IAIT2020

The 11th International Conference on Advances in Information Technology

Book of Abstracts

THEME : Digital Transformation and Innovation for Smart Business and Society

1st – 3rd July 2020, Bangkok, Thailand

ICPS

Published by ACM

KMUTT Celebrating 60 years

Transforming society. Defining the future.

SIT "Empowering Boundless Digital Transformation" 25 YEARS

School of Information Technology
King Mongkut's University of Technology Thonburi

www.iait-conf.org

PREFACE

As the global environment in nowadays is so dynamic that many digital evolutions have continually disrupted and set a new normal in living anywhere. Rapid evolution of digital innovation and transformation becomes a challenge to smartly cope with the volatility in businesses and various needs of the society. The forum that helps to bring together leading researchers, scholars, scientists, and students across the world to exchange and share new developments and research ideas about all aspects of Information and Communication Technologies can be a leverage. This underlies the theme of IAIT2020 "Digital Transformation and Innovation for Smart Business and Society" with the main focus on the recent advancements and breakthroughs in some of the most emerging technologies of the recent times like computational intelligence and machine learning, natural language processing, cyber security, human-centric computing, and smart systems. The wide spectrum of topics, promoting inter-departmental and inter-disciplinary research towards creating a sustainable digital business and a smart society.

The 11th International Conference on Advances in Information Technology: IAIT2020 is being organized during 1st – 3rd July 2020 in a fully online format by School of Information Technology (SIT), at King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand. This is the first time that IAIT is being held as a virtual conference due to the pandemic of COVID-19. IAIT2020 will publish 47 high quality research papers which includes 7 papers belonging to the special session of JSCI (Joint Symposium on Computational Intelligence) arranged in cooperation with the IEEE-CIS (Computational Intelligence Society) Thailand Chapter.

In the review process, a total of 96 papers were received (including JSCI) submitted from 22 countries covering all major continents of Asia, Africa, Europe, Australia, and America. To maintain high standards of publications, each paper has undergone through a stringent reviewing process. For this year, we have reviewers from 17 countries. The acceptance rate is about 48%. All the presented papers will be published in the proceedings managed by ACM/ICPS.

For the success of IAIT2020, our sincere thanks go to all authors for their submissions and everyone who participated in this year's event. Our gratitude also goes to many people who helped make this conference actualized including our reviewers, international advisory committee, international program committee, all chairs, distinguished keynote speakers, invited speakers and the editorial team at ACM. We hope that IAIT will continue to be a prestigious forum for stimulating exchanges of innovative ideas which enable transformation of smart business and society that truly benefit mankind.

Kriengkrai Porkaew, General Chair of IAIT2020

SCHEDULE

July 1, 2020

09.00 – 09.15	Open Ceremony
09.20 – 10.20	Keynote Speaker: Prof. Dr. Chidchanok Lursinsap
11.00 - 11.45	Invited Speaker: Mr. Jirayut Srupsrisopa
11.45 – 13.00	Break
13.00 – 14.00	Keynote Speaker: Prof. Dr. Pascal Bouvry

Session 1: Image Processing and Computer Vision Session Chair: Asst.Prof.Dr. Pornchai Mongkolnam		Session 2: Cybersecurity, Multimedia and Blockchain Session Chair: Dr. Debajyoti Pal		Session 3: Neural Networks and Deep Learning for Smart Decisions Session Chair: Asst.Prof.Dr. Chonlameth Arpnikanondt	
14.15 – 14.40	Digital Rock Core Image Super-Resolution via Deep Convolutional Neural Networks (#53) Yunfeng Bai, Vladimir Berezovsky, Jing Guo and Popov Vasily	14.45 – 15.10	Social Engineering intrusion: A Case Study (#54) Miika Sillanpää and Jari Hautamäki	14.15 – 14.40	Enhancement of Fault Tolerance in Kafka Pipeline Architecture (#22) Thandar Aung, Hla Yin Min and Aung Htein Maw
14.45 – 15.10	Real-time Drowsiness and Distraction Detection using Computer Vision and Deep Learning (#73) Phakawat Pattarapongsin, Bipul Neupane, Jirayus Vorawan, Harit Sutthikulsombat and Teerayut Horanont	15.15 – 15.40	Paying by Your Messaging Application? A Trust Model (#59) Debajyoti Pal, Suree Funilkul and Syamal Patra	14.45 – 15.10	User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder – Insider Threat Detection (#19) Balaram Sharma, Prabhat Pokharel and Basanta Joshi
15.15 – 15.40	Comparative Performance of Various Deep Learning based Models in Fruit Image Classification (#37) Raheel Siddiqi	15.45 – 16.10	A Data Masking Guideline For Optimizing Insights And Privacy Under GDPR Compliance (#47) Chitanut Tachepun and Sotarat Thammaboosadee	15.15 – 15.40	Network Anomaly Detection using Threshold-based Sparse Autoencoder (#45) May Thet Tun, Dim En Nyaung and Myat Pwint Phyu
15.45 – 16.10	Image Recognition for Detecting Hand Foot and Mouth Disease (#76) Mohammad Farhad Aryan, Worarat Krathu, Chonlameth Arpnikanondt and Boonrat Tassaneetrithep	16.15 – 16.40	Online Learning During COVID-19: Students' Perception of Multimedia Quality (#56) Debajyoti Pal, Vajirasak Vanijja and Syamal Patra	15.45 – 16.10	A Comparison of Destination Clustering using Density-based Algorithm on The Trip Planning Optimization for Last-Mile Parcel Delivery (#78) Issaret Prachitmutita, Praisan Padungweang and Olarn Rojanapornpun
16.15 – 16.40	Reversible Data Hiding Using Multi-Layer Perceptron Based Pixel Prediction (#26) Abhishek Bhandari, Shivraj Sharma, Ravi Uyyala, Rajarshi Pal and Mridula Verma	16.45 – 17.10	AcaChain: Academic Credential Attestation System using Blockchain (#24) Kiratijuta Bhumichitr and Songsak Channarukul		
		17.15 – 17.40	Gamifying MOOC's a Step in The Right Direction? A Systematic Literature Review (#9) Rohani Rohan, Debajyoti Pal and Suree Funilkul		

July 2, 2020

09.00 – 09.45	Invited Speaker: Dr. Choochart Haruechaiyasak
09.50 – 10.35	Invited Speaker: Dr. Vijayakumar Varadarajan
10.40 – 13.00	Break

Session 1: User Experience, Interface Design and Related Applications Session Chair: Dr. Vithida Chongsuphajaisiddhi		Session 2: Cyber Physical Systems, and Computer Networks Session Chair: Asst.Prof.Dr. Narongrit Waraporn		Session 3: Machine Learning and Time Series Analysis for Smart Society Session Chair: Assoc.Prof.Dr. Kittichai Lavangnananda	
13.00 – 13.25	ThaiFlick: Flicked Keyboard on Mobile Devices for Thai using Phoneme Matching (#25) Nattachai Saejiem and Songsak Channarukul	13.00 – 13.25	Bio-inspired Routing in DMFB: An Artificial Swarm Propagation Based Application (#81) Pranab Roy, Debajyoti Pal, Tanmoy Biswas, Rupam Bhattacharyay, and Hafizur Rahaman	13.00 – 13.25	Relevant Feature Selection for Predicting the Severity of Motorcycle Accident in Thailand (#75) Sakchai Samerbut, Praisan Padungweang and Worarat Krathu
13.30 – 13.55	Augmented Reality Project Poster - Using Mobile Augmented Reality Application to Enhance Project Poster (#67) Alvin Yong Chie Yew, Haji Mahran Dr Haji Morsidi, and Jonathan H. Chan	13.30 – 13.55	Robust LCSS Beamformer against DOA Mismatch (#18) Raungrong Suleesathira	13.30 – 13.55	The Impact of Data Imputation and Feature Extraction on PM2.5 Forecasting Performance in Bangkok Using Long Short-Term Memory Neural Networks (#43) Panupong Rong-O and Nuwee Wiwatwattana
14.00 – 14.25	Automated Scheduling of Undergraduate Student Advising Reservation using Extended Flower Pollination Algorithm (#21) Ma Shiela Sapul, Deep Kumar, Sai Kham Lao and Rachsuda Setthawong	14.00 – 14.25	Predicting Application Performance in LoRa IoT Networks (#41) Natchaya Chungsawat and Peerapon Siripongwutikorn	14.00 – 14.25	A System to Estimate the Amount and Calories of Food that Elderly People in the Hospital Consume (#23) Pisol Ruenin, Jakramate Bootkrajang and Jakarin Chawachat
14.30 – 14.55	Hybrid Classroom: Designing for the New Normal after COVID-19 Pandemic (#65) Tuul Triyason, Anuchart Tassanaviboon and Prasert Kanthamanon	14.30 – 14.55	Exploration of Hardware Architectures for String Matching Algorithms in Network Intrusion Detection Systems (#11) Muhammad Rashid, Malik Imran and Atif Raza Jafri	14.30 – 14.55	ARMA Model for Revenue Prediction (#30) Thura Zaw, Swe Swe Kyaw and Aung Nway Oo
15.00 – 15.25	Articulated Trajectory Mapping Using Mobile Environments for Location-Based Learning (#40) Iori Sasaki, Masatoshi Arikawa and Akinori Takahashi	15.00 – 15.25	Improved Efficiency of Fingertip Reflective Pulse Oximetry with 2 positions Integrated Optical Biosensor (#71) Supawadee Woraprueksaru and Kitsada Dougjitjaroen		

