The Computer Science Brunel PhD Symposium

2020

Brunel University London

https://brunelcs.github.io/csbps/

MESSAGE FROM THE GENERAL CHAIR

Welcome to our Virtual Computer Science Brunel PhD Symposium (CSBPS 2020).

10th June 2020

Our PhD Symposium celebrates the research accomplishments of our Computer Science PhD candidates. The Symposium is a forum that allows our Computer Science PhD candidates to present and discuss their research and get inspired. It also provides them with the opportunity to share their research with their peers and to gain constructive feedback about their already completed or, more importantly, planned research work from senior and experienced researchers.

This year, the programme has a wide range of contributions from our PhD candidates, our academic staff and our successful PhD graduates. The programme features the contribution of our invited Keynote Speaker, Prof Licia Capra- **<strikethrough> strategies for success in computer science research</strikethrough>,** where she will present career progression opportunities to our PhD candidates.

Our PhD candidates and researchers of the future will present 14 extended abstracts and 9 posters, covering a wide range of topics in Computer Science ranging from software engineering to data analytics, simulation, security and interactive design.

Two of our successful PhD graduates and awardees of previous CSBPS editions will share their experiences of how the CSBPS supported their career progression. We will also have academic staff introducing our Department Research Groups. Finally, a highly requested session entitled, “Acceptance or Rejection: The Reviewing Process in Computer Science” is included to mentor our candidates on how Computer Science Venues make the decision of accepting or rejecting papers.

Many people have worked hard to make this Symposium possible, especially after having to re-organize the event due to Covid-19. We would like to thank our Programme Committee for reviewing the extended abstracts and Session Panellists for providing feedback to the PhD candidates. Special thanks goes to the Computer Science Department for sponsoring the Symposium.

Finally, we hope you enjoy the PhD Symposium and contribute to the future editions.

Dr Nour Ali, Senior Lecturer in Computer Science

Chair of CSBPS 2020

Programme and Organization Chair

Dr Nour Ali

Programme Committee

Nadine Aburumman, Faris Alwzinani, Mahir Arzoky, Andrea Capiluppi, Giuseppe Destefanis, Crina Grosan, Armin Kashefi, Alaa Marshan, Kathy Mcgrath, Alina Miron, Arthur Money, Rumyana Neykova, Alan Serrano, Martin Shepperd and Allan Tucker.

Publicity Chair

Khalipha Nuhu

Website Chair

Bhaveet Nagaria

Proceeding Chair

Fatima Amer Jid Almahri

CSBPS 2020 Virtual Programme

▶ 10:00-10:40	Welcome and Invited Keynote
▶ Invited Keynote	Session Chair: Dr Nour Ali Prof Licia Capra - University College London **<strikethrough> strategies for success in computer science research</strikethrough>**
▶ 10:45-11:35	Main Track – Part One Session Chair: Bhaveet Nagaria • Ashley Mann. An Iterated Local Search Approach to the Software Modularisation Problem. • Thomas Coleman. LogicGate Revised – A Student Centered Digital Game-Based Learning Experience. • Benjamin Evans. Automating Camera Trap Analysis. • Andres Crespo. A Standard Cyber Security Model for Artificial Intelligence Platforms running SMART CITIES. • Afees Odebode. A Model-Based Consensus Clustering For Predicting Number of Partitions in Datasets • Fawziazehra Kara-Isitt. Detecting Online Information Weaponisation.
▶ 11:35 – 12:25	Main Track – Part Two Session Chair: Gabriel Scali • Faisal Maramazi & Ashley Mann. The Software Modularisation Problem, A Meta-Analysis. • Monjur Elahi. Case-based reasoning to minimise the impact of disruptions on public transport. • Toyah Overton. DO-U-Net for Segmentation and Counting. • Zear Ibrahim. Digitising Falls Risk Assessment through Mobile Depth-Sensing. • Fatima Amer Jid Almahri. Acceptance and Use of Chatbots for University Students: An Application of UTAUT2. • Marinela Iuliana Branescu. Detecting the Type of Leukemia with Neuronal Network and TensorFlow. • Mohamad Merhi. An empirical examination of mobile banking use: a cross-national, quantitative study.
▶ 12:25 – 13:00	Lunch and Breakout Rooms
▶ 13:00– 13:15	Graduate Presentations Session Chair: Dr Nour Ali • Dr Diana Suleimenova, "My PhD journey: Peaks and Valleys" - BUL. • Dr Monica Pereira, “Fun and wonder: making friends across all borders” - LMU
▶ 13:15 – 13:45	Research Group Presentations Session Chair: Dr Allan Tucker • Dr Anastasia Anagnostou Simulation & Modelling • Dr Theodora Koulouri Human Computer Interaction (HCI) • Dr Alina Miron Intelligent Data Analysis (IDA) • Dr Giuseppe Destefanis Brunel Software Engineering Laboratory (BSEL) • Dr Alessandro Pandini Bioinformatics • Dr Stasha Lauria Interactive Multimedia Systems
▶ 13:45 -13:55	Special Session Session Chair: Fatima Amer Jid Almahri • Dr Nour Ali. “Acceptance or Rejection: The Reviewing Process in Computer Science”
▶ 13:55 – 14:00	Prize Giving and Closure.
▶ 14:00- Onwards	Celebrating our Research /Breakout rooms

Keynote: Prof Licia Capra

Professor @ UCL in Pervasive Computing

<strikethrough> strategies for success in computer science research</strikethrough>

Abstract

In this talk I will outline my own career journey as a computer scientist, from when it started back in 1996 as a UG student, to 2015 when I was promoted to full professor. Rather than a story of strategy and success, I will cover the many accidents (and failures) that accompanied me through this journey. On the way, I will touch upon some of the research work I have done over the years, spanning different areas of computer science, and different disciplines beyond computer science. I will conclude with a reflection on the choices I am glad I did, and those I regret a little. The aim is not to dispense advise on what is best, but rather to highlight what you might want to think about when the time comes to make choices in your own computer science journey.

Biography

Professor Licia Capra is Professor of Pervasive Computing in the Department of Computer Science at University College London. Licia has an MEng degree in Computer Science from the University of Bologna in Italy, and a PhD in Computer Science from UCL. After a brief post-doctoral experience, she took a Lectureship position at UCL in 2005, and was promoted to Professor in 2015. Licia has supervised 6 PhD students to completion, and has acted as Director of Studies / Departmental Tutor in 2018/19. She has over 170 publications and 8,000 citations in international venues such as the ACM Conference on Computer-Supported Collaborative Work and Social Computing (CSCW) and The Web Conference (WWW). Licia's research now focuses on computational social science, and the development and use of methods to study human behaviors in computer-mediated (collaborative) settings.

Posters

Poster Presenter	Poster Name
Benjamin Evans, Bhaveet Nagaria, Ashley Mann	'An Exploration into Neural Architectures for Learning to Play Pacman.'
Isabel Holmes	'Tracking hate - a data science approach to online racism.'
Nura Tijjani Abubakar	'Cloud-Based Distributed Simulation Framework: A Preliminary Investigation.'
Marinela Iuliana Branescu	'Detecting the Type of Leukemia with Neuronal Network and TensorFlow.'
Mohamad Mehri	'An empirical examination of mobile banking use: a cross-national, quantitative study.'
Monjour Elahi	'Case-based reasoning to minimise the impact of disruptions on public transport.'
Fatima Amer Jid Almahri	‘Augmented Educational Interactions within a Physical Space for University Students Using Conversational Systems'
Yasoda Jayaweera	'Data analytics driven Methodology for identifying alternative solutions in a population-based data generation approach applied to synthetic biology.'
Leila Yousefi	'Opening The Black Box: A Hidden Variable Discovery Methodology to Personalise Patients.'

Table of Contents

Acceptance and Use of Chatbots for University Students: An Application of UTAUT2

Fatima Amer Jid Almahri
Human-Computer Interaction,
Brunel University London
Uxbridge, United Kingdom
Fatima.amerjidalmahril@brunel.ac.uk

David Bell
Human-Computer Interaction
Brunel University London
Uxbridge, United Kingdom
David.bell@brunel.ac.uk

***Abstract*—The purpose of this paper is to investigate the factors that may hinder or facilitate the acceptance and usage of chatbots by university students in higher education institutions (HEIs). For this study, the unified theory of acceptance and use of technology2 (UTAUT2) is adopted for this educational context. A quantitative survey-based approach was used to collect data from 431 undergraduate and computer science masters students at Brunel University London. The data were analysed using partial least squares - structural equation modelling (PLS-SEM) with SmartPLS3. The findings showed that three primary constructs, namely, performance expectancy, effort expectancy and habit significantly predicted students' behavioural intention (BI) and use (USE) of chatbot technology. Our results can guide policymakers in their attempts to explain and encourage the adoption and acceptance of innovative technologies, such as chatbots, by university students. To the best of our knowledge, this study is the first to investigate student acceptance and use of chatbots by university students using UTAUT2. Moreover, the present study will serve as the foundation for the design of features aimed at promoting chatbot acceptance and the enhancement of student engagement.**

Keywords—chatbot; SEM-PLS; structural equation modelling; smartPLS3; technology adoption; UTAUT2

I. INTRODUCTION

A chatbot is defined as a computer application that conducts a conversation with a user using natural language via text or voice [1] or both. They are also referred to as conversational systems or 'artificial conversational entities' (ACEs). One of the earliest created chatbots was called ELIZA, which was developed by Joseph Weizenbaum in the 1960s [2]. To date, a large number of chatbots have been developed, the most popular being: Siri, Google Assistant, Amazon Alexa [3] and Cortana. Chatbots are used for different purposes in different domains. For example, they are used as teaching tools in education, health intervention tools in the health domain, as a source of entertainment, and as customer assistants in business domains [4].

Using chatbots in education can deliver many benefits for students and instructors [5], including 1) supporting continuous learning - they allow learners to view or repeat the exact learning resources several times; 2) making learning more enjoyable - students often prefer conducting a conversation with a chatbot rather than with their peers or instructors; and 3) enhancing students' motivation as they are often attractive innovative learning tool that attracts students attention. Prior to integrating chatbots into an educational environment, it is essential to study student acceptance and use of such technology using one of the available tried and tested technology acceptance models.

Recently, many popular theories and models have been studied to explain the relationship between users' attitudes, beliefs and behavioural intention (BI) to use the technology. [6] proposed The Unified Theory of Acceptance and User of Technology 2 (UTAUT2) model for the consumer context. This came about as a result of reviewing and analysing eight models of technology usage: Innovation Diffusion Theory (IDT); Theory of Planned Behaviour (TPB); Theory of Reasoned Action (TRA); Cognitive Theory; the Motivational Model; Social Cognitive Theory (SCT); the Model of Perceived Credibility (PC); the Technology Acceptance Model (TAM); and a hybrid model combining constructs from TPB and TAM [6].

II. THEORETICAL BACKGROUND AND HYPOTHESIS

The following research model is aimed at explaining the acceptance and use of a chatbot. [6] explained how the construct USE can be utilised to probe the intention and usage of a platform. For the current study, this is adopted for the chatbot in terms of the intention to use it (BI) and six constructs: Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FC), Hedonic Motivation (HM) and Habit (HT). The price value is removed from the originally proposed model because it is not appropriate for the educational context as the chatbot is free for students.

A. Proposed Model (UTAUT2-PERSONA)

Performance Expectancy (PE) is defined as "the degree to which an individual believes that using the system will help him or her to attain gains in job performance" [7, p. 447]. Previous research has elicited that PE is a significant predictor of behavioural intention[7][8]. Hypothesis 1(H1): PE has a positive effect on university students' behavioural intention to use chatbots. Effort Expectancy (EE) is defined as "the degree of ease associated with the use of the system" [7, p. 450]. EE and its latent variable have been found to be significant in several studies and proven to work as a predictor on user intention to adopt new technology [8]. Hypothesis 2(H2): EE has a positive effect on university students' behavioural intention to use chatbots.

Social Influence (SI) is defined as "the degree to which an individual perceives that important others believe he or she should use the new system" [7, p. 451]. SI has emerged as being significant regarding user intention to use certain

technology in a range of studies [8].Hypothesis 3(H3): SI has a positive effect on university students' behavioural intention to use chatbots. The Facilitating Condition (FC) is defined as "the degree to which an individual believes that an organisational and technical infrastructure exists to support the use of the system" [7, p. 453]. Hypothesis 4(H4): FC has a positive effect on university students' behavioural intention to use chatbots.

Hedonic Motivation (HM) refers to "the fun or pleasure derived from using technology" [6, p. 8]. It has been proven that HM plays a positive role in determining technology acceptance and its usage [6][8].Hypothesis 5(H5): HM has a positive effect on university students' behavioural intention to use chatbots. Habit (HT), as a construct in UTAUT2 [6], is defined in the IS and technology context as "the extent to which people tend to perform behaviours (use IS) automatically because of learning"[9, p. 709]. Hypothesis 6(H6): HT has a positive effect on university students' behavioural intention to use chatbots.

Regarding Behavioural Intention (BI), this has been defined as a "function of both attitudes and subjective norms about the target behaviour, predicting actual behaviour" [10, p. 26]. The strength of individual commitment to engage with particular activities can be assessed by their BI [11]. Hypothesis 7(H7): BI has a positive effect on students' behaviour regarding the use of chatbots. The moderator in this research is the persona: age, gender, experience, level of engagement and performance.

III. Research Methodology

The primary data collection was completed through a survey held between July and November 2019. Participation in this study was completely voluntary. In general, less than 8 minutes were required for the completion of questionnaires, which inquired about the participants' perception of chatbots. The scales of this study was adapted from previous studies that used UTAUT2, with all constructs being measured using seven items (7 point Likert-scale). The items of each construct were adapted from [6].

IV. Initial Findings

After performing bootstrapping using SmartPLS3, the results show that four hypotheses are supported: PE and BI (P = 0.00 and T = 4.972, supporting H1), EE and BI (P = 0.018 and T = 2.408, supporting H2), HT and BI (P = 0.00, supporting H6), as well as BI and USE (P = 0.00 and T = 0.6.792, supporting H7). However, three hypotheses are rejected: SI and BI (P = 0.086, T= 1.728, rejecting H3), FC and BI (P = 0.071 and T = 1.848, rejecting H4), along with HM and BI (P = 0.082, rejecting H5). The four supported hypotheses will be used as the basis for designing and developing a chatbot for educational purposes. Moderator results are not covered in this paper due to limited space.

V. Limitations and Future Work

In terms of limitations, the outcomes from this model cannot be generalised, because it is only applicable to the computer science department at Brunel University, London. Hence, future research should involve applying the proposed model across different departments and universities, maybe even different countries. Moreover, given the participants were university students, the model should be utilised for different educational contexts, such as secondary schools. Further, there is still room for improvement of the prediction model. Hence, future work should include other constructors (security, trust and system quality) and moderators (educational level, engagement level)that might increase the use of chatbots in different contexts.

VI. Conclusion

This study was aimed at investigating the factors that affect university students' intention to use chatbots at a UK University using the UTAUT2 model. It involved a quantitative method of surveying university students at Brunel University London about their acceptance and use of chatbots. SEM-PLS was used to analyse the data using SmartPls3. The results show that Performance Expectancy, Effort Expectancy and Habit are the three main predictors of student behavioural intention and use of a chatbot. To the best of our knowledge, this study is the first research investigating student acceptance and use of chatbots in the university context using UTAUT2. Also, this study serves as a basis for designing a chatbot for educational purposes in the context of enhancing student engagement. It is planned that moderating factors will be included in future research.

VII. References

[1] S. A. Abdul-Kader and J. Woods, "Survey on Chatbot Design Techniques in Speech Conversation Systems," *IJACSA) Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 7, pp. 72–80, 2015.

[2] J. Weizenbaum, "ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[3] A. Khanna, M. Jain, T. Kumar, D. Singh, B. Pandey, and V. Jha, "Anatomy and Utilities of an Artificial Intelligence Conversational Entity," *Proc. - 2015 Int. Conf. Comput. Intell. Commun. Networks, CICN 2015*, pp. 594–597, 2015.

[4] F. Almahri, D. Bell, and M. Arzoky, "Personas Design for Conversational Systems in Education," *Informatics*, vol. 6, no. 4, p. 46, 2019.

[5] L. Fryer and R. Carpenter, "emerging technologies Bots as language learning tools," *Lang. Learn. Technol.*, vol. 10, no. 3, pp. 8–14, 2006.

[6] V. Venkatesh, J. Y. L. T. Thong, and X. Xu, "consumer acceptance and use of IT," vol. 36, no. 1, pp. 157–178, 2012.

[7] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "USER ACCEPTANCE OF INFORMATION TECHNOLOGY: TOWARD A UNIFIED VIEW," *MIS Q.*, vol. 27, no. 3, pp. 425–478, 2003.

[8] A. Raman and Y. Don, "Preservice teachers' acceptance of learning management software: An application of the UTAUT2 model," *Int. Educ. Stud.*, vol. 6, no. 7, pp. 157–164, 2013.

