CSCI 2020 BOOK of ABSTRACTS

The 2020 International Conference on Computational
Science and Computational Intelligence (CSCI'20)

https://www.american-cse.org/csci2020/
December 16-18, 2020
Luxor Hotel (MGM Property), 3900 Las Vegas Blvd. South, Las Vegas, 89109, USA

Table of Contents

Message from Program and General Co-Chairs

It gives us great pleasure to introduce this collection of papers to be presented at The 2020 International Conference on Computational Science and Computational Intelligence (CSCI'20), December 16-18, 2020, Las Vegas, Nevada, USA.

The CSCI'20 International Conference includes papers from diverse communities, including researchers from: universities, corporations, and government agencies. Accepted papers are published by Conference Publishing Services (CPS). The full proceedings/book will be published soon after the conference (like prior years). Papers published in the proceedings present solutions to problems in many important areas of computational science and computational intelligence.

Computational Science (CS) is the study of addressing problems that are impossible to solve (or difficult to solve) without computers. CS can be considered to be the bridge between computer science and other sciences. The field is interdisciplinary by nature and includes the use of advanced computing capabilities to understand and solve complex problems. In short, CS is the science of using computers to do science. Computational Intelligence (CI) is the study of computational methods in ways that exhibit intelligence. These methods adapt to changing environments and changing goals. There is a significant overlap between the fields of CI and Artificial Intelligence (AI). However, there is also a difference: in general AI techniques often involve top-to-bottom methods (i.e., methods to the solutions are imposed from the top) whereas CI techniques often involve bottom-up methods (i.e., solutions emerge from unstructured beginnings). An important part of CI includes a set of Nature-inspired computational approaches to address complex problems to which traditional methods are infeasible. Computational Science and Computational Intelligence, both share the same objective: finding solutions to difficult problems. However, as stated earlier, the methods to the solutions are different.

Considering the above broad outline, the International Conference on Computational Science and Computational Intelligence (CSCI'20) is composed of the following topical symposiums: Computational Science (CSCI-ISCS); Computational Intelligence (CSCI-ISCI); Computational Biology (CSCI-ISCB); Cyber Warfare, Cyber Defense, & Cyber Security (CSCI-ISCW); Artificial Intelligence (CSCI-ISAI); Smart Cities and Smart Mobility (CSCI-ISSC); Big Data and Data Science (CSCI-ISBD); Education - CS & CE (CSCI-ISED); Health Informatics and Medical Systems (CSCI-ISHI); Mobile Computing, Wireless Networks, & Security (CSCI-ISMC); Software Engineering (CSCI-ISSE); Internet of Things & Internet of Everything (CSCI-ISOT); Social Network Analysis, Social Media, & Mining (CSCI-ISNA); Cloud Computing and Data Centers (CSCI-ISCC); Parallel & Distributed Computing (CSCI-ISPD); and Signal & Image Processing, Computer Vision & Pattern Recognition (CSCI-ISPC).

The main objective of the CSCI Conference is to facilitate increased opportunities for cross-fertilization across CS and CI. The CSCI Conference is committed to encouraging diversity and eliminating discrimination in both its role as a conference and as a provider of services. CSCI aims to create a culture that respects and values each others' differences, that promotes dignity, equality and diversity, and that encourages individuals to develop and maximize their true potential. We are committed wherever practicable to organizing a conference that broadly reflects the international community. We hope that we have achieved these important objectives.

The Steering Committee and the Program Committee would like to thank all those who submitted papers for consideration. The conference had paper submissions from 56 countries. About 55% of the submissions were from outside the United States. Each submitted paper was peer-reviewed by at least two experts in the field for originality, significance, clarity, impact, and soundness. In cases of contradictory recommendations, a member of the conference program committee was charged to make the final decision; often, this involved seeking help from additional referees. In addition, papers whose authors included a member of the conference program committee were evaluated using the double-blinded review process. One exception to the above evaluation process was for papers that were submitted directly to chairs/organizers of sessions/workshops; in these cases, the chairs/organizers were responsible for the evaluation of such submissions. The overall paper acceptance rate for regular and short papers was 16%; and 20% of the remaining papers were accepted as extended abstract (poster) papers (at the time of this writing, we had not yet received the acceptance rate for three research tracks.)

We are very grateful to the many colleagues who offered their services in organizing the conference. In particular, we would like to thank the members of the Program Committee and the Steering Committee of CSCI'20. The members of the committees will be requested (after the conference) to provide their expertise and services for selecting papers for publication (extended versions) in various research book series (to be prepared for publishers including: Springer, Elsevier, and others). We would also like to thank the main sponsor of the conference: American Council on Science & Education.

We express our gratitude to all speakers and authors - the list of speakers appears in the conference schedules. We would also like to thank the followings: the staff of the Luxor hotel (Conference division); and Conference Publishing Services of IEEE Computer Society production editors and managers: Juan Guerrero, Patrick Kellenberger, Lisa O'Conner, and Lorretta Palagi.

We present the proceedings of CSCI'20.

Professor Hamid R. Arabnia - Editor-in-Chief, The Journal of Supercomputing (Springer); Fellow & Adviser of Center of Excellence in Terrorism, Resilience, Intelligence & Organized Crime Research (CENTRIC); Professor Emeritus of Computer Science, The University of Georgia, Georgia, USA

Professor Leonidas Deligiannidis - Professor, Department of Computer Science and Computer Networking, Wentworth Institute of Technology, USA

Professor Fernando G. Tinetti - Facultad de Informatica, Universidad Nacional de La Plata, La Plata, Argentina; Researcher CIC Prov. Bs. As., Argentina; Recipient of the NSF-TCPP Early Adopter Award, NSF/IEEE-TCPP Curriculum Initiative on Parallel and Distributed Computing

Professor Quoc-Nam Tran – Department of Computer Science, Southeastern Louisiana University, Louisiana, USA

Co-Chairs, Steering Committee, CSCI'20

Conference Organization – CSCI 2020

Steering Committee Co-chairs

Professor Hamid R. Arabnia - Editor-in-Chief, The Journal of Supercomputing (Springer); Fellow & Adviser of Center of Excellence in Terrorism, Resilience, Intelligence & Organized Crime Research (CENTRIC); Professor Emeritus, Department of Computer Science, The University of Georgia, Georgia, USA

Professor Leonidas Deligiannidis - Professor, Department of Computer Science and Computer Networking, Wentworth Institute of Technology, Massachusetts, USA

Professor Fernando G. Tinetti - Facultad de Informatica, Universidad Nacional de La Plata, La Plata, Argentina; Researcher CIC Prov. Bs. As., Argentina; Recipient of the NSF-TCPP Early Adopter Award, NSF/IEEE-TCPP Curriculum Initiative on Parallel and Distributed Computing

Professor Quoc-Nam Tran – Department of Computer Science, Southeastern Louisiana University, Louisiana, USA

Members of Steering Committee

Professor Babak Akhgar - Fellow of the British Computer Society, CITP, Professor of Informatics, Co-Director of CENTRIC (Centre of Excellence in Terrorism, Resilience, Intelligence & organised Crime research), Sheffield Hallam University, Sheffield, United Kingdom

Professor Abbas M. Al-Bakry - University President, University of IT & Communications, Baghdad, Iraq

Professor Emeritus Nizar Al-Holou - Electrical and Computer Engineering; Vice Chair, IEEE/SEM-Computer Chapter, University of Detroit Mercy, Michigan, USA

Professor Rajab Challoo - Professor and Chairman, EECS Department; Director, Center for Controls, Robotics, Intelligent Systems and Sensors Research, CRISS; President, Faculty Senate; Former Chair of Computer Society (for 7 years); Texas A&M University-Kingsville, Kingsville, Texas, USA

Professor Chien-Fu Cheng - Professor of CS, Department of Computer Science and Information Engineering, Graduate Institute of Networking & Communication, Tamkang University, Taiwan, R.O.C.

Professor Hyunseung Choo - Director of Intelligent HCI Convergence Research Center (ITRC) of Ministry of Information and Communication; International Advisory Chair, ACM RACS; College of Information and Communication Engineering, Professor of IS, Department of Interaction Science, Sungkyunkwan University (SKKU), Suwon, Republic of Korea

Professor Emeritus Kevin Daimi - Computer Science and Software Engineering Programs, Department of Mathematics, Computer Science and Software Engineering, University of Detroit Mercy, Detroit, Michigan, USA

Professor Eman M. El-Sheikh - Director, Center for Cybersecurity; Member of AWIS (Association for Women in Science); Professor of CS, Department of Computer Science, University of West Florida, Pensacola, Florida, USA

Professor Mary Mehrnoosh Eshaghian-Wilner – Adjunct Professor of Electrical Engineering, University of California Los Angeles, UCLA, California, USA

Dr. David L. Foster - Associate Professor, Computer Engineering; Kettering University, Michigan, USA

Professor Ching-Hsien (Robert) Hsu - Department of Computer Science and Information Engineering, Chung Hua University, Hsinchu, Taiwan, R.O.C.; and Distinguished Chair Professor, Tianjin University of Technology, P. R. China

Members of International Program Committee (CSCI 2020)

