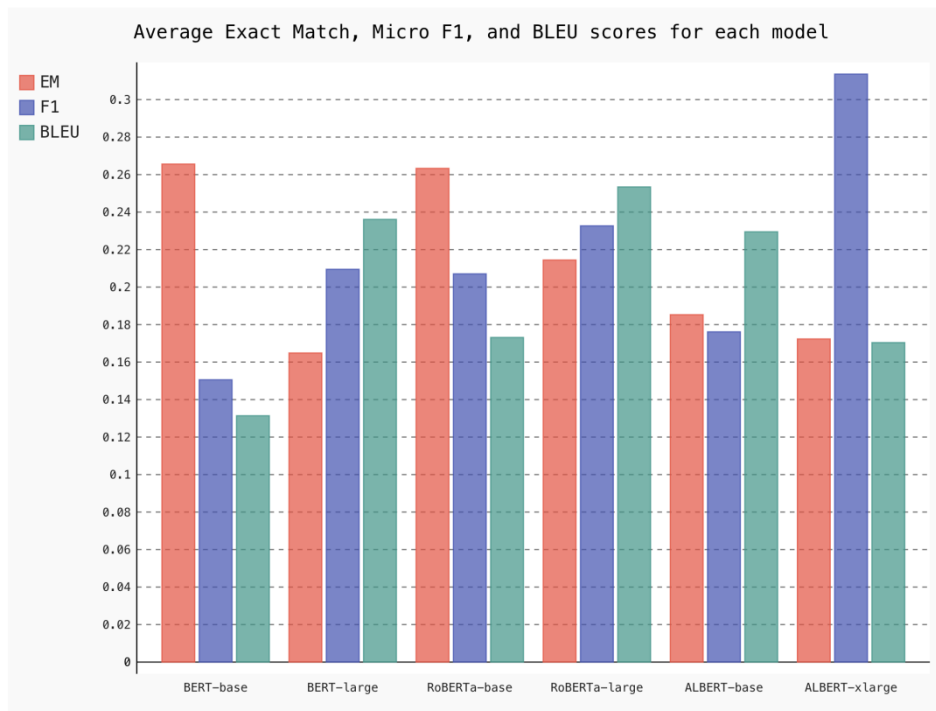


Figure 3: EM, micro-F1, and BLEU-1 scores for each model. The large models are run on a smaller version of the dataset (32%); base models are run on the entire dataset.

Table 2: A table representation of Figure 3.

	EM	F1	BLEU-1
BERT-base	26.561	15.057	13.138
BERT-large	16.484	20.949	23.612
RoBERTa-base	26.333	20.707	17.314
RoBERTa-large	21.445	23.267	25.342
ALBERT-base	18.527	17.613	22.949
ALBERT-xlarge	17.233	31.360	17.039
Average	21.097	21.492	19.899

Comparing our metrics in Figure 3 and Table 2 to the overall BERT scores of the entire DROP dev dataset, we can see that the evaluations of our six models are worse. This is expected, as the majority of the DROP dataset consists of single-span answers. The BERT scores for SQuAD-style Reading Comprehension on the DROP dev dataset are 30.10 (EM) and 33.36 (F1), using the BERT-large model. Additionally, the BERT model used in the DROP dataset was slightly fine-tuned while our BERT models were not.

When we observe the evaluations of the original multi-span subset of DROP, we find that the BERT scores listed in their paper are 0 (EM) and 25.0 (F1). In comparison, our BERT scores are

26.561 (EM) and 15.057 (Micro F1). Our exact match score for BERT-large, and all six models in general, are much higher, while our F1 scores are similar. We reason that this result holds as the multi-span DROP questions only made up around 20% of our dataset while the other 80% came from NewsQA, and EM scores on the NewsQA dataset are much higher. However, BERT-based models were not run on the original NewsQA dataset, so we cannot do a like-for-like evaluation. Our greedy strategy of selecting the multiple top answers has most likely also affected the performance scores of our models.

While NewsQA was not run on a BERT-based model, the alternate models stated in the original paper, mLSTM and BARB, produced EM scores of 34.4% and 36.1% respectively. The F1 scores for both models were 49.6%. Again, the majority of NewsQA consists of multi-span question-answer pairs, so our lower scores are expected. While we cannot definitively conclude the performance of these BERT-based models aren't affected by the change of task, we can see that the models are adapting to multi-span question-answering to a high degree, as the EM and F1 scores are not extremely low.

Table 3: Class-1 metrics, which focus on the true (Class-1) characters rather than words.

	Class-1 F1	Class-1 Precision	Class-1 Recall
BERT-base	18.418	24.892	20.677
BERT-large	28.588	34.509	35.028
RoBERTa-base	24.766	33.690	27.028
RoBERTa-large	34.004	42.586	43.010
ALBERT-base	27.492	31.539	39.406
ALBERT-xlarge	26.199	33.861	28.274
Average	26.578	33.513	32.237

In Table 3, the class-1 metrics are calculated using the previously mentioned methods. Some notable metrics include the high precision and recall of RoBERTA-large, 42.586 and 43.010 respectively, as well as the high recall of ALBERT-base (39.406). The overall recall of the models is the highest of the general metrics, and we judge recall to be the most important metric because it measures the number of true characters that are in the predicted answers. Although the predicted answers may not be exactly the same as the true ones (which the exact match score penalizes), recall checks for the presence of the true characters. Framing our evaluations as a sequence tagging task, we look at the performance of our models as sequence taggers, examining in particular their performance when an answer is the true label for a character (i.e. the character is within the answer). In this regard, class-1 recall is a key statistic, as this would show that the model is able recover the true answers' character tokens, while allowing it to potentially start the span a bit earlier or later than the annotated answer, which does not necessarily affect our recall negatively. Because we are checking for the presence of the true characters and judging the models' abilities to identify them, we do not focus much on the additional characters in the span.

We expect RoBERTa to perform better than BERT, as it is pre-trained for a longer period of time. What is unexpected, however, is that ALBERT does not surpass RoBERTa's performance. There are several potential reasons for this, one being that ALBERT is much smaller in size than RoBERTa, but we leave this exploration to future research.

Generally, the larger models return higher scores than their base counterparts. This is expected, as the larger models usually outperform the base versions, even in non-question-answering tasks. However, one exception to this trend is that the exact match scores of the large models are lower than the bases. Another notable comparison is that despite ALBERT-xlarge being several times larger than ALBERT-base, various ALBERT-xlarge metrics are either close to or lower than ALBERT-base's, like the exact match, class-1 F1, precision, and recall, micro-recall, and BLEU-1 scores. The remaining metrics, the micro-F1 and micro-precision scores, are much higher and match our expected trend that increased size implies improved performance.

Our overall exact match and F1 scores, especially compared to the DROP and NewsQA scores, reveal that the six models are capable of adapting and returning more than one answer span. Because our models can produce scores that are not significantly lower than previous scores and are even better in some cases, we can infer that the models are indeed adjusting to multi-span question-answer pairs.

6. CONCLUSIONS

Because the majority of existing machine question-answering datasets consist of single-span question-answer pairs, we seek to evaluate state-of-the-art model performance on multi-span questions. We have assembled a new multi-span question-answer dataset from the DROP and NewsQA datasets. We have observed the performance of six BERT-based models pre-trained for single-span question-answering when run on this compiled multi-span dataset. We find that RoBERTa has consistently higher scores than BERT and ALBERT, perhaps due to its alternate training strategies, and we also note that the larger variations of each model perform better than the base counterparts, although at the cost of increased runtime. When comparing the EM and F1 scores of our BERT-based models to the BERT-scores of the parent DROP dataset, we find that the EM scores have improved significantly and the F1 scores are similar, although slightly lower. Based on the unbalanced distribution of DROP and NewsQA questions, we also look at the scores of the other parent dataset, NewsQA; although not from BERT, we see that our scores are not drastically lower. Because our BERT models have not been fine-tuned to our multi-span dataset while the BERT model for DROP evaluation has, this difference in scores still allows us to conclude that task formulation does not drastically reduce model performance. We discover that current state-of-the-art models are capable of adapting to multi-span extractive question-answering and are not structurally limited to single-span extraction. We hope that with this information, future research can implement multi-span question-answering into real-world applications to improve efficiency in industry and daily life.

7. FUTURE WORK

Some potential future research directions include fine-tuning a model to answer multi-span questions and conforming to more standard evaluation metrics (such as CIDER [13]) We also suggest exploring alternate, more robust extraction methods that are better than our naive greedy approach. Another prospective project involves more detailed performance evaluation on certain subsets of this multi-span dataset, such as the “who,” “what,” “when,” “where,” “which,” and “how” question subcategories. Additionally, more work can be done in lengthening the extracted spans in order to use the standard BLEU-4 evaluation method. In hopes of furthering MRC and NLP, we also leave this multi-span dataset available for future research.

ACKNOWLEDGEMENTS

We would like to thank the Summer Stem Institute for providing the opportunity to conduct this research.

REFERENCES

- [1] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. “SQuAD: 100,000+ questions for machine comprehension of text”. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas, November 2016. Association for Computational Linguistics.
- [2] Adam Trischler, Tong Wang, Xingdi Yuan, Justin Harris, Alessandro Sordoni, Philip Bachman, and Kaheer Suleman. “NewsQA: A machine comprehension dataset”. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 191–200, Vancouver, Canada, August 2017. Association for Computational Linguistics.
- [3] Wei He, Kai Liu, Jing Liu, YajuanLyu, Shiqi Zhao, Xinyan Xiao, Yuan Liu, Yizhong Wang, Hua Wu, Qiaoqiao She, Xuan Liu, Tian Wu, and Haifeng Wang. “DuReader: a Chinese machine reading comprehension dataset from real-world applications”. In *Proceedings of the Workshop on Machine Reading for Question Answering*, pages 37–46, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [4] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. “MS MARCO: A human generated machine reading comprehension dataset”. *CoRR*, abs/1611.09268, 2016.
- [5] DheeruDua, Yizhong Wang, Pradeep Dasigi, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. “DROP: A reading comprehension benchmark requiring discrete reasoning over paragraphs”. In *Proceedings of NAACL*, 2019.
- [6] Zhuosheng Zhang, Junjie Yang, and Hai Zhao. “Retrospective reader for machine reading comprehension”, 2020.
- [7] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. “Albert: A litebert for self-supervised learning of language representations”. In *International Conference on Learning Representations*, 2020.
- [8] Pranav Rajpurkar, Robin Jia, and Percy Liang. “Know what you don’t know: Unanswerable questions for SQuAD”. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 784–789, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [9] Minghao Hu, Yuxing Peng, Zhen Huang, and Dongsheng Li. “A multi-type multi-span network for reading comprehension that requires discrete reasoning”. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1596–1606, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “BERT: pre-training of deep bidirectional transformers for language understanding”. *CoRR*, abs/1810.04805, 2018.
- [11] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. “Roberta: A robustly optimized BERT pretraining approach”. *CoRR*, abs/1907.11692, 2019.
- [12] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. “Bleu: a method for automatic evaluation of machine translation”. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [13] Ramakrishna Vedantam, C. Lawrence Zitnick, and Devi Parikh. “Cider: Consensus-based image description evaluation”. *CoRR*, abs/1411.5726, 2014.
- [14] BERT-base: <https://huggingface.co/twmkn9/bert-base-uncased-squad2>
- [15] BERT-large: <https://huggingface.co/deepset/bert-large-uncased-whole-word-masking-squad2>
- [16] RoBERTa-base: <https://huggingface.co/deepset/roberta-base-squad2>
- [17] RoBERTa-large: https://huggingface.co/ahotrod/roberta_large_squad2
- [18] ALBERT-base: <https://huggingface.co/twmkn9/albert-base-v2-squad2>

- [19] ALBERT-xlarge: <https://huggingface.co/ktrapeznikov/albert-xlarge-v2-squad-v2>
- [20] Jiahua Liu, Wan Wei, Maosong Sun, Hao Chen, Yantao Du, and DekangLin. “A multi-answer multi-task framework for real-world machine reading comprehension”. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2109–2118, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- [21] Lluís Màrquez, Pere Comas, Jesús Giménez, and Neus Català. “Semantic role labeling as sequential tagging”. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 193–196, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.
- [22] Pradeep Dasigi, Nelson F. Liu, Ana Marasović, Noah A. Smith, and Matt Gardner. “Quoref: A reading comprehension dataset with questions requiring coreferential reasoning”, 2019.
- [23] Subeh Chowdhury, Michael O’Sullivan. “A fuzzy logic-genetic algorithm approach to modelling public transport users’ risk-taking behaviour”. *Transportation Planning and Technology* 41:2, pages 170-185.
- [24] G. Megali, D. Pellicano, M. Cacciola, S. Calcagno, M. Versaci, and F. C. Morabito, "Ec Modelling and Enhancement Signals in Cfrp Inspection," *Progress In Electromagnetics Research M*, Vol. 14, 45-60, 2010.

AUTHORS

Marie-Anne Xu is a student with a keen interest in computational linguistics. She has competed in the North American Computational Linguistics Open Competition twice and qualified for the invitational round in 2020. In her spare time, she has participated in coding events such as hackathons and Google Code-In. She has received the National Center for Women & Information Technology Award for Aspirations in Computing National Certificate of Distinction.



Rahul Khanna is a Masters Student at USC's Viterbi School of Engineering and a Researcher at USC's Intelligence and Knowledge Discovery Research Lab. He is primarily interested in semi-supervised, self-supervised and human-in-the-loop learning for Information Extraction, but also explores the areas of Common Sense Reasoning and Question-Answering. Prior to USC, Rahul earned his B.S. from Columbia's Fu School of Engineering and Applied Sciences and spent three years working as machine learning engineer/datascientist for Vimeo and Impact.



AUTOMATIC GENERATION OF TEXT FOR MATCH RECAPS USING ESPORT CASTER COMMENTARIES

Oluseyi Olarewaju¹, Athanasios V. Kokkinakis¹, Simon Demediuk²,
Justus Roberstson¹, Isabelle Nölle¹, Sagarika Patra¹, Daniel Slawson¹,
Alan P. Chitayat¹, Alistair Coates¹, Ben Kirman², Anders Drachen²,
Marian Ursu², Florian Block² and Jonathan Hook²

¹Weavr, Arena Research Cluster (ARC), University of York, York, UK

²DC Labs, Arena Research Cluster (ARC), University of York, York, UK

ABSTRACT

Unlike traditional physical sports, Esport games are played using wholly digital platforms. As a consequence, there exists rich data (in-game, audio and video) about the events that take place in matches. These data offer viable linguistic resources for generating comprehensible text descriptions of matches, which could, be used as the basis of novel text-based spectator experiences. We present a study that investigates if users perceive text generated by the NLG system as an accurate recap of highlight moments. We also explore how the text generated supported viewer understanding of highlight moments in two scenarios: i) text as an alternative way to spectate a match, instead of viewing the main broadcast; and ii) text as an additional information resource to be consumed while viewing the main broadcast. Our study provided insights on the implications of the presentation strategies for use of text in recapping highlight moments to Dota 2 spectators.