[9] M. Limayem, S. G. Hirt, and C. M. K. Cheung, "How Habit Limits the Predictive Power of Intention: The Case of Information Systems Continuance," *MIS Q. Vol.*, vol. 31, no. 4, pp. 705–737, 2007.

[10] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research.* 1975.

[11] C. C. Lewis, C. E. Fretwell, J. Ryan, and J. B. Parham, "Faculty Use of Established and Emerging Technologies in Higher Education: A Unified Theory of Acceptance and Use of Technology Perspective," *Int. J. High. Educ.*, vol. 2, no. 2, pp. 22–34, 2013.

Detecting the Type of Leukemia with Neuronal Network and TensorFlow

Marinela Branescu
Department of Computer Science
Brunel University
London, UK
Marinela.branescu@brunel.ac.uk

Stephen Swift
Department of Computer Science
Brunel University
London, UK
Stephen.swift@brunel.ac.uk

Abstract--Identifying the right type of Leukemia is very important for proper diagnosis and treatment of the disease, analyzing the cells within the images of the blood smear test will enable hematologists to determine the stages and classification of Leukemia. This study will use 312 already classified Leukemia images and will train the Convolutional Neuronal Network to distinguish the types of the disease using the TensorFlow packages. The system's accuracy was tested by applying Cross-Validation respective K-Fold obtaining a mean value of 56.6%. Future work will include using Otsu Method to improve the accuracy of the experiment .

Index Terms: Convolutional Neuronal Network, Deep Learning Cross-Validation, K-Fold, Otsu Method.

I. INTRODUCTION

Leukemia is evidential by an excess of abnormal cells produced from the bone marrow, can be of more types and is a blood cancer condition that can develop without a specific cause [1]. According to the specified types of cells affected by the disease, the Leukemia will be classified in Acute Lymphocytic Leukaemia (ALL), Acute Myeloid Leukaemia (AML), Chronic Lymphocytic Leukaemia (CLL), Chronic Myeloid Leukaemia (CML) [1]. As the disease progress rapidly, it is very important to diagnose the type and stage of Leukemia fast in order to increase the patient's chances of survival and recovery [4]. The medical specialists can determine Leukemias by observing the images of blood samples trough microscope, as the white blood cell count can be seen trough blood samples they present information about the patient's health [3]. When observing the white blood cells samples lots of impediments can affect the accuracy of the images like poor staining of the cells or inexperienced observer and those will lead to wrong diagnosis [4]. As the disease has to be treated urgently from the time of recognition any misclassification will lead to improper treatment and affect the patient's recovery and survival rate. Recently, the imaging technology has been radically improved by the medical imaging technology, the digital scanners have enabled the studies on pathological image analysis to be more precise and the quantitative analysis tools have decreased the time of evaluating the samples [4]. The Convolutional Neuronal Network (CNN) has become a very useful tool in recognition, detection, retrieval and classification of images, and the medical imagining will benefit greatly from using it [5]. It is estimated that CNN has reached a detection rate of 99.77% and this success is attributed to the advances and development of learning algorithms for deep neural network construction and development [.

TensorFlow is a data mining platform created by Google that can help build the neural network models to classify images, enables the user to define, tune and customize different types of CNN architectures [2].

II. RELATED STUDY

Deep Convolutional Neural Network Based Medical Image Classification for Disease Diagnosis

Image recognition and classification has become a big challenge in machine vision and has a long history with it. The challenges are given by a vast intra-class range of images given by size, shape, color and environmental conditions [5]. The medical image classification is playing an important role in evaluation tasks and clinical treatments, and the deep neural network a machine learning method proves it's potential for different medical classification tasks [4]. Recently in the medical field the Deep Neuronal Networks (DNN), especially the Convolutional Neural Networks (CNNs) are largely used in changing the image classification tasks and have achieved high performance exceeding the human experts [5]. The CNN-based methods are adopting different strategies like data augmentation, data transfer, and capsule network to increase the performance on image classification for small data sets [5].

III. METHODOLOGY

The study was performed on a number of 312 already classified Leukemia images, as can be seen from the Figure 1. The TensorFlow packages were imported into the system and applied on the images, the deep neural network (DNN) had to train all of these sets for the system to recognize any of the images. The classification was performed when the system started to test if each of those images belongs to any of the 4 Leukemia categories.

Total nr of images	M1	M2	M3	M5
312	125	72	25	92

Figure 1. Represents the number of images used and their classification

The total number of Leukemia images are categorized as M1 represents ALL, M2 is AML, M3 is CLL and M5 is CML.

The working environment was set by installing Python, Anaconda and Jupiter Notebook as a working platform for supporting the TensorFlow packages.

A typical process was followed when performing TensorFlow on image classification.

1. The data was pre-processed to generate input of the neural network, the images belonging to the same type of Leukaemia were grouped, labelled, adjusted at the same size, resized and reduced by 4 at each axis.
2. The system was tested by reading and displaying several images on the Python working environment platform Jupiter.
3. The TensorFlow Packages were imported into the system and applied on the images.
4. The accuracy of the system was tested applying Cross-Validation and K-fold.

```
Model: "sequential_74"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_222 (Conv2D)          (None, 238, 318, 32)      896
_________________________________________________________________
max_pooling2d_148 (MaxPoolin (None, 119, 159, 32)      0
_________________________________________________________________
conv2d_223 (Conv2D)          (None, 117, 157, 64)      18496
_________________________________________________________________
max_pooling2d_149 (MaxPoolin (None, 58, 78, 64)        0
_________________________________________________________________
conv2d_224 (Conv2D)          (None, 56, 76, 64)        36928
_________________________________________________________________
flatten_74 (Flatten)         (None, 272384)            0
_________________________________________________________________
dense_148 (Dense)            (None, 64)                17432640
_________________________________________________________________
dense_149 (Dense)            (None, 10)                650
=================================================================
Total params: 17,489,610
Trainable params: 17,489,610
Non-trainable params: 0
_________________________________________________________________
Train on 280 samples, validate on 32 samples
```

Figure 2. CNN Running Steps on Data Set

Figure 2 gives a summary of the experimental parameters of the TensorFlow based CNN. As can be seen from Figure 3, the results are highly variable, with a minimum of 27.3% and a maximum of 81.3%, which means the majority of classify would give an accuracy of 40.0% as can be seen from the Figure 1.

CV Fold	Accuracy
1	27.3%
2	81.3%
3	31.3%
4	62.5%
5	62.5%
6	77.4%
7	63.3%
8	80.0%
9	46.7%
10	33.3%
Min.	27.3%
Max.	81.3%
Mean	56.6%

Figure 3. CNN 10-Fold Cross-Validation Results

As can be seen from Figure 3, the results are highly variable, with a minimum of 27.3% and a maximum of 81.3%, which means the majority classify would give an accuracy of 56.6% as can be seen from Figure 2.

IV. CONCLUSIONS

The Python programming language, it's platform Jupyter and the TensorFlow packages have been used to design the system from the beginning till the end. The study reveals implementing deep learning by using its framework TensorFlow gives good results being able to train, simulate and classify the images with medium accuracy.
The experiment shows that the results after applying Cross-Validation are highly variable with a minimum value of 27.3% and a maximum value of 81.3%. The mean value is 56.6%, as further experiments are needed, perhaps evaluating different techniques that can be applied to the system like Smote Techniques. An increase in the accuracy value is expected trough applying the Otsu Method, which will be used to further preprocess the images to create more specific features.

REFERENCES

[1] H. Alcocer, "6 Strange Facts About Leukemia That TheyDon't Want You to Know", Lulu Press, vol.1, 2015.

[2] M. A. Azlan, N. I. Hazirah, H. A. Rahman, A. Sapiee, I. Ahmad, "A study on Image Classification based on Deep Learning and TensorFlow", International Journal of Engineering Research and Technology, pp. 563-569, vol. 12, 2019.

[3] K. Jiang, Q., M. Liao, Y. Xiong, "A novel white blood cell segmentation scheme based on feature space clustering", Soft Computing - A Fusion of Foundations, Methodologies and Applications: Dl.acm.org. Available at: https://dl.acm.org/doi/10.1007/s00500-005-0458-z.

[4] N. Ahmed, A. Yigit, Z. Isik, A. Alpkocak, "Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neuronal Network", PubMed, 2019.

[5] C. Wang, Y. Xi, (2018). "Convolutional Neural Networks and Image Classification | Content Simplicity". Semantic Scholar. [online] Available at: https://contentsimplicity.com/image-classification.

LogicGate Revised – A Student Centered Digital Game-Based Learning Experience

Thomas Coleman
Brunel University London
Uxbridge, United Kingdom
Thomas.coleman@brunel.ac.uk

***Abstract*—The LogicGate system is a Student Centered Digital Game-Based Learning experience designed to teach Boolean logic. The LogicGate System was deployed in a simulated classroom environment at Brunel University, London and evaluated by counterbalanced randomized trial, comparing it to a traditional student-centered task already in use. A mixed methods approach incorporated quantitative data (surveys, ability testing) and qualitative data (in-game chat, focus groups) to evaluate the system. Strengths included the delivery of SCL tenets, particularly those focused around collaborative and peer learning, with learning outcomes found to equal or exceed those achieved in the traditional task. Areas for improvement and recommendations for practice are included alongside opportunities for further research.**

Keywords—education; student centered; serious games;

I. INTRODUCTION

Student Centered Learning (SCL), a popular educational technique, is deployed to great effect within schools and universities worldwide [1]. SCL has been utilized in combination with digital video games in education, typically referred to as Student Centered Digital Game-Based Learning (SCDGBL) which is an emerging area of research concerned with the deployment of SCL tenets through digital gaming applications [3]. An effective SCL experience may be broken into seven key tenets: Active Learning (AL), Deep Learning and Understanding (DLU), Increased Responsibility (IR), Sense of Autonomy (SA), Teacher and Learner Inter-dependence (TLI), Mutual Respect (MR) and a Reflexive Approach to Teaching and Learning (RA) [2]. Although there are many promising applications within the field of SCDGBL, one drawback of existing SCDGBL offerings is a failure to deliver all seven tenets of student learning. In particular many offerings lack delivery of MR, TLI and RA, along with peer focused aspects of AL, all critical social aspects of SCL [3, 4].

This study presents the LogicGate system which has been designed to deliver all tenets of SCL, with particular focus on the above tenets identified as lacking in other offerings. The LogicGate application delivers a collaborative SCDGBL experience based on popular game Minecraft, designed to be applicable and extensible by education practitioners without need for programming skills. Within the LogicGate system, groups of three students proceed through a series of challenges. After a series of introductory tutorial challenges, students enter a collaborative area with a logic circuit visible on the floor. Each student has access to part of the logic circuit, meaning students must work together discussing what is visible and what inputs may be manipulated. Within the game, AL is realized through the active playing of the game, DLU is realized through the shared experience and the overarching storyline. IR is integrated through each student being crucial to the groups progress, while SA is delivered through the individual areas where a student has sole control. TLI is realized through active teacher participation, where the teacher can join in discussions and can act in the world on a similar level to the students, MR is found in the teamwork and discussion, where students must work with others and RA is realized through both the ability of students to return and practice in other area and the teachers ability to modify the game world and challenges.

To evaluate the LogicGate system in comparison to a traditional paper-based SCL exercise, the following research questions were formulated: **R1**: To what extent does the LogicGate system deliver all aspects of SCL? **R2**: Does the LogicGate system provide a more engaging gameplay and learning experience? **R3**: To what extent does the LogicGate system improve student knowledge?

II. RESEARCH METHODOLOGY

This mixed-methods study utilised a counterbalanced, randomized, within subjects design for effective comparison of the LogicGate system with an equivalent SCL task that is already used in practice to control for order effects. Informed consent was gained from volunteering Brunel computing students (Foundation/First Year) prior to taking part. Upon arrival at a Brunel laboratory used for standard practical teaching, participants (n = 32) were randomly assigned to conditions (A: Traditional task first, B: LogicGate System first) using an online coin flip generator. All participants took a short knowledge test on logic gates. Students in both conditions worked in sequentially assigned groups of three. In the Traditional task, student groups solved a series of logic puzzles, distributing the work between themselves to come to a final group solution. Those using the LogicGate system complete the puzzles together in-game. Post-condition, all students completed the Course Experience Survey (CES) [5] and eGameFlow survey (EGF) [6], rated using a 5-point Likert scale, and a second, identical knowledge test. Participants then crossed over to complete the other condition, completing CES, EGF and knowledge test post-condition. When all experiences had been completed, participants from both conditions (A/B) were invited to take part in a focus group, recorded and transcribed verbatim.

Quantitative data analysis was conducted using SPSS. Whole-group differences in CES and EGF construct scores between the post-LogicGate system and post-Traditional learning experiences were compared by two-tailed paired t-tests (R1, R2), utilized to explore differences in either direction. Knowledge improvement was assessed after each learning experience (R3). This was assessed by comparing test scores pre- and post-experience (one-tailed paired t-test). To compare knowledge improvement between learning experiences, a one-tailed paired t-test was performed on the gap scores (difference in pre- and post-experience test scores). Construct-level reliability was assessed by Mean Inter-Item Correlation to ensure validity of results [7]. Thematic analysis of qualitative data (7 focus group transcripts, recorded voice/text gameplay communications, free-text EGF/CES responses) explored the impact of game revisions upon the learning experience (R2). A hybrid approach utilised deductive a priori themes from qualitative data analysed in the pilot study, as a basis for initial coding.

III. Initial Findings

Of the five CES constructs (Table 1), four presented positive gap scores, indicating participants held positive views on the LogicGate system as compared to the Traditional task. Analysis showed Joint Productive Activity was higher than the Traditional task achieving significance (gap score = 0.245, p=0.018). Five of eight EGF constructs demonstrated positive gap scores, indicating users felt the game-based task was an improvement over the Traditional task. Feedback, Autonomy and Social Interaction constructs achieved statistically significant gap scores of 0.641 (p=0.001), 0.375 (p=0.006), 0.385 (p=0.011) respectively. This indicates that for these constructs, participants felt there was a notable improvement in their learning experience with the LogicGate system. Mean Inter-Item Correlations were within the acceptable range (0.15-0.7) for all constructs, indicating consistency. Constructs achieving significance have been bolded within Table 1.

Table 1: Comparative Survey Results by Construct

Construct	LogicGate system		Traditional Task		Comparison Statistics			
	Mean	S.D.	Mean	S.D.	Gap score	df	*t*	*p*
CES								
Complex Thinking	3.885	0.675	3.844	0.672	0.042	31	0.329	0.745
Contextualisation	2.854	0.742	2.854	0.798	0.000	31	0.000	1.000
Instructive Conv.	3.625	0.837	3.544	0.604	0.081	31	0.489	0.629
Joint Prod. Act.	3.737	0.530	3.479	0.675	**0.258**	31	2.494	**0.018**
Lang./Lit. Dev.	3.593	0.723	3.297	0.868	0.297	31	1.605	0.119
EGF								
Concentration	4.083	0.825	3.927	0.702	0.156	31	1.039	0.307
Goal Clarity	3.727	0.925	3.93	0.821	-0.203	31	-1.187	0.244
Feedback	3.781	0.805	3.141	1.116	**0.641**	31	3.715	**0.001**
Challenge	3.794	0.667	3.963	0.707	-0.169	31	-1.494	0.145
Autonomy	3.833	0.803	3.458	0.949	**0.375**	31	2.958	**0.006**
Immersion	3.227	1.113	2.93	1.078	0.297	31	1.516	0.14
Social Interaction	4.167	0.661	3.781	0.746	**0.385**	31	2.713	**0.011**
Knowledge Imp.	3.994	0.809	3.969	0.614	0.025	31	0.222	0.826

A thematic map (Figure 1) was developed of the five major themes surrounding the LogicGate system: Social Interaction, Knowledge Improvement, Perception, Increased Responsibility and Game Design. Identified subthemes are also shown.

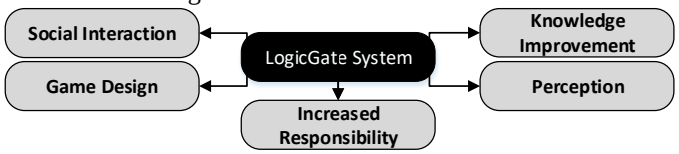


Figure 1: Thematic mind-map showing identified primary themes

Reaction to *Social Interaction* and *Knowledge Improvement* was broadly positive, with participants highlighting increased engagement and deeper learning. Students identified initial frustrations with adapting to teamwork, but expressed surprise at ability of their peers to understand and contribute when offered guidance. Perception covered students' session experiences, particularly in-game immersion and ability to concentrate on tasks, felt to be positive points, although concentration could lapse while waiting for team help to proceed. For *Increased Responsibility*, students felt they all should contribute to group progress and were all responsible. In *Game Design*, students expressed positivity regarding speed of feedback and information made available. Areas for improvement included a more detailed game mechanics tutorial and greater indication of current goals or team mate locations.