- **Professor Afrand Agah** - Graduate Coordinator, Department of Computer Science, West Chester University of Pennsylvania, Pennsylvania, USA
- **Bharat Bhushan Agarwal** - Reader, IFTM University, Moradabad, India
- **Dr. Omaima Nazar Ahmad AL-Allaf** - CIS Deptartment, Faculty of Sciences and Information Technology, Al-Zaytoonah University of Jordan, Amman, Jordan
- **Professor Wasim A. Al-Hamdani** - Professor of Cryptography and InfoSec; Information Security Lab.; Computer Science; Kentucky State University, Frankfort, Kentucky, USA
- **Dr. Ismail Khalil Al Ani** - Dean, College of Engineering and Computer Science, Ittihad University, Ras Al Khaimah, United Arab Emirates (UAE)
- **Professor Mehran Asadi** - Information Technology; Department of Business and Entrepreneurial Studies, Lincoln University, Pennsylvania, USA
- **Dr. Dr. Travis Atkison** - Assistant Professor, Computer Science, The University of Alabama, Tuscaloosa, Alabama, USA
- **Dr. Azita Bahrami** - President, IT Consult, Georgia, USA
- **Dr. Mehdi Bahrami** - IEEE Senior Member; Fujitsu Laboratories of America, Inc., California, USA
- **Dr. Petra Saskia Bayerl** - Department of Organisation and Personnel Management, Rotterdam School of Management (RSM), Erasmus University Rotterdam, Holland, The Netherlands
- **Professor Jane M. Binner** - Chair in Finance, Department of Accounting and Finance, Birmingham Business School, University of Birmingham, Birmingham, West Midlands, United Kingdom
- **Professor Dr. Juan-Vicente Capella-Hernandez** - Universitat Politecnica de Valencia (UPV), Department of Computer Engineering (DISCA), Valencia, Spain
- **Professor Juan Jose Martinez Castillo** - Director, The Acantelys Alan Turing Nikola Tesla Research Group and GIPEB, Universidad Nacional Abierta, Venezuela
- **Dr. Rui Chang** - Lab Director; Department of Genetics and Genomic Science, Icahn Institute for Genomics and Multiscale Biology, Mount Sinai School of Medicine, New York, USA
- **Professor Dongsheng Che** - Director, Che Lab of Bioinformatics at ESU, Department of Computer Science, East Stroudsburg University of Pennsylvania, Pennsylvania, USA
- **Dr. Jianhung Chen** - Department of Computer Science and Information Engineering, Chung-Hua University, Hsin-Chu, Taiwan
- **Professor Mu-Song Chen** - Department of Electrical Engineering, Da-Yeh University, Changhua, Taiwan, R.O.C.
- **Dr. Xin (Thomas) Chen** - Research Corporation of the University of Hawaii, Institute for Astronomy, Honolulu, Hawaii, USA
- **Professor Steve C. Chiu** - Program Director of Electrical Engineering, College of Engineering, Idaho State University, Idaho, USA
- **Professor Mark Yul Chu** - Department of Electrical Engineering, The University of Texas Rio Grande Valley, Texas, USA
- **Professor Ann Leslie Claesson-Vert** (Vice Chair) - School of Nursing, Northern Arizona University, North Valley Campus, Phoenix, Arizona, USA
- **Dr. Jose Alfredo F. Costa** - Member of IEEE Computational Intelligence Society; President, SBIC - Brazilian Computational Intelligence Society; Associate Professor, Department of Electrical Engineering, UFRN - Universidade Federal do Rio Grande do Norte (Federal University - UFRN), Brazil
- **Dr. Arianna D'Ulizia** - Institute of Research on Population and Social Policies, National Research Council of Italy (IRPPS), Rome, Italy
- **Professor Zhangisina Gulnur Davletzhanovna** - Vice-rector of the Science, Central-Asian University, Kazakhstan, Almaty, Republic of Kazakhstan; Vice President of International Academy of Informatization, Kazskhstan, Almaty, Republic of Kazakhstan
- **Professor Noel De Palma** - Associate Vice President for Research, Computer Science; Leader, ERODS Research, LIG Labs.; Board of Directors, OW2 Consortium; Department of Computer Science; University of Grenoble I, Grenoble, France

Note that the title of papers and authors' names that appear in the "Book of Abstracts" were extracted from the papers that were submitted to the EVALUATION web site. The official published proceedings/book will have any and all changes/revisions that authors may have done to the title and/or authors lists in the final version of their manuscripts.

CSCI-ISOT: INTERNET OF THINGS & INTERNET OF EVERYTHING

Proposal of Heartbeat-transmitting Application for Long-distance Communication

Yui Tanaka, Takayuki Fujimoto
Toyo University, Tokyo, Japan

Abstract—In recent years, telecommuting, business trips, and studying abroad have become quite common. Against this backdrop, improvement in ways of long-distance communication such as video calls is needed. Especially between family members and partners, it is desirable to close both physical and emotional distance. With the advancement in information technology, communication became much easier than before. However, the problem of emotional distance still remains unsolved. Having too much communication out of sheer desire to get partner's information can be invasion of privacy. Research efforts are currently being made in various approaches, but it is difficult to have a 'realistic experience' of actually being with the partner. Even a successful tool would require another device in many cases, and it often comes with high initial costs. Therefore, we focus on 'heartbeat' in this study, and propose an application that transmits heartbeat by using the pulse sensor of smartwatch.

Performance Analysis of IoT Physical Layer Security Using 3-D Stochastic Geometry

Hela Chamkhia, Aiman Erbad, Abdullah Al-Ali, Amr Mohamed, Ahmed Refaey, Mohsen Guizani
Computer Science and Engineering, Qatar University, Qatar;
College of Science and Engineering, Hamad Bin Khalifa University, Qatar;
Electrical Engineering, University of Western Ontario, Canada

Abstract—The internet of things (IoT) is becoming part of the infrastructure supporting various services in every day's life. Due to the complex nature of IoT systems with heterogeneous devices, the needed security and privacy aspects are mostly ignored in the initial system design. One of the proposed solutions to address the security threats from the physical layer perspective is physical-layer security (PLS). We propose the use of 3-D stochastic geometry to accurately model IoT systems in a realistic scenarios, where sensors, access points, and eavesdroppers are randomly located in a 3-D space. We use our model with realistic system deployment parameters to conduct rigorous performance analysis for critical security metrics, such as the successful transmission probability (STP) and the secrecy outage probability (SOP) in different potential IoT scenarios. We finally utilize simulation to validate the theoretical analysis.

Key Generation Based Fuzzy Logic and Elliptic Curve Cryptography for Internet of Things (IoT) Authentication

Abderrazak Abdaoui, Ayman Erbad, Abdulla Al-Ali, Amr Mohamed, Mohsen Guizani
College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar;
Computer Science Department College of Engineering, Qatar University, Al Tarfa, Doha, Qatar

Abstract—The security and privacy of the network in the Internet of Things is playing an important challenge for researchers and engineers. Considering that packets are exchanged between the end user and the sensing devices, it is then important to ensure the security, privacy and integrity of the transmitted data by designing a secure and a lightweight authentication protocol for the IoT environment. In this paper, we present a novel method of authentication and encryption based on Elliptic Curve Cryptography (ECC) and on random numbers generated by fuzzy logic for the improvement of the authentication and the encryption in IoT systems. We evaluate our novel key generation method using standard randomness tests such as: Frequency test, Frequency test with mono block, run test, discrete Fourier transform test and advanced discrete Fourier transform test. Our results show superior performance.

An IoT Mutual Authentication Scheme based on PUF and Blockchain

Ore Ndiaye Diedhiou, Cherif Diallo
Laboratoire d'Algebre, de Cryptographie, Codes et Applications (LACCA), Dept Informatique, UFR Sciences Appliquees et Technologie (UFR SAT), Universite Gaston Berger (UGB), Saint-Louis, Senegal

Abstract—Security is one of the major challenges of the Internet of Things (IoT). In an IoT network, data are processed and exchanged without human intervention. Because of this autonomy, objects must authenticate each other as well as to ensure the integrity of the exchanged data. An efficient authentication scheme allows to protect the network against several attacks. Several IoT authentication schemes have already been proposed but they are mostly ineffective and sometimes have limitations. This work proposes a new mutual authentication scheme for IoT based on digital signatures, the Physical Unclonable Functions (PUF) and the blockchain technology. The Global Assessment and analyses show that our new protocol gives more resistance to different types of attacks, and that it also provides a better performance in terms of computing load although it requires fewer storage resources.

Real-time Asset Management and Localization with Machine Learning and Bluetooth Low Energy Tags

Pragya Varshney, Harshveer Saini, Varick L. Erickson
Department of Computer Science, California State University, East Bay, California, USA

Abstract—Effectively managing assets is crucial for businesses, yet asset management can be expensive, time-consuming, and labor intensive. Improving asset management is extremely desirable to increase efficiency and profitability. Nearly 92% of companies invest in technologies to help manage assets. Technologies such as barcodes and Radio Frequency Identification (RFID) tags have long been used for asset tracking. However, these strategies are not able to precisely localize assets and cannot do so in realtime. Other strategies utilizing triangulation have been attempted using radio parameters such as RSSI and time of flight. However, these strategies are highly sensitive to environmental factors. Also, models made from radio parameters often generalize poorly. In this paper, we develop an improved asset tracking method leveraging Bluetooth, RSSI, and Packet Reception Rate. This approach overcomes many limitations of traditional asset tracking methods and enables assets to be accurately tracked in real-time. By training models specific to a deployment, we drastically reduce our sensitivity to environmental factors and improve accuracy. We demonstrate localization accuracies of 85% to 94% and show that these models can be easily trained. Additionally, the models require very minimal training data (only 30-40 samples per room) in order to achieve localization accuracies.