July 3, 2020

Session 1 : NLP, Text Mining and Social analytics Session Chair: Dr. Olarn Rojanapornpun		Session 2 : Cloud Computing, Big Data and Health Analytics Session Chair: Dr. Tuul Triyason		JSCI (Special Session) Session Chair: Prof. Dr. Kiyota Hashimoto	
09.00 – 09.25	Data Mining Methods for Optimizing Feature Extraction and Model Selection (#8) Masha Rouzbahman, Alexandra Jovicic, Lu Wang, Leon Zucherman, Zahid Abul-Basher, Nipon Charoenkitkarn and Mark Chignell	09.00 – 09.25	DeepVix: Explaining Long Short-Term Memory Network With High Dimensional Time Series Data (#85) Tommy Dang, Hao Van, Huyen Nguyen, Vung Pham and Rattikorn Hewett	09.00 – 09.25	Efficient Distributed SNP Selection by a Modified Binary Flower Pollination Algorithm (JSCI05) Wanthanee Rathasamuth and Kitsuchart Pasupa
09.30 – 09.55	Using LSTM for Context Based Approach of Sarcasm Detection in Twitter (#42) Siti Khotijah, Jimmy Tirtawangsa and Arie A Suryani	09.30 – 09.55	Provenance Location-based Authentication in Cloud Computing (#52) Phornsawan Roemsri and Rattikorn Hewett	09.30 – 09.55	1st-degree Atrioventricular (AV-block) and Bundle Branch Block Prediction using Machine Learning (JSCI06) Risul Islam Rasel, Nasrin Sultana, Phayung Meesad, Anupam Chowdhury, and Meherab Hossain
10.00 – 10.25	Question Chunk Extraction for Improving Question Similarity (#87) Chalanun Damrongrat, Chatchawal Sangkeettrakarn and Anocha Rugchatjaroen	10.00 – 10.25	ScagCNN: Estimating Visual Characterizations of 2D Scatterplots via Convolution Neural Network (#86) Vung Pham, Ngan V.T. Nguyen and Tommy Dang	10.00 – 10.25	Effects of Part-of-Speech on Thai Sentence Classification to Wh-Question Categories using Machine Learning Approach (JSCI07) Saranlita Chotirat and Phayung Meesad
10.30 – 10.55	A 3D-CNN Siamese Network for Motion Gesture Sign Language Alphabets Recognition (#61) Hassaluk Jirathampradub, Chakarida Nukoolkit, Kaneungjit Suriyathumrongkul and Bunthit Watanapa	10.30 – 10.55	Health Data Analytics with an Opportunistic Big Data Algorithm (#48) Gantaphon Chalumporn and Rattikorn Hewett	10.30 – 10.55	Correlation-Based Incremental Learning Network with Sliding Window for Perfume Classification (JSCI09) Panida Lorwongtrakool and Phayung Meesad
11.00 – 11.25	A Data Augmentation Technique Based on Text for Vietnamese Sentiment Analysis (#34) Thien Ho Huong and Vinh Truong Hoang	11.00 – 11.25	Privacy Protection in Smart Health (#38) Sevgi Arca and Rattikorn Hewett	11.00 – 11.25	Defect Detection and Quality Level Assignment in Wet Blue Goatskin (JSCI10) Carlos Estevão Bastos Sousa, Cláudio Marques De Sá Medeiros, Renato Francisco Pereira, Mateus Alves Vieira Neto and Alcides Andrade Neto
		11.30 – 11.55	Visual Features for Multivariate Time Series (#39) Bao Dien Quoc Nguyen, Rattikorn Hewett and Tommy Dang	11.30 – 11.55	Car Damage Detection and Classification (JSCI11) Phyu Mar Kyu and Kuntpong Woraratpanya
				12.00 – 12.25	Stock Price Analysis with Natural Language Processing and Machine Learning (JSCI12) Sukanchalika Boonmatham and Phayung Meesad
12.25 – 12.55	Break				
13.00 – 13.45	Invited Speaker: Dr. Mohammad Abdur Razzaque				
13.50 – 14.00	Closing Remarks				

Contents

53

Digital Rock Core Image Super-Resolution via Deep Convolutional Neural Networks

Yunfeng Bai
Northern (Arctic) Federal University
Arkhangelsk, Russia
yumengleme@126.com

Vladimir Berezovsky
Northern (Arctic) Federal University
Arkhangelsk, Russia
v.berezovsky@narfu.ru

Jing Guo
South China University of Technology
Guangzhou, China
z7198185@gmail.com

Vasily Popov
Northern (Arctic) Federal University
Arkhangelsk, Russia
v.popov@narfu.ru

ABSTRACT

Digital Rock is a research hotspot in the field of oil and gas. In practice, oil and gas departments hope to get clearer images of rock core which are helpful for them to understand characteristics of reservoir rock better. And based on that they can develop a better mining strategy. Through literature research, we found that in recent years, deep learning algorithms have achieved great success in processing images, thus we introduced them into the field of processing core images. At the same time, we noticed that core images generally have colourful blocks of different sizes. Based on this feature, we did not directly apply the classic algorithm, but made some improvements. Experiments show that the effect of our proposed method is equivalent to that of the classic algorithms, however, in the pre-processing stage, it saves time significantly.

CCS CONCEPTS

• Computing methodologies ⟶ Reconstruction; • Applied computing⟶Physics.

KEYWORDS

digital rock, rock core analysis, super-resolution, deep learning

73

Real-time Drowsiness and Distraction Detection using Computer Vision and Deep Learning

Phakawat Pattarapongsin
School of ICT
Sirindhorn International Institute of Technology
Pathum Thani, Thailand
5922790167@g.siit.tu.ac.th

Bipul Neupane
Sirindhorn International Institute of Technology
Pathum Thani, Thailand
bipul.n@siitgis.com

Jirayus Vorawan
School of ICT
Sirindhorn International Institute of Technology
Pathum Thani, Thailand
5922792155@g.siit.tu.ac.th

Harit Sutthikulsombat
School of ICT
Sirindhorn International Institute of Technology
Pathum Thani, Thailand
5922770540@g.siit.tu.ac.th

Teerayut Horanont
School of ICT
Sirindhorn International Institute of Technology
Pathum Thani, Thailand
teerayut@siit.tu.ac.th

ABSTRACT

Automated analysis of driver's attention to driving is crucial to reduce traffic accidents. Real-time detection of initial signs of drowsiness or distraction of driver has yet not been most reliable. This study is therefore oriented towards improving this problem by strategically detecting some of the necessary parameters like closing of eye, yawning, and head orientation of driver. We use deep neural network (DNN) to extract face and facial landmarks, followed by calculation of Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR) and pose estimation using 3D location and intrinsic parameters to detect whether or not driver is focused into driving. Our experimental results shows better real-time performance than traditional methods in many aspects, which is why we further deploy the algorithm as a mobile application.

KEYWORDS

Drowsiness Detection, Pose estimation, Face landmarks, Eye Aspect Ratio, Mouth Aspect Ratio

#37

Comparative Performance of Various Deep Learning based Models in Fruit Image Classification

Raheel Siddiqi
Computer Science Department, Bahria University
Karachi, Pakistan
drraheel.bukc@bahria.edu.pk

ABSTRACT

Fruit type classification based on computer vision remains challenging because of shape, color and texture similarity among numerous species of fruits. Recent research results indicate that the Convolutional Neural Network based models have demonstrated substantial improvements in fruit image classification accuracy. This paper proposes fourteen different deep learning models for fruit type classification. Thirteen of these models exploit ImageNet pre-trained feature extractors. In addition to these thirteen models, a simple, lightweight 12-layer model called FruitNet has also been proposed. A large, publicly available fruit images dataset, called Fruits 360, has been used to train and evaluate the proposed models. Among all the models that exploited some pre-trained feature extractor, Xception based model has produced the best validation accuracy rate (i.e. 99.97%) while MobileNet based model has proved to be a very efficient model. Fine tuning has also been found to produce excellent results. In fact, the best validation accuracy of 99.98% is achieved when VGG16 based model is fine tuned. On the other hand, FruitNet has yielded accuracy rate of 99.73% which is excellent given its small size and simple structure. FruitNet is also the most efficient model among all the proposed models.

CCS CONCEPTS

• Computing methodologies • Object recognition • Computer systems organization • Neural networks.

KEYWORDS

Fruit Image Classification, Transfer Learning, Fine Tuning, Convolutional Neural Networks, Computer Vision

#76

Image Recognition for Detecting Hand Foot and Mouth Disease

Mohammad Farhad Aryan
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
aryan200x@gmail.com

Worarat Krathu
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
worarat.kra@sit.kmutt.ac.th

Chonlameth Arpnikanondt
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
chonlameth@sit.kmutt.ac.th

Boonrat Tassaneetrithep
Siriraj Center Research of Excellence in Immunoregulation
Mahidol University
Bangkok, Thailand
boonrat.tas@mahidol.ac.th

ABSTRACT

Hand Foot and Mouth Disease is a common childhood skin infection that could quickly develop into a severe case. It spreads easily, with R0 typically above 2. At school, an individual class could be closed for several days. The school could be closed for clean-up. All these closings become an economic burden, especially in the low-income population, that could be prevented or mitigated by a quick response once the disease is first detected. This paper experimented with various combinations of existing image processing and recognition techniques. A state-of-the-art method was discovered to effectively detect lesions of the Hand Foot and Mouth Disease. The results show that color-space conversion as preprocessing followed by segmentation using the KMeans-Morphological process, GLCM and Mean for feature extraction, and Support Vector Machine classifier performed best for the Hand Food and Mount Disease image recognition.

KEYWORDS

Image recognition, Image processing, Hand Foot and Mouth disease, classification

26

Reversible Data Hiding Using Multi-Layer Perceptron Based Pixel Prediction

Abhishek Bhandari*
Indian Institute of Technology
Jodhpur, India
bhandari.1@iitj.ac.in

Shivraj Sharma†
Indian Institute of Technology
Guwahati, India
shivrajsharma34@gmail.com

Ravi Uyyala‡
ACE Engineering College
Hyderabad, India
uyyala.ravi@gmail.com

Rajarshi Pal
Institute for Development and Research in Banking Technology
Hyderabad, India
iamrajarshi@yahoo.co.in

Mridula Verma
Institute for Development and Research in Banking Technology
Hyderabad, India
vmridula@idrbt.ac.in

ABSTRACT

Reversible data hiding is a branch of data hiding in which cover image can additionally be restored along with the extraction of hidden data. In prediction error expansion based reversible data hiding techniques, pixel values are predicted using the neighborhood pixels. Data bits are hidden in the expanded prediction error. Prediction error expansion based techniques have become popular due to their superior performances over other categories of techniques. In this paper, a novel reversible data hiding technique is proposed based on a multi-layer perceptron based pixel prediction. The multi-layer perceptron is trained to predict a pixel value based on its eight neighbors. Experiments have been conducted by varying the number of hidden layers and the number of neurons in each hidden layer in the multi-layer perceptron. The best architecture for the multi-layer perceptron is chosen for the proposed reversible data hiding technique. Finally, an adaptive embedding strategy is used to insert either one bit or two bits of data in the expanded prediction error of the pixel based on the standard deviation of the pixel values in the local neighborhood.

CCS CONCEPTS

• Computing methodologies⟶Image processing.