In knowledge test performance, the baseline performance was 3.03 (SD 1.68), this was improved upon by both conditions with a post LogicGate performance average of 3.41 (SD 1.27, p=0.036) and a post traditional average of 3.47 (SD 1.41, p=0.014). The improvement over baseline was significant, however the post experience scores were not significantly different from each other. This identifies that both tasks improved subject knowledge and can be considered to be equally effective teaching tools.

IV. Concluding Discussion

The study assessed how effectively the LogicGate system was able to deliver on all aspects of Student Centered Digital Game-Based Learning and how effectively it delivered the learning outcomes in comparison to the traditional task. In answer to the research questions: **R1**: The LogicGate system can form an effective part of a student centered classroom, delivering all aspects of Student Centered Learning. Tenets AL, MR, SA, IR and RA demonstrated notable improvements over the Traditional task. **R2**: Participants felt the game was engaging and provided an interactive experience that made good use of peer engagement, identifying additional areas for improvement. **R3**: The LogicGate system is an effective way to deliver learning, demonstrating equivalence to the Traditional task in knowledge improvement. These findings establish that the deployment of all tenets of SCL through SCDGBL is an achievable goal and that this can be achieved while maintaining educational outcomes. Fulfilling the promise of SCDGBL can lead to a more engaging classroom experience and form part of a modern education system. Future work on the LogicGate system may seek to establish the long term effectiveness of use of the LogicGate system particularly with regard to knowledge retention. Improvements to the game may include extending the game to allow a slower learning curve, and integrating testing into the system to further develop the reflexive learning aspect of the game. Finally the improved system may be deployed in a wider context to assess its practicality outside of close support by the researcher and to gather wider participation data. Such improvements may be addressed in a future revision of the LogicGate system and in future studies.

References

[1] F. S. Din and F. Wheatley, "A literature review of the student-centered teaching approach: National implications," in *National Forum of Teacher Education Journal*, 2007, pp. 1-17.

[2] S. J. Lea, D. Stephenson, and J. Troy, "Higher education students' attitudes to student-centred learning: beyond'educational bulimia'?," *Studies in higher education*, vol. 28, pp. 321-334, 2003.

[3] T. E. Coleman and A. G. Money, "Student-centred digital game–based learning: a conceptual framework and survey of the state of the art," *Higher Education*, July 19 2019.

[4] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," *Theory into practice*, vol. 41, pp. 64-70, 2002.

[5] K. Hyo-Mi, Lois A. Yamauchi, Seong-Ah, Im & Tracy Trevorrow, "Assessing Students' Perceptions of Instruction Designed for Culturally Diverse Classrooms," *Culture, Biography & Lifelong Learning* vol. 4, August 24th 2018.

[6] F.-L. Fu, R.-C. Su, and S.-C. Yu, "EGameFlow: A scale to measure learners' enjoyment of e-learning games," *Computers & Education*, vol. 52, pp. 101-112, 2009.

[7] R. DeVellis, "Factor analysis. Scale development, theory and applications," *Appl. Soc. Res. Method Ser*, vol. 26, pp. 10-137, 2003.

A Standard Cyber Security Model forArtificial Intelligence Platforms running SMART CITIES

Andres Crespo

The Intelligent Data Analysis Research Group,
Brunel University London
Uxbridge, United Kingdom
andres.crespo@brunel.ac.uk

Keywords: *artificial intelligence, cyber security, SMART CITIES, standards*

ABSTRACT —

- SMART cities, whatever their persuasion, are mainly data & technology driven

- Interconnectivity in live systems is now the norm and not the exception. That ranges from personal IoT devices to Transport Management Systems with anything in between

- Artificial Intelligence advances are being introduced replacing or at least upgrading expert systems at a high pace in the last decade.

The increasing use of Artificial Intelligence cannot be stopped or even delayed despite cyber security not having been fully developed and cyberattacks increasing exponentially with some getting through and already causing significant impact

Recent studies and industry solutions have come up with significant ways to protect a specific device or system. But despite the best efforts of governments and scientists and great Industry investment, the risk of cyberattacks posed in the cyber arena is as great as ever with no signs of abating. Cyber security is running behind cyberattacks.

And if scientists are missing a trick that does not allow cyber security to be in total control and ahead of the "game", is there a hypothesis that can be put to the test in order to reverse the currently negative status quo. No matter how protected a key system is, it is only as strong as the weakest device connected to it. With weak security that is difficult to upgrade in most vulnerable IoT devices, ranging from Home devices, wearables, CCTV/Webcams and recently connected vehicles (already hacked), this has been a so-far losing battle. Governments and Industries moving to AI-led solutions have not been offered a homogenous solution to the problem and consequently there are nearly as many cyber security solutions as types of interconnected devices.

Hypothesis: Artificial Intelligence can be better protected if the same security standard is developed and applied by all computer systems and IoT devices that are used to run a SMART CITY.

Research question: Can a standard cyber security model realistically be developed and applied to pervasive digital systems that are part of SMART CITIES?

AIM: This research will **aim** to develop and test a standard model that the industry or legislation can consider using in all their connected systems and IoT devices. This research also aims to offer a stepping-stone model that can be replicated, modified and/or progressed by Researchers and Industry alike.

Methodology:

- To conduct a systematic literature review of current studies in security systems for Artificial Intelligence solutions in the running of SMART cities using Kitchenham guidelines
- To conduct a comparative study of Industry-adopted Security Models for Artificial Intelligence Systems and in the devices and computer systems connected to them in aim of developing a standard model
- To test any suggested model in a real case scenario to be agreed with interested partners e.g. City of Barcelona

Data collection:

- Primary data discussed and agreement in progress and in principle with City of Barcelona and WARP (UK)
- Secondary data that is readily available from and other open sources

REFERENCES –

- Srivastava S., Bisht A., Narayan N. (2017). Safety and Security in Smart Cities Using Artificial Intelligence - A Review. 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence

- Brauna T.,Funga B., Shahb B.(2017) Security and privacy challenges in SMART CITIES. Sustainable Cities and Society Journal vol 39 p499-507 https://doi.org/10.1016/j.scs.2018.02.039

- Harish Kumar, et al (2018)., Moving towards smart cities: Solutions that lead to the Smart City Transformation Framework. Technological Forecasting & Social Change, https://doi.org/10.1016/j.techfore.2018.04.024

- Curzon J, Almehmadi A, El-Khatib K. (2019) A survey of privacy enhancing technologies for 2019 Pervasive and Mobile Computing. Pervasive and Mobile Computing 55 p. 76–95 https://doi.org/10.1016/j.pmcj.2019.03.001

- Zigurat Global Institute of Technology (2019) SMART CITY Series: The Barcelona Experience. https://www.e-zigurat.com/blog/en/smart-city-barcelona-experience/

- Nigon J, Glize E, Dupas D, Crasnier F and Boes J (2017). Use Cases of Pervasive Artificial Intelligence for Smart Cities Challenges. Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress. https://ieeexplore.ieee.org/abstract/document/7816955

Case-based reasoning to minimise the impact of disruptions on public transport

Monjur Elahi
Brunel IDA,
Brunel University London
Uxbridge, United Kingdom
md.elahi@brunel.ac.uk

Theodora Koulouri
Brunel IDA,
Brunel University London
Uxbridge, United Kingdom
theodora.koulouri@brunel.ac.uk

***Abstract* – The research aims to minimise the impact of public transport disruptions through the development of a transparent intelligent system that identifies historically similar disruption scenarios, predicts the impact of such disruptions and recommends optimal action(s). The research follows a Case-based Reasoning (CBR) task-method approach. A tool has been developed to collect real-time disruption and other train network data from TfL to build a library of past scenarios. An exhaustive memory search has been implemented in the Retrieve phase to find matches between a current scenario and past scenarios. Preliminary results show the algorithm can find match(es) of identical scenarios and rank them in order of 1. closest status severity, 2. closest time of day, 3. closest reason of disruption to improve the matches. Next stages of the research will focus on obtaining and ranking action data as well as building an interface for the operator to support transparent interactions with the user. The key contributions of this study are 1. building a transparent model for identifying and predicting the impact of disruption, 2. an active learning recommender system to continually improve the recommendations.**

Keywords— case-based reasoning; recommender system; public transport; active learning

I. INTRODUCTION

Public transport is an integral part of modern society. As cities expand, public transport networks become more complex and disruptions are more likely. Train disruptions have a serious impact on passengers and service providers. Therefore, minimising disruption will produce benefits in terms of quality of life (comfort), economy, and the environment.

The research aims to minimise the impact of public transport disruptions through the development of a transparent intelligent system that can identify historically similar disruption scenarios, predict the impact of such disruptions and recommend optimal action(s). The research follows the CBR approach (particularly the task method framework); this methodology can be generalised and has been successfully been applied to many domains such as healthcare [1], legal [2], cybersecurity, and trading.

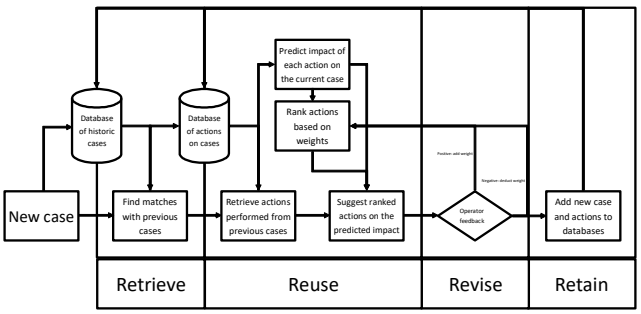


Figure 1: System architecture

II. RESEARCH METHODOLOGY

CBR is a classical artificial intelligence algorithm [3] and has been formalised as a four-step process: in particular, *retrieve* – given target problem, retrieve relevant cases to solving it; *reuse* – map solution from previous case to problem; *revise* – test new solution's effectiveness and adapt if need be; and *retain* – store resulting solution as new case. Within this research, the process is implemented as outlined below. Figure 1 also provides an overview of the process.

Retrieve - As part of the retrieve and reuse steps, a tool has been developed that has collected (and continues to collect) a library of data from TfL which captures disruptions on a minute by minute basis. The data contains every line in the London Underground network, their status severity (for example, good service, minor delay, part suspended, etc.). An exhaustive memory search algorithm has been implemented to find matches of similar disruptions based on a case disruption. To ensure soundness and completeness of retrieval, the memory search algorithm computes global similarity of case interval with every interval in memory.

Reuse – As disruption occurs on the network, historic cases will be retrieved based on the lag time (time difference between now and past time) [4]. The characteristics of the present case at the lag time will be compared to that of the historic cases to find a match, e.g., if lag time is 30 minutes, then two consecutive cases will be compared to find matches. Once a suitable match(es) is found, the characteristics of the actions taken during that case will be suggested along with the impact of that action.

Revise - The user (operator) chooses the action(s) from the recommendations based on the predicted impacts. They can then continue to monitor the situation to see the effects of the recommended action(s) they have just applied. The operator may also choose not to apply any of the recommendations and instead choose to apply their own expertise/judgement. The user can again monitor the situation and provide feedback to the system if their action proved better than (reduced the impact) the recommended actions.

Retain - If the recommended action(s) was of benefit to the user, then they provide feedback (whether useful or not) and the system will use it to learn and adjust future recommendations. If, however, the recommendations were not employed by the user, then their new action rendered by expertise/judgement will be added to the memory as a new case. This new case will then be presented in future recommendations.

For this project, a library of cases is created by aggregating the data into 15 minutes interval. The intervals will contain every line in the London Underground network, their status severity (for example, good service, minor delay, part suspended, etc.), the length of the said status severity, the reason for the disruption (for example, staff shortage, customer incident, trespassing, etc.) and possible areas that were affected. The intervals would be aggregated by frequency and by severity, where frequency refers to the most common status severity in that interval, and severity refers to the highest severity in that interval.

In addition to profiling similar disruptions, the project seeks to profile action taken by the operator to resolve a disruption. The data on actions taken is not yet available so its characteristics are still unknown, therefore the method for profiling the action taken remains to be determined.

The impact will be measured by the difference of changes in the volume of passengers during a disruption against a baseline [5]. The baseline is defined as a day where all services were run as close to a perfect schedule as possible and virtually without any disruption [6].

III. Initial Findings

The work to date includes the implementation of a matching hill climbing algorithm that can, given a case, find matches and rank them in the following order: 1. closest disruption, 2. closest time, 3. closest day and 4. closest reason. Figure 2 provides an example of a case with three other potential matches. The status severity is identical for all cases, suggesting an identical or similar disruption scenario across 3 days. The time difference between the Case interval and Match 1 is the shortest, therefore ranking it above Match 2 and 3. The search filters out any matches that occur on the same case interval day.

IV. Future Work

1. Analysis of past action – Past actions taken by operators will be analysed, summarised, categorised and aggregated to be in line with the cases. This is to aid the matching algorithm return correct matches quickly.

2. Runtime optimisation - Analysis within the project suggest it will seek to reduce the algorithm run time. This is in line to find matches in real-time. As it will aid quicker decision making by the operator and minimise disruptions.

3. Feature optimisation - The matching algorithm will seek to incorporate more variables to find matches that are richer, such as, the reason for the disruption. It is also in line to make the system more transparent.

This may give the user confidence in using the system, knowing that the system had considered several variables before arriving at the recommendations but also able to see the reasons for themselves and eventually build trust. The matching algorithm and its underlying data comprise the Recall and Reuse steps of the CBR methodology.

4. User interface development - The project will design, build and implement the recommender system to create a usable and transparent interface to support the interaction with the user. This will comprise the Revise and Retain aspect of the CBR method.

V. Conclusion

Initial results show a promising case for CBR in the domain of public transport. The matching algorithm in the retrieve phase performs well in returning identical matches for many criteria, which is required for accurate recommendations. The key contributions of this study are expected to include (i) a transparent model for identifying and predicting the impact of disruption and (ii) active learning recommender system for continuous action optimisation. The approach and proposed system have the potential to minimise the impact of public transport disruptions. The experiments also showed that the algorithm works best with a large volume of data, particularly when trying to match more variables, making it suitable for (big) data-intensive fields.

Estimated time of completion is Summer/Autumn 2021.

References

[1] C. L. Chang, "Using case-based reasoning to diagnostic screening of children with developmental delay," *Expert Syst. Appl.*, vol. 28, no. 2, pp. 237–247, 2005.

[2] E. L. Rissland, K. D. Ashley, and L. K. Branting, "Case-based reasoning and law," *Knowl. Eng. Rev.*, vol. 20, no. 3, pp. 293–298, 2005.

[3] C. S. Wang and H. L. Yang, "A recommender mechanism based on case-based reasoning," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 4335–4343, Mar. 2012.

[4] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. Part C Emerg. Technol.*, vol. 66, pp. 61–78, 2016.

[5] C. Malandri, A. Fonzone, and O. Cats, "Recovery time and propagation effects of passenger transport disruptions," *Phys. A Stat. Mech. its Appl.*, vol. 505, pp. 7–17, 2018.

[6] "Project NUMBAT A comprehensive multi-rail demand data set for London."

Case			Match 1			Match 2			Match 3		
10/08/2019 18:15 – 18:30			21/09/2019 15:15 – 15:30			21/09/2019 13:15 – 13:30			08/07/2019 00:45 – 01:00		
68667	Bakerloo	10	98529	Bakerloo	10	65651	Bakerloo	10	78503	Bakerloo	10
68667	Circle	4	98529	Circle	4	65651	Circle	4	78503	Circle	4
68667	District	5	98529	District	5	65651	District	5	78503	District	5
68667	Hammersmith & City	5	98529	Hammersmith & City	5	65651	Hammersmith & City	5	78503	Hammersmith & City	5
68667	London Overground	5	98529	London Overground	5	65651	London Overground	5	78503	London Overground	5
68667	Metropolitan	5	98529	Metropolitan	5	65651	Metropolitan	5	78503	Metropolitan	5
68667	Waterloo & City	10	98529	Waterloo & City	10	65651	Waterloo & City	10	78503	Waterloo & City	10

Figure 2: Algorithm matched and ranked three potential matches to the original case

Automating Camera Trap Analysis

Benjamin Evans
Department of Computer Science
Brunel University London
London, UK
Benjamin.Evans@brunel.ac.uk

Allan Tucker
Department of Computer Science
Brunel University London
London, UK
Allan.Tucker@brunel.ac.uk

Oliver R Wearn
Institute of Zoology
Zoological Society London
London, UK
Oliver.Wearn@ioz.ac.uk

Chris Carbone
Institute of Zoology
Zoological Society London
London, UK
Chris.Carbone@ioz.ac.uk

***Abstract*—Advances in image recognition and the growing interest in the use of camera-traps by ecologists have led to an explosion of ecological data and requires means to scale the image labelling, processing and analysis tasks in accordance. This study evaluates state-of-the-art techniques to assist ecological researchers in minimising human dependency in camera-trap processing and analysis. In particular it suggests an object detector to improve species classifications and bayesian networks to analyse movement patterns.**

***Index Terms*—camera-traps, inter-species behaviour, deep learning, bayesian networks, object detection.**

I. Introduction

In recent years ecologists have increasingly utilised camera-trap survey methods to estimate species abundance. The advancements in camera-trap hardware and a decrease in cost [2] has led to rapid growth in data.