Root Causes of Insecure Internet of Things and Holistically Addressing Them

Christopher A. White
Marymount University, Arlington, Virginia
Washington, District of Columbia

Abstract— The use of Internet of Things (IoT) continues to grow rapidly and with this growth comes increased cybersecurity risks. We continue to observe public cybersecurity incidents caused by insecure IoT and statistics depicting the poor cyber hygiene of IoT devices. This paper identifies many of the root causes for insecure IoT and identifies the need for a holistic approach to IoT security that includes the adoption of secure system development lifecycle practices, prioritizing aspects of organizational and human psychology, improved economic models, and resolution of challenges inherent to the technology and operational use cases for IoT.

An NFC Based Student Attendance Tracking/Monitoring System Using an IoT Approach

Janea Dixon, Abdel-Shakour Abuzneid
Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT, USA

Abstract—Student attendance shapes many aspects of a university; however, it's still widely recorded by hand. While it is common for a professor to pass around a sign-in sheet to enter in attendance to a web-based application later manually, this makes for an unreliable method, leaving many inconsistencies. With today's advances in technology, an automated attendance recording, tracking, and monitoring system will greatly improve the efficiency and reliability of attendance tracking. Several other works have proposed solutions using near field communication and IoT technologies; however, many use a single point at the start of a lecture to record attendance. This is likely to cause bottlenecks and take away time from the course. We implemented a system that incorporates NFC, cloud-based services, and a web interface for end-users. Our design introduces a one-to-one method using a student ID with an RFID tag to RFID reader located at each desk.

Customized Services Using Voice Assistants

Kori Painchaud, Leonidas Deligiannidis
Computer Science and Networking, Wentworth Institute of Technology, Boston, MA, USA

Abstract - Internet of Things (IoT) devices can be defined as a collection of computing devices that communicate and transfer data with one another. As the popularity of IoT devices increases, users want to maximize the functionality of their devices. Voice-enabled IoT devices serve and assist the user by performing functions like playing music, controlling lights, setting alarms and reminders, and much more. The popularity of these devices has grown, and they appeal to consumers because of the available accessories that can be purchased and connected to, such as smart lights, smart shades, etc. These smart accessories can connect to devices such as the Google Home or the Amazon Echo, allowing the user to control multiple common house functions with their voice. This paper demonstrates how to control non-smart LED lights using voice commands with smart home devices. A service known as IFTTT, “If This, Then That”, is utilized to add custom commands to two smart speakers. This paper shows how accessible and simple it is for an individual to control a non-smart device using voice commands, while shedding light on how the use of VPAs can aid people with disabilities. The security risks and threats of using IFTTT are addressed.

Prospectus: An Online Polymorphic Attack Detection Model for Intelligent Transportation Systems

Sultan Ahmed Almalki, Frederick Sheldon
Computer Science Department, University of Idaho, Moscow, Idaho, USA

Abstract—Cooperative Intelligent Transportation Systems (cITSs) is one of the Internet of Things (IoT) applications whose purpose is to improve road safety and traffic efficiency. Within this system, vehicles can communicate with one another by establishing a Vehicular Ad-Hoc Network (VANET) along the road section. Although such connectivity facilitates the exchange of information related to road safety and traffic efficiency, it puts the vehicles at risk in that an attacker could compromise one or more vehicles and use them to share false information causing congestions and/or life-threatening accidents. Although several studies tried to address this issue, they assume that the network topology and/or attack behaviour is stationary, which is not realistic as the cITS is dynamic in nature and the attackers may have the ability and resources to change their behaviour continuously. Therefore, these assumptions are not suitable and lead to low detection accuracy and high false alarms. To this end, this paper proposes a misbehaviour detection model that can cope with the dynamicity of both cITS topology and attack behaviour. The model starts by addressing the issue of missing data that happen at the early stages of the model formation after a topology change. Then, the deep learning approach is used to select the discriminative features used to train. We expect that the proposed model will help to overcome the limitations of related solutions by detecting attacks that change their behaviour continuously.

Applying an Energy-Aware Security Mechanism in Healthcare Internet of Things

Mona Tavakolan, Ismaeel A. Faridi
Computer and Information Sciences, Towson University, Towson, Maryland, USA
Applied Information Technology, Towson University, Towson, Maryland, USA

Abstract—The security risks associated with healthcare Internet of things (HIoT) devices and medical devices includes patient privacy, data confidentiality, and data security. Insecure medical devices allow hackers to infiltrate systems and control medical devices to potentially harm patients and hospitals. HIoT devices have computational and memory limitations, energy limitations, mobility and scalability limitations. Therefore, securing HIoT devices is immensely aggravated due to their resource constrained. Finding a security solution that minimizes resource consumption and thus maximizes security performance is a challenging task in HIoT. This paper proposes adaptive security mechanism that is considering energy level of HIoT devices. This energy-efficient security mechanism considers the residual energy to apply security mechanism and improve lifetime of HIoT devices.

Human Capacity Organizational Roadmap for E-Government Interoperability in the Philippines

Kevin Matthe Caramancion, Pierre Pauline Abesamis
Department of Information Science, University at Albany, State University of New York, New York, USA
College of Computer Studies, De La Salle University - Manila, Philippines

Abstract—This paper offers an assessment on the performance of the Philippines' e-government initiatives particularly in the aspect of interoperability. Benchmarking Canada, Australia, and Singapore—the authors evaluated the successes and failures of the Philippines and created comparative analyses on its initiative performance. Although government funding and human resources on the country is of high magnitude, it is surprising to witness how the Philippines consistently ranks low on UN rankings. Findings revealed that appropriate integration of human factors and dynamics will significantly improve the country's performance. Likewise, this paper affirms a proposition through a roadmap to that possibility.

Metadata Model for Supporting Hierarchical Edge Device Arrangements in an IoT Deployment

Nuwan Jayawardene, Pumudu Fernando
Department of Computer Science, Informatics Institute of Technology, Colombo, Sri Lanka

Abstract— An increasing number of "Smart Devices" are coming online every year. These Smart Devices have the ability to communicate with other Smart Devices across the web, exchange information and make intelligent decisions. Measures to combat the added load brought on by these new connections have inadvertently resulted in increased network complexity. This paper considers why understanding that complexity is necessary, especially on the server side. Existing architectures are thereafter explored to gain insight into a possible solution. A metadata-based model is then proposed that would provide a central server with an overview of the network arrangement. The novelty of this model lies in its simplicity, scalability and technology agnosticism. An implementation of this model is then tested in a mix of realworld and simulated environments and evaluated. The results obtained indicate improvements in benchmarks along with performance gains with regards to network scaling.

On the Application of Machine Learning to Classify Sleep Positions

Brendon Becker, Yazan A. Alqudah
Department of Electrical and Computer Engineering, University of West Flordia, Pensacola, Florida, USA

Abstract- In this work, a low-cost device is developed that enables the monitoring and classification of sleep positions. Driven by the growing problem of sleep disorders, the device can be utilized at the patients' place to record and report their sleep positions. The device can also be used by hospitals and clinics for patients that requires continuous monitoring. Machine learning is used to classify different sleep positions based on sensors data collected from the device.

An IoT Network Coordinated AI Engine to Produce Loading and Delivery Schedules for Capacitated Vehicle Routing Problems

Ernesto Gutierrez-Miravete, Ashok Murthy
Rensselaer at Hartford, Hartford, CT, USA
Devices-Unlimited Corp., Rancho Mission Viejo, California, USA

Abstract - The capacitated vehicle routing problem with weight constraints is investigated using a hybrid simulation-genetic algorithm approach. The computational method is capable of producing near optimal loading and delivery routes for truck fleets delivering products to specific locations. The method is illustrated through its application to a problem consisting of delivering collections of pallets to a number of locations throughout the northern corridor of the United States.

Towards Dynamic Composition of Things in the Internet of Things

Naseem Ibrahim
School of Engineering, The Behrend College, The Pennsylvania State University, Erie, PA, USA

Abstract—The Internet of Things (IoT) is the connection of physical objects that collect and transfer data from one object to another. These objects are embedded with software, and sensors. A large number of IoT platforms have been introduced in the last few years. The major issue with available Internet of Things platforms, that support the communication between Things and their owners, is that they view Things as private property. Only the Thing owner can communicate with the Thing to obtain the service or functionality that is provided by the Thing. We believe a Thing should be viewed as a service or a functionality provider. Our research has introduced an architecture that supports the publication and discovery of Things in the Internet of Things. This architecture enables Thing owners to publish Thing specification and Thing requesters to find Things that best matches their needs. This paper is concerned with the dynamic composition of Things defined using our new architecture.

Automatic Composition of Things in the Internet of Things

Naseem Ibrahim
School of Engineering, The Behrend College, The Pennsylvania State University, Erie, PA, USA

Abstract—The Internet of Things (IoT) added connectivity to devices that traditionally were not connected. In the last view year, a number of IoT platforms for the publication and provision of Things have been introduced. The major drawback with available platforms is that they view Things as private property. Only the Thing owner can communicate with the Thing to obtain the service or functionality that is provided by the Thing. We believe a Thing should be viewed as a service or a functionality provider. Our research has introduced an architecture that supports the publication and discovery of Things in the Internet of Things. This architecture enables Thing owners to publish Thing specification and Thing requesters to find Things that best matches their needs. This paper is concerned with the automatic composition of Things defined using our new architecture.