KEYWORDS

Reversible data hiding, Prediction error expansion, Multi-layer perceptron

54

Social Engineering Intrusion: A Case Study

Miika Sillanpää
Institute of Information Technology
JAMK University of Applied Sciences
Finland
miika.sillanpaa@optimesys.fi

Jari Hautamäki
Institute of Information Technology
JAMK University of Applied Sciences
Finland
jari.hautamaki@jamk.fi

ABSTRACT

Social engineering is a very old method to influence people in their daily actions. The same methods added with new techniques have been implemented to create effective penetration mechanisms against organizations. The goal in this study was to measure employees' security awareness and culture. This is a case study which uses several penetration methods to test an organization's vulnerability against social engineering techniques. The study started with cyber security research questions for all employees in the studied organization Reconnaissance and survey questions together provide use cases to the physical penetration testing phase. When comparing the results of the survey questions with the actual penetration test, a significant difference was found. Even employees understand how to behave in a penetration case; they act differently. This is a problem which can be resolved by increasing the awareness against security engineering attacks. The awareness can be increased by training, education and good security policy.

CCS CONCEPTS

• Security and privacy • Social engineering attacks • Human and societal aspects of security and privacy

KEYWORDS

Social engineering, Information security, Cyber security, Intrusion Detection

59

Paying by Your Messaging Application? A Trust Model

Debajyoti Pal
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
debajyoti.pal@mail.kmutt.ac.th

Suree Funilkul
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
suree@sit.kmutt.ac.th

Syamal Patra
Department of Computer Science & Engineering
Camellia Institute of Technology
Kolkata, India
cit.syamalpatra@gmail.com

ABSTRACT

User adoption of mobile payments is low when compared to the traditional payment methods. At the same time, there has been a surge in the usage of second-generation mobile instant messaging (SMIM) services. Integrating these two services can be a game changer in the market, which needs more research from the viewpoint of end-user adoption. Similarly, for any service, trust is an extremely important factor that affects the adoption. Motivated by this fact, the current work proposes a trust based theoretical model for explaining the user adoption of an integrated mobile payment/SMIM service scenario. By considering both the positive and negative valences of user trust, the current study aims to deeply integrate its mediating role on the adoption intention. The research framework is proposed based on a multi-model approach by combining the concepts from valence framework, information system success model, and the technology adoption model. The proposed framework is empirically tested and validated by using a Partial Least Squares Structural Equation Modelling approach from a sample of 846 people, all having prior experience in using a mobile payment and SMIM service.

CCS CONCEPTS

• Information systems • Information systems applications • Mobile information processing systems

KEYWORDS

Adoption intention, Mobile payment, Multi-model approach, SMIM services, User trust

47

A Data Masking Guideline For Optimizing Insights And Privacy Under GDPR Compliance

Chitanut Tachepun
Information Technology Management,
Faculty of Engineering, Mahidol University
Nakhonpathom, Thailand
chitanut.tac@student.mahidol.edu

Sotarat Thammaboosadee
Information Technology Management, Faculty of Engineering, Mahidol University
Nakhonpathom, Thailand
sotarat.tha@mahidol.ac.th

ABSTRACT

The General Data Protection Regulation (GDPR) has been enforced since May 2019 and became a disruptive issue to every organization due to its severe penalties in the data breaches or use of personal data for illegal purposes, e.g., lack of the consent of data subject. Therefore, the data Pseudonymization and Anonymization are one of the employed techniques to protect and reduce the privacy risks from the data breach. Unfortunately, they also destroy the pattern of the data, which represents the fact that it could be analyzed or monetized to gain useful insights by data analytics or data science approaches. This paper focuses on optimizing the privacy and insight method that the data could be useful for analyzing and also compliance with the GDPR. This paper proposes the guideline consists of three techniques: tokenization, suppression, and generalization to protect personal data by calculating risk scores from two methods: data classification and data uniqueness. The criteria in the guideline are experimented to achieve the optimized classification performance in protected data compared with five original open data by analyzing with three data mining algorithms with the hyperparameter tuning process. The results show that the protected data by the proposed guideline can protect adequate information and achieve insignificant classification performance when compared to the unprotected data.

CCS CONCEPTS

Security and privacy ⟶ Human and societal aspects of security and privacy ⟶ Usability in security and privacy

KEYWORDS

GDPR, Data Pseudonymization, Data Tokenization, Data Masking, Data Analytics, Data protection

56

Online Learning During COVID-19: Students' Perception of Multimedia Quality

Debajyoti Pal
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
debajyoti.pal@mail.kmutt.ac.th

Vajirasak Vanijja
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
vachee@sit.kmutt.ac.th

Syamal Patra
Department of Computer Science and Engineering
Camellia Institute of Technology
Kolkata, India
cit.syamalpatra@gmail.com

ABSTRACT

The COVID-19 pandemic has forced educational institutes worldwide to resort to an "online only" mode of teaching delivery. As a consequence, throughout the globe there has been an increasing trend among the students to use different videoconferencing applications for the purpose of learning online. However, the multimedia quality provided by these different applications provides the key to their success i.e. whether or not the students will be willing to use those for learning online. Consequently, three popular applications (Zoom, Microsoft Teams, and Cisco Webex) are taken up in this work for the purpose of multimedia quality evaluation by using an objective based approach. Results from both the models are in close agreement with each other. Microsoft Teams provides the least experience, whereas those from the others vary depending upon the objective models used. The results obtained are further verified by conducting relevant hypotheses tests.

CCS CONCEPTS

• Networks • Network performance evaluation • Network performance analysis

KEYWORDS

Multimedia quality, Objective models, Online learning, Quality of experience, Videoconferencing

24

AcaChain: Academic Credential Attestation System using Blockchain

Kiratijuta Bhumichitr
Computer Science Department
Assumption University
Bangkok, Thailand
kiratijuta@scitech.au.edu

Songsak Channarukul
Computer Science Department
Assumption University
Bangkok, Thailand
songsak@scitech.au.edu

ABSTRACT

Blockchain has been a part of people's interest in the development of various areas such as digital currencies (e.g., Bitcoin), securities, healthcare, record-keeping and much more. Some of the key features of blockchain technology are immutability, decentralization, distributed ledger, consensus, and smart contract. In this paper, we explore how blockchain technology could possibly be utilized for the academic credential attestation in the education domain. This paper aims to re-engineer the academic credential attestation process to reduce both time and resource consumption and human involvement. It also aims to design a unique way for academic data collection to prevent fraud academic credentials. A blockchain-based system for academic credential attestation has been implemented.

KEYWORDS

Academic Credential Attestation, Blockchain, Hyperledger Fabric

9

Gamifying MOOC's a Step in The Right Direction? A Systematic Literature Review

Rohani Rohan
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
rohani.sabari@gmail.com

Debajyoti Pal
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
debajyoti.pal@mail.kmutt.ac.th

Suree Funilkul
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
suree@sit.kmutt.ac.th

ABSTRACT

Massive Open Online Courses (MOOC) are gaining in popularity in recent years. However, there are certain well-known drawbacks of MOOCs like lack of student motivation, student engagement, and high dropout rates, which cast doubt on the success of MOOCs. Gamification, on the other hand, uses the concept of games but applied to a non-gaming context primarily for improving the motivational level of the users. Gamification is seen to have positive effects in a number of application areas like marketing, tourism, and even education. The primary objective behind this work is to investigate the current state of gamification as applied to MOOCs and check if it is a step in the right direction. For this purpose, a systematic literature review is conducted following the PRISMA standards and results analyzed.

CCS CONCEPTS

• Information systems • Information systems applications
• Computing platforms

KEYWORDS

Dropout, gamification, gaming elements, MOOC, motivation

22

Enhancement of Fault Tolerance in Kafka Pipeline Architecture

Thandar Aung
University of Information Technology
Yangon,Myanmar
thandaraung@uit.edu.mm

Hla Yin Min
University of Information Technology
Yangon,Myanmar
hlayinmin@uit.edu.mm

Aung Htein Maw
University of Information Technology
Yangon,Myanmar
ahmaw@uit.edu.mm

ABSTRACT

Fault tolerance act as the most important factor to improve the reliability of the real-time messaging system. Checkpointing and replication are the main factors to promote the performance of fault tolerance. The paper emphasizes the message recovery scheme based on checkpointing. Apache Kafka is a highly scalable publishing-subscribing messaging system. However, the message recovery process becomes one of the challenges when server failure occurred. Apache Kafka needs to catch up with the lost messages in the failure case. To overcome the problem, Kafka requires to check the failure case by identifying the optimal checkpoint interval in Kafka configuration. The paper focuses on editing the default checkpoint interval (DCI) to recover the lost messages. Kafka needs to estimate the overall total overhead cost after defining the optimal checkpoint interval. The proposed system intends to recover lost messages and reduce the total overhead cost, according to the evaluation of Kafka. The paper measures the total overhead cost by using a fixed checkpoint interval (FCI) method. Conforming to the experimental results, the system achieves the problems of message loss and improve the main factor of fault tolerance in Kafka. The system also measures the total overhead cost by using a fixed checkpoint interval (FCI) method. Experimental results outperform the message recovery and save time approximately 50% than the original system.

KEYWORDS

Apache Kafka, fixed checkpoint interval (FCI), default checkpoint interval (DCI), real-time messaging, fault tolerance

19

User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder – Insider Threat Detection

Balaram Sharma
LogPoint
Lalitpur, Nepal
ab.sharma015@gmail.com

Prabhat Pokharel
LogPoint
Lalitpur, Nepal
pokharelprabhat@gmail.com

Basanta Joshi
Institute of Engineering
Lalitpur, Nepal
joshibasanta@gmail.com

ABSTRACT

Identifying anomalies from log data for insider threat detection is practically a very challenging task for security analysts. User behavior modeling is very important for the identification of these anomalies. This paper presents unsupervised user behavior modeling for anomaly detection. The proposed approach uses LSTM based Autoencoder to model user behavior based on session activities and thus identify the anomalous data points. The proposed method follows a two-step process. First, it calculates the reconstruction error using the autoencoder on the non-anomalous dataset, and then it is used to define the threshold to separate the outliers from the normal data points. The identified outliers are then classified as anomalies. The CERT insider threat dataset has been used for the research work. For each user, the feature vectors are prepared by extracting key information from corresponding raw events and aggregating the data points based on users' actions within respective users' sessions. LSTM Autoencoder has been implemented for behavior learning and anomaly detection. For any unseen behavior or anomaly pattern, the model produces high reconstruction error which is an indication of an anomaly. The experimental results show that in the best case, the model produced an Accuracy of 90.17%, True Positives 91.03%, and False Positives 9.84%. Thus, the results suggest that the proposed approach can be effectively used in automatic anomaly detection.