Once images have been collected from a field site, they must be annotated with labels required for the study at hand. These labels often include the species name sighted amongst other visual clues [1]. Initially, the researcher would embark on the labelling process themselves, however, with growing datasets, it is becoming infeasible and there appears to be a limit on the number of citizen scientists available to assist [3].

Machine learning, specifically Deep Learning has presented promising results in sophisticated image analysis tasks including classification, object detection and segmentation [4] which shows potential for useful application to camera-traps.

II. Related Studies

A. Overview of Image Classification

Prior to Deep Learning approaches, image classification and detection were conducted by extracting features; extractors had to be manually developed by experts in the domain. Deep Learning utilises a convolutional approach to learning the patterns needed to be extracted automatically.

Thus, modern image classification usually requires training a Convolutional Neural Network (CNN), generally consisting of stacked convolution and pooling layers, ending with a fully connected layer. Depending on whether the problem at hand is a single or multi-class problem stipulates which activation function to be used on the final layer. The output layer provides outputs between 0 and 1 where generally anything above 0.5 is considered as containing the class represented.

Intelligent Data Analysis.

Detection takes image classification one step further by predicting a bounding box for each of the objects contained by the image. This provides ability for counting the number of objects of interest.

Detection generally takes the process of first generating a number of region proposals, each region is then passed through a CNN. Once all the CNN activations have been gathered, they are used to identify which regions might represent specific classes. The union of strong regions is then passed forward as the detections.

B. Review on Existing Methods

The majority of the prior work in camera-trap automation has been conducted using whole image classification. While this provides fair results, it lacks in the ability to count multiple individuals within an image or utilise the sequence of images camera-traps are often set to capture.

III. Methods

A. Datasets

The datasets we currently have available have been collected from Borneo, Royal Parks around London (Hampstead Heath and Home Park) and Costa Rica. The Borneo dataset consists of 500627 images taken from 605 locations. The Hampstead dataset includes 133273 images from 50 locations; Home Park consists of 481476 images from 147 locations and the Costa Rica dataset consists of 2119 images from 26 locations. Each image is about 5 megabytes. All datasets have had each image labelled with the species sighted and camera location. The Borneo dataset includes a subsection with bounding boxes.

B. Benchmarks

To provide a benchmark to compare future work with, multiple state-of-the-art CNN architectures, matching prior studies in the field were trained on a sub-sample of the Borneo dataset. Each class contained 1000 images of each species divided up by 60% training, 20% validation, 20% testing.

It should be noted that these benchmarks did not utilise cross-location training/test splitting which has now become the de facto method to test that the model can generalise between locations to identify species.

Due to the limited size represenation a CNN can intake on common hardware, an operation is required to reduce the number of pixels in an image passed to the network.

TABLE I
WHOLE IMAGE CLASSIFICATION BENCHMARKS

Architecture	Accuracy (Crop)	Accuracy (Resize)
Squeezenet 1.1	0.8118	0.7789
AlexNet	0.7451	0.7460
ResNet 18	0.8391	0.8259
ResNet 50	0.8620	0.8533
ResNet 152	0.8623	0.8586
VGG 16	0.8407	0.8471
Average	0.8268	0.8183

We found that the difference between crop and resize methods, shown in Figure 1, do not provide a significant increase in accuracy over the other as seen in Table I. We also found a lot of camera-trap imagery contains an animal on the outer region of an image that the crop method cuts out. Therefore, we continue by using the resize method.

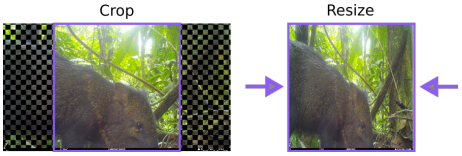


Fig. 1. Example Crop and Resize Operations.

By utilising Grad-CAM [5] to accredit which pixels the model is basing its prediction on, it is often seen that the background is focused on more than the animal. This suggests the model is overfitting to the background/location where there is a high frequency of one species.

C. Object Detection

To assist in negating any overfitting to the background at a location, object detection has been explored.

The MegaDetector is an object detection model for detecting 'animal', 'human' and 'vehicle' classes [6]. With these predictions, we are provided with the class and bounding box. Any bounding box containing an 'animal' is then extracted from its initial image and used for species classification.

Again, multiple state-of-the-art CNN architectures were tested.

TABLE II
DETECTION APROACH ACCURACY

Architecture	Accuracy (Resize)
Squeezenet 1.1	0.8138
AlexNet	0.7718
ResNet 18	0.9192
ResNet 50	0.9420
ResNet 152	0.9540
VGG 16	0.9215
Average	0.8870

As seen in Table II, the detection pipeline increases the classification accuracy and by analysing the predictions using Grad-CAM it is evident that the model has reduced its focus on the background and instead further utilises the pixels containing the animal.

D. Analysing Behaviour

In each of the datasets, every image is labelled with the location the camera was placed. Combining a location and species pair to create nodes, we learn a structure forming a Bayesian Network.

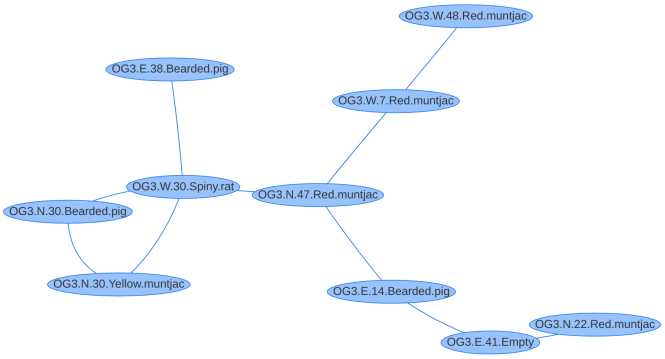


Fig. 2. Initial Bayesian Network of Location/Species in Borneo.

With a Bayesian Network learnt, it provides a potential for predicting movement patterns i.e. one species tracking another or general migration pattern over a region. An example part of a learnt Bayesian Network can be seen in Figure 2.

IV. CONCLUSION

The use of an Object-Detector provides superior results over whole image classification when generalisability between locations is also taken into consideration. This is due to minimising the background region which a whole-image classifier tended to overfit to in camera-trap imagery.

Given the output of an artificial neural network is a set of activations, these can be taken as a representation of the species sighted. When paired with locational data, this provides us with further opportunities to automatically identify patterns through the use of Bayesian Networks which can assist in achieving higher certainty in later sightings and open new possibilities in ecological research.

REFERENCES

[1] M. A. Tabak et al., Machine learning to classify animal species in camera trap images: applications in ecology. 2018.
[2] Rowcliffe, J.M. and Carbone, C. (2008), Surveys using camera traps: are we looking to a brighter future?. Animal Conservation, 11: 185-186. doi:10.1111/j.1469-1795.2008.00180.x
[3] GloverKapfer, P., SotoNavarro, C.A. and Wearn, O.R. (2019), Camera-trapping version 3.0: current constraints and future priorities for development. Remote Sens Ecol Conserv, 5: 209-223. doi:10.1002/rse2.106
[4] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436444 (2015). https://doi.org/10.1038/nature14539
[5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, Int. J. Comput. Vis., 2019, doi: 10.1007/s11263-019-01228-7.
[6] S. Beery, D. Morris, and S. Yang, Efficient Pipeline for Camera Trap Image Review, arXiv Prepr. arXiv1907.06772, 2019.

Computational Modelling of Quorum Sensing Driven Biofilm Formation in Bacteria: Representing Chemotaxis, Duplication and Death

Leila Ghanbar

Department of Computer Science, Brunel University London
Leila.Ghanbar@brunel.ac.uk
PhD supervisor: Professor David Gilbert

Abstract

We developed a detailed multilevel and multiscale dynamic model of intracellular and intercellular mechanisms of biofilm formation and bacterial communication. Then in an abstract model, we added movement, duplication and death *in silico*. To develop this model we then altered the random movement, to aimed movement toward food (chemotaxis). The question is can we model the communication between bacteria and the host using computational methods, Petri nets in particular, to describe the behaviour of bacteria, in different populations, when they move, reproduce and die. In this study we present the modelling of these three essential parts of a bacterial life in Petri nets and the results of these models.

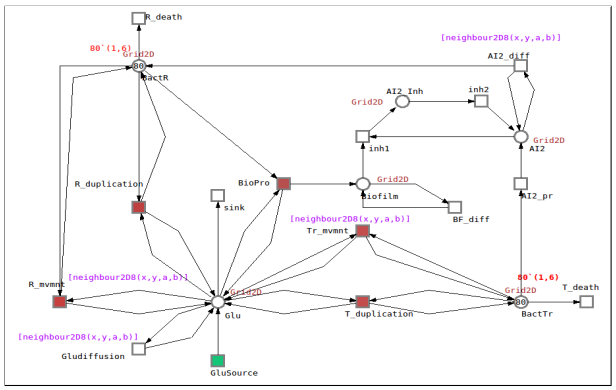


Figure 1: The final model of TR and chemotaxis, death and duplication.

1 Introduction

Studies have shown that bacteria communicate with each other and create societies in order to improve their life quality through quorum sensing and its consequences, changes in gene transcription, cell physiology and cell behaviour, such as biofilm [13, 7]. Biofilm is a dense matrix that bacteria produce to attach to the surface and/or each other to create a society in order to communicate and survive by making them resistant to antibiotics and other environmental changes [6, 12]. Biofilm a serious problem from medical and industrial perspective as it impacts human health in different ways. It can cause a problem as simple as plaque formation on the teeth or as complex as cancer in different parts of the body [9, 14, 11].

Modelling the structure of a biological system makes it possible to describe, analyse and predict the behaviour of the system spending less time and money while providing better understanding of the details behind bacterial interactions which can suggest new therapeutic methods for different diseases [9].

The importance of biofilm made us to consider designing a model which can describe, analyse and predict the process of quorum sensing and biofilm formation while presenting the other aspects of Bactria life such as movement, duplication and death, all together in one place for the first time. Our attempt is to design a multilevel, multiscale, multi-dimension and dynamic model of all the mentioned mechanisms in bacteria. After studying different methods and techniques of modelling for biological systems[1, 8, 3], we decided to use Petri nets as our platform in Snoopy Software considering that Snoopy can model space and provide location for a model in Coloured Petri nets[10, 5]. Our study is the first model which has used diffusion, space and location to represent these features of bacterial life. Before this, there had been some mathematical or computational models of the same biological data. But none of them present all the aspects we show in our study in one place.

2 Models

The Li-Ghanbar model is a complex and detailed model [2] of quorum sensing and biofilm formation which was published in 2019 in BMC Bioinformatics journal. To develop the model further, first we needed to make an abstract models, easy to study and computationally less expensive to analyse. We designed a library of spatial 2D and 3D dynamic models of required bacterial behaviour. These describe: movement – both random and directed (chemotaxis), duplication, and death while representing intracellular signalling via AI-2 and the formation of biofilm derived by quorum sensing in an abstract form. As quorum sensing refers to a society, we designed homo- and heterogeneous societies of bacteria, some just producing AI-2 and some receiving it to form biofilm, considering that not all the bacteria produce and receive AI-2 [13]. Also, we added glucose source, “Glu” in the model, which is representing the human cell providing food for bacteria which is mainly located on one side of the grid, far from the initial location of the bacteria.

After combining the transmitter and receiver bacteria models into the combination of these three to the result was an abstract model of two bacteria who move, duplicate and die while communicating via AI-2 and producing biofilm. The model could be seen in 1.

By analysing our Data we have got 3D plots which describe the amount of biofilm or other components of the model based on automatised functions.

3 Methodology

Petri nets are bipartite graphs used to study multilevel, dynamic and multi component biological networks with two different nodes: places and transitions, which can represent respectively metabolites and reactions. Each place can contain tokens which take a data value and in a biological system they can represent the number of molecules. This method can be used for

modelling and simulation of qualitative and quantitative analysis [10].

Coloured Petri nets are a type of Petri net in Snoopy which give Petri nets the asset of adding space and location to the model. It is designed for modelling in a multi-dimensional and multi-level environment and to do so, it gives every token in places a colour, or in other words a specific location. Each colour has its own specific properties of the programming language [4]. Coloured Petri nets also make modelling complicated repeated networks easier. In coloured Petri nets, one network is displayed as an abstract and folded version of the whole model and then it can be unfolded, in Snoopy by choosing the unfolding engine and simulate it [2]. We also used the time series data to generate 3D plots which are shown in Figure 2.

4 Results

Using R, and plot3D package and scatter3D function, we analysed the combined model which can be seen in figure 1. In this model the transmitter and receiver communicate with each other via AI-2 and produce biofilm while moving toward the food source, dividing as they get closer to the food source and dying on a 2D grid.

Using R we were able to show not only the amount, but also the location from a side view, see Figure 2. A general look at the plots shows that the concentration of bacteria and their products is around where Glucose is placed, on one side of the grid, which should happen in a biological lab as well. The bacteria move toward the food and accumulate there. Considering that this is the last time frame of the data, the bacteria still exist in other locations of the grid, because they have been duplicated. Yet the maximum number of bacteria in both receiver and transmitter are located by the food.

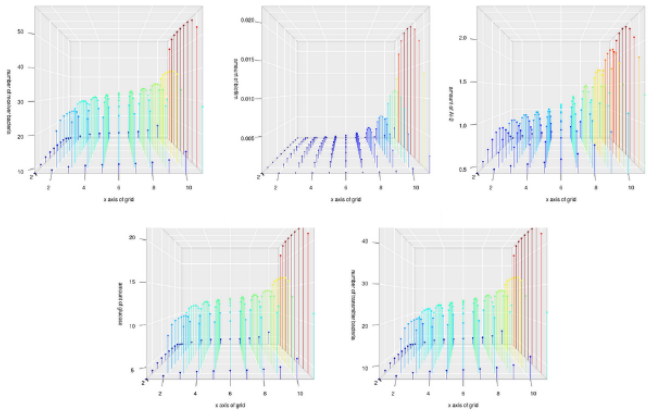


Figure 2: The 3D plots made by Rstudio shows the location and the number of bacteria while presenting the amount of biofilm, AI-2 and food in the last time frame in a 2D grid. the starting point of the bacteria, see in figure 2, was (1,6). Based on this visual analysis we can verify that the model is working as expected: The bacteria which started at (1,6) moved forward to the food which is placed like a strip at (11,1) to (11,11) at the end of the grid. The amount of biofilm is not much, since it is the last product being produced in the process and it will increase if we carried out the simulation for a longer time.

5 Conclusion

In our study we combined the models of movement, duplication and death and an abstract of quorum sensing and biofilm formation together in Petri nets in order to design a model which can predict the behaviour of bacteria. We showed that bacteria, driven by the food source, move toward it while communicating. the bacteria accumulate around the food source where they are settled and start production of biofilm in order to adhere to the surface. The next step shall be taking these models into details while presenting the intracellular mechanisms in a multi level and multi scale model.

These models need to be validated biologically, but when they are then they could be used in different studies such as drug discovery in order to predict the bacteria behaviour for example, when a specific drug is used on it, how the bacteria would react to it.

References

[1] J. Fisher and D. Harel. On statecharts for biology. *Symbolic Systems Biology: Theory and Methods*, 2010.

[2] D. Gilbert, M. Heiner, L. Ghanbar, and J. Chodak. Spatial quorum sensing modelling using coloured hybrid Petri nets and simulative model checking. *BMC Bioinformatics*, 20(4)(173):1–23, 2019.

[3] J. Hartmanis and J. V. Leeuwen. *Membrane Computing*, volume 9. Springer, Vienna, 6 edition, 2005.

[4] F. Liu, M. Heiner, and D. Gilbert. Coloured Petri nets for multilevel, multiscale and multidimensional modelling of biological systems. *Briefings in Bioinformatics*, 2017.

[5] L. Loewe and J. Hillston. Computational models in systems biology. *Genome Biol.*, 9(12):328, 2008.

[6] D. López, H. Vlamakis, and R. Kolter. Biofilms., 2010.

[7] L. McLandsborough, A. Rodriguez, D. Pérez-Conesa, and J. Weiss. Biofilms: At the interface between biophysics and microbiology. *Food Biophysics*, 1(2):94–114, 2006.

[8] R. Rojas. A Tutorial Introduction to the Lambda Calculus Raul Rojas FU Berlin, WS-97/98 Abstract. *Blackwell*, 2015.

[9] C. J. Seneviratne, C. F. Zhang, and L. P. Samaranayake. Dental plaque biofilm in oral health and disease. *The Chinese journal of dental research : the official journal of the Scientific Section of the Chinese Stomatological Association (CSA)*, 14(2):87–94, 2011.

[10] o. J. Shaw. *COMPUTING Modelling Bacterial Regulatory Networks with Petri Nets*. PhD thesis, 2007.