"Touch a paper" System Design for Reading Utilizing Physical Touch

Yulana Watanabe, Takayuki Fujimoto
Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan

Abstract—Every environment and tool in our lives has been increasingly digitized: watches, cameras, books, magazines, and other stationery goods. Most basic necessities are integrated into a smartphone. The digitization of our live has greatly contributed to the increased convenience. Yet, there are also lost attractions due to digitalization. For instance, people choose analog records for the pursuit of the sound quality and its design, rather than convenience. There are still so many unwelcoming the digitization of living environments and tools. A typical example is “books”. While in recent years e-books are rapidly becoming widespread, many people choose “paper books” to enjoy reading. Part of the reason is assumed that reading books is not just about absorbing the information written in them, but it is about enjoying a sense of touch when turning the pages. In this research, we focus on “a sense of physical touch”, the feeling of actually turning the pages of a paper book. We also design software and devices that provide users with a “sense of physical touch”, often lost when reading an e-book on a smartphone.

Visualized Model using a Tree Structure for a Transmedia Storytelling Project Design

Shunsuke Aoki, Takayuki Fujimoto
Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan

Abstract—According to Henry Jenkins, Transmedia storytelling is defined as “Transmedia storytelling (TMS) represents a process where integral elements of a fiction get dispersed systematically across multiple delivery channels for the purpose of creating a unified and coordinated entertainment experience. Ideally, each medium makes its own unique contribution to the unfolding of the story.” (Jenkins 2007) In recent years, this technique is becoming popular in Europe and America. A creator needs to combine “Flow of story” and “Story experience” at a stage of planning a TMS project. Also, it is hard to build an elaborate plan; the world view of the story, time base, and how to invite the audience to participate in the experience. In this study, we propose the method of visualizing a structure of the entire story with a possible tree-structure model assumed from content elements that users are thinking about when they start to work on a TMS project.

Visual verification of the Use of Educational Space as Advertising Medium in Consideration of Display Position

Kanata Itoh, Takayuki Fujimoto
Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan

Abstract—In this study, we conduct a visual verification experiment and an evaluation regarding the use of educational space as advertising medium. The 'use of educational space as an advertising medium' means to incorporate the commercial advertisement to the lecture slides in university classes. There has been much debate about the business use of education. On the other hand, the benefits such as the reduction of tuition fees and the improvement in the quality of education / research environment by the sponsorship could be considered as merits. In particular, we examine the optimized position to display advertisement by analyzing the movement of the experiment participant's point of view in the educational space with eye trackers and discuss advertisement effectiveness based on the results

AWS IoT and the Interconnected World - Aging in Place

Justin Waterman, Hyeongjun Yang, Fadi Muheidat
Computer Science and Engineering, California State University, San Bernardino, California, USA

Abstract— This project is considered the startup component of a framework for sensor data fusion, data mining, and analytics in real-time. The long-term goal is to build a dynamic, efficient, scalable, and secure runtime of parsing and handling multi-sensor data. The outcomes of the project are meant to design a multi-sensors data fusion model. During this research, some concepts are used and studied for future use, such as data imperfections and inconsistencies, misled data association, data fusion portability, machine independence, security vulnerabilities, etc. We will build the sensor network, acquire data from multiple sensors, and use AWS IoT SDK (Software Development Kit) to handle the data fusion for basic algorithms. The core of the multiple sensor network will be a Raspberry Pi 4 to decrease system cost and increase portability. The outcomes of this project will help build a comprehensive, real-time data monitoring, and support system for assistive living and care providers.

Data-Saving office System That Can Be Stored on a Floppy Disk

Miki Sunakawa, Takayuki Fujimoto
Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan

Abstract—Since the 1980s, when computers began to spread in ordinary households, various recording media for storing data have appeared. Initially, the data capacity was less than 1MB. In recent years, data storage capacity has been rapidly increasing, so the data capacity is not concerned as much as before. In this research, we examine the possibility of a sufficiently practical system by carefully selecting the necessary functions for general use with the method of not- over-decorating. Production applies a 3.5-inch floppy disk to limit the amount of data. By carefully selecting necessary functions, we believe that the visibility and usability will be improved, and the functionality will be enhanced.

CSCI-ISCW:
Cyber Warfare, Cyber Defense, & Cyber Security

Optimizing Cyber Security Education: Implementation Of Bloom's Taxonomy for the Future Cyber Security Workforce

Nageswaree Kodai Ramsoonder, Selvamanee Kinnoo, Craig Valli, Nicola F. Johnson
Security Research Institute, Edith Cowan University, Perth, Australia

Abstract—This research paper determines how education can be optimized by mapping curricula onto Bloom's taxonomy. The cyber security industry has disparate ways of recruiting its workforce owing to the prevailing educational system and diverse work roles in the field of cyber security. Comparison is drawn between the old and revised models of Bloom's taxonomy to depict which model best represents the dynamic learning objectives and outcomes of cyber security education. A rubric is designed from the Dreyfus et al model of skill acquisition based on the iterations of Bloom's taxonomy to assess cyber security competencies to achieve mastery. To bridge the skills gaps in the industry, Bloom's taxonomy provides the impetus to align the needs of industry with educational programs [1].

Detection and Defense from False Data Injection Attacks in Aviation Cyber-Physical Systems Using Artificial Immune Systems

Abdulaziz A. Alsulami, Saleh Zein-Sabatto
Department of Electrical and Computer Engineering, Tennessee State University, Nashville, Tennessee, USA

Abstract—In recent years, there has been a rapid expansion in the development of Cyber-Physical Systems (CPS), which allows the physical components and the cyber components of a system to be fully integrated and interacted with each other and with the physical world. The commercial aviation industry is shifting towards Aviation Cyber-Physical Systems (ACPS) framework because it allows real-time monitoring and diagnostics, real-time data analytics, and the use of Artificial Intelligent technologies in decision making. Inevitably, ACPS is not immune to cyber-attacks due to integrating a network system, which introduces serious security threats. False Data Injection (FDI) attack is widely used against CPS. It is a serious threat to the integrity of the connected physical components. In this paper, we propose a novel security algorithm for detecting FDI attacks in the communication network of ACPS using Artificial Immune System (AIS). The algorithm was developed based on the negative selection approach. The negative selection algorithm is used to detect malicious network packets and drop them. Then, a Nonlinear Autoregressive Exogenous (NARX) network is used to predict packets that dropped by the negative selection algorithm. The developed algorithm was implemented and tested on a networked control system of commercial aircraft as an Aviation Cyber-physical system.

Moving Target Defense Discrete Host Address Mutation and Analysis in SDN

Charan Gudla, Andrew H. Sung
School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, Mississippi, USA

Abstract - Moving Target Defense is a technique focused on disrupting certain phases of a cyber-attack. Reconnaissance is the preliminary phase of the attack in the cyber kill chain. The static nature of the existing networks gives an adequate amount of time to the adversaries in gathering enough data concerning the target and succeed in mounting an attack. Randomization of the host addresses is well known MTD technique that hides the actual network configuration from external scanners. Although random host mutation techniques are investigated extensively, the limitations such as less availability of unused public address space for mutation and host unavailability due to mutation time interval deteriorate the network's stability. Due to address space unavailability, each host address's mutation is not feasible according to the time interval, or the address space is repeated multiple times. When the host establishes a session of transmitting or receiving data, due to mutation interval, the session is interrupted, leading to the host's unavailability. In this paper, we propose a moving target defense technique to achieve the following objectives:

(1) using mutation technique, randomization of IP addresses is achieved to create high uncertainty in adversary scanning; (2) the mutation time interval is separated from each host to preserve network performance and stability; (3) the mutation scheme is adapted by analyzing the data stats from the individual host (4) the analyzed data stats are used to manage the available unused address space.

Quantitatively Examining Service Requests of a Cloud-Based On-Demand Cybersecurity Service Solution for Small Businesses

Landon McLilly, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract— With the rise of cybercrimes, limited by the funding and in-house technical resource, the small business sector is already known for struggling with cybersecurityrelated issues. Regardless of 65% acknowledging they have been the target of a cyberattack and 86% believing digital risk will upsurge, only 4% of small and medium-sized business owners have executed all of the U.S. Small Business Administration's cybersecurity best practices, rendering to a survey by Nationwide [1]. To meet this challenge, we have proposed a cloud-based on-demand cybersecurity service solution for small businesses (CODCSSSB) to provide a cost-effective cybersecurity resolution for small businesses. This paper has explored how to apply a quantitative examination approach to validate the security service requests sent to and processed by the CODCSSSB to discover the weakness of design with low cost in terms of the time of development and accuracy of identifying the root cause of the design issues.

Explore the Relationship between Authentication Factor Multiplicity and Composite Vulnerability Exposures

Adam English, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract— With the continued expansion and adoption of multi-factor authentication technologies within organizations across industries, there exists a need to determine the relationship between composite vulnerability exposures and the multiplicity of authentication factors. To meet this need, we propose a power curve formula for the relationship between these variables, demonstrating the wellness of fit to generalized data sets. We provide an extension to the Common Vulnerability Scoring System (CVSS) v3 calculator, allowing the characterization of combined Common Vulnerability Exposure (CVE) in an objective and repeatable means, demonstrating the capabilities of that extent. This paper explores the potential for future work derived from the proposed mathematical formula representation of the relationship between authentication factor multiplicity and composite vulnerability scores.