CCS CONCEPTS

• Security and privacy • Intrusion/anomaly detection and malware mitigation • Intrusion detection systems

KEYWORDS

LSTM Autoencoder, Deep Learning, Anomaly Detection, Behavior Analytics, Insider Threat, UBA

45

Network Anomaly Detection using Threshold-based Sparse Autoencoder

May Thet Tun†
University of Information Technology
Yangon, Myanmar
maythettun@uit.edu.mm

Dim En Nyaung
University of Information Technology
Yangon, Myanmar
dimennyaung@uit.edu.mm

Myat Pwint Phyu
University of Information Technology
Yangon, Myanmar
myatpwintphyu@uit.edu.mm

ABSTRACT

Nowadays, cyber-attacks have been dramatically increased due to the rapid development of Internet-based services. The current network anomaly detection solutions such as Firewall, Snort, honeypots are useful, but they are insufficient to deal with zero-day attacks. At present, unsupervised deep learning technologies focus on increasing the performance of anomaly-based network attack detection in the recent research area. The normal and attack transactions can be distinguished by the threshold in the network anomaly detection using a sparse autoencoder. The threshold value estimation is not a non-trivial task. The key factor of the threshold value which draws a line between normal and attack highly depends on the accuracy of the network anomaly detection. In this paper, the network anomaly detection based on a sparse autoencoder with a percentile-based threshold selection algorithm is proposed. The results of the proposed system have been validated concerning the accuracy, precision, recall and false positive rate. As a result, the experimental results on UNSW-NB15 and NSL-KDD datasets have shown that the proposed system provides higher accuracy and outperforms the previous related works in recent years.

CCS CONCEPTS

• Network Security • Network anomaly detection • Deep learning • Sparse autoencoder

KEYWORDS

Network anomaly detection, Deep learning, Sparse autoencoder, Percentile threshold approach, Chi-squared feature selection

78

A Comparison of Destination Clustering using Density-based Algorithm on The Trip Planning Optimization for Last-Mile Parcel Delivery

Issaret Prachitmutita
King Mongkut's University of Technology Thonburi
Bangkok Thailand
Issaret.m@gmail.com

Praisan Padungweang
King Mongkut's University of Technology Thonburi
Bangkok Thailand
praisan.pad@sit.kmutt.ac.th

Olarn Rojanapornpun
King Mongkut's University of Technology Thonburi
Bangkok Thailand
olarnr@sit.kmutt.ac.th

ABSTRACT

This study examines the pipeline for real-world delivery trip planning optimization. The study focuses on the capacitated vehicle routing problems. Density-based clustering algorithms were applied prior to resolving the capacitated vehicle routing problem (CVRP) to reduce processing time and to maintain acceptable efficiency. The experimental results with 25 CVRP pipelines were compared. The results showed that the hierarchical density-based spatial clustering of applications with noise (HDBSCAN) method achieved the highest performance. It could reduce the time of trip planning by 30-40% and total distance by 2.1% compared with the traditional method.

KEYWORDS

Logistics, Last-Mile Delivery, Clustering, Route Optimization, Capacitated Vehicle Routing Problem

#25

ThaiFlick: Flicked Keyboard on Mobile Devices for Thai using Phoneme Matching

Nattachai Saejiem
Computer Science Department
Assumption University
Bangkok, Thailand
nattachai.saejiem@gmail.com

Songsak Channarukul
Computer Science Department
Assumption University
Bangkok, Thailand
songsak@scitech.au.edu

ABSTRACT

Composing messages with virtual keyboard is a challenging task especially for Thai messages. There are too many Thai characters to fit a mobile keyboard. Moreover, screens of smartphones are so small that those having big fingers frequently tap on undesired keys. This paper introduces a Thai keyboard layout that helps improve typing performance on smartphones. It adopts Japanese Flicked layout along with intuitive key arrangement based on character arrangement in Thai words. In addition, it provides word suggestion based on the sound of the entered characters. Users can just type characters having the same sound of the desired alphabet without correctly typing all required characters. Our experiments show that ThaiFlick outperforms the existing Ketmanee layout on mobile devices.

CCS CONCEPTS

• Human-centered computing • Human computer interaction (HCI) • Interaction devices • Keyboards

KEYWORDS

Virtual Keyboard, Flicked Layout, Thai Virtual Keyboard Layout

67

Augmented Reality Project Poster
Using Mobile Augmented Reality Application to Enhance Project Poster

Alvin Yong Chie Yew
Creative Multimedia
Universiti Teknologi Brunei
Tutong,
Brunei Darussalam
alvincreativeworks@gmail.com

Haji Mahran Dr Haji Morsidi
Computer Network and Security
Universiti Teknologi Brunei
Bandar Seri Begawan,
Brunei Darussalam
mahran.morsidi@gmail.com

Jonathan H. Chan
School of Information
Technology King Mongkut's
University of Technology Thonburi
Bangkok, Thailand
jonathan@sit.kmutt.ac.th

ABSTRACT

This work presents a process to create an Augmented Reality (AR) smartphone application which aims to enhance the way that contents of posters are delivered to viewers. Background and literature reviews provide context for the development of this project. The project consists of two parts: the poster and the mobile application. The app is created in Unity using the Vuforia engine. User will be able to download and install the app. This will allow any viewer to interact with the poster via AR technology.

CCS CONCEPTS

• Computing methodologies ⟶ Computer graphics ⟶ Graphics systems and interfaces ⟶ Mixed / augmented reality.

• Human-centered computing ⟶ Web-based interaction.

User interface toolkits; Virtual reality; Mixed/augmented reality.

KEYWORDS

Augmented Reality, Poster, Unity Engine, Vuforia, Mobile device, Interactive media.

21

Automated Scheduling of Undergraduate Student Advising Reservation using Extended Flower Pollination Algorithm

Ma. Shiela C. Sapul
Computer Science Department
Assumption University
Bangkok, Thailand
p5919398@au.edu

Deep Kumar
Computer Science Department
Assumption University
Bangkok, Thailand
u5818029@au.edu

Sai Kham Lao
Computer Science Department
Assumption University
Bangkok, Thailand
u5748124@au.edu

Rachsuda Setthawong
Computer Science Department
Assumption University
Bangkok, Thailand
rachsuda@scitech.au.edu

ABSTRACT

Academic advising is an important process to assist undergraduate students in their academic life because it provides students the direction about personal, social, and educational concerns. However, the traditional process that an advisor manually generates an advising schedule often ends up with some conflicts with student schedules. To address this problem effectively and efficiently, this work introduces the eXtended Flower Pollination Algorithm (XFPA), which extends the Adapted Flower pollination Algorithm (AFPA) to automatically generate an optimal academic advising schedule that conforms to availability and/or preferences of both advisor and students. In particular, this work redefines a representative of a candidate schedule (solution), adapts the operators so that candidates' solutions can exchange information during evolution steps, and also accepts the hard and soft constraints when generating an optimal solution. Moreover, to further improve the proposed algorithm, the remedy function is also introduced to enhance the nearly optimal solution generated in the complex datasets. The experiment results show that XFPA is capable of automatically and efficiently generating optimal solutions in both simple and complex datasets.

KEYWORDS

Academic Advising, Advising Schedule, Constraints, Flower Pollination Algorithm, Remedy Function.

65

Hybrid Classroom: Designing for the New Normal after COVID-19 Pandemic

Tuul Triyason
School of Information Technology,
King Mongkut's University of
Technology Thonburi,
Bangkok, Thailand.
tuul.tri@sit.kmutt.ac.th

Anuchart Tassanaviboon
School of Information Technology
King Mongkut's University of
Technology Thonburi,
Bangkok, Thailand.
anuchart@sit.kmutt.ac.th

Prasert Kanthamanon
School of Information Technology,
King Mongkut's University of
Technology Thonburi,
Bangkok, Thailand
prasert@sit.kmutt.ac.th

ABSTRACT

In the end of December 2019, there was a reported of novel coronavirus in Wuhan, China. Later, this virus named COVID-19 became one of a major outbreak of the world. Since the virus can be spread through the air, the social distancing is essential. Many reports indicated that this will make the world change into a new normal. Working or learning from home by using information and communication technologies will become a new standard of the future. This paper proposed the ideas of the King Mongkut's University of Technology Thonburi to design and implement the hybrid classroom for teaching and conducting university's business after the pandemic crisis. The hybrid classroom concept aims to reducing the number of involved people in each activity by offloading some group of people to online from their home. The hybrid room has also been designed to allow online and physical attendees interact during operating sessions. The conclusion of this paper describes the problems and obstacles that arise during the implementation of prototype rooms, as well as the solutions and future direction of this work.

CCS CONCEPTS

Applied Computing • Education • Distance Learning

KEYWORDS

Distance Learning, Hybrid Classroom, Synchronous Learning, Online Learning

40

Articulated Trajectory Mapping Using Mobile Environments for Location-Based Learning

Iori Sasaki
Graduate School of Engineering Sci., Akita University
Akita, Japan
m8020506@s.akita-u.ac.jp

Masatoshi Arikawa
Graduate School of Engineering Sci., Akita University
Akita, Japan
arikawa@ie.akita-u.ac.jp

Akinori Takahashi
Graduate School of Engineering Sci., Akita University
Akita, Japan
akinori@ie.akita-u.ac.jp

ABSTRACT

This paper addresses how to enhance post-tour activities with digital mobile devices. Walkers sometimes use GPS loggers to record, review, and share their walking stories. However, the raw trajectory data recorded can be difficult to browse through on a map. To resolve this, based on our observations, we illustrated chunks of messy trajectories, inaccurate indoor points, and unstable trajectory lines as problems encountered when mapping raw trajectory data. Then, we proposed a new framework that focuses on GPS horizontal accuracy to find indoor location points and on an accelerometer to find stopping points. We also applied the Douglas-Peucker algorithm to make the trajectory cleaner and then integrated the extracted points with the clean trajectory line. An implementation test demonstrated that the detection of the indoor location points and outdoor stopping points was done correctly, and the proposed system gave users a better experience for their review. The paper contributes to the research on cleaning up map-based displays of raw trajectory data by using not only location data but also other spatial information that smartphones can collect.