[11] M. G. Surette, M. B. Miller, and B. L. Bassler. Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: A new family of genes responsible for autoinducer production. *Microbiology*, 96:1639–1644, 1999.

[12] M. Toscano, R. De Grandi, M. Bortolin, L. Drago, and L. Drago. biofilm production and adhesion to intestinal cells: characterization of six lactobacillus strains isolated from italian probiotic products. *International Journal of Probiotics and Prebiotics*, 12(3):153–158, 2017.

[13] M. Walters and V. Sperandio. Quorum sensing in Escherichia coli and Salmonella. *International Journal of Medical Microbiology*, 296(SUPPL. 1):125–128, 2006.

[14] J. Washio and N. Takahashi. Metabolomic studies of oral biofilm, oral cancer, and beyond, 2016.

OT-Vision: Digitising Falls Risk Assessment through Mobile Depth-Sensing

Mr. Zear Ibrahim
Human Computer Interaction,
Brunel University London
Uxbridge, United Kingdom
zear.ibrahim@brunel.ac.uk

***Abstract* – This study presents OT-Vision, a mobile 3D depth-sensing enabled measurement guidance application for home falls risk assessments. A trial was conducted with trainee and registered Occupational Therapists to evaluate the accuracy of measurements recorded using OT-Vision compared with those recorded using current 2D state-of-the-art paper-based measurement guidance tools. The relative accuracy, efficiency and usability in terms of user satisfaction and attitudes were measured through a mixed methods approach. Results indicate that the paper guidance was accurate in seven out of 11 cases versus six out 11 cases for OT-Vision. In terms of accuracy consistency, the booklet outperformed the OT-Vision app in four of the 11 cases. Task completion and usability in terms of user-satisfaction reveal significant performance gains over the paper-based equivalent for OT-Vision. Collectively, this study has also identified opportunities to improve the future performance of OT-Vision, by incorporating computer-vision and passive-parallax approaches into the application.**

Keywords – Occupational Therapy, Depth Sensing, Healthcare

I. INTRODUCTION

In recent years it has been well understood that the UK population is growing and ageing rapidly. It now is the most significant driver for the ever-increasing social care needs [1]. Innovations in ICT applications to assist in healthcare delivery, have been affirmed to be a key strategy in addressing this growth [2]. A recent literature survey exploring the current state of the art in ICT applications for heath practitioner interventions found there is significant a lack of ubiquitous Mobile Health Sensing (mHealth, mSensing) systems specifically for supporting Occupational Therapy (OT) practice with the older population.[3]. OTs provide care to individuals of all ages whom suffer from the effects of disability caused by illness, ageing or accident in order to remain independent. To achieve this, Assistive Devices (AD) are prescribed as part of the Home Environment Falls-risk Assessment Process (HEFAP) where OTs gather information on the patient's functional abilities and measure key items of furniture to ensure correct fitment of ADs and prevent falls. The state of art current relies on a 2D paper-based guidance leaflet [4]. A set of 2D illustrations with annotated measurement arrows serve as prompts to indicate the precise points of measurement for OTs upon which further assessment is formulated to prescribe the necessary ADs. Despite the provision of detailed paper-based measurement guidance, there has been a 30% abandonment rate of prescribed ADs for patients, largely due to a 'poor fit' and as a result of measurement inaccuracies [5]. Whilst some promising depth sensing research has been presented in the remote patient monitoring domain, there is a distinct lack of research that explores and exploits the use of depth-sensing technologies specifically for HEFAP [6]. The intramural relationship between the measurement guidance and resulting measurement accuracy ostensibly plays an important role in reducing the risk of falls and helping older adults and persons with disabilities to remain living in their communities. Empirical evidence shows depth-sensing technologies can deliver increased measurement accuracy albeit in different health-contexts [7]. In response to the lacking depth-sensing measurement tools that can assist clinicians to better guide the HEFAP, a trial was conducted with a custom built mobile depth-sensing measurement guidance application that explored the clinical utility of its performance in terms of measurement accuracy and consistency, efficiency, usability and user satisfaction, compared with the state of the art 2D paper based equivalent.

II. RESEARCH METHODOLOGY

The aim of this pilot study was two-fold. First, to present the OT-Vision application. Second, through trainee and registered OTs, the indoor measurement accuracy of the app was evaluated in comparison with a 2D state of the art paper-based guidance leaflet, currently used in practice.

A. Research Questions

This study addresses the following Research questions (RQ):

RQ1	Does the OT-Vision app, on average, enable more accurate recording of measurements, compared with the paper-based measurement guidance booklet?
RQ2	Does the OT-Vision app enable more consistently accurate recording of measurements, compared with the paper-based measurement guidance booklet?
RQ3	Does the OT-Vision app enable measurements to be recorded more efficiently, compared with the paper-based measurement guidance booklet?
RQ4	How satisfied, in terms of usability, are users of the OT-Vision app, compared with the paper-based measurement guidance booklet?
RQ5	What are the OTs view of the OT-Vision app's perceived usefulness, challenges and opportunities and their intention on adopting this technology in practice?

B. Protocol and Instrumentation

This research has taken a within subjects counterbalanced design through a mixed methods experimental approach to collect data that can verify the accuracy and accuracy consistency of the measurements recorded from the depth-perception enabled system compared to the paper-guidance booklet. Five items of furniture were measured (bath, bed, chair, toilet, stairs) using both the OT-vision and the 2D paper-based equivalent measurement guidance. Upon completion, participants filled in a System Usability Scale (SUS) which included 10 standard questions about the clarity of the guidance they feel the respective measurement tools provided for the task of taking measurements. Further semi-structured post-task interviews were conducted with each participant to capture the user's views about the perceived usefulness, challenges and opportunities. Measurements recorded during the trials were compared against the true items measurements using Wilcoxon paired samples test, one-sampled and paired t-tests. SUS questionnaire responses were calculated using acceptability range, the adjective and school grading scales. A thematic analysis was applied to interview data.

III. SYSTEM ARCHITECTURE

The OT-Vision system has been deployed on a commercially available depth-perception enabled tablet utilizing active range laser sensors. Measurement capability has been enabled through Open-sourced Visual Inertial Odometry unit API's which are ubiquitously available [8], [9] and a physics engine by touching and placing 2D measurement markers on the screen. Interpolation between 2D measurement markers and the point-cloud data has been achieved via the implementation of a naïve

1NN (Nearest-Neighbour) Fixed Radius Linear Search algorithm labelled as NNFRLS presented in Table 1.

TABLE 1: PSEUDO-CODE: NNFRLS ALGORITHM

INPUT: S <PointCloudMatrix>, **inputVector** <x, y>, **PixelRange** <int>	
OUTPUT: An integer index of the PCM closest to the inputVector	
1	**SET** bestPCMIndex = -1;
2	**SET** bestSqrDistance = 0;
3	**FOR** i = 0 TO s.Count () - 1 **DO**
4	**SET** screenPos3D = **Dehomogenize** (S[i]);
5	**SET** screenPos2D = vector [screenPos3D.x, screenPos3D.y]
6	**SET** sqrDistance = **SquareMag** (screenPos3D-screenPos2D);
7	**IF** (sqrDistance > PixelRange * PixelRange)
8	**EXIT;**
9	**END IF;**
10	**IF** (bestPCMIndex == -1) **OR** (sqrDistance < bestSqrdistance)
11	**SET** bestPCMIndex = i;
12	**SET** bestSqrDistance = SqrDistance;
13	**END IF;**
14	**END FOR;**
15	**RETURN** bestPCMIndex;

Numbers on the left represent lines of code; The set 'S' refers to a Point Cloud Matrix

Key features of the algorithm pertain to performing a naïve (linear) iteration of each point cloud vector whereby 4-D Homogeneous coordinates, are de-homogenized to provide 3D spatial mapping in the local coordinate system. The input touch vector (x, y) is subjected to a pixel distance δ (delta) such that $\| x, y \| \leq \delta$ equals to all pairs $(x, y) \in S$ by which the distance between x and y is no more than δ and is an element of the set S. Upon completion, an index is returned that is closest to the input vector that satisfy $\| x, y \| \leq \delta$. The pixel threshold is an adjustable search range that is based on the average size of the pointer finger set to 16-20mm (45-57 pixels) [10]. Accordingly, OT-Vision has opted for an unobtrusive GUI overlay, which is visible at all times irrespective of the device's POV and positioning in the physical world with relation to object arrangement and depth presented in Figure 1.

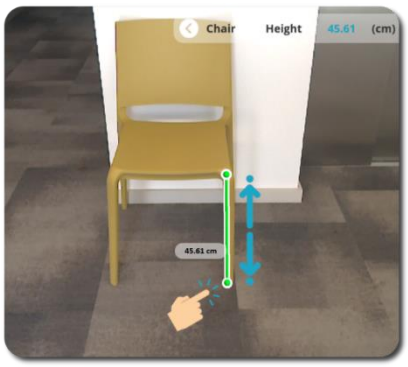


Figure 1: NNFRLS Algorithm Deployed on a Mobile Tablet

IV. RESULTS

The results of RQ1 indicate that in seven of the 11 cases, the paper-guidance produced accurate measurements, whereas OT-vision produced accurate measurements in six of the 11 cases. The results of RQ2 revealed that, when considering absolute median error differences by means of statistical significance, the booklet outperformed OT-Vision in four of the 11 cases. Although the differences in the remaining seven cases are not statistically significant, two cases lead to the booklet generating larger error differences. RQ3 revealed that the OT-Vision facilitated participants to capture individual measurements items significantly faster when compared to that of the booklet (Mean diff: 113.6s, p = 0.046). RQ4 revealed that OT-Vision achieved a significantly higher overall SUS score versus the booklet (73.7 vs 54.6 respectively). RQ5 revealed, in terms of Performance Expectancy, participants reported digitisation of the current guidance process and the resulting app is a crucial tool for OTs to engage with throughout the pre-assessment protocols. In terms of Effort Expectancy, participants were satisfied with the ease of use of the OT-Vision and that the intuitiveness of placing the start and end markers, were clear for the intended purpose of measuring items. Factors that affect practice and relating to Social Influence included OTs commenting on the OT-Vision having the ability to pave the way for inducing a gradual yet steady cultural shift within the community to purge negative habitual practices and implicit bias when performing measurements.

V. CONCLUDING DISCUSSION

This study presents an interactive mobile depth-sensing enabled digital measurement application (OT-Vision) utilizing active range sensors in conjunction with its system architecture for carrying out home falls risk assessments. Empirical mixed methods evaluations of the performance of OT-Vision revealed that in terms of accuracy and accuracy consistency, current paper-based 2D measurement guidance was marginally superior to that of OT-Vision. However, significant performance gains of using OT-Vision over current paper-based methods were delivered in the task completion, usability and perception measures in terms user satisfaction and attitudes towards adopting and using this new technology in practice. This research makes contributions in several areas, it demonstrates that mobile 3D depth-sensing technologies are a promising alternative to existing paper-based measurement practices as OTs appear to prefer the tablet-based system and that they are able to take measurements more efficiently. Although, it is evident that more work is to be done on improving accuracy and accuracy consistency, if it is to be used as a realistic alternative. Future versions of OT-Vision will look to exploit advanced computer vision and passive parallax techniques to improve accuracy and accuracy consistency.

REFERENCES

[1] AgeUK, "Briefing: Health and Care of Older People in England 2017," 2017.

[2] Department of Health, *Equity and excellence: Liberating the NHS (White Paper)*. London, 2010.

[3] Z. Ibrahim and A. G. Money, "Computer mediated reality technologies: A conceptual framework and survey of the state of the art in healthcare intervention systems," *J. Biomed. Inform.*, vol. 90, p. 103102, Feb. 2019.

[4] A. Atwal, A. Luke, and N. Plastow, "Evaluation of Occupational Therapy Pre-Discharge Home Visit Information Leaflets for Older Adults," *Br. J. Occup. Ther.*, vol. 74, no. 8, pp. 383–386, Aug. 2011.

[5] J. K. Martin, L. G. Martin, N. J. Stumbo, and J. H. Morrill, "The impact of consumer involvement on satisfaction with and use of assistive technology," *Disabil. Rehabil. Assist. Technol.*, vol. 6, no. 3, pp. 225–242, May 2011.

[6] J. Hamm, A. Money, and A. Atwal, "Guidetomeasure-OT: A mobile 3D application to improve the accuracy, consistency, and efficiency of clinician-led home-based falls-risk assessments," *Int. J. Med. Inform.*, vol. 129, pp. 349–365, Sep. 2019.

[7] G. Rogati, A. Leardini, M. Ortolani, and P. Caravaggi, "Validation of a novel Kinect-based device for 3D scanning of the foot plantar surface in weight-bearing," *J. Foot Ankle Res.*, vol. 12, no. 1, Sep. 2019.

[8] Google, "ARCore: Dynamic Depth," *Android Developer*, 2019. [Online]. Available: https://developer.android.com/training/camera2/Dynamic-depth-v1.0.pdf. [Accessed: 26-Sep-2019].

[9] Huawei, "HUAWEI AR Engine," *Huawei Developer*, 2019. [Online]. Available: https://developer.huawei.com/consumer/en/devservice/doc/20301.

[10] K. Dandekar, B. I. Raju, and M. A. Srinivasan, "3-D Finite-Element Models of Human and Monkey Fingertips to Investigate the Mechanics of Tactile Sense," *J. Biomech. Eng.*, 2003.

Detecting Online Information Weaponisation

Fawzia Zehra Kara-Isitt,
Intelligent Data Analysis Group,
Department of Computer Science,
Brunel University London
Uxbridge, United Kingdom
FawziaZehra.Kara-Isitt@brunel.ac.uk

Dr Stephen Swift,
Intelligent Data Analysis Group,
Department of Computer Science,
Brunel University London
Uxbridge, United Kingdom
Stephen.Swift@brunel.ac.uk

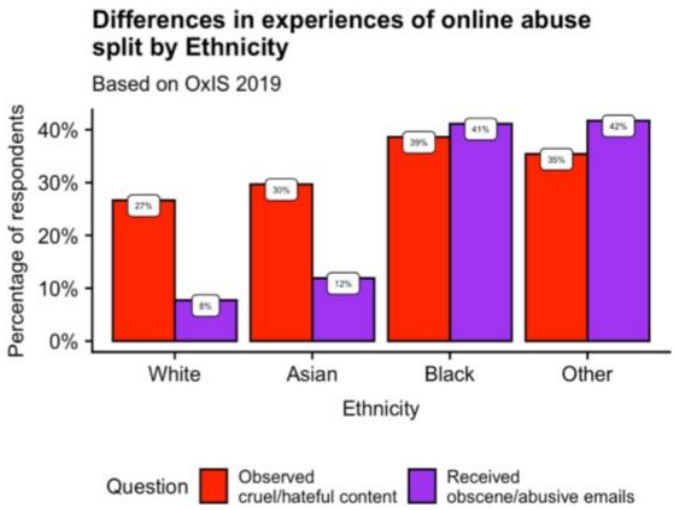
***Abstract*— This paper presents a review on sentiment analysis and related topics applied to detecting online harm within social media text. Social Media interactivity has led to a staggering amount of data being generated by the minute. Under the pretext of freedom of speech, some users have been given a whole new level of ammunition to harm to other people or victims behind the shield of a screen. The objective here is to spot attacking slurs, hate-based negative, bigoted and dangerous text. Intelligent algorithms are applied to captured content data, and after quantifying the context, to catch trends or localization of hate-speak. Findings have so far focused on the fact that part of the challenge is that, at present, the data, tools, processes and systems needed to effectively and accurately monitor online abuse are not fully available and the field is beset with terminological, methodological, legal and theoretical challenges [1]. This research will be tackling only tools, processes and the methodological systems as a challenge. Different approaches of classifying negative social media text into either 'Hateful', 'Offensive', and 'Neutral' are discussed using sentiment analysis and text mining via machine learning and data sciences tools. The inferred information from the research can then be applied to a practical predictive analytical service, by using the source or location of the speech and it's spread over a time span, to spot a possible trend coming on, based on specific requirements.**

Keywords—Classification; Detection ; Online Abuse; Online Harm; Information Weaponisation; Cumulative Extremism; Sentiment Analysis; Text Mining

I. INTRODUCTION

Online abuse, which includes both interpersonal attacks, such as harassment and bullying, and verbal attacks against groups (usually called 'hate speech'), is finally receiving more attention in the UK. [2], [3]. It poses a myriad of problems, including inflicting harm on victims who are targeted, creating a sense of fear and exclusion amongst their communities, eroding trust in the host platforms, toxifying public discourse and motivating other forms of extremist and hateful behaviour through a cycle of 'cumulative extremism' [4]. Assessing the prevalence of online abuse is a difficult task. Nonetheless, the UK public has expressed concern about its harmful effects: according to a 2019 survey by Ofcom, 18% of UK adults are concerned about hate speech on the internet; and a survey of 500 women in the UK showed that 61% of respondents believed online abuse/harassment of women is common and 47% stated that current laws are inadequate. A recent report from the Commission for Countering Extremism found that 56% of the public believe 'a lot more' should be done to counter extremism online [5]. Analysis of Oxford Internet Survey shows that experiences of online abuse vary considerably across demographics. eg: Ethnicity: Black people and those of 'Other' ethnicities are far more likely to be targeted by, and exposed to, online abuse than White and Asian people. See Figure 1 below [6].