Methodological Proposal for the Optimization of the Installation Times of Hardenized Linux Operating Systems through the Spacewalk Solution in Critical Infrastructures

Ivan Ortiz, Aaron Echeverria, Roberto Andrade
Universidad de Las Americas Escuela Politecnica Nacional, EPN, Ecuador

Abstract—Critical infrastructures are physical or virtual assets that provide a variety of continuous services through various networks like LAN, MAN, WAN, and cloud services. Any interruption in critical infrastructures would have a great impact on the services they provide, preventing the continuity of various business models. The purpose of this methodology is to increase the levels of hardening in critical infrastructures that use the Linux operating system, allowing the continuity of the services they provide and reducing the risk of network attack vectors, adjacent and local. The hardening level will increase during the installation of the Linux operating system by means of a kickstart file with the hardening settings and the Spacewalk server that hosts this file. The proposed methodology consists of four phases. The first phase will be carried out in the installation and configuration of the Spacewalk server. The second phase is to establish the hardening configurations based on the CIS Benchmark in the kickstart file,

which will be hosted on the Spacewalk server. The third phase is the deployment of the kickstart file when performing a hardenized installation. In the last phase, the percentage of hardening and the optimization achieved in the installation time will be verified.

Cyber-Security Strategy for Internationally-dispersed Industrial Networks

Ralf Luis de Moura, Alexandre Gonzalez, Virginia N. L. Franqueira, Antonio Lemos Maia Neto
Operational Technology Architecture, Vale S.A., Vitoria, Brazil
Enterprise Architecture, Vale S.A., Rio de Janeiro, Brazil
School of Computing, University of Kent, UK
Foundation Architecture, Vale S.A., Belo Horizonte, Brazil

Abstract— Globalization implies geographically dispersed supply chains composed of facilities strategically located in several countries and regions of the world. These structures commonly involve several Operational Technology (OT) and Information Technology (IT) infrastructures and integration to enable accurate and useful information processing. Such integration (also called Cyber-Physical Systems) transforms the industry and facilitates massive data volumes' systematic transformation into valuable information. Security risks posed by such integration may be substantial and, depending on the size of the company, and the number of integration points, dealing with them could easily cost millions of dollars. With the main objective of studying available strategies to manage security risks in companies with dispersed supply chains, this paper reviews international cyber-security standards and regulations and proposes a more comprehensive strategy. The strategy includes IT services, optimized perimeter segregation, and data flow policies among OT and IT networks to balance a high level of protection and cost-effectiveness.

Vehicle Security Learning Tools and Scenarios

Guillermo Francia, III, Eman El-Sheikh, Hongmei Chi
Center for Cybersecurity, University of West Florida, Pensacola, Florida, USA
Department of Computer Science, Florida A&M University, Tallahassee, Florida, USA

Abstract—The rapid pace with which connected and autonomous vehicles is evolving presents security challenges that are prevalent on communication technologies. Although it is universally accepted that tremendous benefits can be derived from this emerging technology, we need to make sure that this critical infrastructure is secured and protected. Recent attacks on vehicle networks have validated the urgent need for a robust and sustained effort to stem the tide of these debilitating incursions. Our ever-increasing dependence on this type of transport system brings us to new crossroads and challenges that are confronting our economic security, privacy protection, and well-being. One major challenge is the education and training of the current and future workforce in this emerging technology. This paper explores key curriculum issues in securing modern automobiles, including the essential tools necessary to implement meaningful hands-on laboratory experiments and learning scenarios.

Mitigating Interleaving Jamming of IEEE 802.11

Benjamin Davis, Bruce Debruhl
California Polytechnic State University, San Luis Obispo, California, USA

Abstract—Recent jamming attacks against IEEE 802.11 (WiFi) have targeted the interleaving mechanism to greatly improve power efficiency, getting as good as 95% packet loss using 0.1% transmitter power. In this paper, we propose a novel method of modifying the interleaver to prevent this attack vector through using a shared secret to deterministically interleave the data in such a way that is not susceptible to targeted subcarrier jamming. We implement and test this approach using software defined radio. We show it entirely negates the impact of interleaving targeted jamming, without making the signal more susceptible to more traditional jamming methods or additive Gaussian white noise. We show that our modified interleaver makes these attacks less effective than whole-channel jamming by as much as 15% lower packet loss.

Effectiveness of Real-Time Network Monitoring for Identifying Hidden Vulnerabilities inside a System

Haydar Teymourlouei, Vareva E. Harris
Department of Technology & Security, Bowie State University, Bowie, Maryland, USA
Department of Criminal Justice Administration and Center for Cybersecurity, Benedict College, Columbia, USA

Abstract - Computer systems will continue to be the target of cybercriminals as they become an integral part of society. Without being able to monitor network traffic in real-time, systems might be in danger of losing sensitive data to hackers. Knowing the vulnerability allows cache hackers that tinker with data to be detected in real-time. Normally, attackers can use vulnerabilities to exploit system weaknesses and create a backdoor. Identifying hidden vulnerabilities provides accurate results that prevent hackers from accessing the system. This research presents the effectiveness of using real-time monitoring by using event logs, network monitoring, and management tools for identifying vulnerabilities. The monitoring techniques offer a method of detection for any unusual behavior in a network system.

Nonproliferation of Cyber Weapons

Jacob Benjamin, Michael Haney
Dragos, Murrells Inlet, USA
University of Idaho, Idaho Falls, Idaho, USA

Abstract—The Treaty on the Nonproliferation of Nuclear Weapons has been the basis for international cooperation on stopping the spread of nuclear weapons. However, in the last twenty years, the world has seen the emergence of a new type of weapon, cyber weapons. The impacts of these attacks in the physical realm is possible in part to the convergence of informational and operational technology. This paper examines examples of cyber weapons and their impacts on critical infrastructure. Additionally, the paper discusses the differences between nuclear weapons and cyber weapons, identifies the challenges of disarmament, and postulates ways to overcome them.

A Survey of Artificial Intelligence in Cybersecurity

Katanosh Morovat, Brajendra Panda
Department of Mathematics and Computer Science, Western Carolina University, Cullowhee, USA
Department of Computer Science and Computer Engineering, University of Arkansas, Faytteville, Arkansas, USA

Abstract – During the last decades, not only the number of cyberattacks have increased significantly, they have also become more sophisticated. Hence designing a cyber-resilient approach is of paramount importance. Traditional security methods are not adequate to prevent data breaches in case of cyberattacks. Cybercriminals have learned how to use new techniques and robust tools to hack, attack, and breach data. Fortunately, Artificial Intelligence (AI) technologies have been introduced into cyberspace to construct smart models for defending systems from attacks. Since AI technologies can rapidly evolve to address complex situations, they can be used as fundamental tools in the field of cybersecurity. AI-based techniques can provide efficient and powerful cyber defense tools to recognize malware attacks, network intrusions, phishing and spam emails, and data breaches, to name a few, and to alert security incidents when they occur. In this paper, we review the impact of AI in cybersecurity and summarize existing research in terms of benefits of AI in cybersecurity.

Reproducible Software Vulnerability Testing With Iac

Kohei Akasaka, Akihito Nakamura
Graduate School of Computer Science and Engineering, University of Aizu, Fukushima, Japan
Computer Science Division, University of Aizu, Fukushima, Japan

Abstract—Cybersecurity is more important now than ever. Attackers maliciously use vulnerabilities to exploit IT systems. Software developers are responsible for fixing the reported vulnerabilities in their own products, while system administrators should assess the impacts of attacks against their working systems. However, they are wasting time to make the problem appear because provisioning is a complex and troublesome process. This paper presents a model of vulnerability testing and our original system which enables efficient and automatic reproduction of the vulnerable environments and replicable execution of attacks. We utilize the Infrastructure as Code (IaC) principle. Vulnerability testing is represented as code and the system automates the process of provisioning vulnerable environments and executing attacks through machine-readable definitions, rather than manual configuration and execution. A simple human-readable YAML language is used for writing testing definitions. If the definitions are maintained in publicly accessible repositories, security experts have a chance to contribute their knowledge and expertise to the community and an open-source style ecosystem may be created.

Indistinguishability of Biometric Honey Templates: Comparing Human Testers and SVM Classifiers

Edlira Martiri, Bian Yang, Muhammad Ali Fauzi
Norwegian University of Science and Technology, NTNU, Gjovik, Norway

Abstract—In high level security environments, data protection and leakage prevention remains one of the main challenges. In biometric systems, its most sensitive piece of information, the template, is constantly being exchanged between its building blocks. instead of having one template, in this paper we generate a set of synthetic templates to camouflage the genuine one. To test their indistinguishability, we suppose an attack and compare two different classifications results of reconstructed faces: humans and SVM classifier. For the former, we built a platform where testers could classify a set of random preimages reconstructed from real or synthetic (honey) templates. From an attacker point of view, we noticed that, compared to the SVM classifier, human testers showed better results in terms of classification distinguishability.

Anomalous Detection System in Crowded Environment using Deep Learning

Dorcas Oladayo Esan, Pius Adewale Owolawi, Chuling Tu
Department of Computer Systems Engineering, Tswhane University of Technology, Shoshanguve, South Africa

Abstract— In recent years, surveillance systems have become very important due to security concerns. These systems are widely used in many applications such as airports, railway stations, shopping malls, crowded sports arenas, military etc., [1]. The wide deployment of surveillance systems has made the detection of anomalous behavioral patterns in video streams to become increasingly important. An anomalous event can be considered as a deviation from the regular scene; however, the distribution of normal and anomalous events is severely imbalanced, since the anomalous behavior events do not frequently occur, hence it is imperative to accurately detect anomalous behavioral pattern from a normal pattern in a surveillance system. This paper proposes a Convolutional Neural Network and Long Short-Term Memory (CNN-LSTM) technique. The CNN is used to extract the features from the image frames and the LSTM is used as a mechanism for remembrance to make quick and accurate detection. Experiments are done on the University of California San Diego dataset using the proposed anomalous behavioral pattern detection system. Compared with other existing methods, experimental analysis demonstrates that CNN-LSTM technique has high accuracy with better parameters tuning. Different analyses were conducted using the publicly available dataset repository that has been used by many researchers in the field of computer vision in the detection of anomalous behavior. The results obtained show that CNN-LSTM outperforms the others with overall F1-score of 0.94; AUC of 0.891 and accuracy of 89%. This result shows that the deployment of the proposed technique in a surveillance detection system can assist the security personnel to detect an anomalous behavioral pattern in a crowded environment.