CCS CONCEPTS

• Applied computing • Computers in other domains • Cartography

KEYWORDS

GPS, Trajectory, Mobile Mapping, Location-based learning

81

Bio-inspired Routing in DMFB: An Artificial Swarm Propagation Based Application

Pranab Roy
School of VLSI Technology
I.I.E.S.T, Shibpur
West Bengal, India
ronmarine14@yahoo.co.in

Debajyoti Pal
School of Information Technology
KMUTT
Bangkok, Thailand
debajyoti.pal@kmutt.ac.th

Tanmoy Biswas
Department of Computer Science
West Bengal State University
Barasat, West Bengal, India
tan25biswas@gmail.com

Rupam Bhattacharya
School of VLSI Technology
I.I.E.S.T, Shibpur
West Bengal, India
th_rup@yahoo.co.in

Hafizur Rahaman
School of VLSI Technology
I.I.E.S.T, Shibpur
West Bengal, India
rahaman_h@yahoo.co.in

ABSTRACT

Digital Microfluidic Biochips (DMFB) has been evolved as a new generation of Lab-on–Chip device. This appears as a viable alternative to conventional benchtop laboratory process in the fields of point of care diagnostics, environmental toxicity monitoring and other biomedical applications. With increase in complexity, scalability and requirements of more precise control in microfluidic operations - simultaneous execution of complex bioassays within the same substrates poses a major challenge. Transportation of multiple droplets comprising complex bioassays within a 2D array faces the major issue of cross contamination. Poorly estimated route path and bad pre-scheduling may result in unwarranted cross overs resulting in faulty detection and dire wrong diagnostic consequences. In this work we used a swarm propagation based technique to resolve the cross contamination issues for parallel routing within a single 2D DMFB array. The proposed technique shows encouraging improvements in terms of arrival time as well as optimal contamination for enhanced routing performance.

KEYWORDS

Digital microfluidics, Biochips, Routing, Droplet actuation, Cross contamination, swarm propagation

18

Robust LCSS Beamformer against DOA Mismatch

Raungrong Suleesathira
Department of Electronic and Telecommunication Engineering, Faculty of Engineering
King Mongkut's University of Technology Thonburi
Thung Khru, Bangkok, Thailand
raungrong.sul@kmutt.ac.th

ABSTRACT

To cancel interference signals that have fluctuated orientation of directions, the LCSS (Linear Constrained Sector Suppressed) approach is a null broadening beamforming based on the correlation matrix of the steering vectors on the predefined sector of direction fluctuation. However, the LCSS beamformer is still sensitive to mismatch between the presumed and actual desired signal direction. In this paper, a robust LCSS (RLCSS) beamformer is presented not only to remedy the mismatch situation but also to support the case of small size of snapshot samples available in any input SNR values. It requires a constraint matrix and a gain vector to create the LCSS beamformer. To make LCSS beamformer robust against the DOA mismatch, an algorithm to construct the constraint matrix and the gain vector is proposed. The RLCSS algorithm exploits the presumed direction and its region in order to create the constraint matrix. For the gain vector, the steering vector in the created constraint matrix that is correlated to the received signal vector the most is selected to be used to generate the gain vector. Consequently, utilizing the modified constraint matrix and the new gain vector into the LCSS beamformer can provide the robustness against the DOA mismatch of the desired signal steering vector. Simulation results demonstrate that the RLCSS approach can enhance the LCSS beamformers when the mismatch of the DOA of the desired signal exists.

CCS CONCEPTS

• Applied computing • Mathematics and statistics

KEYWORDS

Beamforming; DOA mismatch; Null Broadening Beamforming

41

Predicting Application Performance in LoRa IoT Networks

Natchaya Chungsawat
Department of Computer Engineering
King Mongkut's University of Technology Thonburi
Bangkok Thailand
natchaya.chu@mail.kmutt.ac.th

Peerapon Siripongwutikorn
Department of Computer Engineering
King Mongkut's University of Technology Thonburi
Bangkok Thailand
peerapon.sir@mail.kmutt.ac.th

ABSTRACT

Network planning is critical to satisfactory application performance and efficient network resource utilization especially in an outdoor IoT network such as LoRa. As IoT devices may locate at different distances from the gateway and use different transmission rates and data generation rates, determining if a given set of devices with heterogenous configurations can be feasibly deployed is a challenging task. Unlike previous works which mostly focus on the effect of transmission techniques in MAC and physical layer parameters, this paper develops prediction models of various IoT application performance metrics in a single-gateway LoRa IoT network given a set of heterogeneous device configurations in terms of distance from gateway, data rate, and packet generation rate, which are more relevant inputs to network provisioning. Performance data of packet loss, average packet delay and high-percentile delays are first obtained from simulation experiments over a wide range of factor values and the data is fitted to binomial regression, linear regression, and neural network models. Our results show that with appropriate model tuning, a standard technique like neural network regression is able to give high prediction accuracy, with prediction errors around 1.5 - 5.3% on the test dataset depending on the application performance metrics.

KEYWORDS

Application performance, Internet of thing, LoRa, Packet loss, Average delay, High-percentile delays

11

Exploration of Hardware Architectures for String Matching Algorithms in Network Intrusion Detection Systems

Muhammad Rashid
Computer Engineering Department
Umm Al-Qura University
Makkah, Saudi Arabia
mfelahi@uqu.edu.sa

Malik Imran
Science and Technology Unit (STU)
Umm Al-Qura University
Makkah, Saudi Arabia
mlk.imran88@gmail.com

Atif Raza Jafri
Electrical Engineering Department,
Bahria University
Islamabad, Pakistan
atifraza.buic@bahria.edu.pk

ABSTRACT

An intrusion detection system monitors and analyzes all the incoming packets, on a given network, to detect any corresponding vulnerabilities and intrusions. It consists of four major modules: packet capturing, packet decoding, packet pre-processing and string/pattern matching. Among these, the string matching is computationally the most intensive part and a number of hardware architectures/designs have already been proposed to accelerate its performance. Consequently, an exploration of existing hardware architectures for string matching algorithms is critical. This paper identifies the most frequently used string matching algorithms and techniques, utilized for the hardware implementation. Subsequently, an exploration of various hardware architectures is provided for the identified algorithms and techniques. Finally, the implementation details of explored architectures are discussed in terms of the used device, consumed hardware resources, operational clock frequency and throughput.

KEYWORDS

Network intrusion detection system, hardware implementations, string matching algorithms, throughput

71

Improved Efficiency of Fingertip Reflective Pulse Oximetry with 2 positions Integrated Optical Biosensor

Supawadee Worapruekjaru
Darunsikkhalai Science School
Bangkok Thailand
supawadee.wor@mail.kmutt.ac.th

Kitsada Dougjitjaroen
Darunsikkhalai Science School
Bangkok Thailand
kitsada.dou@gmail.com

ABSTRACT

Currently, the reflected-light fingertip blood oxygen meter has been used for medical purposes to diagnose the initial physical condition of patients due to the advantages of this equipment designed able to measure the amount of oxygen in the blood at other locations of the patient's body aside from finger positions such as the forehead, chest, wrists, etc. But a user still found the equipment limitations in measurement accuracy at these body positions compared with the measurement at the finger. This work proposes to develop a low cost newly designed reflected-light blood oxygen meter system to improve efficiency and accuracy of this equipment type when using at another position of the body and evaluate the system performance with comparison to commercial transmissive pulse oximetry. The newly designed system evaluated result by using student's t-test analysis observed prob > |t| greater than 0.05% at the 95% confidence intervals and gauge repeatability and reproducibility is 4.31% measurement variations to ensure that the newly designed system performance is comparable with a commercial pulse oximetry system and capable to accept in clinical condition.

KEYWORDS

pulse oximetry, SpO2, Optical Biosensor

#75

Relevant Feature Selection for Predicting the Severity of Motorcycle Accident in Thailand

Sakchai Samerbut
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok Thailand
dev2life@gmail.com

Praisan Padungweang
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok Thailand
praisan@gmail.com

Worarat Krathu
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok Thailand
worarat.kra@sit.kmutt.ac.th

ABSTRACT

Thailand is the 5th ranked for road accident death toll in the world and the first ranked in Asia, which is considered a major problem in Thailand. Road accidents are an important problem that affects the quality of life of people and the country's economy. In Thailand, most of road accidents happening with motorcycles. There is the need of study for factors effecting these accidents. In this work, we applied feature selection and classification techniques for analyzing important factors causing road accidents on the dataset of motorcycle accidents. In particular, the experiment compared the performance of K-Nearest Neighbor classification models trained from (i) dataset with all features and (ii) dataset with selected features from the Wrapper technique. It was found that there was no significant difference, so the selected features could represent the models that were similar to the original ones. These selected features are the main contribution of this work since there are potential factors that can cause road accidents. This finding yield an insight information that can be incorporated as a future prevention plan for Thailand and other neighboring countries with similar environment.

CCS CONCEPTS

•Applied computing~Operations research~Transportation

KEYWORDS

Road accident, Thailand, Motorcycles, Classification, Feature Selection

43

The Impact of Data Imputation and Feature Extraction on PM2.5 Forecasting Performance in Bangkok Using Long Short-Term Memory Neural Networks

Panupong Rong-O
Master of Science in Information Technology
Srinakharinwirot University
Bangkok, Thailand
panupong.rongo@g.swu.ac.th

Nuwee Wiwatwattana
Department of Computer Science
Srinakharinwirot University
Bangkok, Thailand
nuwee@g.swu.ac.th

ABSTRACT

The last few years have seen a dramatic increase in PM2.5 air pollution in Thailand's major cities. Various works have tried to develop efficient Long Short-Term Memory (LSTM) deep neural network models for PM2.5 concentration forecasting. However, little has been studied about the impact of data imputation and feature extraction on the model performance in this context. In this paper, we imputed missing values using Kalman Smoothing and Linearly Weighted Moving Average. We utilized the LSTM Autoencoder (LSTM AE) for feature extraction. Using the Chokchai Police station in Bangkok as a case study to predict PM2.5 in the next 24 hours, we demonstrated that the performance gain from training LSTM models with imputed data is more than 7 percent overall with respect to the root mean square error (RMSE) and more than 10 percent overall with respect to the mean absolute error (MAE). Improvement with LSTM AE varies according to time steps. Forecasting 22 to 24 hours ahead tends to favor the use of LSTM AE.

CCS CONCEPTS

• Computing methodologies • Neural networks; • Applied Computing • Environmental sciences.