Figure 1: Experiences of online abuse, split by ethnicity [6].



II. RESEARCH METHODOLOGY OF RELATED WORK

A. Data

Bag-of-words approaches tend to have high recall but lead to high rates of false positives since the presence of offensive words can lead to the misclassification of tweets as hate speech [7], depending on user's native slang perchance. *(eg. the noun 'burger' to eat in England is an adjective for a 'wannabe' in Pakistan or the word 'nigga' amongst the black ethnicity themselves is not insulting.)*

The research constructs scorable linguistic features from tweets and used to train a classifier. A random sample of 25k tweets containing terms from the lexicon were selected and had then manually coded by CrowdFlower moderators. Workers were asked to label each tweet as one of three categories: extreme hate speech, just simply offensive or neutral. The intercoder-agreement score provided by CrowdFlower is 92%. This results in a sample of 24,802

labeled tweets used for model training [8].

B. Features and Model

An automatic sentiment lexicon was designed for social media to assign sentiment scores to each tweet [10] to be used against the manual score. Scores were then tested with a variety of models that had been used in prior work: logistic regression, Naive Bayes, and linear SVMs being the main ones suggested [12]. Each model would be tested holding out 10% of the sample for evaluation to help prevent over-fitting. After using a grid-search to iterate over the models and parameters, Davidson et al. [9] had found that logistic regression with L2 regularization performed well for the final model as it more readily allowed to examine the predicted probabilities of each class membership. Probability across all classifiers was assigned to each tweet. All modeling was performed using Python scikit-learn [11]. This research will mimic Davidson et al. [9] to get the ambitious similar results of about overall precision 0.91, recall of 0.90, and F1 score of 0.90.

C. Results

Seeing as this research is only at its beginning, true findings are yet to be established for this study, however while reading Wiley Open Access journal article, *'Systematic literature review of sentiment analysis on Twitter using soft computing techniques'*, [12], it was stated that Machine Learning's sub category Deep Learning's Convolutional Neural Network technique yielded the best accuracy when applied to textual datasets over other soft computing techniques Naïve Bayes, Support Vector Machines and Logistic Regression. It is planned to apply that over the LogR with L2 model of this research.

III. +Expected Contribution and Initial Findings of The Experiment

Comparing and studying the vast online communication forums, like social media, which have enabled users to express themselves freely and behind the shield of the screen, attack others quite viciously, thus abusing the power of the freedom of speech. As such, many online forums such as Facebook, YouTube, and Twitter consider hate speech harmful, and have policies to remove hate speech content because of how widespread hate speech is becoming on the Internet [13], thus there is strong motivation to study automatic detection of hate speech. By automating its detection, the spread of hateful content and its negative tangible impact (suicide via bullying, terrorism, crime, genocide) can be reduced. Detecting hate speech is still a challenging task and literature reviews and journal publishing show its research is still needs more unique solutions and helpful findings to counteract the actual spread of online harm.

IV. Conclusions and Objectives

Evaluations using Java, Python scikit-learn and the R environment libraries with their classifiers will be used for the experiment on the initial recommended models. Online negativity location source by providing IP addresses in a study in 2019 [14] can also be replicated in the experiment. Sequential Pattern Mining techniques for predicting hate-speech paths within forum based social media discussions will also be used to harness a possibility of a viral harmful trajectory trend. Research timeline includes literature review, scraping and studying data with initial results with a published paper by summer 2020, followed by more deep learning and temporal abstraction experiments [15] with another published paper in 2021. Next, the final implementation will be applied to real time data and another publication beginning 2022 and finally, further actualized research findings with project and thesis completion deadline by November 2022.

References

[1] B. Vidgen, A. Harris, D. Nguyen, R. Tromble, S. Hale, and H. Margetts, "Challenges and frontiers in abusive content detection," no. Section 2, pp. 80–93, 2019.

[2] A. Brown, "What is hate speech? Part 1: The Myth of Hate," *Law Philos.*, vol. 36, no. 4, pp. 419–468, 2017.

[3] Z. Waseem, T. Davidson, D. Warmsley, and I. Weber, "Understanding Abuse: A Typology of Abusive Language Detection Subtasks," pp. 78–84, 2017.

[4] R. Eatwell, "Community Cohesion and Cumulative Extremism in Contemporary Britain - ProQuest Central - ProQuest." [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-923X.2006.00763.x. [Accessed: 28-Dec-2019].

[5] "Challenging Hateful Extremism," https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/836538/Challenging_Hateful_Extremism_report.pdf - Report [2019].

[6] G. Blank and W. H. Dutton, "Perceived threats to privacy online: The internet in Britain," pp. 1–23, 2019. The Oxford Internet Survey. https://oxis.oii.ox.ac.uk/

[7] I. Kwok and Y. Wang, "Locate the Hate: Detecting Tweets against Blacks," 2013.

[8] P. Burnap and M. L. Williams, "Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making; Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making."

[9] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," in *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, 2017, pp. 512–515.

[10] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 2014, pp. 216–225.

[11] F. Pedregosa FABIAN PEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," 2011.

[12] A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on Twitter using soft computing techniques," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, 2020, doi: 10.1002/cpe.5107.

[13] M. Mondal, L. A. Silva, and F. Benevenuto, "A measurement study of hate speech in social media," in *HT 2017 - Proceedings of the 28th ACM Conference on Hypertext and Social Media*, 2017.

[14] M. L. Williams, P. Burnap, A. Javed, H. Liu, and S. Ozalp, "Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime," Br. J. Criminol., Jul. 2019.

[15] D. Paul, F. Li, M. K. Teja, X. Yu, and R. Frost, "Compass: Spatio temporal sentiment analysis of US election what twitter says!," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. Part F129685, pp. 1585–1594, 2017.

An Iterated Local Search Approach to the Software Modularisation Problem

Ashley Mann, Mahir Arzoky, Stephen Swift
Brunel University, London, UK
{ashley. mann2, mahir. arzoky, stephen. swift}@brunel. ac. uk
https://brunel. ac. uk

James Cain
Grass Valley, Newbury, UK
{james. cain}@grassvalley. com
https://grassvalley. com

***Abstract*—The complexity of the software increases exponentially as software engineers continue development until a threshold is reached in which maintenance becomes unsustainable. As a result, code must be rewritten or restructured otherwise code becomes unmanageable. Software Modularisation is the problem of clustering dependencies with strong relationships to better understand and maintain software architecture. The motivation of this paper is to investigate whether Iterated Local Search (ILS) can be used to improve Modularisation. This investigation is part of a larger study aimed at mapping out the modularisation of a large collection of Industrial and Open Source data-sets. Our aim is to demonstrate evidence of the effectiveness and efficiency of ILS compared to common Hill Climbing strategies aimed at solving the modularisation problem.**

I. Introduction

As large software systems develop over time, documentation becomes increasingly unsuitable for maintenance with scarce information regarding the software architecture. Software engineers must therefore perform modifications to the code with possibly minimal understanding of the relationships between systems. As this pattern progresses large systems will either need to be redesigned or discontinued. The common technique to avoid complexity is to cluster code into individual subsystems. While segmenting software systems can be beneficial for software design; the representation of software clustering does not translate effectively in source-code when attempting to map out relationships. First defined by Mancoridis et al in 1998 [5], the Modularisation Problem is described as the exponential complexity of interconnected software module relationships within software systems as they increase in development overtime. To map out the relationships within a software system, Mancoridis et al [5] proposed a directed graph representation of the relationships between subsystems known as a Module Dependency Graph (MDG). Within an MDG, we refer to the nodes as the classes, and the edges are represented as the relationships between classes. An example of a relationship would be the call of a method from one class inside another. The creation and observation of an MDG does not inherently allow software engineers to greater understand the software architecture. On the contrary, MDG's can also can focus on specific classes rather than entire software systems. This allows for MDG's to be used to observe software structure over-time, or to be used to segment extremely large software systems for better understanding.

II. Related Studies

Since Mancoridis et al [5] first outlined the Software Modularisation Problem, there have been a variety of studies focused on it's effects within software systems. Our research focuses on two related studies which both focus on designing software solutions to help understand the software structure of software systems.

A. Bunch

To map out relationships within a Software System, Mancoridis and Mitchell et al designed Bunch [6]. Bunch is a Search Based Software Engineering inspired Software Visualisation tool which utilises Steepest Ascent Hill Climbing (SAHC) and Genetic Algorithms to map out the relationships within software systems. While this strategy is beneficial for accurate mapping of software systems, the SAHC has to search every neighbouring solution before continuing it's exploration of the search space. This therefore means Bunch excels at mapping smaller systems however, becomes ineffective at mapping large software systems. The quality of the produced MDG's is measured through a fitness function called Modularisation Quality (MQ). MQ works by finding the best partitions of an MDG by measuring all the interconnected components within clusters.

B. Munch

Inspired by Bunch, Arzoky et al [1] designed Munch. Munch similar to Bunch however Munch focuses on the mapping of software systems over time rather than singular systems. Munch uses a Random Mutation Hill Climbing (RMHC) Algorithm with a much more robust and faster fitness function allowing for much faster optimisation of larger software systems. Munch is also unique due to its modular design which allows for experimentation of other optimisation algorithms. To measure the quality of MDG's Munch uses two fitness functions known as EVM and EVMD. Designed by Tucker et al [7], EVM acts similarly to the MQ metric however EVM focuses on measuring the relationships between nodes compared to measuring the interconnected clusters. Designed by Arzoky et al [1], EVMD performs the same measuring score however the score is calculated with a "Try Swap" small-change operator. EMVD therefore scores a solution before performing the swap, reducing complexity.

III. The Experiment

A. Lunch

Our experiments are based on a custom built version of Munch called Lunch. Lunch is designed for mapping the relationships of software systems in the same way as Munch however our experiments will focus on single systems rather than over time. Lunch uses a Meta-Heuristic optimisation strategy known as Iterated Local Search (ILS). ILS aims to improve upon the effectiveness of optimisation algorithms by designing the next starting solution based on all the best components of the previous solutions attempted by the selected optimisation strategy. Our ILS uses RMHC as it's optimisation strategy therefore making it a Random Restart Hill Climb (RRHC). Our random starting solutions are created using Robust Clustering (RC). All former solutions form an Agreement Matrix of nodes within clusters which are in the same locations overtime therefore becoming robust. Unallocated nodes are then placed into clusters where agreement has acknowledged the node has improved the solution within that cluster previously. This enables us to focus on un-allocated nodes which have not been explored within the search space, thus widening our understanding of the entire search space. We use the same fitness metrics as Munch to score our final solutions.

B. Data-sets

The data-sets for our experiments provided by Grass Valley contain 6 industrial software systems and 14 open source software systems. Each data-set has been converted to an MDG for experimental purposes. EVM and the Homogeneity and Separation (HS) Metric will be our main metrics for scoring each optimisation strategy. HS was designed by Tucker et al [7] based on Chidamber and Kemerer's Coupling Between Object metric [4]. It is a measurement between -1 and 1 which represents the coupling (HS$<$0) and cohesion (HS$>$0) of a software system.

C. Experimental Setup

For our experiments, we ran all 20 MDG's 100 times on the three different strategies (RMHC, RRHC and ILS). RRHC and ILS experiments were split into three sets of 100 experiments to investigate whether changing the repeats can affect overall effectiveness. To keep the experiments consistent and fair, all experiments were run for 100,000,000 iterations each. In addition, our experiments included the testing of two different variants of Iterated Local Search with minor changes to the clustering algorithm. ILS-Random (ILS-R) is an ILS strategy with RC which randomly creates n clusters to add all un-allocated nodes with minimal agreement. ILS-Single (ILS-S) is similar to ILS-R however each un-allocated node with minimal agreement is placed into its own singleton cluster until the optimisation finds more robust agreements. For our initial experiments, we focused on the same fitness functions as Arzoky et al [1]. Our Experiments also document the runtime, the number of improved small changes (Updates) and finally, the end convergence point and duration of convergence (Flat-line).

IV. Results

TABLE I
Total Average and Standard Deviation of EVM and HS

Strategy	Repeats	Avg EVM	Avg HS	StDev EVM	StDev HS
RMHC	1	470.904	-0.454	320.729	0.180
ILS-R	5	479.082	-0.448	328.625	0.178
ILS-S	5	479.014	-0.448	328.547	0.179
RRHC	5	478.105	-0.447	327.308	0.180
ILS-R	10	478.392	-0.447	327.587	0.178
ILS-S	10	478.228	-0.447	327.480	0.179
RRHC	10	476.375	-0.448	325.128	0.180
ILS-R	25	473.244	-0.450	321.508	0.179
ILS-S	25	472.931	-0.450	321.197	0.180
RRHC	25	468.344	-0.452	315.933	0.181

TABLE II
Total Average statistics per experiment across all data-sets

Strategy	Repeats	Updates	Convergence	Flat-line	Time
RMHC	1	1020.86	5617941.249	4382057.752	01:21:15.780
RRHC	5	1042.4245	10118397.78	9881601.22	02:06:17.534
ILS-R	5	389.718	8413668.969	11586330.03	02:03:34.076
ILS-S	5	193.148	8260800.61	11739198.4	02:00:36.221
RRHC	10	1048.01	5785193.947	4214805.054	01:04:01.269
ILS-R	10	434.6935	5021854.937	4978144.064	01:01:00.729
ILS-S	10	172.2425	5035968.725	4964030.275	01:00:24.747
RRHC	25	1144.038	2608474.981	1391524.019	00:25:25.159
ILS-R	25	525.502	2411700.861	1588298.139	00:24:21.552
ILS-S	25	196.914	2435838.617	1564160.384	00:24:07.399

V. Summary and Future Work

Our results show that ILS is a competitive strategy with potential to improve over popular alternatives. Tables 1 and 2 suggest ILS scores higher EVM while maintaining similar HS with less updates and a smaller convergence. This suggests ILS can achieve similar, if not better results with less small-changes which can reduce computation. Our future work is focuses finding optimal parameters for running ILS. Our future work also includes implementing and experimenting with Consensus Clustering - a potential improvement over RC.

References

[1] Arzoky, M. , Swift, S. , Tucker, A. , Cain, J. : 'Munch: An Efficient Modularisation Strategy to Assess the Degree of Refactoring on Sequential Source Code Checking's', IEEE Forth International Conference on Software Testing, Verification and Validation Workshops (2011), pp. 422-429, 10. 1109/ICSTW. 2011. 87

[2] Arzoky, M. , Swift, S. , Tucker, A. , Cain, J. :'A Seeded Search for the Modularisation of Sequential Software Versions', Journal of Object Technology vol 11, no 2, 2012

[3] Arzoky, M. , Swift, S. , Counsell, A. , Cain, J. : 'An approach to controlling the runtime for search based modularisation of sequential source code check-ins', Lecture Notes Computer Science (2014), 25–36. 10. 1007/978-3-319-12571-8

[4] Chidamber, S. R. , Kemerer, C. F. :'A Metric Suite for Object Oriented Design', IEEE Transactions on Software Engineering, Vol 20, No 6, pp. 476-493

[5] Mancoridis, S. , Mitchell, B. S. , Rorres, C. , Chen, Y. , Gansner, E. R. : 'Using Automatic Clustering to produce high-level system organisations of source code', Proceedings. 6th international Workshop on Program Comprehension (1998), 10. 1109/WPC. 1998. 693283

[6] Mitchell, B. S. :'A Heuristic Search Approach to Solving the Software Clustering Problem (2002)

[7] Tucker, A. , Swift, S. , Liu, X. : 'Variable Grouping in Multivariate Time Series Via Correllation', IEEE Transactions on Systems, Man, Cybernetics, Part B: Cybernetics, 31(2), 2001, pp. 235-245.

The Software Modularisation Problem A Meta-Analysis

Faisal Maramazi, Ashley Mann, Stephen Swift, Mahir Arzoky
Brunel University, London, UK
{faisal.maramazi, ashley.mann2, stephen.swift, mahir.arzoky}@brunel.ac.uk
https://www.brunel.ac.uk

***Abstract*—**

The Software Modularisation Problem has been well documented in Search-Based Software Engineering (SBSE) since its conception emerged, however few publications have performed an in-depth meta-analysis of the current studies so far. Our study aims to perform a systematic literature review on previous publications focused on the Software Modularisation Problem to answer three key research questions with high relevance to research into the NP-Hard problem. Our study is determined to investigate the impact of the Software Modularisation Problem, the proportion of proprietary and open source datasets used to investigate the problem and the techniques investigated while researching the Software Modularisation Problem.