KEYWORDS

Esport, Data-Driven Storytelling, Dota 2, Game Analytics, Broadcasting, social viewing, Linguistic Resources, Natural Language Generation.

1. INTRODUCTION

Advances in Natural language generation (NLG) and data-to-text tools have found relative success outside esport (e.g. weather forecasting, robojournalism, traditional sport summaries and fantasy football). Bringing the richness of automated storytelling to audiences of esport using NLG and data-to-text techniques by generating text-based recaps creates a unique selling point for content producers who may want to appeal to a wider audience beyond expert esport players and viewers. Automatically generated text can also be easily customized to reflect preferences that are particular to players' style and level of experience of the viewers. This short form of highlight personalization can help gaming operators maintain an active player base who otherwise may not have the time or resource to commit to viewing an entire esport game. In addition to the proposed support by [1] for amateur casters such as pairing casters for live games in a matchmaking queue. Automatically generated text recap offer opportunity to create professional level text recaps similar to those that are written by humans for lower leagues that have fewer resources.

One of the most significant differences between esports and traditional sports is the level of complexity of the gameplay. Complexity has been known to create significant barriers to sport enjoyment and may take away the initial appeal that the aesthetic aspect of the games may have made to players/spectator [2]. As esports popularity grows, content creators need to present the complexities in the games played in understandable ways without comprising on the quality of insights they provide. This motivates the research presented here, we study the use of text-based recaps generated by Natural Language Generation (NLG) systems for presenting highlight moments in esports games (e.g. Dota 2). Can content producers communicate non-linguistic esports data to better engage their audience and make watching esports more meaningful and enjoyable experience? While the potential of these approaches is enticing, it is not yet understood whether current state of the art techniques for NLG can be used to create text summaries that are accurate and engaging for audiences and what value these summaries can have for audiences during and around the spectating of matches. Our work aims to contribute to knowledge in this area by addressing these questions using a generation grammar derived from parsed live commentaries to automatically generate text-based recaps of most important actions and highlights of Dota 2 games.

Text-based recaps are translations of non-linguistic performance data of players in a match to linguistic data stories. Using production rules from the generation grammar, inputs obtained from game data are mapped to parsed commentaries to generate recaps for moments of highlights in matches. On one hand, we investigate if spectators perceive automatically generated text from the grammar as accurate recaps of the highlight moments in Dota 2 matches. On the other hand, we asked if text recaps generated by the derived grammar support understanding of highlight moments in Dota 2 matches by considering two use cases. In the first case, we considered the potential use of automatically generated text as an alternative way to spectate a match, instead of viewing the main broadcast. In the second case we considered a combination of linguistic and visual information for recapping highlight moments and investigate using text as an additional information resource to be consumed while viewing the main broadcast.

In many real life applications, it is essential that text outputs from Natural Language Generation (NLG) systems accurately inform end users about input data that is being communicated. According to [3], hallucination and other forms of inaccuracy are unacceptable in NLG application contexts such as journalism, financial reporting, and medical patient information. Esports, although an emerging industry is without exception. User analytic are important for improving skill level and all non-linguistic data need to be communicated correctly. In this work we used handcrafted features with grammar rules derived from commentary corpus to support the linguistic communication of user performance data in easy-to-interpret text formats. Commentaries in esports games are one of the ways that enjoyment is sustained by casters who provide a hyped narrative of play or events of the game. Entertaining and informing audiences about different aspects of a live game are two of the roles casters have to undertake. To keep audiences engaged, the comments that casters make during games are often articulate and colourful. These characteristics of esports commentaries make them an interesting resource for inducing rule-based grammars that can be used to generate text-based recaps which closely resemble human-written summaries.

In Section 2 we discuss previous attempts to using text corpus for generating text. We describe our approach to deriving a generation grammar and detail the process of generating text recaps in Section 3. In section 4, we provide details of our lexical resource and the corpus creation process. In Sections 5 and 6 we present our study design and the results of evaluations of the NLG System. The implications of our results for esports domain is discussed in section 7. Finally, we present our conclusion and future ideas in section 8.

2. RELATED WORK

According to [2] and [4], Esports are video games played by professional gamers that are broadcast online. Esport attracted a global viewership of 454 million in 2019 and has been projected to have a year on year growth of 15% over the next four years [5]. Esports popularity has also drawn the attention of national governments as potential means of job creation and revenue generation and has prompted large government research into exploring Esports as a new immersive data-driven experience. A point in case is the Weavr project in [6] which is government-sponsored consortium project bringing together partners with extensive expertise in a range of field relating to the creation of data-driven experiences. In traditional sports such as football, data-driven contents have been shown to enable content producers in reporting and explaining things that were not previously explicable [7]. Research in content production in Esports by [8] confirmed that data-driven content can be an effective tool to make gameplay more transparent to viewers. Using simple graphical overlays of data-driven insights, they found measurable effects on the commentary and quality of Esport coverage.

The large volume of publicly available data from Esport production has enabled the extraction of insight from esport games using novel machine learning techniques where model performance is proportional to the size of the training data. A considerable number of literature have used Esport data and published work on different insights extracted from the data using different machine learning and analytical approaches. These include classification and detection of player roles using unsupervised learning techniques for more accurate analysis [9] and [10]. A hero recommendation engine using regression and nearest neighbour algorithms was proposed by [11]. Researcher in [12] reported the performance of an encounter-based algorithm in encounter detection and win probability predictions for Dota 2 games. A few research efforts have also used visual features in video streams to identify and forecast moments of highlights within games. These include the work of [13] where they used a psycho-physiological approach and the data-driven approach to construct highlight models. Esport games use distinct visual effects to highlight key moments in a game, [14] used Convolved Neural Networks to learn filters of those visual effects for detecting highlights. Highlights were also detected by [15] using deep learning methods on a combination of game footage, facial expressions and speech data.

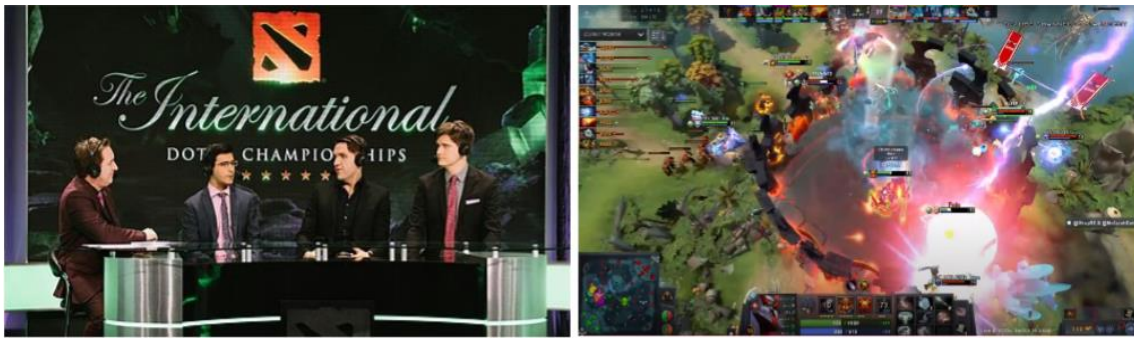
Natural language and narrative storytelling researchers have used techniques in computational and theoretical linguistics to gain a better understanding of player behaviour and creating novel game experiences. Gameplay was modelled by [16] as an online mediation search and used a text-based implementation to show that the search space can embed all traversals through a game world. Insights on how player behaviour affected players' collaboration in teams were extracted by [17] and [18] by analysing texts from player discussion forums and chat logs specifically for the Esport game -League of Legends. A growing number of linguistic resources have also been obtained from crowdsourcing games or by using games with a purpose (GWAP) approaches. A dataset of sentimentally classified tweets was built by [19] using GWAP and [20] used a crowdsourcing game to create language-independent sentiment lexicons. A subject of equal importance to extracting insights for enhancing the quality of gaming experience is the communication of insight to players. In game analytics, researchers have leveraged the simplicity of text to communicate insights from otherwise complex game telemetry to viewers of Esport games. To facilitate the use of their hero recommendation engine, [11] used a text auto-complete interface to present suggestions to users. A simple text annotations was used by [8] to wrap the core message of insights on player performances derived from historic and real-time game data to esport audience.

Closely related to effective communication of insight is the task of maintaining viewership and active engagement of esports audience. One way currently explored by esports projects is to make consumption of esports content more enjoyable by providing different ways of viewing contents. Although studies such as [6] has shown that presenting viewers with cross-platform options for consuming content through mobile applications, Virtual Reality or via interactive overlays on live stream platforms can elicit engagement with esports content. Research has yet to systematically investigate what merit the contents and the different types of format in which they are delivered provide to audiences during and around the way they view esports games. For example, will spectators who are interested in viewing a particular esports match but unable to follow a live stream benefit from textual updates in a similar way to watching the event live. Extending the narrative components of the systems that deliver textual content to include NLG techniques for outputting summaries of in-game telemetry provided us with the opportunity to design experiments that mimic different viewing scenarios for viewers.

Over the last decade, there has been a sustained drive towards the improvement of representation learning for creating language models that can be used for text generation task. A growing body of research work such as [21] and [22] have shown that domain-specific natural language tasks benefit substantially from fine-tuning on massive pre-trained corpus of text. There is an agreement in these literature that the size of the training dataset required to fine-tune language models while comparatively fewer than end-to-end approaches still requires thousands or tens of thousands of examples. The data requirements of neural-based language models have constrained their usefulness to domains where labelled data are readily accessible and reduced adoption in niche real-world applications where labelled data may be limited data. Although Neural Language approaches have become more popular in current natural language research, they are typically characterized by unexplainable outcomes with a high computational cost.

In this work, we focus on text generation approaches that build grammar rules from text corpora. Our approach aligns closely with that of [23] who built surface realizers and grammar-based NLG from corpora. In addition to interpretability and traceability of the generated textual contents offered by these approaches, the need for massive labelled data and resource-dependent tools required by neural-based models were avoided. One drawback reported in the work of [23] was the huge number of mismatches of lexical items that they recorded due to the use of different corpora and general-purpose lexicons for their grammar-based NLG system. According to [23], this type of mismatch is one of the largest problems with the use of general-purpose knowledge resources created by different researchers for different purposes.

Using grammar rules derived from a corpus of professional Dota 2 commentaries and post-match analysis, we mapped non-linguistic game data to parsed commentary texts to generate recaps in Dota 2 matches. Corpus curated from commentaries have the distinct characteristic of preserving the choice of words and professional style of commentating by professionals like the group of analysts shown in figure 1(a). Consequently, the syntactic and lexical rules derived from such a corpus can potentially enhance the readability of text generated text to recap intense moment in the game such as figure 1(b).



(a) Dota 2 The International, CC BY 2.0, via Wikimedia Commons.

(b) A frame of a highlight moment in a Dota 2 match with video effect.

Figure 1: Esport Commentators and Analysts discussion interesting moments of matches

3. AUTOMATIC TEXT GENERATION

We posed the research questions in earlier sections to investigate the accuracy of generated text from a commentary corpus and how the text may support understanding of highlight moments. Central to this investigation is a rule-based NLG system that generates the text. A general overview of the system is presented in figure 2, with a text preprocessing corpus creation component, a grammar inducing component and a text generation component with surface realization engine that facilitates the text generation. Our method induces grammar in a similar way to the technique by [23] who derived Probabilistic Lexicalized Tree-Adjoining Grammars(PLTAG) from corpora and handcrafted examples.

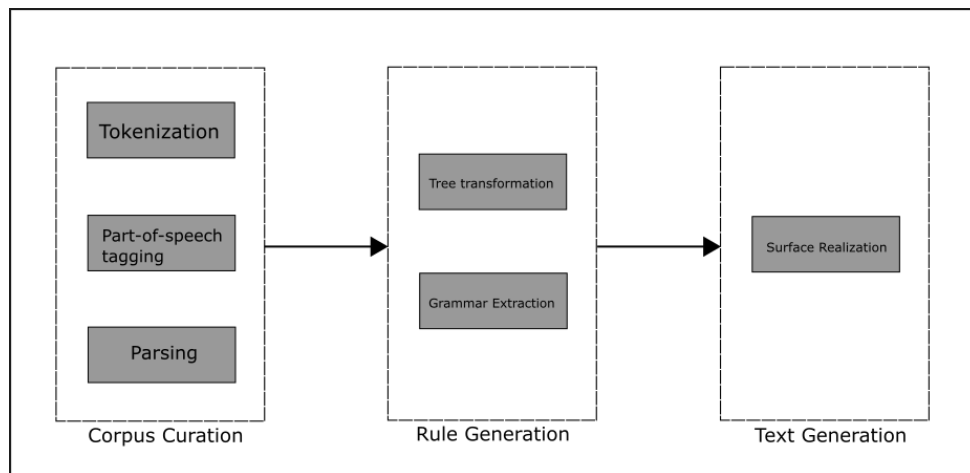


Figure 2: Overview of corpus-based NLG system

A major advantage of using PLTAG over other more popular grammars such as Context-Free Grammar (CFG) is that PLTAG has the property of being re-writable at non-terminals nodes [24]. This is particularly useful for augmenting -with more functional features- the grammars that are induced from corpora whose texts may not all be well-formed. Using Natural Language Tool Kit (NLTK), the transcribed commentaries are first pre-processed by splitting into sentences, annotated with part-of-speech tags and then parsed. The trees from the parsed sentences are then used to obtain probabilistic lexicalized tree-adjoining grammar rules for text generation. The task of an NLG system is to map from some input data to output text. To complete the text generation

task, input data from in-game events during the moment of highlights are first extracted. The trees that cover the input data are combined using unification and substitution techniques. The production rules that derive the combined trees containing all or most part of the input data are used to generate the text. Finally, we used simpleNLG [25] an off the self NLG toolkit to create and enforce all linguistic rules for the final text that describes the moment of highlight.

4. CORPUS CREATION

In the absence of naturally occurring corpus of human-written match highlights, since casters and analysts normally run commentaries orally we collected our own corpus of human transcribed audio post-match analysis. Researcher in natural language such as [26] argued that in the absence of linguistic rules both template-based and rule-based generation system can profit from the use of statistical information derived from corpora. Corpus curated from commentaries preserve the lexical and syntactic properties of words and sentences used by the casters and are natural to read and understand in contrast to other sources such as event logs or other human-written templates. To curate high-quality text from commentaries and post-match analysis of highlights moments, we considered only matches from professional tournaments under the Major championships category. We provide the details of the number of comments made by analysts and casters covering the selected matches in table 1. We selected 55 different professional Dota 2 matches across 5 tournaments in 2019 and 2020. First, we randomly selected the publicly available Dota 2 match highlights and powerplays from internet-based video on demand systems. We note here that powersplay analysis are short moments when analysts give post-match analysis on few selected moments of highlights. Comments made in powerplays analysis are considerable fewer than the comments made by casters during the live match.