KEYWORDS

Data imputation, feature extraction, LSTM, LSTM autoencoder, PM2.5 forecasting

23

A System to Estimate the Amount and Calories of Food that Elderly People in the Hospital Consume

Pisol Ruenin
Data Science Research Center,
Department of Computer Science,
Faculty of Science,
Chiang Mai University
Chiang Mai, Thailand
pisol.ruenin@gmail.com

Jakramate Bootkrajang
Data Science Research Center,
Department of Computer Science,
Faculty of Science,
Chiang Mai University
Chiang Mai, Thailand
jakramate.b@cmu.ac.th

Jakarin Chawachat*
Data Science Research Center,
Department of Computer Science,
Faculty of Science,
Chiang Mai University
Chiang Mai, Thailand
jakarin.c@cmu.ac.th

ABSTRACT

Malnutrition in the elderly is an important issue, especially for the elderly in hospitals. In general, nutritionists calculate energy from foods that the elderly need each day. If the elderly eat less food and do not receive enough energy, supplementary food must be provided. Currently, the problems in hospitals are the lack of nutritionists and the calculation of the amount and energy of food takes a long time. In this research, we present a system to estimate the amount and calories of the food that the elderly in the hospital consume by taking pictures in a divided food tray before and after consumption. Our system has three main parts: food detection and classification, food weight estimation, and calories estimation application. In the first part, we use the Faster R-CNN technique and select ResNet-50 as a pre-trained model. Our prediction model was trained on a Suandok Hospital Food Images Dataset (SH-FID) of 16,067 food images from 39 different classes. The result with 4017 food images showed that mAP = 73.354. In the second part, we use the CNN technique, which uses a pre-trained model as InceptionResNetV2. From the experiment with the 4017 food images found that MAPE = 16.9729, which is considered a good prediction. Finally, we created a web application to display the number of calories consumed by the elderly by converting the calculated food weight into calories using the hospital reference table.

CCS CONCEPTS

• Computing methodologies • Computer graphics • Image manipulation • Computational photography

KEYWORDS

Food estimation, Food classification, Image recognition, Machine learning, CNN

30

ARMA Model for Revenue Prediction

Thura Zaw†
University of Information Technology
Yangon, Myanmar
thurazaw@uit.edu.mm

Swe Swe Kyaw
Faculty of Computing,
University of Information Technology
Yangon, Myanmar
sweswekyaw@uit.edu.mm

Aung Nway Oo
Faculty of Computer Science,
University of Information Technology
Yangon, Myanmar
aungnwayoo@uit.edu.mm

ABSTRACT

For every country in over the world, tax revenues appear to be the main engines contributing to the growth momentum. The prediction of tax revenues is one of the main challenges of the Myanmar Internal Revenue Department. It is not easy to get an accurate prediction of the tax revenues of the coming financial year. This is an important issue because the obtained results are used in the decision making of the target budget for the coming year's revenues. In this paper, a model of time series analysis based on Autoregressive Moving-Average (ARMA) for the forecast of the tax revenue collection is introduced. The results were more accurate in comparison to the outcome of the IRD that is estimated with the traditional estimation method. ARMA models are constructed for the prediction of each of four different tax revenues, income tax, commercial tax, lottery tax, and stamp duties. Although the proposed method is applied to four main types of tax revenues, only forecasting of commercial tax revenues is found out in the experiments of this paper. The results show that the error rate reduces to one-third of the traditional forecasting method of the Internal Revenue Department.

CCS CONCEPTS

• Mathematics of Computing • Probability and Statistics • Time Series Analysis • Statistics

KEYWORDS

Time Series Analysis, Forecasting, tax revenues, Auto-Regressive Moving Average (ARMA) Model, Error Rate

8

Data Mining Methods for Optimizing Feature Extraction and Model Selection

Masha Rouzbahman
Microsoft,
Canada.
masha.bahman@gmail.com

Alexandra Jovicic
Sun Life Financial,
Toronto, Ontario,Canada.
alexandra.jovicic@sunlife.com

Lu Wang
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada
wanglu.wang@mail.utoronto.ca

Leon Zucherman
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada
lzucherm@gmail.com

Zahid Abul-Basher
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada
zahid@mie.utoronto.ca

Nipon Charoenkitkarn
School of Information Technology, King Mongkut University of Technology Thonburi, Bangkok, Thailand.
nipon@sit.kmutt.ac.th

Mark Chignell
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada
chignel@mie.utoronto.ca

ABSTRACT

How can we carry out on-the-fly data mining on massive amounts of data, to make relevant predictions, based on data for similar observations to the one currently under consideration? In this paper we show the benefit of using large numbers of computationally efficient analyses to tune the feature extraction, and prediction, steps in data mining, using cross-validated prediction accuracy as the evaluative criterion. Different feature extraction strategies are also compared in terms of their predictive effectiveness in this context. While the research reported here focused on clinical prediction of healthcare outcomes, the results should have broader implications for large scale data mining in general.

CCS CONCEPTS

• Applied computing⟶Health informatics.

KEYWORDS

Data Mining, Feature Extraction, Health Informatics, Information Engineering, Clustering, Regression.

42

Using LSTM for Context Based Approach of Sarcasm Detection in Twitter

Siti Khotijah†
School of Computing
Telkom University
Bandung, Indonesia
sitikhotijah@student.telkomuniversity.ac.id

Jimmy Tirtawangsa
School of Computing
Telkom University
Bandung, Indonesia
jimmytirtawangsa@telkomuniversity.ac.id

Arie A Suryani
School of Computing
Telkom University
Bandung, Indonesia
ardiyanti@telkomuniversity.ac.id

ABSTRACT

In this research, we propose a sarcasm detection by taking into consideration its many varying contexts, related to the word or phrase in a tweet. To get the related context, we extract the information with paragraph2vec to simplify the process of finding the contextual meaning. The result paragraph2vec will provide the features to help classification in Long Short Term Memory (LSTM). Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. We applied a sarcasm detection method to identify sarcasm in two different languages: English and Indonesian and classification with balanced and imbalanced data. It aims to measure the reliability of the proposed approach and how effective the method is in detecting sarcasm. The result of the experiment shows that in Indonesian, balanced data has a good accuracy of 88.33 % and imbalanced data of 76.66 %, whereas in English the balanced data has an accuracy of 79% and imbalanced data of 54.5%.

CCS CONCEPTS

• Computing methodologies • Artificial intelligence • Natural language processing

KEYWORDS

Sarcasm detection, lstm, paragraph2vec, context, deep learning

87

Question Chunk Extraction for Improving Question Similarity

Chaianun Damrongrat
National Electronics and Computer Technology Center (NECTEC)
Pathumthani, Thailand
chaianun.damrongrat@nectec.or.th

Chatchawal Sangkeettrakarn
National Electronics and Computer Technology Center (NECTEC)
Pathumthani, Thailand
chatchawal.san@nectec.or.th

Anocha Rugchatjaroen
National Electronics and Computer Technology Center (NECTEC)
Pathumthani, Thailand
anocha.rugchatjaroen@nectec.or.th

ABSTRACT

Intention classification is one of the challenging issues of question understanding. The same intention could be described in different styles e.g., “when”, “what time.” It is more challenging for a language that has no word separator, flexible grammar and lack of annotated corpus. In Thai, a word “when” could be described as “เมื่อไหร่”, “ตอนไหน”, “กี่โมง”, “เวลาไหน.” The underlining text are question identifier i.e., “when” “which” “how many” and “which”, respectively. With above limitations, traditional word segmentation could not tokenize question words properly especially on question tokens. As a result, word similarity or word suggestion services using word embedding, which heavily relies on word segmentation could not provide a decent result when it is a question. We proposed a question chunk extraction method since question should be tokenized into a chunk instead of a word. This can improve quality of word segmentation and can improve question similarity results. A question chunk (QC) is a combination of question word (QW) and focus word (FW.) We compared our question chunk based word embedding (QCWE) using QC and traditional word embedding (TWE) using traditional word segmentation, the results showed that our QCWE outperformed the TWE in all aspects.

KEYWORDS

Thai question chunk extraction, word embedding, question understanding, natural language processing

61

A 3D-CNN Siamese Network for Motion Gesture Sign Language Alphabets Recognition

Hassaluk Jirathampradub
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
hassaluk.j@mail.kmutt.ac.th

Chakarida Nukoolkit
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
chakarida@sit.kmutt.ac.th

Kaneungjit Suriyathumrongkul
Cititel Co., Ltd.
Bangkok, Thailand
kaneungjit@hotmail.com

Bunthit Watanapa*
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
bunthit@sit.kmutt.ac.th

ABSTRACT

Sign language recognition is developed for assisting hearing- and speech-impaired people who communicate with others by using hand gestures as a nonverbal communication. The system aims to translate still images or video frames of sign gestures to a corresponding message, word or alphabet of a designated language. For the American Sign Language (ASL), there are alphabets of J and Z out of 26 letters which require motion gesture and cannot be recognized using static 2-dimension images as usually possible for the others. Their starting gesture also causes a confusion with the other alphabets, esp. I and D. Samples of these two alphabets J and Z, are scarce. This paper developed a n-shot 3D-CNN deep learning Siamese network model to recognize J and Z and address the said problems. Two experiments were conducted to give insight into the performance of the proposed 3D-CNN Siamese network and the improvement using regularization and dilation techniques. The effect of the number of shot in the Siamese model to the performance of the recognition is also provided in the second experiment.

CCS CONCEPTS

• Computing methodologies • Machine learning • Applied computing

KEYWORDS

3D-Convolutional Neural Network, American Sign Language, Sign Language Recognition

34

A Data Augmentation Technique Based on Text for Vietnamese Sentiment Analysis

Thien Ho Huong
Ho Chi Minh City Open University
Ho Chi Minh City, Vietnam
thienhh.188i@ou.edu.vn

Vinh Truong Hoang
Ho Chi Minh City Open University
Ho Chi Minh City, Vietnam
vinh.th@ou.edu.vn

ABSTRACT

Online opinions are used as a data source that contains relevant information about customer sentiments toward a product or service. This can be used to make a specific decision for customers and management. In order to achieve the good models for sentiment analysis, we require a large humanlabeled data which is costly to obtain. This paper proposes an approach based on text data augmentation based on product reviews in Vietnamese language. Several basic techniques are applied to generate more comments by random insertions, substitutions. The experimental results demonstrate the efficiency of the proposed approach.