***Index Terms*—Software Modularisation, Systematic Literature Review, SBSE, EBSE,**

I. Introduction

A. What is the Software Modularisation Problem?

The Software Modularisation Problem refers to the problem of mapping software sub-systems based on the relationships between classes. The problem of clustering sub-systems together based on high cohesion and low coupling makes the Software Modularisation problem an NP-Hard optimisation problem. The problem was first defined by Mancoridis et al in the 1998 publication 'Using Automatic Clustering to Produce High-Level System Organizations of Source Code'. [1] In order to map within source code, Mancoridis et al [1] represented software systems as directed relationship graphs known as Module Dependency Graphs (MDG). These directed graphs can be represented as clustering arrangements which can be optimised using common heuristic optimisation algorithms. This technique of mapping software sub-systems allows practitioners to better understand the underlying architecture of software systems thus allowing for better maintenance or optimisation. The conception of the Software Modularisation Problem has lead to further research into optimisation algorithms and clustering techniques to better understand and structure software systems.

B. Why research the Software Modularisation Problem?

Since it's initial conception in 1998 by Mancoridis et al [1], research into the Software Modularisation Problem has provided a variety of research into areas such as optimisation heuristics and data clustering techniques. Due to the nature of software architecture, the Software Modularisation Problem focuses on clusters of clusters or, packages of classes. This makes the Software Modularisation Problem unique for research into better optimisation algorithms and clustering techniques, as the overall search space of a given software system would contain approximately $(2^{n-1})n!$ possible solutions. As an example, a small, open-source system such as a library with 20 classes, can be re-organised into approximately 1.28 x 10^{24} solutions. The importance of finding the optimal structure of software systems given such incomprehensible values may seem insignificant however, research into the problem has seen popularity due to further developments in larger, more complex proprietary and open source software systems. By segmenting code into areas of high cohesion, maintenance and debugging can be made more efficient.

II. Evidence Based Software Engineering

A. Systematic Literature Review

In 2004, Kitchenham et al, published a conference paper in the International Conference on Software Engineering which focused on the idea of Evidence-Based Software Engineering (EBSE). The primary aim of EBSE is to provide Software Engineering an infrastructure similar to common practices within the Medicinal Research which would allow for the empirical observation of research within Software Engineering. The main focus of Evidence-Based Software Engineering is to provide a Systematic Literature Review procedure for providing insight into specific areas of Software Engineering. Kitchenham et al (2004) state that the "goal of evidence based software engineering should be to provide the means by which current best evidence from research can be integrated with practical experience and human values in the decision making process regarding the development and maintenance of software". Thus, putting into consideration the results gained from best research evidence

B. Methodology

Our study follows the main methodology of an EBSE Systematic Literature Review. The first stage of a Systematic Literature as demonstrated in Figure 1 is the definition of a Research Question. Our study will be focusing on three main research questions with importance and relevance to the Software Modularisation Problem. Our study's three primary research questions are as follows:

- RQ1 - What are the types of techniques that are employed to solve the software Modularisation problem?
- RQ2 - What is the proportion of Open and Industrial datasets used in Software Modularisation Problem Research?
- RQ3 - What has been the impact of Software Modularisation Research for practitioners?

After the research questions have been defined; the objective of the systematic literature review is to find all relevant literature, narrow down the more relevant literature, then draft the review and reflections of the literature. To find the most relevant literature, criteria needs to be established first.

C. Inclusion / Exclusion Criteria

For our meta-analysis on the Software Modularisation Problem, we will be searching for relevant literature based on certain criteria as demonstrated in Stage 2 of Figure 1. Our main criteria is to focus on publications which focus on the Software Modularisation Problem. The phrase "Software Modularisation" does not inherently focus on the problem which we are focused on. This can affect the validity and relevance of our meta-analysis. This narrows down our research to publications relevant to our research questions. Due to the nature of the problem, our search will not be limited on conference papers and journals, to allow for relevant publications in the form of white papers. This further narrows our search for relevant literature associated with our research questions with the inclusion of industrial publications of relevance. In order to perform an extensive search for our Meta-Analysis study, the search strategy include manual search of major conference papers and journals and relevant online digital libraries. We have manually searched for all journals and conference papers relevant to Search-Based Software Engineering (1990-2019) and Evidence-Based Software Engineering repository. The following online databases were searched to find the most relevant papers: ACM Digital Library, SpringerLink, Science Direct, IEEE Xplore and Google Scholar. Based on our question research we came across different search criteria by using various combinations of Strings. The following search string were considered to be most appropriate for our study: Software Modularisation, Software Modularisation Problem AND Clustering, Software Modularisation AND maintenance, Software Modularisation AND optimisation. All search terms were also combined by using the Boolean 'OR' operator in order to include any one of the terms. Search string such as Software Modularisation, Software Modularisation Problem OR Clustering, Software Modularisation OR maintenance, Software Modularisation OR optimisation.

D. Ontology

The critical appraisal of the relevant literature associated to our primary research questions will be measured through an ontology which will enable us to measure the validity and relevance of the literature in comparison to our relevant literature. The ontology that our study will focus on will be a Mapping Study which will categorise literature based on relevance to our research questions.

III. Research Questions

A. What are the types of techniques that are employed to solve the software Modularisation problem?

There are several optimisation techniques that are deployed to resolve the software Modularisation problem. Techniques such as Stochastic Hill Climbing, Genetic Algorithms, Simulated Annealing and Random Mutation Hill Climbing will be considered.

B. What is the proportion of Open and Industrial datasets used in Software Modularisation?

We would like to concentrate on both proprietary (industry) and open source datasets / source code, and investigate factors such as the type, size and format of the datasets.

C. What has been the impact of Software Modularisation Research for practitioners?

We look to investigate the impact that the software modularisation research had on practitioners from different aspects such as feedback from practitioners themselves as well the validity and impact of the results. the Evidence-based Software Engineering steps are as follows: Step 1: Establish the Search Questions to be answered by selected information. Step 2: track the most suitable evidence with strong evidence supporting the Research Questions. Step 3: critically appraise evidence based on validity, impact and capability. Step 4: integrate the critically appraisal evidence with our expertise relevant to the Research Questions. Step 5: evaluate out efficiency and effectiveness in executing step 1-4.

IV. CONCLUSION:

This paper presents a meta-analysis about the Software Modularisation Problem. The Software Modularisation attempts to assign modules of the system with low coupling and high cohesion. The Author conducted comprehensive search and comparison among the selected papers and found out the proportion of open and industrial dataset and techniques been used for software modularisation problem solving.

References

[1] Mancoridis, S., Mitchell, B.S., Rorres, C., Chen, Y., Gansner, E.R. (1998) 'Using Automatic Clustering to produce high-level system organisations of source code', Proceedings. 6th international Workshop on Program Comprehension, doi: 10.1109/WPC.1998.693283

[2] Kitchenham, B. et al. (2009) 'Systematic literature reviews in software engineering - A systematic literature review', Information and Software Technology. Elsevier B.V., 51(1), pp. 7–15. doi: 10.1016/j.infsof.2008.09.009.

[3] Mitchell, B. S. (2003) 'A Heuristic Approach to Solving the Software Clustering Problem', IEEE International Conference on Software Maintenance, ICSM, pp. 285–288. doi: 10.1109/icsm.2003.1235432.

An empirical examination of mobile banking use: a cross-national, quantitative study

Mohamad Merhi
Department of Computer Science,
Brunel University London
Uxbridge, United Kingdom
mohamad.merhi@brunel.ac.uk

Prof. Kate Hone
Department of Computer Science,
Brunel University London
Uxbridge, United Kingdom
kate.hone@brunel.ac.uk

***Abstract*— Subpar mobile banking adoption rates persist despite the advantages of this platform. A cross-national examination of adoption intentions and usage behaviour and their variation across two contexts (UK and Lebanon) was undertaken in this study to address existing gaps in the mobile banking acceptance literature. An extended UTAUT2 conceptual framework was proposed, and partial least squares-structural equation modelling revealed that habit, performance expectancy, and perceived security consistently predicted behavioural intention to adopt mobile banking across both examined contexts. Country-specific variations in user behaviour were also demonstrated on the level of the remaining model factors and moderators, while certain previously established factors (i.e. social influences, hedonic motivations, effort expectancy, system quality and information quality) failed to explain adoption behaviour in both countries. Our findings suggest the importance of establishing secure and useful mobile banking applications integrating habit-forming elements, irrespective of the setting. Careful consideration and specific targeting of potential mobile banking users according to age groups and income range could favour its adoption in England, while a gender- and experience-oriented approach seems to be more appropriate in Lebanon.**

Keywords*— *Mobile banking, Technology adoption, UTAUT2, Developing countries, PLS-SEM

I. Introduction

Ubiquitous technology usage is a hallmark of the modern world, along with increasing reliance on digitised services. Mobile banking is one example of the latter, providing users with a seamless, simple platform to conveniently complete routine financial activities. Regardless, subpar adoption rates have been frequently observed, with distinct variations in mobile banking acceptance noted across different settings.

Mobile banking is a promising innovative service presenting with many benefits to banks and customers alike [1]. Essentially, mobile banking confers constant availability of banking services through handheld mobile devices which are not limited to working hours, while offering a higher degree of privacy and accuracy in processing financial transactions [2].

Despite its benefits, customers continue to regard mobile banking services with caution and reluctance [3]. Frequent deterrents to Mobile Banking adoption by consumers were related to the technology's perceived ease of use and effectiveness, the skills necessary for its use, as well as security risks associated with this novel technology [4][5].

The antecedents of mobile banking adoption behaviour and actual usage have been described in IS literature within the conceptual framework of various models. The extended unified theory of acceptance and use of technology (UTAUT2) successfully incorporated relevant concepts from previous models (e.g. Examples include the Theory of Reasoned Action (TRA), and the Technology Acceptance Model (TAM)) into its framework. It emphasised the importance of accounting for cultural, social as well as technological factors, which are invariably different in developed and developing countries [6]. Context-dependent variations in construct relevance can be traced back to the study setting, country development level, nature of technology, as well as individual factors [7][8]. However, the role and interplay between different adoption constructs remain poorly understood, particularly in developing countries. Moreover, the UTAUT2 neglects to account for important quality, security and privacy considerations that remain critical for technology success and acceptance. To that end, this study undertook the examination of key UTAUT2 and IS success factors that may hinder or facilitate the adoption of mobile banking services in the cross-national context of Lebanon and the UK, while incorporating security privacy and trust.

II. Research Methodology

A positivist quantitative approach was deemed appropriate for the purposes of this study. The latter undertook the objective interpretation of generated quantitative data for the validation or rejection of the formulated hypotheses within the structural model [9]. A pilot study reflected the validity and suitability of the designed questionnaire and allowed its amendment [10][11]. Using convenience sampling, a total of 478 and 419 completed questionnaires were collected from Lebanon and the UK, respectively. Data was then analysed by partial least squares-structural equation modelling.

III. Findings

As shown in Table I, findings revealed that habit, performance expectancy and perceived security consistently predicted behavioural intention to adopt mobile banking across both examined contexts. Lebanese respondents were additionally concerned about the availability of facilitating conditions and trust when formulating their intentions to adopt this channel, while perceived privacy, price value and service quality concerns were evident in their British counterparts

TABLE I. Results Summary

	Constructs	Lebanon		United Kingdom	
		P Values	*Sign Levels*	*P Values*	*Sign Levels*
Behavioural Intention	Effort Expectancy	0.98	N/S	0.811	N/S
	Facilitating Conditions	0.01	*	0.287	N/S
	Habit	0	***	0	***
	Hedonic Motivations	0.107	N/S	0.208	N/S
	Social Influence	0.764	N/S	0.389	N/S
	Performance Expectancy	0	***	0.002	**
	Price Value	0.687	N/S	0.021	*
	Perceived Privacy	0.23	N/S	0.015	*
	Perceived Security	0	***	0.0005	**
	Trust	0.007	***	0.15	N/S
	Information Quality	0.784	N/S	0.09	N/S
	System Quality	0.22	N/S	0.112	N/S
	Service Quality	0.078	N/S	0	***
Moderators	Age		*		***
	Gender		***		N/S
	Education		N/S		N/S
	Experience		***		N/S
	Income		*		***

[a.] Note: *p < 0.05; **p < 0.01, ***p < 0.001. N/S: not significant

However, contrary to the proposed expectations, social influences, hedonic motivations, effort expectancy, system quality and information quality failed to significantly explain adoption behaviour in both Lebanon and the UK. Notable differences were observed on the level of moderator influences between Lebanon and the UK, with the exception of education, which remained insignificant in both (Table I). Age and income-mediated variations in adoption intentions were primarily evidenced in the UK, while gender and experience only played a significant role in the developing country context of Lebanon.

IV. Conclusion

Our findings suggest the importance of establishing secure and useful mobile banking applications integrating habit-forming elements regardless of the examined context. In fact, the improvement of available offerings is expected to encourage increased mobile banking usage, which in turn would contribute towards habit building. Similarly, frequent push notifications and reminders of available features in mobile banking applications, such as the ability to remotely check monthly statements, transfer funds and pay bills, would promote habitual mobile banking use. Age, gender, experience and income emerge as notable moderators of adoption behaviour, while the role of well-established constructs such as social influence and effort expectancy seems to wane in modern-day settings. Theoretically, security, privacy and trust concerns remain critical variables in financial dealings and their addition notably improved the UTAUT2's explanatory power in a cross-cultural setting, while achieving good model fit. This study also demonstrated the applicability of an extended UTAUT2 model in both a developed and developing context. Moreover, it reflected distinct country-specific variations in technology adoption, further highlighting the importance of accounting for cultural factors when examining mobile banking adoption.

However, larger samples could provide better insights into mobile banking adoption in different cultural contexts. Cultural factors could also be further examined through the integration of Hofstede's cultural dimensions into the research model. Our findings indicate the necessity of accounting for country-specific variations in mobile banking adoption intentions when designing platforms and campaigns. Careful consideration and specific targeting of potential mobile banking users according to age groups and income range could favour its adoption in the UK, while a gender- and experience-oriented approach seems to be more appropriate in Lebanon.

References

[1] A. A. Shaikh, P. Hanafizadeh, and H. Karjaluoto, "Mobile Banking and Payment System: A Conceptual Standpoint," vol. 13, 2017.

[2] P. Ahluwalia and U. Varshney, "Composite quality of service and decision making perspectives in wireless networks," vol. 46, no. 2, pp. 542–551, 2009.

[3] A. A. Alalwan, Y. K. Dwivedi, and N. P. Rana, "Factors influencing adoption of mobile banking by Jordanian bank customers: Extending UTAUT2 with trust," vol. 37, no. 3, pp. 99–110, 2017.

[4] S. A. Raza, A. Umer, and N. Shah, "New determinants of ease of use and perceived usefulness for mobile banking adoption," vol. 11, no. 1, pp. 44–65, 2017.

[5] J. Mullan, L. Bradley, and S. Loane, "Bank adoption of mobile banking: stakeholder perspective," vol. 35, no. 7, pp. 1154–1174, 2017.

[6] V. Venkatesh, J. Y. Thong, and X. Xu, "Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology," pp. 157–178, 2012.

[7] A. Burton-Jones and G. S. Hubona, "The mediation of external variables in the technology acceptance model," vol. 43, no. 6, pp. 706–717, 2006.

[8] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," vol. 46, no. 2, pp. 186–204, 2000.

[9] W. Chen and R. Hirschheim, "A paradigmatic and methodological examination of information systems research from 1991 to 2001," vol. 14, no. 3, pp. 197–235, 2004.

[10] U. Sekaran and R. Bougie, Research methods for business: A skill building approach. John Wiley & Sons, 2016.