Table 1: Summary of Dota 2 matches selected for corpus creation

	Tournament	Games	Teams	Comments
Powerplays	ESL One Mumbai 2019	6	several	154
	ESL One Katowice 2019	15	several	369
	ESL One Birmingham 2019	5	several	117
	ESL One Hamburg 2019	8	several	171
Commentaries	ESL One LA 2020	3	Nigma vs VP Prodigy	596
	ESL One LA 2020	2	Ehome vs RNG	605
	ESL One LA 2020	3	Fnatic vs Adroit	661
	ESL One LA 2020	2	OG vs Secret	766
	ESL One LA 2020	2	OG vs Vikin	330
	ESL One LA 2020	2	OG vs VP Prodigy	485
	ESL One LA 2020	4	RNG vs Newbee	542
	ESL One LA 2020	3	Secret vs Alliance	482

We selected powerplay analysis for all the 2019 tournaments and commentaries for all the 2020 tournaments. Five transcribers were paid to transcribe audio files extracted from the highlight moments into text documents with timestamps. One researcher with Dota 2 expert knowledge further improved the transcription quality by filling in all missing terms in the transcriptions.

5. EXPERIMENT DESIGN

Recent studies using quantitative and qualitative data from experiments on data-driven experiences for esports have shown that users want a holistic cross-platform experience [6]. We contribute to this finding by investigating if spectators that use text recaps to enhance their viewing experience benefit more than users that use the generated text recap in-place of viewing the highlights. To start our investigation, we asked *if text generated from the grammar induced from commentary corpus can accurately recap highlight moments in Dota 2 matches*. As a follow up question to study the utility of the generated text, we asked *if the text recaps generated by the derived grammar support understanding of highlight moments in Dota 2 matches*.

For our experimental setup, we collected quantitative data from 127 participants to address our question about linguistic quality of text generated from a corpus of commentaries and how the text generated support understanding of highlight moments in dota 2 matches. To test for evidence of highlight understanding, we created three different groups from the participants with three viewing condition to simulate different experiences. In the following section, we detail the choices we made for our design and the general procedure for our experiment.

5.1. Highlight Moment Selection

Esport tournament organisers typically organize the online broadcast of games into two streams to cover all the games leading to the final game of the competition. In order to evaluate generated text for high-quality moments of highlight, we selected the first game of the final event of the most recent professional competition at the time of conduction this experiment. Dota 2 is a multiplayer online battle arena (MOBA) game and as many other games in this genre players strategically build their heroes through the cause of a game in order to increase their teams win probabilities. Generally, there are three loosely defined phases of the game and many casters and analyst talk about highlight moments in each game based on how far it has progressed. The highlights in the first (early) phase typically progress for about 15 minutes according to [27], and it includes events where players are gaining experiences (XP) and golds mostly from NPC to unlock levels for their heroes. During the second phase of the game, highlights shifts to groups of players as they begin to purchase items from gold earned and engage in fights as a team. Highlights in the third phase are more coordinated team fights with mostly all players fully contributing towards defeating their opponents. In selecting highlight moments for evaluating our NLG system, we focus on the second phase of the game where events that have to be described in the text are within reasonable size for participants judging the generated text. We selected the first two highlights starting at minute 15:45 and 17:07 after the first phase of the game for investigating our research question on the accuracy of generated text and how they support understanding of the highlight moment.

5.2. Evaluating Generated Text

To evaluate the accuracy of generated text and their potential to support understanding of highlight moments, we align our approach with current research such as [28] and [3] who emphasized using human evaluation over automatic metrics. Human evaluation using Likert scale has been shown to give a better evaluating of text accuracy of NLG systems designed to generate short texts. Furthermore, we ensured that the quality of evaluation we get from the participants is not reduced by restricting our crowdsourcing platform to Prolific. Prolific has been reported by [29] to have a pool of participants that is more diverse and honest than other crowdsourcing alternatives such as Amazon Turk. We recruited 127 participants from Prolific after sorting

responses according to the primary inclusion criteria of being active Dota 2 players and the ability to read and write in English language.

We divided our online survey into 2 parts, with each part corresponding to the two research questions we were investigating. In the first part, all 127 participants were first asked to watch a video clip of approximately two minutes with a Dota 2 highlight moment. After seeing the video clip, all the participants were shown the automatically generated text recap for the highlight moment in the video clip. Multiple-choice questions asking participants to rate the text on a 5-point Likert scale were then presented to participants to provide quantitative data on accuracy and fluency. For example, on the linguist quality of accuracy of the text recap we asked “*How accurately do you think the generated text recaps the highlight in the video that you watched previously?*”

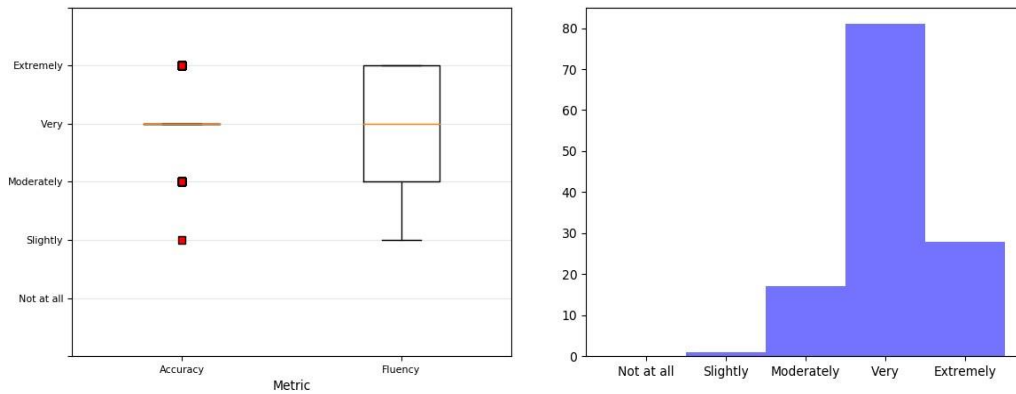
To evaluate the utility of the generated text in terms of support for the understanding of highlight, we divided the participants into three groups and presented each group with one of three conditions. The first group were shown a video clip only with a Dota 2 highlight moment and asked a set of multiple-choice questions about some of the events in the highlight. One of the questions for example was “*Who was the first hero to be killed in the highlight?*”. Results on data collected from participant response in this first group served as the baseline for assessing if there has been a gain in understanding of the highlight moment. The second group were shown the generated text only, while the third group were first shown the video clip followed by the generated text recapping the highlights. Both the second and third group were asked the presented with the same set of questions as the first group. In addition to the set of multiple-choice questions, a rank-based question was also asked of participants in all the three groups. We included a rank-based question based on the growing body of evidence such as in [30] and [28] that they provide a better way to access useful of texts generated from NLG systems.

6. RESULTS

In addressing our questions about automatically generated texts that recap esports highlights, our expectations from the results of our experiment were in two-fold. First, we expected our result on the linguistic quality of the generated text to align closely with outcomes of other work that reported high accuracy for templates and rule-based NLG approaches. Secondly, we expected that participants in our baseline group (Video only group) to perform better at recall tests than those in the group that received text only. Video recaps give extra cues such as visual and audio effects that may aid understanding. We expect that our result shows whether groups that read a text recap to enhance the viewing of video clip had a better performance on the recall test than the groups that either read the text recap only or viewed the video clip only.

6.1. Intrinsic Evaluation

In the first set of questions, our goal was to evaluate the NLG system using conventional intrinsic techniques where our human subjects read and rate the text generated after been shown a video highlight and its corresponding text recap. Debate continues about which property of text generated by an NLG system should be measured for linguistic quality. We followed best practices suggested by [28] for evaluating real-life of NLG system which placed text accuracy and fluency as top criteria for measuring the quality of the automatically generated text.



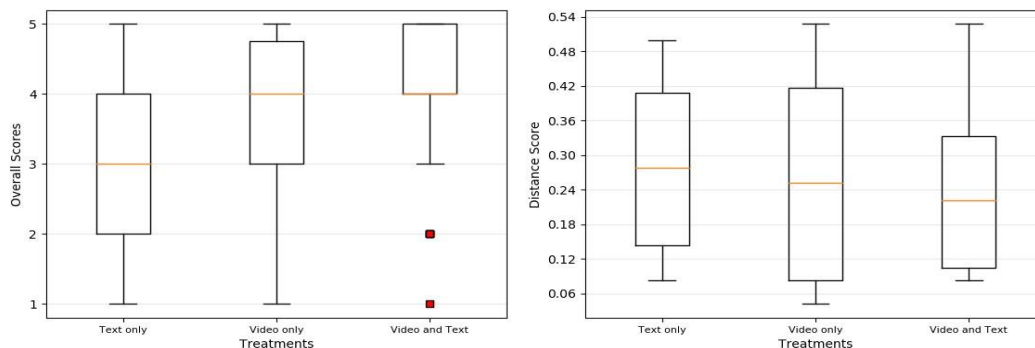
(a) Boxplot of responses to accuracy and fluency (b) Distribution of responses to questions on accuracy

Figure 3: Linguistic quality of generated text recaps by human evaluators

Looking at figure 3(a), it is clear from the small range in the boxplot for accuracy of generated text that the responses were more consistent and largely focused around the “*very accurate*” option. The data in figure 3(a) can be contrasted with data in chart of figure 3(b) where 64% of the 127 participants recruited selected the option of very accurate to indicate their perception of the accuracy of the generated text. The remaining 36% expressed their perception on the accuracy of text as “*extremely accurate 22%*”, “*moderately accurate 13%*” and “*slightly accurate*” at just about 1%. None of the participants selected the option of “*not accurate at all*”. On the question about fluency, the boxplot of figure 3(a) shows that the spread of participant response was not as consistent as responses for accuracy. The responses were skewed towards options that indicate the generated text have a higher degree of fluency.

6.2. Extrinsic Evaluation

In the second set of questions, we used extrinsic NLG evaluation techniques to measure the utility of the text generated in terms of how much they help support understanding of the match highlight event. Specifically, we presented the five task-based questions to the three groups of our human subjects and performed statistical analysis on the responses that we collected from the participants. First, we report the descriptive statistics of extrinsic methods in the boxplot of figure 4 and followed by statistical inferences using factorial design.



(a) Boxplot of recall test responses

(b) Boxplot of recall test responses with ranking

Figure 4: Recall test responses for tasks and ranking question

The range in the scores in the boxplots of figure 4(a) shows different responses to five task presented to participants in the three different groups. The spread of responses from the group that read the text recap only shows an even distribution around a median score of 3. Responses from the groups that received video only and those that received video and text both had a median score of 4 and a spread of responses skewed towards higher scores. The range of responses for the group that received both video and text was smaller than those that received video only and therefore more consistent. A factorial experiment (One-way ANOVA) was conducted to validate the significance of the differences between the three groups. There was a significant difference at $p < 0.05$ in responses of the participants. Our experiment gave a F-statistic ($F_{2,123}=3.535$) and a p-value ($p = 0.0322$) for the three groups. A post hoc Tukey test showed that the score for the group that watched the video and read a text recap was significantly higher than the group of participants that read the text recap only. According to the results of group differences obtained from the Tukey test, there was no significant difference between the baseline group of participants that watched the video highlight only and the other two groups.

To further test if the understanding of participants who read a text recap only differed significantly from those who watched the video highlight and read the text recap, we presented a task-based test that required participants to rank their response. A distance score was calculated based on how different responses of the group participants was from the expected answer. Specifically, we used the Jaro–Winkler distance as metric for measuring the difference between participants' rank sequence and the correct rank sequence. The lower the score the more similar the ranking is between the response and the correct sequence. The range of distance score is shown in figure 4(b), while the boxplot showed some difference in the median of distance scores for participants in the three groups, a one-way ANOVA could not validate the difference shown in figure 4(b).

7. DISCUSSIONS

Our focus in investigating if grammar rules from casters' commentary corpus can create texts that accurately recap moments of highlight in Dota 2 matches was not to establish the general validity. We wanted to extend prior findings such as [23] and [24] on the general use of rule-based systems to domain-specific (esport) text generation task and explore the use of curated commentary corpus as a viable linguistic resource for such purpose. The high rating in the quantitative data of our study for accuracy and fluency from the first part of our experiment addresses the question about the linguistic quality of generated text recap from commentary corpus.

The result of our study did not find any evidence to validate our second hypothesis, where we asked if text-based recaps can supports understanding of highlight moment in Dota 2 match. The lower performance of participants who received text only in the recall tests is interesting, but not surprising. While the performance of the “*text-only*” group was less than the groups that viewed video highlights, we recognise a value for text-only recaps in the same way large-scale esport project like [6] identified that there are some circumstances where people are unable to view video and therefore may benefit from text recaps (e.g. during transit or when working and trying to keep up with a tournament). For these cases, the text recaps which participants perceived to be accurate in the first part of the study offer good value to esport spectators. In the following section, we discuss the findings of our study on NLG systems that generate text recaps with respect to their implications for real-life applications in esport.

7.1. Text Recap and Highlight Accuracy

One prospective benefit of generating text-based recaps using an NLG system is that they make match data more explicable and therefore more promotable. This means that text-based recaps can contribute to the democratisation of the performance data that esports content producers want to communicate to all their audiences regardless of skill level. According to [31], in many commercial NLG applications, it is essential that the system produce reasonable texts in all scenarios. A system that generates a text summary of a player's performance/skill from non-linguistic data should not produce incorrect text that may negatively impact their skill level. Building applied NLG systems should involve testing processes that evaluate the systems for performances in both expected outcomes and edge-case scenarios. In addressing the first question about linguistic quality of the generated, the high concentration of participant choice (64%) around the "very accurate" option on our 5-point Likert scales aligns with results in rule-based NLG literature such as [32] and [33] where templates and hand-engineered systems generally resulted in semantically accurate text than systems that use machine learning approaches like neural text generation. The finding on the accuracy of the generated text recaps contrast with our finding on participant response for the fluency question, where the largest response was also for "very fluent" but only by 38% of the participants. This is however consistent with recent large-scale NLG system evaluation by [34], involving 20 different NLG systems. The evaluation of the shared NLG task found that systems that produce semantically accurate text often suffer from lack of naturalness (fluency and readability) in the text if controls are not applied before the generation process. Investigation of NLG metric by [32] suggested that correct semantic accuracy is regarded by users as more important than fluency and should be prioritised when training the models.