OctoBot: Human Activity Orchestration System for Cybersecurity Experiment and Exercise

Aris Cahyadi Risdianto, Ee-Chien Chang
School of Computing, National University of Singapore (NUS), Singapore

Abstract—Cybersecurity experiment and exercise in a testbed environment required normal traffic to ensure a better result, a learning experience, and the effectiveness of the whole process. The traffic camouflages the attack traffic from the attacker by generating different types of activities and tasks that are usually generated by the human. Unfortunately, in a large-scale experiment and exercise, it is impractical to employ a large number of humans to generate normal traffic. There are several methods available to produce a single human activity from an agent, but automating it from multiple agents with multiple and different activities is very challenging. By learning and adopting the concept of BotNet, we designed an orchestration system for a human agent (i.e., bot), called OctoBot, to generate human activity that is easy to be deployed and utilized. This paper discusses primary considerations in designing and implementing our system with an example of unique bot development. It also presents deployment experience in the production testbed environment with an example use case for a virtual environment to show the potential usability of our system.

Cloud Incident Response: Challenges and Opportunities

Murat Ozer, Said Varlioglu, Bilal Gonen, Victor Adewopo, Nelly Elsayed, Selcuk Zengin
School of Information Technology, University of Cincinnati, Cincinnati, Ohio, USA

Abstract—Many organizations migrate their on-premise infrastructure to the cloud for its known advantages such as low cost and flexibility. However, a dilemma appears that despite the increased usage of cloud computing among small and midsize businesses, the reported number of cyber attacks does not reduce and even astronomically increases. Recent studies explain the source of this dilemma as having scarce familiarity with cloud computing. Scholars basically stress that cloud computing offers better malware and DDOS attack protection; however, limited knowledge of cloud computing infrastructure and the complex nature of cloud systems contribute to the high number of cyber attacks. In this context, this study attempts to add one more step to the traditional incident response plan to adapt it to cloud computing. In addition, the current study introduces certain solutions to gather dispersed data from complex cloud computing infrastructure which past studies cited as the main challenges of effective cloud incident response.

Cyber as a Service: Automating First Responders' Service in the Cyberspace

Matthew Blair, Davis Jeffords, Eric Lilling, S. Banik
Department of Cyber & Computer Sciences, The Citadel, The Military College of South Carolina, Charleston, SC, USA

Abstract—Due to increasing number of attacks in the cyberspace that deals with different types of users, it is imperative that an automated responder service will be efficient to help the users detect and mitigate different types of attacks in their systems. In this research, we propose to replicate the framework of emergency responder service (911) of the physical space to the cyberspace. Towards this we propose a framework for Cyber-as-a-Service for end-users. In our proposed model, we have three entities: the Dispatch Center, the Guard, and the Client Software. These entities will communicate with each other to detect and extinguish any malicious activity on the host computer. The host machine will run a software that scans and detects any abnormal or malicious activity and communicates this activity to the Guard, which then replies with an executable resolution back to the host. Meanwhile, the Dispatch Center manages connections between hosts and Guards to ensure that hosts are connected to the optimal Guard. We propose algorithms that will place and distribute the Dispatch Center and the Guards. These algorithms allow for fair distribution of Guards, as well as balance the workload among the Guards. We propose the communication protocol that will take place between the Client software, Guards and the Dispatch Center. Our goal is to design the framework for Cyber-as-a-Service for everyday users in the cyberspace who do not have sufficient technical skills to manage tools to detect different attacks.

Analytical Framework for National Cyber-security and Corresponding Critical Infrastructure: A Pragmatistic Approach

Marc Wright, Hassan Chizari, Thiago Viana
School of Computing and Engineering, University of Gloucestershire, Cheltenham, United Kingdom, UK

Abstract—Countries are putting cyber-security at the forefront of their national issues. With the increase in cyber capabilities and infrastructure systems becoming cyber-enabled, threats now have a physical impact from the cyber dimension. This paper proposes an analytical framework for national cyber-security profiling by taking national governmental and technical threat modeling simulations. Applying thematic analysis towards national cybersecurity strategy helps further develop understanding, in conjunction with threat modeling methodology simulation, to gain insight into critical infrastructure threat impact.

The Use of Runtime Verification for Identifying and Responding to Cybersecurity Threats Posed to State Actors During Cyberwarfare

Jeremy Straub
Institute for Cyber Security Education and Research, North Dakota State University, Fargo, North Dakota, USA

Abstract— This paper considers the utility of the use of runtime verification techniques for detecting and responding to cybersecurity threats. To this end, it considers two questions: First, it evaluates the efficacy of runtime verification for identifying zero-day threats and threats that are not otherwise widely known based up system operations. Second, it considers the particular use of these techniques by state actors (i.e., nation states engaged in declared or undeclared cyberwarfare) who are likely to encounter a greater level of such vulnerability exploits than individuals or private businesses during the normal operations. Drawing on the analysis in the two foregoing areas, the paper concludes by identifying key areas of needed future work to support runtime verification's application in this area.

CSCI-ISNA:
Social Network Analysis, Social Media, & Mining

Understanding Violence Against Women in Digital Space from a Data Science Perspective

Gregorio Arturo Reyes Gonzalez
Tecnologico de Monterrey, School of Engineering and Sciences, Monterrey, NL, Mexico

Abstract—Research has shown that Violence Against Women is a pervasive problem which has been increasing. Until a few years ago, it took place both in public and private spaces, but it has now broken into Digital Space, adopting more symbolic expressions. There has been some important related work from Data Science approaches. Mainly about cyberbullying, detection of language patterns through supervised algorithms, a few works on unsupervised learning, and some on violence against women. This article is part of a larger investigation related to the Violence Against Women in Digital Space phenomena in Mexico from Spanish-language interactions in microblogging social network, Twitter. We present the framework to address these phenomena from a Data Science perspective and go through some first stage results. The goal is to give Data Science perspectives and insights related to the understanding of Digital Violence Against Women, with the purpose of generating awareness about this problem by, modifying manners and implementing public policies that could counteract this social problem.

Scoring Popularity in GitHub

Abduljaleel Al-Rubaye, Gita Sukthankar
Department of Computer Science, University of Central Florida, Orlando, Florida, USA

Abstract—Popularity and engagement are the currencies of social media platforms, serving as powerful reinforcement mechanisms to keep users online. Social coding platforms such as GitHub serve a dual purpose: they are practical tools that facilitate asynchronous, distributed collaborations between software developers while also supporting passive social media style interactions. There are several mechanisms for "liking" content on GitHub: 1) forking repositories to copy their content 2) watching repositories to be notified of updates and 3) starring to express approval. This paper presents a study of popularity in GitHub and examines the relationship between these three quantitative measures of popularity. We introduce a weight-based popularity score (W T P S) that is extracted from the history line of GitHub repositories popularity indicators.

Evaluation of Elements of a Prospective System to Alert Users to Intentionally Deceptive Content

Matthew Spradling, Jeremy Straub
Department of Computer Science, Engineering and Physics, University of Michigan Flint, Michigan, USA
Institute for Cyber Security Education and Research & Challey Institute Faculty Fellow, North Dakota State University, Fargo, North Dakota, USA

Abstract— This paper presents a system designed to alert users to potentially deceptive online media. It reviews existing alert mechanisms and labeling systems which provide inspiration for the system's components. It also discusses how a new mechanism could be developed that uses elements of these common alert mechanisms to warn users about potentially deceptive content on web pages. This paper evaluates the benefits, risks, and limitations of the proposed system. Sources of the alert components, and their integration into a single combined alert system, are discussed as are planned areas of future work.

Optimizing Global Processing Time in the Detection of Patterns Related to Suicide in Social Networks

Damian Martinez Diaz, Francisco Javier Luna Rosas, Julio Cesar Martinez Romo, Marco Antonio Hernandez Vargas, Ivan Castillo Zuniga
Departamento de Sistemas y Computacion, Instituto Tecnologico de Aguascalientes (ITA), Mexico

Abstract— There is a suicide every 40 seconds in the world and it is the third cause of death for young people between 15 and 19 years old worldwide. For every suicide, many more attempt it, which is why suicide prevention remains an universal challenge and has been recognized by the World Health Organization (WHO) as a public health priority. Experts say that one of the best ways to prevent suicide is for people who are going through this urge to take their own lives to listen to people who are close to them and social networks such as Twitter or Facebook are in a unique position to help these people connect in real time in difficult situations that people with these suicidal tendencies are going through, but also represents a potential risk to receive information that could later prove harmful, either by stressing the same information or by taking some suicidal ideas. In this research we propose a model to optimize the global time processing in the detection of patterns related to suicide in the social network Twitter. Our results show that the proposed model can be a good alternative when it comes to optimizing the response time in this type of problems.