[11] J. W. Creswell and J. D. Creswell, Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications, 2017

A Model-Based Consensus Clustering For Predicting Number of Partitions in Datasets*

Afees Adegoke Odebode
Department of Computer Science
Brunel University,
London, UK
afees.odebode@brunel.ac.uk

Stephen Swift
Department of Computer Science
Brunel University,
London, UK
stephen.swift@brunel.ac.uk

***Abstract*—The availability of extensive data set from the web and social media sources to machine-generated content and advances in storage methods, present a new set of challenges in data mining, especially clustering and anomaly detection. This study is an ongoing examination of algorithms: Proximity-based approach, Density-based methods, Nearest-neighbour approach, Ensembles Methods, Principal Component Analysis and Singular Value Decomposition towards adapting them to the requirement of the new data sources, (Medical, Educational, and Financial data set) volumes and varieties using specific examples to illustrate challenges and solutions. This aspect of the research is concerned with determining the number of clusters in datasets and sample results obtained is quite promising.**

***Index Terms*—Consensus Clustering, Ensemble, Sampling.**

I. Introduction

Clustering is the process of determining the intrinsic structure of a dataset when there is no other information other than the values of the observation [1]. It is an application that determines partitions in a dataset based on distance and correlation metrics [2]. In determining the appropriate clustering methodology, researchers need to understand or acknowledge the effect of the underlying data analytic characteristics exhibited by the clusters they are aiming at identifying. A major problem with earlier methods such as K-means and other partition-based algorithm is that the number of clusters is included in the cluster identification process. This may result in generating varied partitions which depends on the formation and interpretation of the cluster patterns. As a result of this variation, there is a need to harmonise or aggregate the results to obtain a better result that summarises the input clustering(s). One approach to this is Ensemble clustering. Ensemble clustering is a clustering technique that seeks to improve clustering performance by generating partitions from multiple clustering algorithm or multiple iteration of a clustering algorithm and combining them to create a summary clustering solution. Ensemble clustering, like all other ensemble analysis [3], is a method that is commonly used in clustering dataset to reduce the dependence of the model on the specific dataset or data locality. Consensus Clustering(CC), a heuristic version of Ensemble clustering, integrates solutions from numerous distributed sources of data or attributes to generate presumably better grouping. In this paper, we explored the ability of CC to integrate both weak and strong solutions into a representative solution to correctly estimate the number of clusters in datasets. We used a K-Mean-based CC to generate models for predicting the number of clusters in datasets. The majority of studies use metrics that are based on earlier methods such as Gap and Average Silhouette, and this is oftentime data dependant (Spherical data) or distribution dependent. This work is an attempt to unify the process of finding the clustering arrangement in datasets using the consensus from the input clustering, which CC approach represents well. Two-hundred and ninety-two (292) publicly available datasets of different sizes, volumes and varieties downloaded from UCI and other machine learning repository with a known number of clustering arrangement were used as test-bed for the experiment. The datasets were later reduced to one-hundred and thirty-seven (137) during the initial clustering and cleaning process to remove those without clear clustering structure or those that fail to cluster. We choose to estimate the number of partitions using the CC approach for two reasons. First, CC has been shown to correctly aggregate all possible input solution into one representative solution that accurately represent the final solution. Second, CC provides a natural way to define the similarity measure between the different clustering from the base algorithm(s). Although we have not eliminated the requirement for the number of clusters(represented by *r*) in the clustering process, our approach has reduced the problem to guessing another unknown parameter - the best number of input that correctly predicts the number of clusters in the dataset. The rest of the paper is organised as follows: Section 2 is a short description of the methods. Section 3 details the results of all of the experiments and compares them with existing methods. Section 4 contains the conclusion and recommendations for future work.

II. Description of the Method

The definition and description of the clustering arrangement used in this work is adapted from [4]. The clustering arrangement G is defined as $G = [g_1, ..., g_m]$ for a list of variables $x_1, ...x_n$, : $x_i = x_j \iff i = j$. A cluster is defined as $g_i \subseteq x$. $i = 1, ..., m$ is a partition of x and the union of all g_i is x; additionally $g_i \cap g_j \neq \phi\{\forall i \neq j\}$. The cardinality of g_i is defined as s_i.

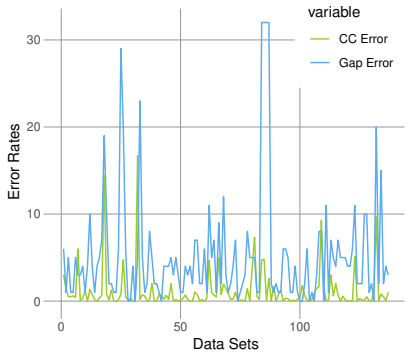


Fig. 1. Comparison of Error rate between Gap Statistics and CC.

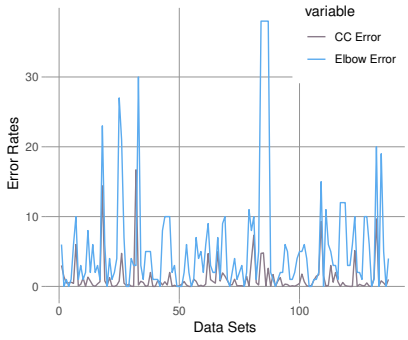


Fig. 2. Comparison of Error rate between Elbow and CC..

CC consists of two stages: a pre-processing stage and an optimisation stage. The pre-processing involves constructing an agreement matrix from the input clustering. In our case, we only use the K-means algorithm to generate the input clusterings, in the subsequent approach, we might want to create a data specific input to improve the input clustering. The second stage is the optimisation stage which searches for optimal clustering arrangement using a fitness function applied to the agreement matrix. We use Simulated Annealing to perform the optimisation because it has the potential to determine the global optimum.

III. Results

This section discusses the results and the derived models from the CC algorithm results. The results section is divided into three parts. The first[Figs(1,2,3)] part shows the performance of the CC compared to each of the three algorithms in terms of predicting the number of clusters in the dataset and the error rate (the difference between the predicted value and the actual cluster number predicted) is used to measure the accuracy of the prediction. The second aspect of the result Fig4 is on the performance of the derived models from CC compared with the three standard algorithms based on the number of right prediction across all of the datasets.

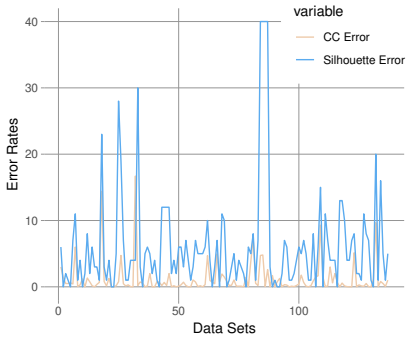


Fig. 3. Comparison of Error rate between Silhouette and CC.

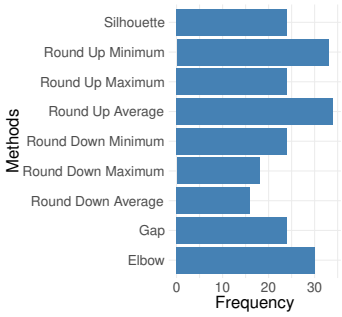


Fig. 4. Average Performance:Models,Gap, Silhouette and Elbow

IV. Conclusions and Recommendations

This paper has introduced a model-based heuristic to predict the number of partitions in datasets. This new approach, rather than require an input parameter or a range of input parameter offered by CC, have reduced the problem to determine the best number of inputs. Hence the model we present. The result shows that a better estimate of the number of clusters contained in a dataset can be determined using the models generated from CC results and with a better model, the results are promising. However, the result of the average error rates of the dataset was not substantially different from the previous methods. The new models were much better when individual dataset studied were taken into consideration. The result indicates that with an improved model, the predicted results may be more promising than the ones currently obtained.

References

[1] C. Fraley and A. E. Raftery, "How many clusters? which clustering method? answers via model-based cluster analysis," *The computer journal*, vol. 41, no. 8, pp. 578–588, 1998.

[2] S. Swift, A. Tucker, J. Crampton, and D. Garway-Heath, "An improved restricted growth function genetic algorithm for the consensus clustering of retinal nerve fibre data," in *Proceedings of the 9th annual conference on Genetic and evolutionary computation*. ACM, 2007, pp. 2174–2181.

[3] C. C. Aggarwal and S. Sathe, *Outlier ensembles: An introduction*. Springer, 2017.

[4] S. Swift, A. Tucker, V. Vinciotti, N. Martin, C. Orengo, X. Liu, and P. Kellam, "Consensus clustering and functional interpretation of gene-expression data," *Genome biology*, vol. 5, no. 11, p. R94, 2004.

DO-U-Net for Segmentation and Counting

1st Toyah Overton
Department of Computer Science
Brunel University
London, England

2nd Allan Tucker
Department of Computer Science
Brunel University
London, England

***Abstract*—Many image analysis tasks require the automatic segmentation and counting of objects. Current approaches use bounding boxes as a method of locating and counting objects, however, these methodologies struggle to distinguish between co-located and overlapping objects. In this paper, we propose Dual-Output U-Net (DO-U-Net), a novel approach for tackling this issue. DO-U-Net is an Encoder-Decoder style, Fully Convolutional Network (FCN) for the segmentation and counting of densely overlapping object. Our proposed architecture achieves precision and sensitivity superior to other, similar models by producing two target outputs: a segmentation mask and an edge mask. Here, we discuss an application to the segmentation and counting of erythrocytes in blood smears, on which DO-U-Net achieved a sensitivity of 99.07%.**

I. Introduction

In recent years, there has been a large increase in the number of methodologies to fully automate data analysis. The most common approaches to large scale data analysis rely on the use of supervised and unsupervised Machine Learning, and increasingly, Deep Learning. Using only a small number of human-annotated data samples, we can train models to rapidly analyse vast quantities of data without sacrificing the quality or accuracy compared with a human analyst. This work focuses on object segmentation of erythrocytes, or red blood cells, in blood smear images, followed by a further step to accurately count the segmented cells. Whilst easily performed by humans, albeit slowly, this task is often non-trivial in Computer Vision, especially in the cases where the objects exist in complex environments or when objects are closely co-located and overlapping. An average blood smear, imaged using a microscope, contains several hundred erythrocytes of varying size, many of which are overlapping, making an accurate manual count both difficult and time-consuming.

II. Data

We use blood smear images from the Acute Lymphoblastic Leukemia (ALL) Image Database for Image Processing[1]. These images were captured using an optical laboratory microscope, with magnification ranging from 300-500×, and a Canon PowerShot G5 camera. We used 10 images from the ALL_IDB1 dataset for training, all of which were taken from both ALL and non-ALL patients. The training images contained a total of 4,360 erythrocytes. Each erythrocyte in the training data has been marked with a polygon, using a custom GUI tool.

[1]Provided by the Department of Information Technology at Università degli Studi di Milano, https://homes.di.unimi.it/scotti/all/

III. Methodology

The classical U-Net architecture [1], has revolutionised the field of biomedical image segmentation. Similarly to other encoder-decoder Fully Convolutional Networks (FCN), U-Net is capable of producing highly precise segmentation masks. What differentiates it from Mask R-CNN, SegNet and other similar networks is its lack of reliance on large datasets [1]. This is achieved by the introduction of a large number of skip connections, which reintroduce some of the early encoder layers into the much deeper decoder layers. This greatly enriches the information received by the decoder part of the network, and hence reduces the overall size of the dataset required to train the network.

Due to the overlaps between erythrocytes, the classical U-Net struggles to produce a segmentation mask that distinguishes between overlapping cells. This makes counting erythrocytes with existing methods impossible. Driven by the need to reduce overlap in segmentation masks, we modified the U-Net architecture to produce dual outputs, thus developing the DO-U-Net. With the aim of counting closely co-located and overlapping objects, we are predominantly interested in the correct detection of individual objects as opposed to the exact precision of the segmentation mask itself.

An examination of the hidden convolutional layers of the classical U-Net showed that the penultimate layer of the network extracts information about the edges of our objects of interest, without any external stimulus. We introduce a secondary output layer to the network, targeting a mask segmenting the edges of our objects. The edge masks are generated for the training images by rasterising the edge of the mark-up polygons. By subtracting this "edge" mask from the original segmentation mask, we can obtain a "reduced" mask containing only non-overlapping objects.

We minimised the complexity of the model by looking at smaller input regions of the images which minimises the memory footprint of the model. We follow the approach of Ronneberger et. al. by using unpadded convolutions throughout the network, resulting in a model with smaller output masks (196×196px) corresponding to a central region of a larger (380×380px) input image region. DO-U-Net uses two

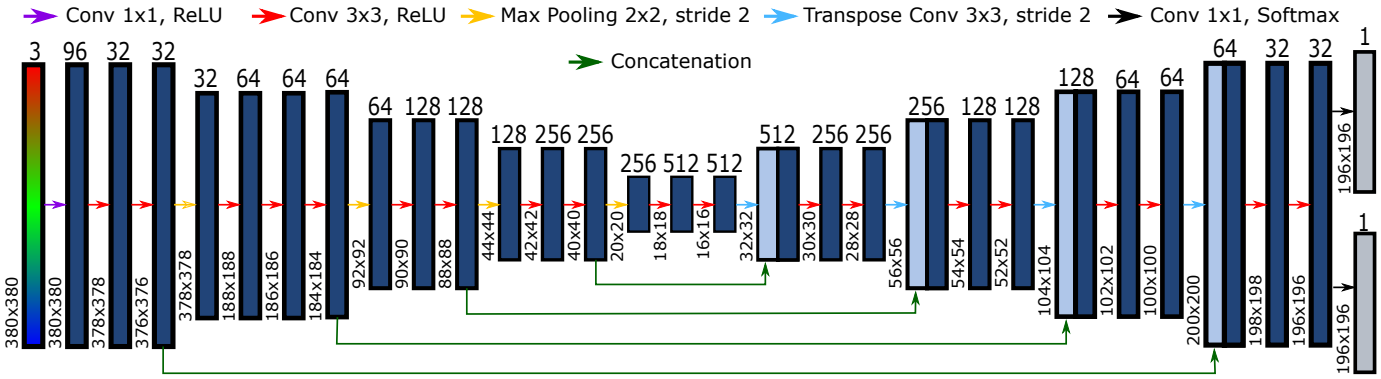


Fig. 1. The DO-U-Net architecture, showing two output layers that target the segmentation and edge masks corresponding to the training images.

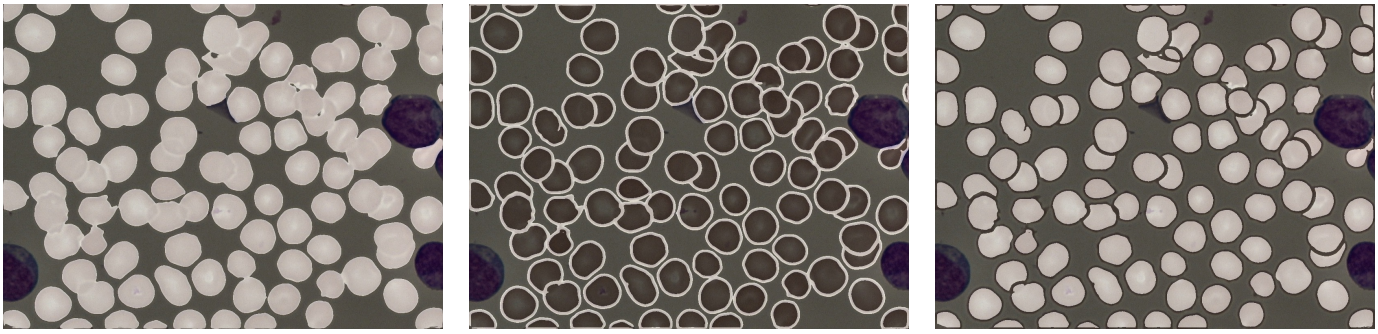


Fig. 2. *Left:* Segmentation mask produced for a sample blood smear. *Centre:* Edges mask produced for the same image. *Right:* The final output mask.

independently trained output layers of identical size. Figure 1 shows our proposed DO-U-Net architecture. Examples of the output edge and segmentation masks, as well as the final "reduced" mask, can be seen in Figure 2.

The 10 large training images were divided into smaller non-overlapping image chips. We train our model against both segmentation and edge masks. The edges of the mark-up polygons, annotated using our custom tool, were used as the "edge" masks during training. The edges were taken to span 4px wide. The polygons acted as segmentaions for training.

As the resulting "reduced" masks produced by our approach do not contain any overlaps, we can use simple counting techniques, relying on the detection of the bounding polygons for the objects of interest. We threshold the outputs to remove all negative values from the image, which may occur due to the subtractions. We then use the Marching Squares Algorithm, implemented as part of `Python`'s `skimage.measure` image analysis library [3].

IV. Results

To validate the performance of DO-U-Net at counting erythrocytes, we use 3 randomly selected blood smear images from ALL patients and a further 5 selected images from non-ALL patients. On a total of 2,775 erythrocytes, as found in these 8 validation images, DO-U-Net achieved an average precision of 98.31% and an average sensitivity of 99.07%.

Whilst our proposed DO-U-Net is extremely effective at producing image and edge segmentation masks, as demonstrated in Figure 2, we do note that the obtained erythrocyte count may not always match the near-perfect segmentation. Mistakes typically arise due to the translucent nature of erythrocytes and the difficulty in differentiating between a cell which is overlapping another and a cell which is overlapped. While these cases are rare, this demonstrates that further improvements can be made to the architecture.

V. Future Work

DO-U-Net has been designed to segment only one type of object, which is a clear limitation of our solution. The blood smear images from the ALL-IDP1 dataset contain normal erythrocytes as well as two clear types of morphology: burr cells and dacryocytes. These morphologies may be signs of disease in patients, or simply artefacts. It is therefore important to not only count all erythrocytes, but to also differentiate between their various morphologies. While our general theory can be applied to identifying different types of object, further modifications to our proposed DO-U-Net would be required.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, LNCS vol. 9351, pp. 234-241. Springer (2015).

[2] N. Abraham, N.M. Khan, "A Novel Focal Tversky Loss Function with Improved Attention U-Net for Lesion Segmentation," IEEE International Symposium on Biomedical Imaging, ISBI 2019. arXiv:1810.07842.

[3] Scikits-image, "Contour Finding," https://scikit-image.org/docs/0.5/auto_examples/plot_contours.html, last accessed: 11 Nov. 2019.