Accuracy of text is particularly important in building applied systems for domains where texts are generated to reduce complexity and to provide simple and clear information to audiences. Consider esports where a large-scale case study of a production tool called Echo was deployed. Echo uses live and historic match data to detect extraordinary player performances in the popular esports Dota2 and dynamically translates interesting data points into audience-facing graphics superimposed with text. Echo was deployed at one of the largest yearly Dota 2 tournaments, which was watched by 25 million people. The creator of Echo [8] analysed 40 hours of video, 46,000 live chat messages, and feedback of 98 audience members and showed that Echo measurably affected the range and quality of storytelling, increased audience engagement, and invoked rich emotional response among viewers. In one instance [8] reported that a small number of twitch users picked part of the text in the graphic and expressed concern with a phrase that casters had relayed as a global record instead of a personal record that referenced a single player's history. [8] reported that there were only three total instances among 5091 chat posts and their evidence suggests that a small number of users do inspect little details in the small texts, the query of the validity of the details. According to [8], they included records of matches after a major patch (update that changes the game) because the patch significantly impacted many performance features that were tracked by Echo. This made it questionable to compare data collected before and after the patch.

Patch updates and how they impact on the performance of players is often an issue that is discussed by casters during the game and by analysts during the post-match analysis. Extracting grammar rules from corpus curated from transcripts of previous caster and analyst commentaries in a similar way to our methodology can help produce a textual explanation of player performance that account for the type of edge case that appeared in Echo's graphics. Esports data analytic tools such as Echo can be developed into second-screen companion apps with a functional NLG module that handles accurate processing of historic and real-time non-linguistic performance data and generation of textual messages for esports audience.

7.2. Text Recap and Highlight Understanding

According to [2], the video games being played in eSports e.g. Dota 2, are usually complex and require a considerable amount of concentration to comprehensively follow the game. While the simplicity that comes with the descriptions of the text may bring to focus the outcomes of key events during a highlight moment, they lack the visual and audio signals that help the audience understand more of what is happening. As reflected in the better recall test scores of participants that saw both video highlight and text recap, a case to improve audience understanding of Dota 2 can be made to take advantage of both the simplicity of the generated text and additional cues from video by bundling them together when recapping highlight moment. One possible application scenario of this finding is the addition of an explainer feature to already existing systems such as the WEAVR app in [6]. The WEAVR app annotates Dota 2 maps with simple explanations of the highlights and stats from in-game data. Enhancing these annotations with engaging text that explains the context of these highlights to new players of the game makes the application of our findings in this part of our study an interesting one.

The continuous improvements in both hardware and middleware used for delivering online broadcasting remain the driving force for esports and video game streaming as a new form of online media consumption. Acquisition of knowledge related to an esports game has been shown by [2] to contribute the most to the frequency of watching esports. The audience of esports who are presented with simple information about intense moments in a game in a way that supports acquiring knowledge such as text recaps of highlight moments are therefore less likely to be part of the churn that typically occur at the onset of mastering a game. In the face of competition for viewership, the esports content producers trying to stand out and attract audiences to their platform can take advantage of the knowledge acquisition capability that generated text recaps can enable when superimposed on a video highlight. An NLG system that is well integrated and adequately recap highlights of visually complex moments in esports games can help increase audience engagement and loyalty.

In addition to the growing online viewership, the current decade-leap projected by [35] and [36] in the consumption of digital media due to the global pandemic may provide the spotlight that esports needs to encourage broader participation from mainstream television broadcasting. One of the ways the result of this study will benefit adoption by mainstream TV is the ease of integration with the way mainstream broadcasting deliver headlines and breaking news. Many major Mainstream TV channels have decades of experience at relaying details of breaking news by presenting the information as new tickers scrolling across the area at the bottom of the screen. Automatically generated text recapping complex in-game events can be scrolled as breaking stories for the time the video recap is being broadcast on the screen.

7.3. Limitation

As an alternative approach to a neural text generation, our grammar-based approach has avoided the need for large labelled data required by neural network to generate accurate text. However, the process of obtaining the generation grammar from the commentary corpus meant that all audio recordings of the commentaries had to be transcribed first. Gameplay in Dota 2 is extremely complex [8] and as reported by [37], Dota 2 lies on the threshold between excitement and difficulty. Dota 2 require a lot of domain knowledge and expertise beyond the skill of a proficient transcriber. As result, the process of curating a commentary corpus was a resource-intensive task requiring a two-tiered (non-expert and expert) approach to improving the transcribed commentaries. It will be interesting to see if this task can be less intensive when state-of-the-art speech-to-text machine learning models such as reported in [38] are used for transcribing commentaries.

8. CONCLUSIONS

In this paper, we have induced rules from a collection of annotated esports commentary for the task of generating accurate text that recaps highlight moments in Dota 2 matches. The experiments in our study have shown that viewers perceive rule-based text recaps as accurate and fluent accounts of highlight moments in Dota 2 matches. Comparisons between three groups of participants were made to test the utility of text-based recaps implemented using an NLG system. Using recall tests, we observed significant improvements in the score of participants that viewed video highlights and its text recap over the groups that viewed the text recap only. Consequently, we can conclude from our results that esports commentaries are a viable linguistic resource for generating accurate text recaps that can support understanding of the complex actions in Dota 2 video highlights. Further to recapping video highlights, the NLG system used for the text-based recaps can also add to spectator experiences by making lexical choices according to the skill and preference of the viewer. We intend to explore this variation of the NLG system for creating personalised text recaps in future work.

ACKNOWLEDGEMENTS

This work has been created as part of the Weavr project (weavr.tv) and was funded within the Audience of the Future programme by UK Research and Innovation through the Industrial Strategy Challenge Fund (grant no.104775) and supported by the Digital Creativity Labs (digitalcreativity.ac.uk), a jointly funded project by EPSRC/AHRC/Innovate UK under grant no. EP/M023265/1.

REFERENCES

- [1] Lucas Kempe-Cook, Stephen Tsung-Han Sher, and Norman Makoto Su, (2019) “Behind the voices: The practice and challenges of esports casters”. In Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos, editors, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019*, pp565.
- [2] Juho Hamari and Max Sjoblom, (2017) “What is esports and why do people watch it?”, *Internet Research*, Vol. 27, No. 2, pp211-232.
- [3] Craig Thomson and Ehud Reiter, (2020) “A gold standard methodology for evaluating accuracy in data-to-text systems”, *ArXiv preprint arXiv:2011.03992*.
- [4] Jason G. Reitman, Maria J. Anderson-Coto, Minerva Wu, Je Seok Lee, and Constance Steinkuehler, (2020) “Esports research: A literature review”, *Games and Culture*, Vol. 15, No. 1, pp32–50.
- [5] Douglas Heaven, (2014) “Rise and rise of esports”, *New Scientist*, Vol. 223, No. 2982, pp17.
- [6] Athanasios Vasileios Kokkinakis, Simon Demediuk, Isabelle Nolle, Oluseyi Olarewaju, Sagarika Patra, Justus Robertson, Peter York, Alan Pedrassoli Pedrassoli Chitayat, Alistair Coates, Daniel Slawson, Peter Hughes, Nicolas Hardie, Ben Kirman, Jonathan Hook, Anders Drachen, Marian F Ursu, and Florian Block, (2020) “Dax: Data-driven audience experiences in esports”. In *ACM International Conference on Interactive Media Experiences, IMX '20*, pp94–105.
- [7] Thomas Horky and Philipp Pelka, (2017) “Data visualisation in sports journalism”, *Digital Journalism*, Vol. 5, No. 5, pp587–606.
- [8] Florian Block, Victoria Hodge, Stephen Hobson, Nick Sephton, Sam Devlin, Marian F. Ursu, Anders Drachen, and Peter I. Cowling, (2018) “Narrative bytes: Data-driven content production in esports”. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video, TVX '18*, pp29–41.
- [9] Christoph Eggert, Marc Herrlich, Jan Smeddinck, and Rainer Malaka, (2015) “Classification of player roles in the team-based multi-player game dota 2”. In Konstantinos Chorianopoulos, Monica Divitini, Jannicke Baalsrud Hauge, Letizia Jaccheri, and Rainer Malaka, editors, *Entertainment Computing - ICEC 2015*, pp112–125.
- [10] Simon Demediuk, Peter York, Anders Drachen, James Alfred Walker, and Florian Block, (2019) “Role identification for accurate analysis in dota 2”. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 15, No. 1, pp130–138.

- [11] K. Conley and D. Perry, (2013) “How Does He Saw Me? A Recommendation Engine for Picking Heroes in Dota 2”, Available: <http://cs229.stanford.edu/proj2013/PerryConley-HowDoesHeSawMeAREcommendationEngineForPickingHeroesInDota2.pdf>, [Accessed: 10-Oct-2020].
- [12] Tobias Mahlmann, Matthias Schubert, and Anders Drachen, (2016) “Esports analytics through encounter detection”, In *Proceedings of the MIT Sloan Sports Analytics Conference. MIT Sloan*.
- [13] Wei-Ta Chu and Yung-Chieh Chou, (2015) “Event detection and highlight detection of broadcasted game videos”. In *Proceedings of the 2nd Workshop on Computational Models of Social Interactions: Human-Computer-Media Communication, HCMC '15*, pp1–8.
- [14] Yale Song, (2016) “Real-time video highlights for yahoo esports”, *ArXiv*, Vol. abs/1611.08780.
- [15] Charles Ringer and Mihalis A. Nicolaou, (2018) “Deep unsupervised multi-view detection of video game stream highlights”. In *Proceedings of the 13th International Conference on the Foundations of Digital Games, FDG '18*.
- [16] Justus Robertson and Michael Young, (2014) “Gameplay as on-line mediation search”, *Experimental AI in Games*.
- [17] Shaowen Bardzell, Jeffrey Bardzell, Tyler Pace, and Kayce Reed, (2008) “Blissfully productive: Grouping and cooperation in world of warcraft instance runs”. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW '08*, pp357–360.
- [18] Yubo Kou and Xinning Gui, (2014) “Playing with strangers: Understanding temporary teams in league of legends”, *CHI PLAY*, In *Proceedings of the 2014 Annual Symposium on Computer-Human Interaction in Play*, pp161–169.
- [19] Marco Furini and Manuela Montangero, (2016) “TsentiMENT: On gamifying twitter sentiment analysis”. In *2016 IEEE Symposium on Computers and Communication (ISCC)*, pp91–96.
- [20] Yoonsung Hong, Haewoon Kwak, Youngmin Baek, and Sue Moon, (2013) “Tower of babel: A crowdsourcing game building sentiment lexicons for resource-scarce languages”, In *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*, pp549–556.
- [21] Zhiyu Chen, Harini Eavani, Wenhui Chen, Yinyin Liu, and William Yang Wang, (2020) “Few-shot nlg with pre-trained language model”, *CoRR*, Vol. abs/1904.09521.
- [22] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei, (2020) “Language models are few-shot learners”, *ArXiv*, Vol. abs/2005.14165.
- [23] Huayan Zhong and Amanda Stent, (2005) “Building surface realizers automatically from corpora using general-purpose tools”. In *Proceedings of UCNLG'05*, pp49–54.
- [24] Trevor Cohn, Sharon Goldwater, and Phil Blunsom, (2009) “Inducing compact but accurate tree substitution grammars”. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp548–556.
- [25] Albert Gatt and Ehud Reiter, (2009) “Simplenlg: A realisation engine for practical applications”, In *Proceedings of the 12th European Workshop on Natural Language Generation, ENLG '09*, pp90–93.
- [26] Kees Van Deemter, Emiel Krahmer, and Mariet Theune, (2005) “Real versus template-based natural language generation: A false opposition?” *Comput. Linguist.*, Vol. 31, No. 1, pp15–24.
- [27] Kevin Godec, (2011) “Dota 2 guide: Welcome to dota, you suck”, *Purge gamers*, Available: <https://purgegamers.true.io/g/dota-2-guide>, [Accessed: 05-Nov-2020].
- [28] Chris Van der Lee, Albert Gatt, Emiel van Miltenburg, Sander Wubben, and Emiel Krahmer, (2019) “Best practices for the human evaluation of automatically generated text”. In *Proceedings of the 12th International Conference on Natural Language Generation*, pp355–368.
- [29] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti, (2017) “Beyond the turk: Alternative platforms for crowdsourcing behavioral research”, *Journal of Experimental Social Psychology*, Vol. 70, pp153 – 163.
- [30] Jekaterina Novikova, Ondřej Dusek, and Verena Rieser, (2018) “RankME: Reliable human ratings for natural language generation”. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 2, pp72–78.

- [31] Ehud Reiter and Robert Dale (2002). Building applied natural language generation systems. *Natural Language Engineering*, Vol. 3, No. 1, pp57-87.
- [32] Ehud Reiter and Anja Belz, (2009) “An investigation into the validity of some metrics for automatically evaluating natural language generation systems”, *Computational Linguistics*, Vol 35, No. 4, pp529–558.
- [33] Sam Wiseman, Stuart Shieber, and Alexander Rush, (2017) “Challenges in data-to-document generation”. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp2253–2263.
- [34] Ondřej Dusek, Jekaterina Novikova, and Verena Rieser, (2020) “Evaluating the state-of-the-art of end-to-end natural language generation: The e2e nlg challenge”, *Computer Speech & Language*, Vol. 59, pp123 – 156.
- [35] Bidit L. Dey, Wafi Al-Karaghoul, and Syed Sardar Muhammad, (2020) “Adoption, adaptation, use and impact of information systems during pandemic time and beyond: Research and managerial implications”, *Information Systems Management*, Vol. 37, No. 4, pp298–302.
- [36] Netta Iivari, Sumita Sharma, and Leena Venta-Olkkonen, (2020) “Digital transformation of everyday life – how covid-19 pandemic transformed the basic education of the young generation and why information management research should care?”, *International Journal of Information Management*, Vol. 55, pp102183.
- [37] Nick Yee, (2016) “Game Genre Map: The Cognitive Threshold in Strategy Games”, *Quantic Foundry*, Available: <https://quanticfoundry.com/2016/01/20/game-genre-map-the-cognitive-threshold-in-strategy-games/>, [Accessed: 05-Oct-2020].
- [38] Alexei Baevski, Henry Zhou, Abdelrahman Mohamed and Michael Auli, (2020) “wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations”, *ArXiv preprint arXiv:2006.11477*.

AUTHORS

Oluseyi Olarewaju is a Postdoctoral Research Associate in the Department of Theatre, Film, Television and Interactive Media at the University of York. UK. He is a member of Weavr within Digital Creativity Labs. Oluseyi completed his PhD in Computer Science at University of Brighton focusing on Personalization Systems. His research interests are in Cross-domain Recommender Systems, Predictive Modelling and Natural Language Processing. Oluseyi is currently exploring the application of computational linguistics for generating personalized stories from game.