On the Accuracy Evaluation of Access Control Policies in a Social Network

Jedidiah Yanez-Sierra, Arturo Diaz-Perez, Victor Sosa-Sosa
Cinvestav - Tamaulipas, Cd. Victoria, Mexico
Cinvestav - Guadalajara, Guadalajara, Mexico

Abstract—Access control policies are mandatory for organizations whose operation involves sharing resources that must be kept private. In this paper, we address the problem of evaluating the accuracy of access control policies distributed in an interaction network modeled from a social network. Such network models granted access to documents owned by a large set of users. Since denied accesses are not included in the input network, we discuss a method based on Network Science to include complementary edges to have an approximate evaluation of ACPs' accuracy. The synthetic interactions allow the evaluation of ACPs by assessing the explicit and implicit intentions of the owners. We present an evaluation strategy to measure the accuracy of the generated ACPs. The results will be of interest to academics who want to synthesize information for similar phenomena.

Rules for Optimal Perpetual Gossiping Schemes

Ivan Avramovic, Dana S. Richards
Department of Computer Science, George Mason University, Fairfax, Virginia, USA

Abstract—Perpetual gossiping is an all-to-all communication problem on social networks, or any coordinated distributed system in general. In perpetual gossiping, a state of universal reachability is maintained by a continuous sequence of communications between participants. Unlike the well-understood static case, perpetual gossiping is a difficult problem, with some NP-complete classes of solutions. A basic question which remains open is whether an optimal scheme of contiguous calls is guaranteed to exist for a tree. This paper presents a series of theoretical tools directed towards answering the question.

Shapley based Interpretable Semi-supervised Model for Detecting Similarity Index of Social Media Campaigns

Aman Framewala, Aum Patil, Faruk Kazi
Centre of Excellence in Complex & Nonlinear Dynamical Systems, VJTI, Mumbai, India

Abstract—Although significant research is done on fake news detection, the definition of a news being fake has always been ambiguous. This problem can be addressed by comparing such campaigns of fake news or harmful content with previously detected similar campaigns on social media. Detecting similar campaigns also helps in structured analysis of social media content useful for various applications including comparison of different marketing campaigns. In this paper authors showcase a novel semi-supervised learning approach for detecting similar campaigns and giving human-like explanations for those predictions. This is done by using a modified version of Artificial Immune System (AIS) for unsupervised detection of similar campaigns. We then use eXtreme Gradient Boosting (XGBoost) model for extracting rules from predictions which are further fed to Shapley Additive exPlanations (SHAP) to generate payoffs of each contributing feature. We demonstrate the approach using three different social media campaigns covering public outrage, marketing and public awareness.

Emotion Detection in Twitter Posts: A Rule-based Algorithm for Annotated Data Acquisition

Maria Krommyda, Anastasios Rigos, Kostas Bouklas, Angelos Amditis
Institute of Communication and Computer Systems, Athens, Greece

Abstract—Social media analysis plays a key role to the understanding of the public's opinion regarding recent events and decisions, to the design and management of advertising campaigns as well as to the planning of next steps and mitigation actions for public relationship initiatives. Significant effort has been dedicated recently to the development of data analysis algorithms that will perform, in an automated way, sentiment analysis over publicly available text. Most of the available research work,

focuses on binary categorizing text as positive or negative without further investigating the emotions leading to that categorization. The current needs, however, for in-depth analysis of the available content combined with the complexity and multidimensional aspects of the human emotions and opinions have rendered such solutions obsolete. Due to these needs, currently, research is focusing on specifying the emotions and not only the sentiment expressed in a given text. This is, however, a very challenging effort due to not only the lack of annotated datasets that can be used for emotion detection in text but also the subjectivity infused in datasets that have been created based on manual annotations. A hybrid rule-based algorithm is presented in this paper, that supports the creation of a fully annotated dataset over the Plutchik's eight basic emotions. The presented algorithm takes into consideration the available emoji in the text and utilized them as objective indicators of the expressed emotion thus efficiently tackling both identified challenges. This is a full regular paper submitted to the CSCI-ISNA Symposium.

An Evaluation of Tweet Sentiment Classification Methods

Lihua Yao, Hassan Alam, Jerry Li, Oleg Melnokov
Department of Defense, Office of People Analytics, Seaside, California, USA
AI Assessment Center
DeepHole AI Consulting
Stanford University, California, USA

Abstract - In this paper, we present the result of our research in predicting sentiment from Twitter data derived from a Kaggle competition. Our goal was to determine the efficacy of different supervised classification methods to predict Twitter sentiment to be Positive, Neutral or Negative. We evaluated four different classification statistical models: 1. Logistic Regression (LR), 2. Linear Support Vector Machine (LSV), 3. Multinomial Naïve Bayesian (NB), and 4. Random Forest (RF). We also evaluated two different tokenization methods 1. Document Term Matrix (DTM) and 2. Term Frequency–Inverse Document Frequency (TF-IDF). We combined this with three text extraction methods 1. Original Tweet Text, 2. Rapid Automatic Keyword Extraction (RAKE) and 3. Hand curated "Selected Text". Furthermore, various Neural Networks were applied to the Tweet Text and BERT extracted data that reduced the original 1000 features to be 768 that were applied to different models. Our experiment shows RF and LR gives the best results and there is little difference between DTM and TF-IDF. Fully connected neural network (FCNN) performed the best for the BERT extracted data with a test score of 0.75.

Cyberbullying Detection Through Sentiment Analysis

Jalal Omer Atoum
Department of Mathematics and Computer Science, East Central University, Oklahoma, USA

Abstract - In recent years with the widespread of social media platforms across the globe especially among young people, cyberbullying and aggression have become a serious and annoying problem that communities must deal with. Such platforms provide various ways for bullies to attack and threaten others in their communities. Various techniques and methodologies have been used or proposed to combat cyberbullying through early detection and alerts to discover and/or protect victims from such attacks. Machine learning (ML) techniques have been widely used to detect some language patterns that are exploited by bullies to attack their victims. Also. Sentiment Analysis (SA) of social media content has become one of the growing areas of research in machine learning. SA provides the ability to detect cyberbullying in real-time. SA provides the ability to detect cyberbullying in real-time. This paper proposes a SA model for identifying cyberbullying texts in Twitter social media. Support Vector Machines (SVM) and Naïve Bayes (NB) are used in this model as supervised machine learning classification tools. The results of the experiments conducted on this model showed encouraging outcomes when a higher n-grams language model is applied on such texts in comparison with similar previous research. Also, the results showed that SVM classifiers have better performance measures than NB classifiers on such tweets.

Multi-modal Deep Learning Based Fusion Approach to Detect Illicit Retail Networks from Social Media

Anamika Paul Rupa, Aryya Gangopadhyay
University of Maryland Baltimore County, Baltimore, Maryland, USA

Abstract—Illicit drug trafficking poses a significant threat in today's society in general and in the younger population in particular. Drug abuse has been known to be correlated with car-accidents, crimes, diseases, deaths and many other negative aspects of society. With the advancement of social media being an open platform to express all kinds of social activities, drug use can be encouraged on this platform to lead other people towards drug abuse. Where people abuse drugs on social media, it is a prominent platform for drug dealing as well. Drug dealers post different types of drug images along with their contact information on social media. Tracking drug dealers among millions of social media users can be very challenging for law enforcement agencies. Therefore, automatic detection of drug dealers (along with the type of drugs they sell and their contact information) is crucial. In this article we have presented a state-of-the-art social media analytic algorithm which does multi modal analysis in order to detect drug-related posts and drug dealers from social media. We propose to detect different types of drugs from social media posts which include: pills, mushrooms, LSD, cannabis, cocaine, syrup, hookah and cigars using our drug type detection model. We also propose to extract drug dealer's information from social media and create novel AI algorithms to improve understanding of the operations of Illicit Retail Networks (IRN) that will help detect, disrupt and ultimately dismantle these networks. Our approach is generalizable to detect different illicit networks from social media such as human trafficking, illegal gun sales, money laundering and others.

Social Network Influencers' Data Augmenting Recommender Systems

Ashrf Althbiti, Xiaogang Ma
Department of Computer Science, University of Idaho, Moscow, Idaho, USA

Abstract— The ever-increasing use of social network sites and the availability of Internet services have introduced opportunities for users to communicate, interact, and connect with one another. Social network sites can identify the similarity between users in order to recommend new friends to a user. Therefore, users might not only communicate with their actual offline friends, but also they are motivated to communicate with strangers and friends of their actual friends. This paper introduces a model that integrates and incorporates social network influencers' data for augmenting recommender systems. This model is developed based on three main techniques. The first one is the contentbased technique used to recommend items to an active user based on his/her previously collected data or interests. The second technique is a Bayesian classifier used to learn an active user's profile and to determine social network influencers who can contribute to augment the quality of recommendations. The third technique is used to identify social media influencers. This model can be generalized to other domains that collect a side information from external sources.

Information Security Attacks on Mobile Messaging Applications: Procedural and Technological Responses

Lokesh Saravanan Ramamoorthi, Gabrielle Peko, David Sundaram
Department of Electrical and Computer Engineering, University of Miami, Coral Gables, Florida, USA
Information Systems and Operations Management, The University of Auckland, Auckland, New Zealand

Abstract—In today's digital world mobile phones have a significant impact on our day-to-day lives including the use of internet chat applications designed for smartphone users. Generally, these mobile messaging apps claim they protect the user's information using encryption techniques. Yet, information security attacks that exploit the apps' vulnerabilities are increasingly common. These vulnerabilities are the main gateway for hackers to access information. Considering the four most popular messaging apps, a taxonomy of attack targets of messaging applications is introduced that consists of three broad categories of attacks. Each of these categories is discussed and analyzed in order to propose several combinations of technological and procedural solutions to mitigate the vulnerabilities. Further, it is envisioned that these solutions provide the foundation for building prevention and protection mechanisms against such attacks.