CCS CONCEPTS

• Applied computing ⟶ Document preparation.

KEYWORDS

sentiment analysis, product reviews, text augmentation, Vietnamese language, natutal language processing, text mining

85

DeepVix: Explaining Long Short-Term Memory Network With High Dimensional Time Series Data

Tommy Dang
Texas Tech University
Lubbock, Texas, US
tommy.dang@ttu.edu

Hao Van
Texas Tech University
Lubbock, Texas, US
hao.van@ttu.edu

Huyen Nguyen
Texas Tech University
Lubbock, Texas, US
huyen.nguyen@ttu.edu

Vung Pham
Texas Tech University
Lubbock, Texas, US
vung.pham@ttu.edu

Rattikorn Hewett
Texas Tech University
Lubbock, Texas, US
rattikorn.hewett@ttu.edu

ABSTRACT

Machine learning automates the process of analytical model building by means of the computing power of machines. Visual analytics couples interactive visual representations and underlying analysis, putting the human at the center of the analytics and decision makingprocess. This paper aims to combine the strengths of both data science fields into a unified system, called DeepVix, which focuses on the visual explainability of the multivariate time-series predictions using neural networks. Within our DeepVix system, a visual presentation of the neural network explains the intermediate steps, as well as the temporal weights of various gates of the entire learning process. The relationships between input variables and the target variable can also be inferred automatically from the trained model. Interactive operations allow users to explore the neural network, to gain understandings of the model and essential features with layers and nodes, and finally to customize the neural network configurations to fit their needs. We demonstrate our approach with Recurrent Deep Learning on various real-world time series datasets, including the multivariate measurements of a medium-size High-Performance Computing Center, the S&P500 stock data over the past 39 years, and the US employment data retrieved from the Bureau of Labor and Statistics.

CCS CONCEPTS

• Human-centered computing⟶Information visualization.

#52

Provenance Location-based Authentication in Cloud Computing

Phornsawan Roemsri
Department of Computer Science
Texas Tech University
Lubbock Texas USA
phornsawan.roemsri@ttu.edu

Rattikorn Hewett
Department of Computer Science
Texas Tech University
Lubbock Texas USA
rattikorn.hewett@ttu.edu

ABSTRACT

Cloud computing will continue flourishing. Like any technology that is widely used and advancing with many benefits, it also comes with issues, some of which are unavoidable. Security is one of the major concerns for cloud computing. To protect user information leakages or system intrusion, the first simple step is to make sure that only legitimate users can access the computing resources. This makes user authentication an important safeguard. There are many approaches to authentication for users in the cloud environments. The password-based authentication is one such technique that is old but is still in use because it is simple. Unfortunately, it is also vulnerable for attack. To alleviate this drawback, recent research has applied provenance to authentication but most do not deal with location. Using a location as part of authentication increasingly becomes necessary, as mobile and wearable technology become a big part of our daily life. This paper presents a new approach to authentication on cloud computing environments. The approach uses data provenance as well as a location of user request. The paper describes the approach, illustrates its use in use case scenarios, and compares its performance with that of the machine-learning model. Using a data synthesized from simulated authentication system, it is shown that the decision tree model has an average of 83.5% accuracy while ours, by relying on theoretical ground, results in 100% accuracy.

CCS CONCEPTS

• Security and Privacy • System security • Storage security

KEYWORDS

Provenance, authentication, location-based, cloud computing

86

ScagCNN: Estimating Visual Characterizations of 2D Scatterplots via Convolution Neural Network

Vung Pham
Texas Tech University
Lubbock, Texas, US
vung.pham@ttu.edu

Ngan V.T. Nguyen
Texas Tech University
Lubbock, Texas, US
ngan.v.t.nguyen@ttu.edu

Tommy Dang
Texas Tech University
Lubbock, Texas, US
tommy.dang@ttu.edu

ABSTRACT

Scagnostics is a set of visual features that characterizes the data distribution of a 2D scatterplot and has been used in a wide range of applications. However, calculating the scagnostics scores involves computationally expensive algorithms. Moreover, the algorithms are sensitive to the slight changes in the underlying data distribution within the scatterplot. Therefore, this work provides a machine learning model, called ScagCNN, to estimate the scagnostics scores. This model aims to improve the scagnostics computation time and to reduce the sensitivity to the small shifts in the data distribution. This work also provides a web prototype to explore the predictive performance of the model and to give a visual explanation about whether a prediction is accurate. Furthermore, we test the performance of our solution on datasets of various sizes.

CCS CONCEPTS

• Human-centered computing⟶Visual analytics.

KEYWORDS

Visual features, Scagnostics, Convolution Neural Network

48

Health Data Analytics with an Opportunistic Big Data Algorithm

Gantaphon Chalumporn
Department of Computer Science
Texas Tech University, TTU
Lubbock, USA
g.chalumporn@ttu.edu

Rattikorn Hewett
Department of Computer Science
Texas Tech University, TTU
Lubbock, USA
rattikorn.hewett@ttu.edu

ABSTRACT

In data-driven society, health data can lead to profound impacts on public safety policies, epidemic modeling, and advancement of health science and medicine. This paper presents an approach to automatically elucidating useful information from "Big" health data. In particular, we analyze manufactured cosmetic products containing chemicals that are known or suspected to cause cancer, birth defects, or developmental and reproductive harm. Our analysis is based on the Apriori algorithm, the heart of the popular Association Rule Mining to discover associations among sets of influencing factors. However, with rapid growth of huge amount of data, including ours, existing data analytics algorithms designed for in-memory data are not adequate. Most Big data analytics algorithms are implemented on MapReduce framework for execution in parallel and distributed environments. Unlike traditional implementation, our approach employs an opportunistic MapReduce-based Apriori algorithm to fully exploit parallelism. The paper describes the algorithm and presents our findings, from 113,179 data instances, both in terms of the execution times and the discovered associations among product profiles. For a support threshold of 10% (5%,), 20 (53) association rules are obtained with an improved execution time over that of the traditional MapReduce-based algorithm by 14.6% (40.3%) on the average over three machines.

CCS CONCEPTS

CSS ⟶ Distributed algorithms ⟶ MapReduce algorithms

KEYWORDS

Association rules mining, MapReduce, Big Data Algorithms

38

Privacy Protection in Smart Health

Sevgi Arca
Department of Computer Science
Texas Tech University
Texas, USA
sevgi.arca@ttu.edu

Rattikorn Hewett
Department of Computer Science
Texas Tech University
Texas, USA
rattikorn.hewett@ttu.edu

ABSTRACT

Smart health exploits smart health devices (e.g., fitness trackers, heart rate or glucose monitoring units) and Internet of Things technologies to improve users' health and wellness. By enabling self-monitoring and data sharing among users and healthcare professions, smart health can increase healthy habits, timely treatments, reduce hospital visits/re-admissions and even save lives. While smart health comes with great benefits, it also poses a privacy threat to the re-identification of users and their personal data. This paper presents an approach to protecting users' privacy by generalizing critical data so that they belong to multiple users as a way to anonymize user identity. Unlike existing anonymization techniques, our approach efficiently produces shared data that satisfy user-specified anonymity requirements while keeping the data as informative as possible. The approach is based on an Artificial Intelligence search technique using two proposed heuristics. The paper describes and illustrates the approach with experiments to compare its effectiveness with other techniques. The results show that, given a trade-off of privacy preserving, data retention and computational cost, our approach gives the most effective solution for data sharing as expected.

CCS CONCEPTS

• Security and privacy • Human and societal aspects of security and privacy • Privacy protections

KEYWORDS

Anonymization, privacy protection, smart health

#39

Visual Features for Multivariate Time Series

Bao Dien Quoc Nguyen
Department of Computer Science
Texas Tech University
Lubbock, Texas, USA
Bao.D.Nguyen@ttu.edu

Rattikorn Hewett
Department of Computer Science
Texas Tech University
Lubbock, Texas, USA
rattikorn.hewett@ttu.edu

Tommy Dang
Department of Computer Science
Texas Tech University
Lubbock, Texas, USA
tommy.dang@ttu.edu

ABSTRACT

Visual analytics combines the capabilities of computers and humans to explore the insight of data. It provides coupling interactive visual representations with underlying analytical processes (e.g., visual feature extraction) so that users can utilize their cognitive and reasoning capabilities to perform complex tasks effectively or to make decisions. This paper applies successfulness of visual analytics to multivariate temporal data by proposing an interactive web prototype and an approach that enables users to explore data and detect visual features of interest. A list of nonparametric quantities is proposed to extract visual patterns of time series as well as to compute the similarity between them. The prototype integrates visualization and dimensional reduction techniques to support the exploration processes. Many different temporal datasets are used to justify the effectiveness of this approach, and some remarkable results are presented to show its value.

CCS CONCEPTS

Human-centered computing ⟶ Visual analytics.

KEYWORDS

visual features extraction, clustering method, dimension reduction

#JSCI05

Efficient Distributed SNP Selection by a Modified Binary Flower Pollination Algorithm

Wanthanee Rathasamuth
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
rathasamuth.wan@gmail.com

Kitsuchart Pasupa*
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
kitsuchart@it.kmitl.ac.th

ABSTRACT

Porcine Single Nucleotide Polymorphisms (SNPs—certain pieces of nucleotide in a DNA sequence) can be indirectly associated with traits of an individual pig, like its meat quality or resistance to common diseases. It is most desirable to obtain a smallest number of most significant SNPs in genomics research, and several computer classification algorithms have been used to do so. For instance, for breed classification, one needs to obtain a set of a much smaller number of significant SNPs than that of the entire SNP data set. This study attempted to find such significant porcine SNPs by using computational feature selection and classification methods. In a preliminary trial, a binary flower pollination algorithm (BFPA) was used and shown not to able to reduce the number of selected SNPs to a sufficiently low number. Therefore, to achieve our objective, we developed a vertically distributed feature selection method incorporating a modified BFPA and a support vector machine classifier for selecting significant porcine SNPs. The developed method was evaluated and compared against four baseline methods. It provided the smallest average number of significant SNPs (128.40) that resulted in 94.57% classification accuracy. This and other findings in this study may directly benefit researchers in the bioinformatics field in their effort to map SNPs.

CCS CONCEPTS

• Applied computing⟶Bioinformatics; • Computing methodologies⟶Search methodologies.