Jon Hook is a Lecturer in Interactive Media at the University of York’s Theatre, Film, TV and Interactive Media department. He has a background in computing, having done a PhD in Human-Computer Interaction at Newcastle University. Jon’s research explores the design and development of novel interactive technologies for a broad range of artistic and everyday creative practices. In the past he’s investigated the design of technologies that help live performers, digital artists, design educators and people with disabilities be creative in new ways. In DC Labs he’s continuing this work through collaborations with high-profile partners in interactive television, theatre performance and digital heritage.



PROFILING NVIDIA JETSON EMBEDDED GPU DEVICES FOR AUTONOMOUS MACHINES

Yazhou Li¹ and Yahong Rosa Zheng²

¹School of Computer Science and Engineering, Beihang University, Beijing, China

²Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, 18015, USA

ABSTRACT

This paper presents two methods, tegrastats GUI version jtop and Nsight Systems, to profile NVIDIA Jetson embedded GPU devices on a model race car which is a great platform for prototyping and field testing autonomous driving algorithms. The two profilers analyze the power consumption, CPU/GPU utilization, and the run time of CUDA C threads of Jetson TX2 in five different working modes. The performance differences among the five modes are demonstrated using three example programs: vector add in C and CUDA C, a simple ROS (Robot Operating System) package of the wall follow algorithm in Python, and a complex ROS package of the particle filter algorithm for SLAM (Simultaneous Localization and Mapping). The results show that the tools are effective means for selecting operating mode of the embedded GPU devices.

KEYWORDS

Nvidia Jetson, embedded GPU, CUDA, Automous Driving, Robotic Operating Systems (ROS).

1. INTRODUCTION

NVIDIA Jetson is a complete embedded system-on-module (SoM) device that integrates the CPU, GPU, PMIC, DRAM, and flash storage on a small-form-factor platform. The current Jetson series include Jetson Nano, Jetson TX2, and Jetson Xavier (NX and AGX), which are commonly used for edge computing, autonomous machines, machine learning, and artificial intelligence. An example application is the toy race car F1/10 which is of 1/10 of a real car size (<https://f1tenth.org>) and uses a Jetson TX2 as its computing hardware, as shown in Fig. 1, where the Connect Tech Obitty Carrier board is on top of the Jetson TX2 module, the power distribution board is on top of the VESC motor controller, and the Lidar is in the front of the car. All electronics are mounted on a plate above the brushless motor and battery compartment, making it a compact and powerful platform for testing autonomous driving algorithms. The Jetson device can connect to host computer via WiFi antennas and a dedicated WiFi router with static IP address assignment is recommended.

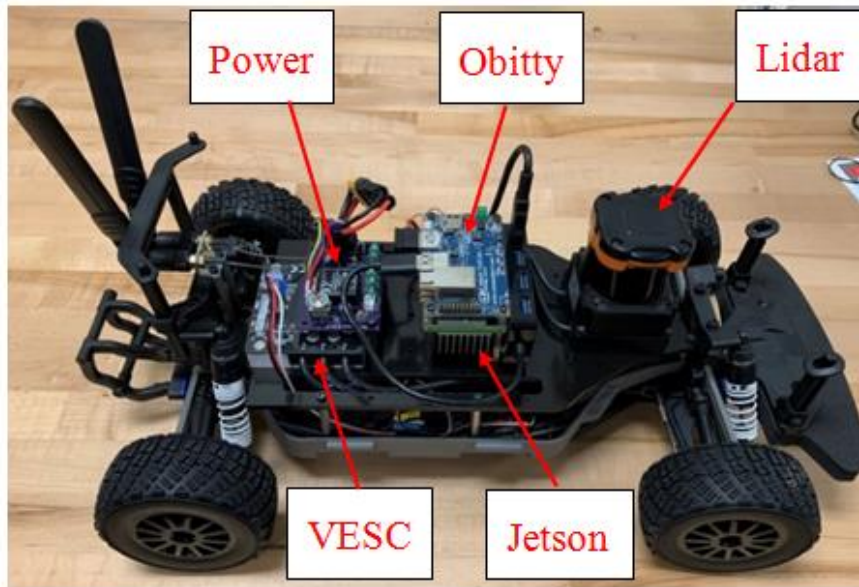


Figure 1. The F1/10 race car with Nvidia Jetson TX2, Connect Tech Obitty Carrier board, Hokuyo UMT-30 Lidar, and VESC 6 MK-III motor controller.

With the embedded GPUs, the NVIDIA Jetson systems provide the performance and power efficiency to run autonomous machines software faster and with less power than CPU-only embedded systems. However, how do we profile the performance of the Jetson devices? How do we use the profiling results to help improve programs written for Jetson devices? Answering these questions requires some serious effort because Jetson embedded devices require special versions of profilers than the commonly used `nvidia-smi` utility and Nsight Systems for desktop or workstations.

In this paper, we explore two methods for profiling the performance of Jetson TX2 8GB: one is the `tegrastats` utility and its graphical APIs which can be used directly on the Jetson device; one is Nsight Systems for Tegra target systems which is used on a host computer to remote access the Jetson device.

Several coding examples are used to bench mark the performance of Jetson TX2 under different working modes: a vector-add program in C or CUDA C, a wall-follow python program for f1/10 race cars [1], and a particle filter algorithm for Simultaneous Localization and Mapping (SLAM) [2]. The profilers measure the power consumption, the run time of the CUDA C threads, and the CPU/GPU utilization under five operating modes of the Jetson TX2. The program files used in this paper can be found at <https://github.com/li630925405/jetson-profile>.

1.1. Jetson TX2 Working Modes

The Jetson TX2 consists of a 256-core Pascal GPU along with a CPU cluster. The CPU cluster consists of a dual-core NVIDIA 64-bit Denver-2 CPU and a quad-core ARM Cortex-A57. By configuring the 6 CPU cores and the GPU, a Jetson TX2 typically runs in five different modes, as shown in Table 1, where the GPU is enabled in all five modes, but at different clock speeds. The two types of CPUs are enabled or disabled in different configurations. Different modes show different speed-power performance trade-offs. It is clear that the Max-N mode is the fastest and consumes the most power as all CPUs and GPU are enabled and they run at their maximum speeds.

Table 1. Working modes of typical Jetson devices configurable by the NVPmodel utility [3].

Mode	Mode Name	Denver 2 CPU Core		ARM A57 Core		GPU Frequency
		# of Cores	Frequency	# of Cores	Frequency	
0	Max-N	2	2.0 GHz	4	2.0 GHz	1.30 GHz
1	Max-Q	0		4	1.2 GHz	0.85 GHz
2	Max-P Core-All	2	1.4 GHz	4	1.4 GHz	1.12 GHz
3	Max-P ARM	0		4	2.0 GHz	1.12 GHz
4	Max-P Denver	1	2.0 GHz	1	2.0 GHz	1.12 GHz

The Jetson operation modes are enabled by the Dynamic Voltage and Frequency Scaling (DVFS) technology and can be configured at run time by a command-line tool NVPModel [3]. For example, we use 'sudo nvpmode -m 2' to change the working mode of a Jetson to Mode 2 Max-P Core-All mode. The configuration of the five modes are saved in the file /etc/nvpmode.conf which can also be customized to produce user flexible modes. The current mode of the Jetson is queried by 'sudo nvpmode -q -verbose'.

1.2. NVIDIA Profiling Tools

NVIDIA profiling tools help the developers to optimize their programs and applications. The newly announced NVidia Nsight Systems [4], [5] supersedes the command-line nvprof and visual profiler NVVP tools and combine them into one unified tool. To apply it to profile Jetson devices, NSight Systems for Tegra target system package has to be used on the host computer which remote accesses the jetson device profilee. The package is part of the JetPack 4.3 installation [6] and is installed on the host computer. Alternatively, a command-line utility 'tegrastats' and its graphical APIs [7], [8] are available to profile Jetson devices directly. In addition, user-defined functions or python wrappers may utilize the system APIs such as timer to profile the performance.

```
robot-1@robot-1:~$ tegrastats
RAM 2268/7861MB (1fb 49x4MB) SWAP 0/3930MB (cached 0MB) CPU [6%@345,off,off,2%@345,6%@345,5%@345] EMC_FREQ 0% GR3D_FREQ 0%
PLL@39C MCPU@39C PMIC@100C Tboard@36C GPU@38C BCPU@39C thermal@38.6C Tdiode@36.25C VDD_SYS_GPU 143/143 VDD_SYS_SOC 286/286
VDD_4V0_WIFI 305/305 VDD_IN 1909/1909 VDD_SYS_CPU 95/95 VDD_SYS_DDR 229/229
```

Figure 2. profile jetson GPU status using tegrastats

The tegrastats utility reports memory usage and processor usage for Jetson-based devices [7] similar to the Nvidia-smi utility which is not supported on Jetson. An example of tegrastats is shown in Figure 2, which means CPU2 and CPU3 are off, CPU1, CPU5 and CPU6 are using 6% of their loads, CPU4 is using 2% of its load and their current running frequency is 345 MHz. The details of the tegrastats output is explained in [7], but it is rather user unfriendly. The tegrastats output can be visualized by a tool called jetson_stats [8], a package that combines both tegrastats and NVPmodel into a GUI to profile and control Jetson. It contains five different tools, among which the jtop tool is the most useful for profiling.

Nsight Systems for Tegra targets is a part of the JetPack SDK [9] for Jetson devices and is installed on a host computer to remotely profile the Jetson target device. Nsight Systems is a low-overhead sampling, tracing and debugging tool for C or CUDA C programs and may not be effective to profile a python program.

Therefore, we rely on the system APIs such as high resolution clock in the C++ library and `time.time()` in the python library and write custom functions and evaluate the performance of python code on Jetson devices.

2. HARDWARE PLATFORM AND SAMPLE PROGRAMS

The Jetson TX2 device in the F1/10 race car platform [1] is used to profile the example programs. The Jetson TX2 is flushed with the Ubuntu 18.04.4, ROS Melodic, and Jetpack 4.3 L4T 33.3.1 packages. The F1/10 race car ROS simulator is also installed. We profile three examples on the f1/10 race car platform: Example 1 is the vector-add program in two versions [10]: `vector_add.cpp` and `vector_add.cu`. The `cpp` version uses the CPU cores only, while the `cu` version explicitly utilizes the GPU parallel programming APIs and memory management via CUDA C/C++.

Example 2 is a wall follow python program [11] running on the f1/10 simulator platform [1]. As shown in Fig. 3, the simulator takes the map and drive information to generate the Lidar sensing data. It also interfaces the joy pad or keyboard inputs for control and runs the autonomous algorithms via the New Planners module and the `/nav` and `/brake` topics. The RViz package is used to visualize the race track, the car motion, and the Lidar sensing data.

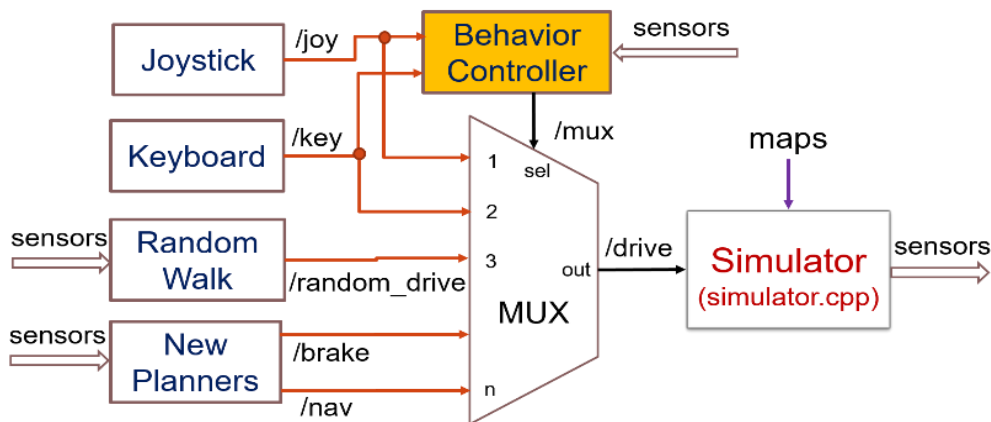


Figure 3. Block diagram of the F1/10 race car ROS simulator.

The wall follow python program subscribes to the LidarScan data and computes the distance of the race car to the left or right wall, then determines the PID control parameters according to a pre-set distance, and publishes the Ackermann steering parameters to the `/drive` topic to control the race car. When the car receives data from the lidar, the function `lidar_callback` will be called to calculate the control signal and send it to control the car. So the speed of wall follow program can be represented by time consumed by `lidar_callback`.

Example 3 is a particle filter algorithm for SLAM and is available at github [2]. The python code implements a fast particle filter localization algorithm and uses the RangeLibc library for accelerated ray casting. By specifying the range method "rmgpu", the program explicitly use the GPU to achieve fast ray casting, while other range methods, such as "bl", "rm", and "glt", etc., use the CPUs only. The bench mark performed in C++ is reprinted in Table. 2. The range method options can be specified in `localize.launch` of the particle filter package.

Table 2. Different ray casting algorithms in the particle filter example [2].

Method	Init time (sec)	Random queries Thousand/ sec	Grid queries Thousand/ sec	Memory (MB)
BL	0.02	297	360	1.37
RM	0.59	1,170	1,600	5.49
RMGPU	0.68	18,000	28,000	5.49
CDDT	0.25	2,090	4,250	6.34
PCDDT	7.96	2,260	4,470	4.07
LUT	64.5	2,160	4,850	296.6

3. TEGRASTATS PROFILING PROCEDURES AND RESULTS

The performance of Example 1 is profiled by three methods: the average time to run the program is extracted by the `clock()` function in `cpp` and `cu` versions of the `vector_add` programs; the `jetson-stats` is used to extract the `tegrastats` results on CPU/GPU utilization and power consumption and graph them; and `Nsight Systems` is used to trace the APIs of GPU synchronization and memory management in details.

The performance of Example 2 is evaluated by two methods: the first is to utilize a python wrapper with python library function `time.time()` to calculate time spent for each function in the python program; the second is to use `jetson-stats` to graph the CPU usage and power consumption. The wall-follow program runs for minutes and the python wrapper outputs more than 60,000 calls of each function and the time of each function run is averaged over the total numbers of the calls. The python wrapper is disabled when using `jetson-stats` to profile Example 2.

Example 3 has a built-in performance calculator and it outputs the average number of iterations of ray casting in the particle filter algorithms. The `jetson-stats` is used to profile its power consumption and CPU/GPU utilization.

3.1. Example 1: Vector Add

The two versions of `vector-add` programs are run on Jetson TX2 in modes 0 -- 4, respectively. Although all modes have the GPU enabled, only the `vector_add.cu` makes use of the GPU, while `vector_add.cpp` utilizes the CPU cores only. The results of the execution time is shown in Table 2, where the time is measured by high resolution clock from the C++ library "`chrono`". Without utilizing the GPU, Modes 0, 2 and 4 run in similar speeds which are faster than Modes 1 and 3. This is because Modes 0, 2 and 4 use the Denver-2 and Modes 1 and 3 use ARM cores only. The example code is a single-threaded kernel and Denver 2 CPU has better single-threaded performance than the ARM core.

Table 3. Power consumption of Example 1 in different Jetson modes.

Mode	GPU	CPU power	GPU power	IN power
0	No	1983	95	7151
0	Yes	567	2315	8229
1	No	524	95	2527
1	Yes	190	1045	5155
2	No	1426	95	4660
2	Yes	379	1706	6568
3	No	1237	95	4187
3	Yes	331	1707	6663
4	No	1900	95	5108
4	Yes	521	1706	6682

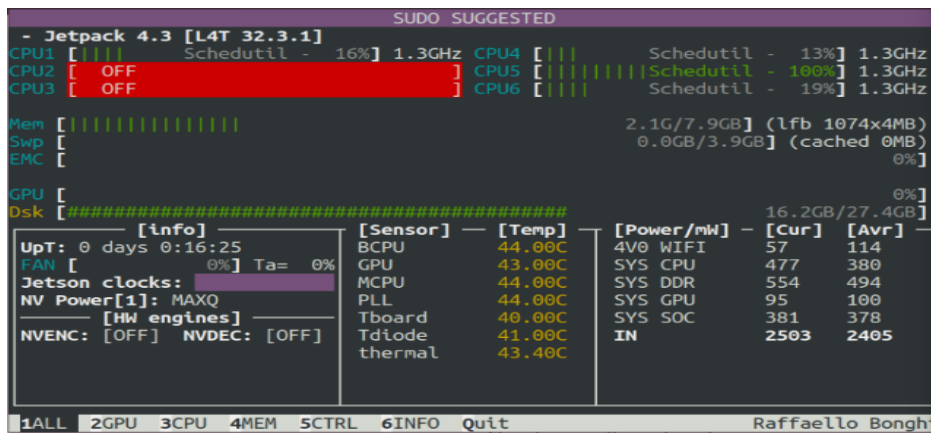
Utilizing the GPU, all five modes run at the similar speed as most of the processing is offloaded to the GPU running at similar speeds in all modes. Comparing the performance between the CPU+GPU and the CPU only modes, performances vary significantly in different GPU execution configurations, as shown in Table 3. Note that the streaming multiprocessors in Jetson TX2 is 2 and the maximum number of thread per block is 1024. Also note that the size of the vectors to be added is 224. With a small number of thread per block and a small number of blocks per grid, the performance of the GPU version is worse than the CPU only version, as the potential of the GPU is under utilized and the overhead of memory management is relatively large. Let the number of threads per block and the number of block per grid be N_t and N_b , respectively. If we increase N_t or N_b such that $N_t N_b \geq 211$, then the GPU performs faster than the CPU only version, as shown in Table 4. Therefore, the try and error method is used to find out the best combination of the numbers of threads and blocks.

Table 4. Time needed to run Example 1 with different CUDA execution configurations

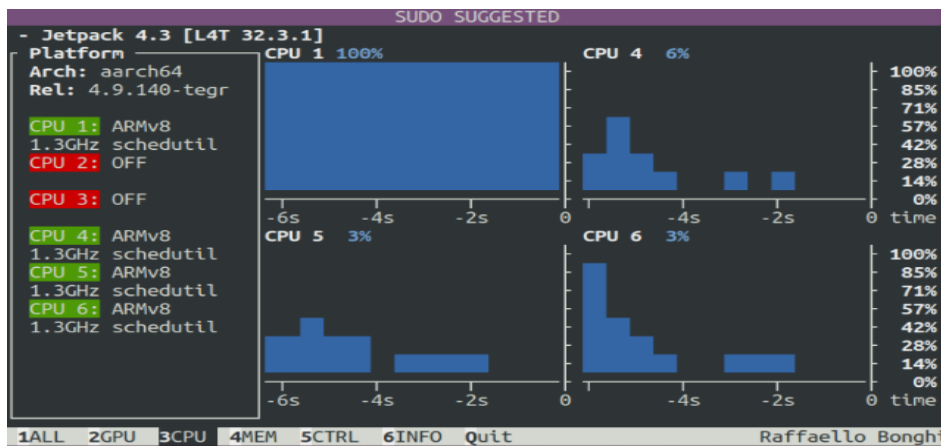
# threads / block	# blocks /grid	run time (s)
32	2	240
128	2	71
1024	2	15
32	64	15

The two versions of `vector_add` programs are also profiled by `tegrastats` via `jtop` and the results for Mode 1 (Max-Q) are shown in Fig. 4 and 5. Mode 1 uses all four ARM cores at 1.2 GHz clock rate while the GPU can run at 0.85 GHz speed. The CPU version of the program utilizes all four ARM cores and its GPU utilization is 0%, as shown in Fig. 4a. CPU 2 and 3 are off as they are the Denver 2 cores which are disabled in Mode 1. The average power consumption of the program in Mode 1 is 2405 miliwatts. It is interesting to note that the GPU still consumes 100 mW power on average even though the program does not utilize the GPU. This is because the GPU is used by Jetson to support graphics in the operating system.

The CPU utilization is further detailed by going to menu 3CPU at the bottom of the `jtop` window, as shown in Fig. 4b, where the time snap shots of the CPU utilization is captured. All four ARM cores run at the 1.3 GHz clock rate and CPU 1 is utilized fully at 100% at the time of the capture, while other three ARM cores are used occasionally. Observing the CPU graphs over time reveals that the four ARM cores are utilized uniformly by the "schedutil".



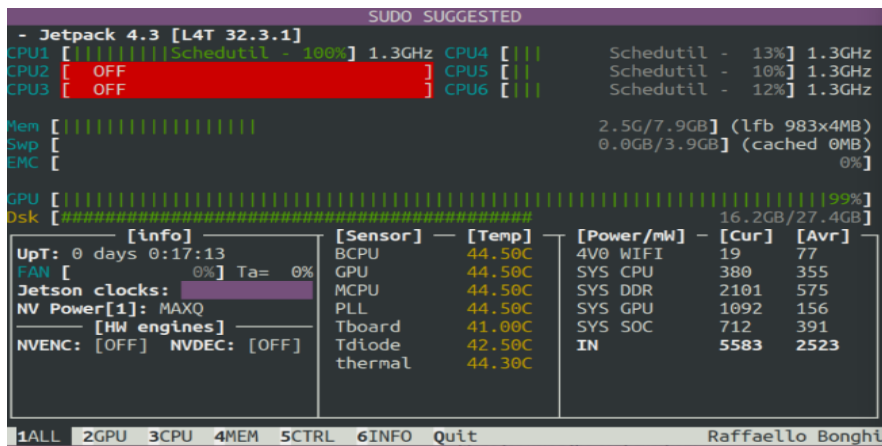
(a) Overall performance of vector_add.cpp w/o using the GPU



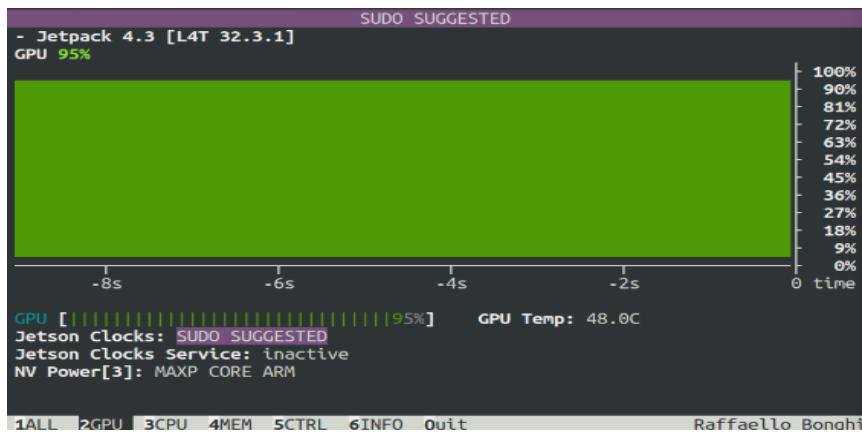
(b) CPU utilization of vector_add.cpp w/o using the GPU

Figure 4. Results profiled by jtopfor vector_add.cpp program which uses the CPU only. Jetson mode 1 was used to run the program.

In comparison, vector_add.cu utilizes the GPU almost fully at 99%, as shown in Fig. 5a. The CPU1 ARM core is also fully utilized, while other ARM cores are used by 10% -- 13%. The power consumption of the CPU+GPU mode is 2523 milliwatts which is slightly higher than the CPU only version. Note, jtop menu 5CTRL also provides a user-friendly interface for the NVModel utility so the user can change the Jetson operation mode and CPU/GPU clock rates easily.



(a) Overall performance of vectoradd.cu utilizing the GPU



(b) GPU utilization of vector_add.cu using the GPU

Figure 5. Results profiled by jtop for vector_add.cpp program which uses the CPU only. Jetson mode 1 was used to run the program.

The profiling results of vector_add.cu using Nsight Systems is shown in Fig. 6. The host computer runs the Nsight Systems while the Jetson TX2 is running the vector_add.cu. The host computer remote accesses the Jetson via SSH to profile the performance of each CPU core and GPU thread. It is clear that the utilization of CPU is decreased when the GPU kernel function is being executed.

3.2. Example 2: Wall Follow Python Code

The wall-follow python code is run on the f110 simulator with time.time() to calculate the time consumed in each function. The car is placed at the right bottom corner of the track when the wall-follow program starts. In all Jetson modes except Mode 4, the car completes the whole track without collision. The time spent on function lidar_callback is shown in Table 5. It is clear that Mode 4 runs the slowest because it only uses 2 CPUs, thus the car reacts slowly and collides to the wall easily; No differences are observed in car behaviors in other modes, although the times spent by modes 0 - 3 are slightly different and Mode 1 is notably slower than Mode 0, 2 and 3.



Figure 6. Nishgt System profiling timeline results for Example 1

Table 5. Time spent by the lidar_callback function of Example 2 in five working modes of Jetson TX2

Function	Mode 0	Mode 1	Mode 2	Mode 3	Mode 4
lidar_callback	626us	967us	511us	541us	1010us

The jtop profiler results are shown in Fig. 7 with Mode 4 as an example. Since only one Denver2 core and one ARM core are enabled in Mode 4, both of the CPU cores are utilized at 100% and the GPU is utilized at 10% mainly to support Rviz graphical display.

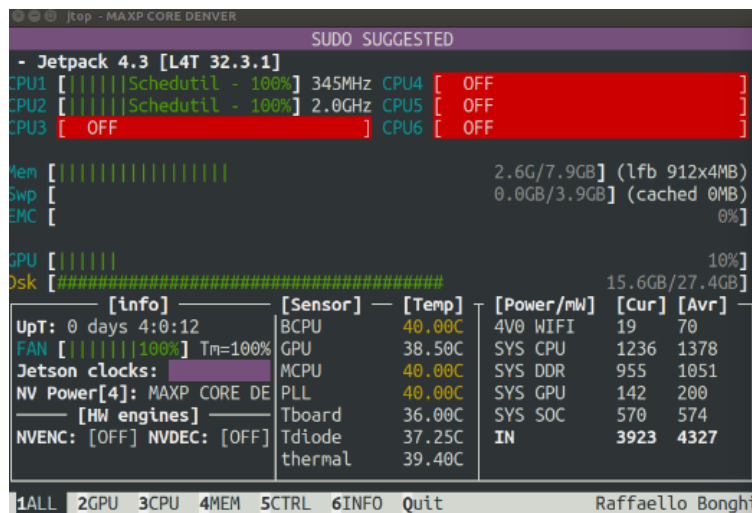


Figure 7. Performance of Example 2: Wall follow python program run in Mode 4

3.3. Example 3: Particle Filters

The particle filter algorithm is run in the f110 simulator on the Jetson TX2 device first, and then run in ROS directly with the hardware. The operating Mode 0 is used for both cases. In the f110 simulator, the car runs the wall-following algorithm through one lap of the track while the particle filter runs simultaneously with a way-point logger, so that the localized waypoints are logged in to a file as the car running through the track. When the range method is set to "rm" (ray marching), the points in the logged file are sparse, as shown in Fig. 8b, because the algorithm runs slow with CPU cores only. When the range method is set to "rmgpu" (ray marching using GPU), the particle filter performance is faster than the "rm" option, and the number of way points logged in the track is much higher, as shown in Fig. 8a.

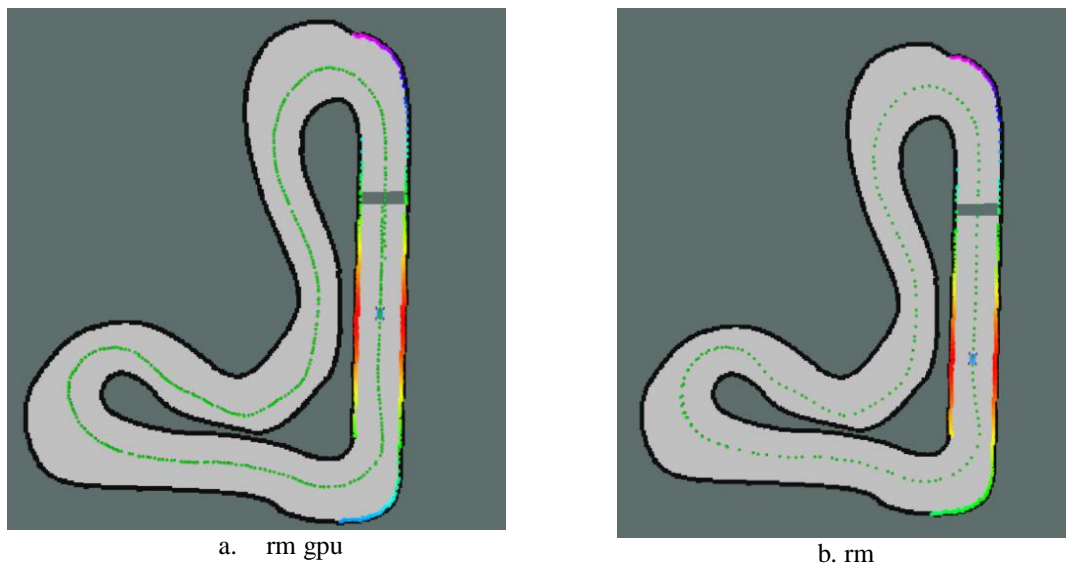
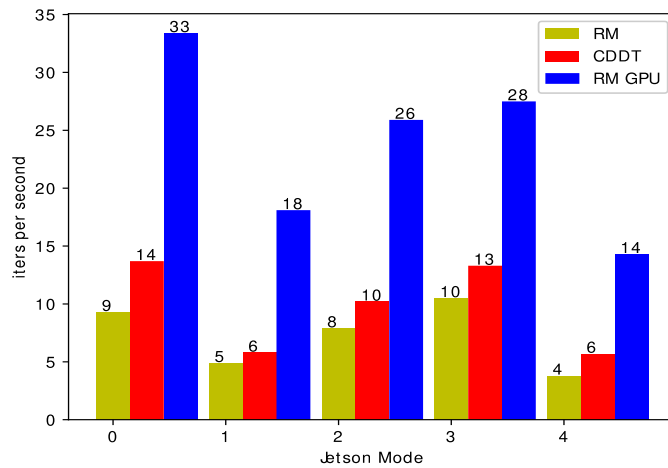


Figure 8. Way points generated by ray casting methods: ray marching with or without GPU. The number of way points is proportional to the processing speed of the algorithm.