KEYWORDS

Single nucleotide polymorphism, Feature selection, Flower pollination algorithm, Support vector machine

1st-degree Atrioventricular (AV-block) and Bundle Branch Block Prediction using Machine Learning

Risul Islam Rasel†
Department of CSE
Chittagong Independent University
Chittagong, Bangladesh
rasel.kmutnb@gmail.com

Nasrin Sultana
Department of CSE
University of Chittagong
Chittagong, Bangladesh
nasrin1219@email.com

Phayung Meesad
Faculty of IT
King Mongkut's University of Technology North Bangkok
pym@kmutnb.ac.th

Anupam Chowdhury
Department of CSE
International Islamic University Chittagong
Chittagong, Bangladesh
anu8chy@gmail.com

Meherab Hossain
Department of CSE
International Islamic University Chittagong
Chittagong, Bangladesh
h.meherabhossain@gmail.com

ABSTRACT

Heart block occurs when the flow of electricity interrupted or partially delayed between the top and bottom chambers of the heart. People are now more often affected by this kind of disease. However, early prediction of heart block can reduce the diagnosis complexity and treatment cost. In this study, a data mining and machine learning model is proposed to predict three types of heart blocks, such as 1st-degree A-V block, Left Bundle Branch Block (LBBB), and Right Bundle Branch Block (RBBB). Experiment data samples are collected from the cardiology department of Chittagong Medical College Hospital (CMCH), Bangladesh. The dataset contains 32 types of numeric and categorical features about the patient's ECG report, daily activities, and food habits. The prediction model has been designed, trained, and tested with some empirical machine learning algorithms namely Decision Tree, Random Forest, K-Nearest Neighbor, and Support Vector Machine. Finally, the experimentation shows that Decision Tree and Random Forest models outperform the other algorithms in overall heart block prediction with an accuracy of more than 92%.

CCS CONCEPTS

• Computing methodologies ~ Machine learning ~ Machine learning approaches ~ Classification and regression trees

KEYWORDS

Heart block prediction, 1st-degree A-V block, Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB)

Effects of Part-of-Speech on Thai Sentence Classification to Wh-Question Categories using Machine Learning Approach

Saranlita Chotirat †
Information Technology Program,
Department of Information Technology,
Faculty of Information Technology and Digital Innovation
King Mongkut's University of Technology North Bangkok,
Thailand
saranlita.c@email.kmutnb.ac.th

Phayung Meesad
Department of Information Technology Management,
Faculty of Information Technology and Digital Innovation
King Mongkut's University of Technology North Bangkok,
Thailand
pym@kmutnb.ac.th

ABSTRACT

In the last decade, question classification is a strong signal for answer selection and help to find the structure of question sentences from sentences. For this paper, we evaluated the proposed pre-processing method for classifying the simple sentence to wh-question categories ("What", "When", "Who", "Where", and "How") on Thai texts by considering Part-of-Speech tagging (POS). The performances are evaluated using classification accuracy obtained from traditional classification models including Naïve Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbors, and neural networks which employs Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). We compared traditional models and neural networks; the experimental results showed that the result of the neural networks models better than the traditional model. The accuracy of the proposed model using the LSTM model with pre-trained word embedding is improved with an average F1 of 79.60%.

CCS CONCEPTS

• Computing methodologies • Machine learning • Machine learning algorithms • Feature selection

KEYWORDS

Question classification, Sentences classification, Natural language processing, Part-of-Speech tagging

Correlation-Based Incremental Learning Network with Sliding Window for Perfume Classification

Panida Lorwongtrakool
Department of Information Technology
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
panida.l@email.kmutnb.ac.th

Phayung Meesad
Department of Information Technology Management
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
pym@kmutnb.ac.th

ABSTRACT

Contamination inspection or quality inspection of raw materials or products is a very important task, especially in the perfume industry that requires an expert for inspection. However, the human nose has limitations such as fatigue, which affects the accuracy. Therefore, an electronic nose or sensor array has been developed to assist in the inspection. The signal data from electronic nose is fed into machine learning models to learn and process. Since the data change over time, the input data fluctuate according to the changing environment. In addition, when there are new data with features that change from the original patterns, the classification outcome may not be correct and the model will not be able to classify as effective as the original model. Therefore, to solve the problem mentioned this research proposes Correlation-Based Incremental Learning Network with Sliding Window (CILNS), which learns automatically by adapting to new data while maintaining the existing knowledge. The experiments were conducted on classifying perfumes. The experimental data were divided into 4 batches.Batch 1 wasused asthe training data and the otherbatches were used asthe testing data.The proposed algorithm was compared with other well-knownclassifiers. The results showed that the proposed CILNS algorithm model 4-1 (Window size = 4, step size = 1) provides the highest accuracy of 95.16%.

CCS CONCEPTS

•Computing methodologies •Machine learning •Online learning setting

KEYWORDS

Correlation, Incremental learning, Sliding window, e-nose, Perfume classification

Defect Detection and Quality Level Assignment in Wet Blue Goatskin

Carlos E. B. Sousa
Postgraduate Program in Science Computer
Federal Institute of Education,
Science and Technology of Ceará
Fortaleza - CE, Brazil
estevao.sousa@ppgcc.ifce.edu.br

Cláudio M. S. Medeiros
Postgraduate Program in Science Computer
Federal Institute of Education,
Science and Technology of Ceará
Fortaleza - CE, Brazil
claudiosa1965@gmail.com

Renato F. Pereira
Postgraduate Program in Science Computer
Federal Institute of Education,
Science and Technology of Ceará
Fortaleza - CE, Brazil
renato.francisco.pereira@gmail.com

Mateus A. V. Neto
Postgraduate Program in Science Computer
Federal Institute of Education,
Science and Technology of Ceará
Fortaleza - CE, Brazil
mateus.vieira@ifce.edu.br

Alcides A. Neto
Postgraduate Program in Science Computer
Federal Institute of Education,
Science and Technology of Ceará
Fortaleza - CE, Brazil
alcides.andrade@ifce.edu.br

ABSTRACT

Tanneries acquire hides, in most cases, from rural workers, so, due to the informality of the creation, slaughter and extraction of the animal's skin, they receive them with different types and different defect levels. That said, classifying acquired and processed skins become very complex and tiring activity. The leather discrimination process is completely handmade and subjective, too dependent on the experience of the professional responsible for this step, which, due to tiredness, stress, and other factors, end up generating several errors in this process. Currently, there are several studies in the literature related to the detection of leather flaws, however, few studies go further and qualify the skins based on the detected problems. In view of this factor, a system based on computer vision and artificial intelligence are proposed in which it obtains an accuracy rate of 95.9 % in the detection of defects in wet blue goatskin and 93.3 % in the identification of the quality level of these parts.

CCS CONCEPTS

• Computing methodologies → Artificial intelligence •Computing methodologies → Computer vision • Applied computing → Enterprise computing

KEYWORDS

Goatskin, Pattern Recognition, Quality Level, Textile industry

Car Damage Detection and Classification

Phyu Mar Kyu[1]
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
62606003@kmitl.ac.th

Kuntpong Woraratpanya
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
kuntpong@it.kmitl.ac.th

ABSTRACT

Nowadays, the proliferation of automobile industries is directly related to the increasing number of car incidents. So, insurance companies are facing many simultaneous claims and solving claims leakage. The sense of Artificial Intelligence (AI) based on machine learning and deep learning algorithms can help to solve these kinds of problem for insurance industries. In this paper, we apply deep learning-based algorithms, VGG16 and VGG19, for car damage detection and assessment in real-world datasets. The algorithms detect the damaged part of a car and assess its location and then its severity. Initially, we discover the effect of domain-specific pre-trained CNN models, which are trained on an ImageNet dataset, and followed by fine-tuning, because some of the categories can be fine-granular to get our specific tasks. Then we apply transfer learning in pre-trained VGG models and use some techniques to improve the accuracy of our system. We achieve the accuracy of 95.22% of VGG19 and 94.56% of VGG16 in the damaged detection, the accuracy of 76.48% of VGG19 and 74.39% of VGG16 in damage localization, the accuracy of 58.48% of VGG19 and 54.8% of VGG16 in damage severity with the combination of transfer learning and L2 regularization. From their results, the performance of VGG19 is better than VGG16. After analyzing and implementing our models, we find out that the results of using transfer learning and L2 regularization can work better than those of fine-tuning.

CCS CONCEPTS

• Computing methodologies • Machine learning • Machine learning approaches • Neural networks

KEYWORDS

Damage assessment, Deep learning, Machine learning, Transfer learning, Pre-trained VGG

Stock Price Analysis with Natural Language Processing and Machine Learning

Sukanchalika Boonmatham†
Department of Information Technology,
Faculty of Information Technology and Digital Innovation,
King Mongkut's University of Technology North Bangkok,
Bangkok, Thailand
sukanchalika.b@email.kmutnb.ac.th

Phayung Meesad
Department of Information Technology Management,
Faculty of Information Technology and Digital
Innovation, King Mongkut's University of Technology
North Bangkok, Bangkok, Thailand
pym@kmutnb.ac.th

ABSTRACT

Finding stock price classification based on Thai news corporate is a challenging task. In this research, we try to build machine learning models that capture the relationship of news and stock prices of several companies. In this work, eight companies were selected randomly from Industry Group Index and Sectoral Index. Corporate news articles from the eight selected companies were collected along with their stock prices. Two of traditional machine learning models and two deep learning models were used in this study for comparison purpose. The models were based on Support Vector Machine (SVM), Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). Using news articles as inputs, the models were trained to classify stock prices into two classes: Up and Down of the stock closing price. For classification performance, Accuracy, Precision, Recall and F1 were used. The results showed that GRU had highest average accuracy, precision, recall and F1 higher than other model values with 0.79, 0.79, 0.79, 0.79, respectively.

CCS CONCEPTS

• Computing methodologies • Machine learning

KEYWORDS

Text Mining, Deep Learning, Classification Model, Natural language processing

www.iait-conf.org

IAIT2020

SPONSOR

Professional Computer Company Limited
บริษัท โปรเฟสชั่นนัล คอมพิวเตอร์ จำกัด

3BB BROADBAND

Microsoft

CDG Systems Ltd.

Softever

FWD
เอฟดับบลิวดี ประกันชีวิต

SUPPORTER

KMUTT

SIT 25 YEARS
"Empowering Boundless Digital Transformation"
School of Information Technology
King Mongkut's University of Technology Thonburi

NECTEC
a member of NSTDA

acm Thailand

ส.ก.พ.ท.

depa

สคทส. CITT

SOFTWARE PARK THAILAND