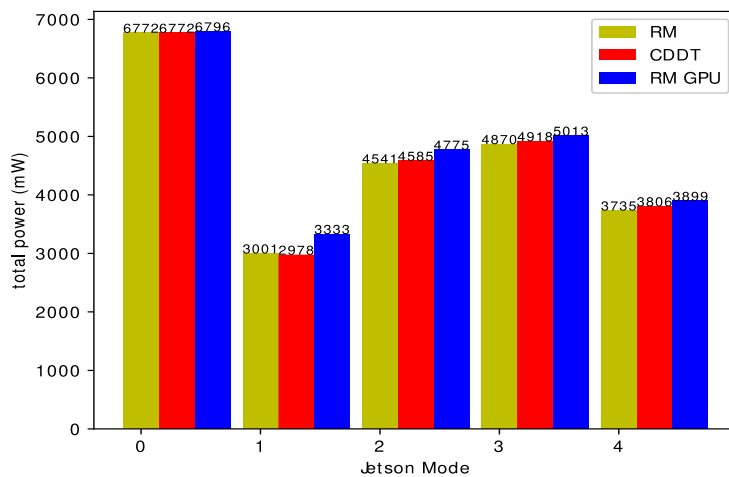
The speed and power efficiency of Example 3 in different modes are shown in Fig. 9. Utilizing the GPU, the rmgpu method gains about 2x to 4x speedup over the CDDT and rm methods in all Jetson modes, as shown in Fig. 9a. Meanwhile, the power consumption of the rmgpu method is slightly more than the rm and CDDT methods running in the same operation mode, as shown in Fig. 9b. This demonstrates that it is beneficial to utilize the GPU when ever is possible.

4. CONCLUSION

This paper has investigated two profiling tools for Jetson devices using three examples that involve CUDA C, python or ROS packages. The two profiling tools are tegrastats and Nsight Systems. While tegrastats and its GUI version jtop can run directly on the Jetson devices, the Jetson version of Nsight Systems has to run on a host computer to remotely profile a Jetson devices. Jtop provides summary of the Jetson CPU and GPU utilization, power consumption and dynamic flow in coarse time scale, while Nsight Systems can provide detailed CPU/GPU and memory activities of individual threads in fine time scale. Therefore, Nsight Systems is a better tool for performance optimization. Performance of five working modes of Jetson TX2 has been compared to illustrate the capabilities of the two profiling tools.



(a) iters per second



(b) power consumption

Figure 9. Example 3 Performance in different modes: RM and CDDT run w/o GPU, rmgpu runs with GPU fully utilized.

REFERENCES

- [1] University of Pennsylvania. F1/10 race car. [Online]. Available: <https://f1tenth.org>
- [2] MIT-racecar. Particle filter localization. Retrieved Feb. 2020. [Online]. Available: https://github.com/mit-racecar/particle_filter/blob/master/README.md
- [3] J. Benson. NVPMModel - NVIDIA Jetson TX2 DevelopmentKit. [Online]. Available: <https://www.jetsonhacks.com/2017/03/25/nvpmmodel-nvidia-jetson-tx2-development-kit/>
- [4] NVIDIA. (2020, Feb.) Nvidia releases nsight systems 2020.2. [Online]. Available: <https://news.developer.nvidia.com/nvidia-announces-nsight-systems-2020-2/>
- [5] S. McMillan. (2019, Aug.) Transitioning to nsight systems from nvidia visual profiler/nvprof. [Online]. Available: <https://devblogs.nvidia.com/transitioning-nsight-systems-nvidia-visual-profiler-nvprof/>

- [6] NVidia. Profiler user's guide. [Online]. Available: <https://docs.nvidia.com/cuda/profiler-users-guide/index.html>
- [7] NVIDIA. Tegrastats utility. [Online]. Available: <https://docs.nvidia.com/jetson/archives/14t-archived/14t-231/index.html#page/Tegra%20Linux%20Driver%20Package%20Development%20Guide/AppendixTegraStats.html>
- [8] R. Bonghi. Jetson stats. [Online]. Available: <https://github.com/rbonghi/jetsonstats>
- [9] NVidia. Jetpack SDK: Jetpack 4.3. [Online]. Available: <https://developer.nvidia.com/embedded/jetpack>
- [10] M.Harris. (2017, Jul.) Unified memory for CUDA beginners. [Online]. Available: <https://devblogs.nvidia.com/unified-memory-cuda-beginners/>
- [11] fltenth team. fltenth racecar wall follow. [Online]. Available: https://github.com/fltenth/fl10_ros/tree/master/wall_follow

SMARTAJWEED

AUTOMATIC RECOGNITION OF ARABIC QURANIC RECITATION RULES

Ali M. Alagrami¹ and Maged M. Eljazzar²

¹Department of Computer Science, University of Venice, Italy

²Faculty of Engineering, Cairo University, Egypt

ABSTRACT

Tajweed is a set of rules to read the Quran in a correct Pronunciation of the letters with all its Qualities, while Reciting the Quran. which means you have to give every letter in the Quran its due of characteristics and apply it to this particular letter in this specific situation while reading, which may differ in other times. These characteristics include melodic rules, like where to stop and for how long, when to merge two letters in pronunciation or when to stretch some, or even when to put more strength on some letters over other. Most of the papers focus mainly on the main recitation rules and the pronunciation but not (Ahkam AL Tajweed) which give different rhythm and different melody to the pronunciation with every different rule of (Tajweed). Which is also considered very important and essential in Reading the Quran as it can give different meanings to the words. In this paper we discuss in detail full system for automatic recognition of Quran Recitation Rules (Tajweed) by using support vector machine and threshold scoring system.

KEYWORDS

SVM, Machine learning , speech recognition , Quran Recitation, Tajweed

1. INTRODUCTION

The correct pronunciation of Al-Quran is based on the "Tajweed rules" which are essential to follow while recitation of the Al-Quran [1]. "Jawwada" is the root word from which the Arabic word Tajweed is derived. Jawwada means to make better or improve the pronunciation and it's accuracy during recitation [2].

Tajweed is obligatory for all Muslims as following Tajweed rules properly: it becomes easy to read the Al-Quran accurately [3-4]. Because if any mistake happened in Tajweed's rules during the recitation of Al-Quran, it can alter the definite meaning of that word, which is mispronounced. Many of the Islamic education studies showed that reciting The Al-Quran has always been a complex issue. First studies demonstrate that the students face many problems to use the correct Tajweed rules during reciting and reading Al-Quran [5-7]. Secondly, many Muslims are non-Arabic native speakers. At last, non-native Arabic speakers may face issue understand the difference in meaning for the same words; as low regularity in the reciting of Al-Quran is an issue that majorly contributes to the misreading or not properly reciting the Al-Quran among Muslims [8].

One of the traditional methods in learning Tajweed rules is "Talaqqi Musyafahah's" [9] and like any traditional learning process, it is supervised by the teacher, who you may not be completely

sure about his skills and abilities. on the other hand, problems as the teaching environment, lack of attention, or even surroundings noises which may contribute to the possibility of non-accurate hearing and prone to errors. Moreover, the hours given for learning are also limited. Currently, several applications utilize smart phones to simplify the Tajweed learning process. In a previous work, we introduced Imam [24], an application to learn Al-Quran with gamification, and make it simplify the Tajweed learning process between Sheikh and student.

Based on the problem mentioned above, it's been observed that it is necessary to advance the existing way of Tajweed learning. In this paper we propose a system capable of automatic recognition of four different rules of Tajweed with very high Accuracy 99% which is very promising results for further improvements, also the whole system is implemented in IMAM Application in its beta version.

2. LITERATURE REVIEW

Computer-Aided Pronunciation Learning (CAPL) has been presented and attained special consideration in past years. In second language teaching, several research works have been made to develop these systems [15, 16]. After facing many challenging issues and complications in knowing Quranic letters, specifically in checking the Tajweed's rules, the researchers took many serious actions in these past years.

While researching on the Quranic verse recitation recognition, a group of study headed by H.Tabbal, et al. (2006) [17] observed the delimitation system of Quranic verse in the form of audio files via this speech recognition system. This project's approach concerns more with the computerized delimiter that can excerpt verse through audio files. Examination techniques for every phase were elaborated and assessed using several reciters techniques, which recite surah "Al-Ikhlās". The most significant rules of Tajweed and tarteel were also the part of the discussion that can Impact the recognition of a particular recitation in Surah Al-Ikhlās.

To assist the learner in order to recite Al-Quran through a collaborative learning method, computerized Tajweed Inspection Rules Engine for the Learning the quran have been presented [18]. This proposed system will help the students to recite the Al-Quran in the absence of any instructor. The recitation done by the user will get processed by using the system, and revision of recitation is done. Still, the system's engine can only be verified based on certain Tajweed rules of surah Al-Fatihah.

The system of E-Hafiz functions similarly to a Hafiz that can help in the recitation of learning Al-Quran by reducing errors and inaccuracies during practice of the recitation. The Hafiz is a proficient person who performs functions similar to a teacher, i.e., first to listen and then teach Tajweed's correct recitation [19]. But still, this model concerns to only those who are already familiar with the Tajweed. In offline mode, the system does work. This system does not point any of the user's mistakes concurrently during recitation [13].

With the use of speech recognition technology, another Tajweed rule checking has been recently purposed to assist students in learning and reviewing correct recitation of Al-Quran by on their own [3]. This system can identify and highlight the incompatibility among the students recitations with that of the experienced and expert instructors stored in the database. The feature classification system has implemented the HMM algorithm, and for feature extraction, the system adopted the MFCC algorithm.

Moreover, another study using the image processing technique based on Tajweed's automatic rules is proposed [20]. The method of this research is constrained only by the laws of Idgham.

Next, input image passed via process of pre-processing, consisting of four sub processes: binary conversion, thinning and flip, grayscale conversion, and the word segmentation. Hereafter, six characteristics of shape descriptors were extracted from each input image: minor axis length, major axis length, filled area, eccentricity, perimeter and solidity. To understand the two forms of Idgham Laws that are Idgham Maal Ghunnah and Idgham Bila Ghunnah, a method of k-Nearest Neighbor (k-NN) is used. To assess the proposed analysis's efficiency, 180 test images were analyzed, which demonstrated the classification accuracy of 84.44%. The research outcome is supposed to instantly understand Tajweed's rules and allow the speakers to correctly recite the Al-Quran.

3. PROPOSED APPROACH

Our main objective in this paper is to build a complete system capable of recognizing the different rules of Tajweed in an audio. And determine whether it was pronounced correctly or not in a percentile metric. moreover, build an application over this system. which will enable hundreds of millions of Muslims, and all non arabic speakers all around the globe to learn how to read the Holy Quran Correctly. In our System we considered four different rules (Edgham Meem, Ekhfaa Meem, takhfeef Lam, Tarqeeq Lam) for each rule we collected a dataset from universities, Expert Volunteers and paid Experts. The dataset contains the right way for the pronunciation and different wrong ways in which it can be pronounced. taking in consideration Arabic natives and non-natives. Then we built a system that can work not only on specific Verses but a system able to recognize the rule in General that can Recognize the Rule in any verse in the holy Quran. In the first section we will discuss the overall system pipeline after that in the second section we will talk more about the Feature Extraction methodology and the Machine learning technique and at last we will show the results and conclusion and the future work.

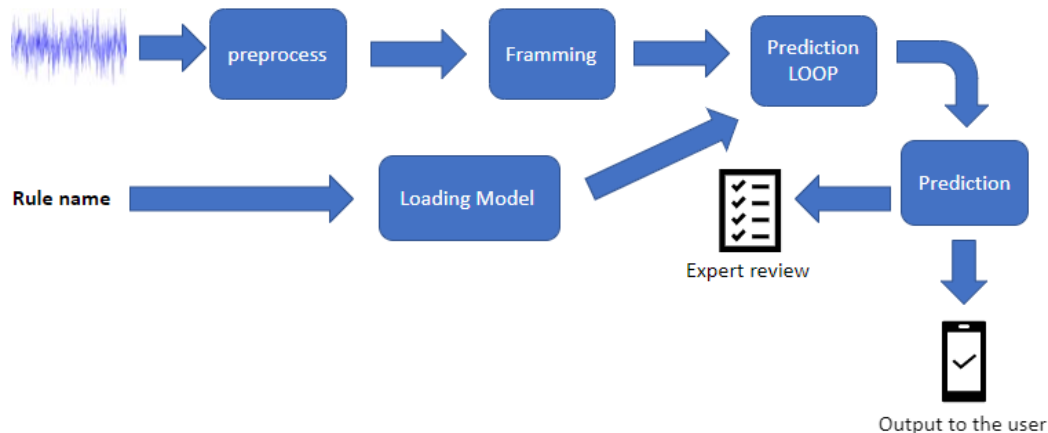


Figure 1. Tajweed Abstracted System Architecture

3.1. Abstract System Design.

As mentioned earlier the system main goal is to recognize the different rules of “Tajweed” and to indicate whether it's correct or not with a focused objective to make it easier for all Muslims whatever their native language to learn how to read the Quran probably. The input to the system is composed of two main Parameters, first the audio file which is supposed to contain the “Tajweed“ rule, along with the name of the rule which should be found in the audio file. By knowing the name of the rule, we load the model built specifically for this rule and send it to the prediction loop (Figure 1). At the same time the input audio is preprocessed, framed into 4 seconds with stride 0.5 seconds. The frames and the loaded model are sent to the prediction loop

which loops over the frames and uses the model loaded to predict the part of the audio which contains the rule. The prediction on each frame is compared to a set of specific thresholds for each recognition (Right/Wrong) for this specific rule. if the prediction is higher than that threshold the prediction will be taken in consideration otherwise it will be discarded.

The prediction loop sends the list of the predictions to the prediction module which compares the predictions and chooses the one to be shown to the user based on a certain rule. (Right pronunciation/Wrong Pronunciation and how close is the user from the right pronunciation) Moreover it saves all the records and the prediction which is then sent to an expert to review correctly so as to be used later to refine the model and the system periodically.

3.2. Dataset

First, we used a limited dataset we have to build the beta system. The dataset contained about roughly 80 record for each Rule name and type with a total of 657 recordings of 4 different rules each has two different types (4 rules, 8 classes). To Collect more dataset and refine the model we launched beta feature in our application by which we received about 600 more records from different app users in only 2 months and more is coming every week with 30+ new users ever month. which will help us a lot in refining the Model. all the recordings we receive is reviewed and labelled by an expert.

3.3. Processing

before this stage each audio file in the dataset is manually cut so as to contain only the part in which the rule is pronounced (Right or Wrong), all the audios sampling rate was changed to 8000 Hz, then we took the average length for all the audio files which was about 4 seconds , then we converted the length of all the other audio files to be 4 seconds. if it is less than 4 seconds we add silence, if longer randomly truncate is done.

3.4. Feature Extraction

Our approach employs filter banks as a baseline method for feature extraction. Filter banks were motivated by the nature of the speech signal and the human perception of such signals[21]. First the signals are cut into frames each of size 25 ms with stride 10 ms. hamming window is applied for minimizing the disruptions at the starting and at the end of each frame , then we perform FFT to obtain the magnitude frequency response to get the spectrum which is then subjected to 70 triangular band pass filters in order to get smooth spectrum, and to reduce the size of the features involved.

3.5. Classification Model

For the Classification task we used Support Vector Machine (SVM) which was first Introduced in 1992 by Boser, Guyon and Vapnik [22] . SVM is considered a supervised machine learning method that can be used for classification, regression, and Outliers Detection. However, it is mostly used in classification problems. SVM also belongs to the general category of Kernel methods [23] which are methods that depend on the Data through only the dot products. kernels use this fact and introduce this dot product in a possibly higher dimensional feature space which in turns make it easier to find nonlinear decision boundaries using linear classifier methods and allow us to use the classifier on data that does not have fixed dimension this trick used here is called the “kernel trick”.

As mentioned The support vector machine is a binary classifier Algorithm tries to find the Maximum marginal hyperplane which is the plane with maximum distance from the two classes introduced. Let $\{x_i, y_i\}$ be sample of the dataset in which x_i is an input features and y_i is the corresponding class value (-1,+1) “P” and “N” . so that we have

$$y_i(w^T x_i + b) \geq 1 \text{ for all } x_i$$

To make computations easier and without loss of generality the pair (w, b) can be rescaled such that:

$$\min_{i=1 \dots l} |(w^T x_i + b)| = 1$$

To restrict the hypothesis space, the support vector machine tries to find the simplest solution that classifies the input data correctly. The learning problem is therefore can be written as: Minimization $\|W\|^2 = W^T W$ subjected to the constraints of linear separability. This is equivalent to maximizing the margin between the convex Envelope of the two classes we have. The optimization is now a convex quadratic programming problem.

$$\text{Minimization } \Phi(W) = \frac{1}{2} \|W\|^2$$

$$\text{Subject to } y_i(W^T x_i + b) \geq 1, \text{ for all } i$$

And as the input data appears only in the form of dot product $x^T x'$ the Kernel $K(x, x')$ can be introduced to map the data into higher dimensional space. Also one of the main attributes that affect the performance of the SVM is the penalty parameter “C”, where lower value of “C” encourages greater margin for a lower accuracy. On the other hand, the “gamma” parameter can be seen as the inverse of the radius of influence of the data points selected by the model as support vectors.

To build the model we used a radial basis kernel and tuned the hyperparameters “C” and “gamma” through a grid search algorithm where the best performance was achieved when the “C” is tuned to 1 and “gamma” to 0.1

4. TESTING AND RESULTS

Each model in the system was tested against 30% of the Data recorded with a validation accuracy of 99%. Then the system as one entity is then tested against full verses to extract the rules, a sample of the testing table can be seen in (Table 2) where each row is a verse in Quran in which certain rule should be found. The system was not only tested to recognize the rule but also extract the exact timing in which the rule is recognized and how close it is to the threshold set earlier for this specific rule name and type. In (Figure 2) you can see the results of the system on two test audios one on “Edgham meem” and the other on “Tarqeeq lam” where the green line depicts the rule starting time recognized by the system while the red line is the true line set by the expert. The thresholds for each rule name and type (Right, Wrong) are selected so as to minimize the false positive as much as possible in the testing Data-set. But the thresholds are subjected to be changed in general as a response to the overall performance after deploying the system.

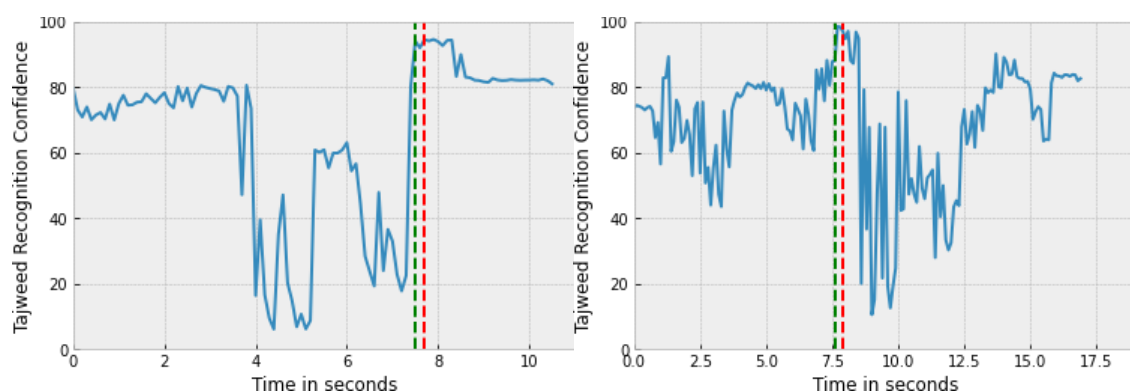


Figure 2. the figure on the left shows the output result of the model on edgham meem test sample and on the right output results on tarqeeq lam test sample

Table 1. Confusion Table for the 4 SVM models

Rule Name	True Positive	False Positive	True Negative	False Negative
Edgham Meem	30	0	30	0
Tafkheem Lam	30	0	30	0
Tarqeeq Lam	29	0	30	1
Ekhfaa Meem	30	0	30	0

Table 2. Samples of the verses used in the testing phase

Rule Name (Right)	Chapter	Verse
Ekhfaa Meem	41	52
Ekhfaa Meem	7	45
Ekhfaa Meem	40	75
Ekhfaa Meem	38	46
Tafkheem Lam	2	9
Tafkheem Lam	2	10
Tafkheem Lam	8	16
Tafkheem Lam	2	149
Tafkheem Lam	14	51
Tarqeeq Lam	2	67
Tarqeeq Lam	14	5
Tarqeeq Lam	18	26
Tarqeeq Lam	16	41
Edgam Meem	38	62
Edgam Meem	38	4
Edgam Meem	6	164
Edgam Meem	4	1
Edgam Meem	37	151
Edgam Meem	39	60

4. CONCLUSION AND FUTURE WORK

In this paper, we proposed the new approach that we used to build the Tajweed System in IMAM Application. By treating the problem as a normal binary classification problem. The input Audio Signal is first preprocessed then goes through the features extraction in which we used 70 Filter banks. And for the classification, we used SVM with the threshold scoring method which allowed us to have full control over the system rule recognition capability and allowed us to tune it to get the best results. The Work done here will be continued to include even more rules with main target goal to include all the rules for Recitation in the Quran and deploy it in our application free to use for anyone, anywhere.

REFERENCES

- [1] Y. O. M. Elhadj, M. Aoun-Allah, I. A. Alsughaiyer, and A. Alansari, "A New Scientific Formulation of Tajweed Rules for E-Learning of Quran Phonological Rules," Edited by Anderson Silva, Elvis Pontes, p. 197, 2012.
- [2] A. Alfaries, M. Albahlal, M. Almazrua, and A. Almazrua, "A rule-based annotation system to extract tajweed rules from quran," in 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences, 2013: IEEE, pp. 281- 286.
- [3] I. Ahsiah, N. Noor, and M. Idris, "Tajweed checking system to support recitation," in 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2013: Ieee, pp. 189-193.
- [4] N. J. Ibrahim, M. Y. I. Idris, M. Z. M. Yusoff, and A. Anuar, "The Problems, Issues and Future Challenges of Automatic Speech Recognition for Quranic verse Recitation: A Review," *Al-Bayan: Journal of Qur'an and Hadith Studies*, vol. 13, no. 2, pp. 168-196, 2015.
- [5] S. S. B. Hassan and M. A. B. Zailaini, "Analysis of Tajweed errors in Quranic recitation," in *Procedia-Social and Behavioral Sciences*, 2013, vol. 103, no. Tq 1000: Elsevier, pp. 136-145.
- [6] A. Awang, A. Che Mat, and A. N. Yaakub, "Kebolehpayaan membaca al-Quran di kalangan pelajar diploma UiTM Terengganu: satu kajian kes," 2010.
- [7] A. R. Ali Hasan, S. Sabarudin, and C. B. Che Mat, "Tahap penguasaan bacaan Al-Quran dikalangan pelajar-pelajar UiTM Sarawak," 2010.
- [8] N. Kamarudin and S. Salam, "Tajweed understanding among Malaysian muslim (a preliminary findings)," in *Proc. 1st International Conference on Mobile Learning, Applications and Services (Mobilecase2012)*, Kuala Lumpur, 2012, 2012, pp. 1-4.
- [9] A. H. Tamuri, A. M. I. Amal, H. M. Noor, and M. I. M. Pisol, "Penilaian Guru terhadap Pelaksanaan Model-Model Tilawah al-Quran Program j-QAF/Teachers Evaluation on the Implementation of j-QAF Quranic Recitation Models," *International Journal of Islamic Thought*, vol. 3, p. 1, 2013.
- [10] A. N. Akkila and S. S. Abu-Naser, "Rules of Tajweed the Holy Quran Intelligent Tutoring System," 2018.
- [11] S. M. Abdou and M. Rashwan, "A Computer Aided Pronunciation Learning system for teaching the holy quran Recitation rules," in 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), 2014: IEEE, pp. 543-550.

- [12] A. A. Y. AlZoubi, "Use of Information Technology in the Teaching of Quran Recitation (Qira'at)-Electronic Miqrah as a Model," in 2013 Taibah University international conference on advances in information technology for the Holy Quran and its sciences, 2013: IEEE, pp. 593-621.
- [13] W. M. Muhammad, R. Muhammad, A. Muhammad, and A. Martinez-Enriquez, "Voice content matching system for quran readers," in 2010 Ninth Mexican International Conference on Artificial Intelligence, 2010: IEEE, pp. 148-153.
- [14] E. Mourtaga, A. Sharieh, and M. Abdallah, "Speaker independent Quranic recognizer based on maximum likelihood linear regression," in Proceedings of world academy of science, engineering and technology, 2007, vol. 36: Citeseer, pp. 61-67.
- [15] H. Franco, L. Neumeyer, M. Ramos, and H. Bratt, "Automatic detection of phone-level mispronunciation for language learning," in Sixth European Conference on Speech Communication and Technology, 1999.
- [16] S. M. Witt, "Use of speech recognition in computer-assisted language learning," 1999.
- [17] H. Tabbal, W. El Falou, and B. Monla, "Analysis and implementation of a " Quranic" verses delimitation system in audio files using speech recognition techniques," in 2006 2nd International Conference on Information & Communication Technologies, 2006, vol. 2: IEEE, pp. 2979-2984.
- [18] R.-J. Raja-Yusof, F. Grine, N. J. Ibrahim, M. Y. I. Idris, Z. Razak, and N. N. A. Rahman, "Automated tajweed checking rules engine for Quranic learning," *Multicultural Education & Technology Journal*, 2013.
- [19] A. Muhammad, Z. ul Qayyum, S. Tanveer, A. Martinez-Enriquez, and A. Z. Syed, "E-hafiz: Intelligent system to help Muslims in recitation and memorization of Quran," *Life Science Journal*, vol. 9, no. 1, pp. 534-541, 2012.
- [20] F. A. A. R. Shafaf Ibrahim, Zaaba Ahmad, "Automatic Tajweed Rules Recognition using k-Nearest Neighbour (k-NN)," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. Volume-8, no. Issue-2S11, 2019.
- [21] Vetterli, M., & Herley, C. (1992). Wavelets and filter banks: Theory and design. *IEEE transactions on signal processing*, 40(ARTICLE), 2207-2232.
- [22] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Fish Annual Workshop on Computational Learning Theory*, pages 144- 152. ACM, June 1992.
- [23] Scholkopf, B., & Smola, A. J. (2018). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Adaptive Computation and Machine Learning series.
- [24] Alargrami, A. M., & Eljazzar, M. M. (2020, October). Imam: Word Embedding Model for Islamic Arabic NLP. In *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)* (pp. 520-524). IEEE.

AUTHOR INDEX

<i>Alan P. Chitayat</i>	117
<i>Ali M. Alagrami</i>	145
<i>Alistair Coates</i>	117
<i>Anders Drachen</i>	117
<i>Anna Georgiadou</i>	71
<i>Athanasios V. Kokkinakis</i>	117
<i>Ben Kirman</i>	117
<i>Boureima Zerbo</i>	81
<i>Daniel Slawson</i>	117
<i>Désiré Guel</i>	81
<i>Dimitrios Askounis</i>	71
<i>Evan Gunnell</i>	47
<i>Evan R.M. Debenham</i>	01
<i>Florian Block</i>	117
<i>Gilbert Busolo</i>	19
<i>Hager Ali Yahia</i>	97
<i>Hatem Awad Khater</i>	97
<i>HuiHui He</i>	61
<i>Isabelle Nölle</i>	117
<i>Jacques Palicot</i>	81
<i>John Hawkins</i>	37
<i>Jonathan Hook</i>	117
<i>Justus Roberstson</i>	117
<i>Kennedy Ogada</i>	19
<i>Lawrence Nderu</i>	19
<i>Maged M. Eljazza</i>	145
<i>Marian Ursu</i>	117
<i>Marie-Anne Xu</i>	107
<i>Marisabel Chang</i>	47
<i>Mohammed Rizk Mohammed</i>	97
<i>Mohammed Zakaria Moustafa</i>	97
<i>Oluseyi Olarewaju</i>	117
<i>Oumarou Sié</i>	81
<i>Rahul Khanna</i>	107
<i>Roberto Solis-Oba</i>	01
<i>Sagarika Patra</i>	117
<i>Shiyuan Zhang</i>	47
<i>Simon Demediuk</i>	117
<i>Spiros Mouzakitidis</i>	71
<i>Yahong Rosa Zheng</i>	133
<i>Yazhou Li</i>	133
<i>YongJun Wang</i>	61
<i>Yu Sun</i>	47