A Dataset for the Detection of Fake Profiles on Social Networking Services

Samuel Delgado Munoz, Edward Paul Guillen Pinto
Systems Engineering Department, El Bosque University, Bogota, Colombia

Abstract—The use of multiple social media platforms is a common practice on more than two-third of all Internet users, according to OurWorld In Data. From this perspective, the verification of a real profile is a matter of growing interest, because false virtual identity could trigger problems such as spoofing, bots, grooming, sextortion, just to name a few. This paper presents a method to detect fake profiles on social media platforms by deploying some machine learning detection methods over a novel dataset. The dataset was designed with 17 metadata features from real and fake profiles and it was tested on Instagram profiles. After deploying different machine learning algorithms, the obtained detection rate was near to 96% with good false positive rates.

An Attention-based Deep Learning Method for Text Sentiment Analysis

Thanh Le
School of Business Information Technology, University of Economics HCMC, Vietnam

Abstract— Text sentiment analysis is target-oriented, aiming to identify the opinion or attitude from a piece of natural language text toward topics or entities, whether it is negative, positive or neutral using natural language processing and computational methods. With the growth of the internet, numerous business websites have been deployed to support shopping products, booking services online as well as to allow online reviewing and commenting the services in forms of either business forums or social networks. Use of text sentiment analysis for automatically mining opinion from the feedbacks on such emerging internet platforms is not only useful for customers seeking for advice, but also necessary for business to study customers' attitudes toward brands, products, services, or events, and has become an increasingly dominant trend in business strategic management. Current state-of-the-art approaches for text sentiment analysis include lexicon based and machine learning based methods. In this research, we proposed a method that utilizes deep learning with attention word embedding. We showed that our method outperformed popular lexicon and embedding based methods.

A New Ensemble Method for Classifying Sentiments of COVID-19 Related Tweets

Meng Hsiu Tsai, Yingfeng Wang
Department of Computer Science and Engineering, University of Tennessee at Chattanooga, Tennessee, USA

Abstract—Twitter sentiment analysis enables scientists to monitor people's attitudes to public health policies and events in the era of COVID-19. Although the pre-trained model can conduct sentiment analysis since the beginning of a pandemic or a global emergency, this model is not optimized for a specific topic. Furthermore, the way people express opinions on the topic may quickly change during the emergency. Therefore, the early-trained model may not work consistently well during the emergency. Unfortunately, the late-trained model will not be available until months after the emergency begins. In this paper, we propose an ensemble method to combine the pre-trained model and early-trained model for achieving better analysis performance on COVID-19-related tweets. The effectiveness of this method has been verified by the experiments.

Finding Pseudo-Cliques with Core Nodes Based on Formal Concept Analysis

Yoshiaki Okubo
Faculty of Information Science and Technology, Hokkaido University, Sapporo, Japan

Abstract—In this paper, we present an algorithm for finding τ -pseudo cliques, a kind of dense communities, in a given network/graph. A τ -pseudo clique is defined as a union of several maximal cliques in the graph which has a certain degree of ovarlapness as its core. Although an algorithm for detecting those pseudo cliques has already been proposed, its empirical

performance would not be acceptable particularly for large graphs because the computational procedure of the algorithm is rather complicated. In order to make this kind of pseudo-cliques more practical, we design a new algorithm for detecting them with the help of formal concept analysis, a method for analyzing relational data. A relationship between τ -pseudo cliques and formal concepts shows that pseudo cliques can be found as formal concepts satisfying a certain constraint w.r.t. their cores. In our FCAbased algorithm, we can enjoy simple pruning rules based on the definition of τ -pseudo cliques. From our preliminary experimental results for several benchmark and real networks, it is expected that the proposed method would be a promising approach to detecting τ -pseudo cliques.

Bullying and Hazing in Computer Science

Jeremy Straub
Department of Computer Science, North Dakota State University, Fargo, North Dakota, USA

Abstract—To determine how prevalent hazing is in tenuretrack computer science positions at United States research universities, a survey was sent to all individuals holding a position with a tenure-track title at these universities. This article reports that 26% of respondents reported one or more activities commonly associated with bullying / hazing occurring to them and 20% reported being aware of it happening to a colleague in a tenure-track position. Even more problematic, 71% of respondents reported having one or more symptoms commonly associated with hazing or bullying themselves and 56% of respondents reported observing these symptoms in a tenure-track colleague. Despite this, the vast majority of respondents indicated that they did not consider themselves to be hazed or bullied.

CSCI-ISBD: BIG DATA AND DATA SCIENCE

Discovering a Learning Module for Poker's Rule through Data Mining Algorithms

Tsenguun Tsogbadrakh, Amal Alhosban
Department of Computer Science, Engineering and Physics, University of Michigan - Flint, Michigan, USA

Abstract—Data mining classification methods can be a powerful tool when it comes to learning card game rules such as Poker. There are millions of possible combination in the game and making a decision tree to cover all the rules is not desirable. We used the J48 decision tree model of data mining software Weka and made parameter analysis. Then we show experimentally how the number of instances is affecting the correctness of the classification, and propose an equation to determine accuracy based on the number of instances in a data set. We examine several different attributes and the experiment shows high performance.

Impact of Weather Conditions on the COVID-19 Pandemic in the United States: A Big Data Analytics Approach

Farid Ghareh Mohammadi, Farzan Shenavarmasouleh, M. Hadi Amini, Hamid R. Arabnia
Department of Computer Science, University of Georgia, USA
School of Computing and Information Sciences, College of Engineering and Computing Sustainability, Optimization, and Learning for InterDependent Networks Laboratory (solid lab), Florida International University, Miami, Florida, USA

Abstract—It is 2020 and the world is dealing with a unique pandemic situation as the infamous Coronavirus (COVID-19) has caused an emergency in many countries. Assessing the death rate and confirmed cases of this epidemy with respect to external

causes can be extremely beneficial; however, it comes with its own challenges. In this study, we aim to model and investigate the correlation between meteorology factors and the COVID-19 situation in the United States. To that end, we make use of two datasets for Spatiotemporal analysis. We first gather the surveillance data of COVID-19 released by Johns Hopkins University for the past few months and then fuse it with the meteorology data that we collect from the National Oceanic and Atmospheric Administration accordingly. The meteorology dataset provides different useful factors including but not limited to the minimum and maximum temperature of each day, all in Fahrenheit. We seek to find out the relationship between the weather condition and the virus spread and the number of death rates in different states of the United States.

Cross-Compiled Plotting with SwiftVis2

Nicholas J. Smoker, Mark C. Lewis
Department of Computer Science, Trinity University, San Antonio, Texas, USA

Abstract—This paper explores adding support to SwiftVis2, a plotting library written in Scala, for compiling and running across the JVM, web browsers, and native compiled code. We find that it is possible to use a single core library written in standard Scala across these various compile targets with little additional code. The quality of the output and ability to handle large datasets varies significantly for different implementations. The use of Scala, or other languages that have compilers targeting multiple platforms, makes it possible to easily target many platforms with relatively little effort.

Discovery of Burglary Hotspots and Extraction of their Features

Andrew Little, Ray R. Hashemi, Jeffrey Young
Department of Computer Science, Georgia Southern University, Savannah, Georgia, USA
Department of Computer Science, Clemson University, Clemson, South Carolina, USA

Abstract- Investigation of the underline contributing factors to burglaries within a geographical area, is essential in finding solutions. Such investigation for a geographical area in the southeast of Georgia is the subject of this paper. Our goal is bifold: (a) discovering the burglary hotspots for the area and (b) extracting the features associated with each given hotspot. Such features are the underlying contributing factors for each hotspot, which differ from one hotspot to the next. Data sources used for meeting the goal are cleansed city Police reports, United States Census, Google maps, and the City Ordinances. The discovery of the hotspots is accomplished by using the Optics clustering technique and feature extraction is performed by an association analysis.

Predicting Large-scaled, Cloud-hosted Virtual World Resource Demands for Automated Load Balancing

Nikki Sharma, Kelly Rivera, Anastasia Angelopoulou, Sean Mondesire
College of Science, St. Thomas University, Miami, Florida, USA
College of Science, Columbus State University, Columbus, Georgia, USA

Abstract—The complexity and computational resource demands of large-scale virtual world simulations are increasing with advancements in CPU and GPU computing power, availability of broadband internet connectivity, and reduction in hardware costs. From these areas of growth, cloud-based services are gaining in popularity for hosting simulations due to their costeffectiveness, ease in resource management and scale, and data center availability in virtually every region in the world. The next phase of cloud-based simulation hosting will address the fallacies of over/under-utilization of compute resources by incorporating big data analytics and processing, software containerization, automated allocation schemes, and load-balancers. To meet these demands, this study performs the preliminary work of profiling virtual-world simulation loads and then determine which form of analysis (linear or non-linear methods) produces the lowest error prediction of resource needs. As a result of this study, future automated load-balancers will be able to estimate upcoming simulation demands and modify current resource allocations to better match actual needs.

A Mediated Multi-RNN Hybrid System for Prediction of Stock Prices

Ray R. Hashemi, Omid M. Ardakani, Azita A. Bahrami, Jeffrey A. Young
Department of Computer Science, Georgia Southern University, Savannah, Georgia, USA
Department of Economics,Georgia Southern University, Savannah, Georgia, USA
IT Consultation, Savannah, Georgia, USA
Department of Computer Science, Clemson University, Clemson, South Carolina, USA

Abstract - A multi-recurrent neural network (RNN) hybrid system made up of three RNNs is introduced to predict the stock prices for 10 different companies (five selected from the Dow Jones Industrial Average and five from the Standard and Poor's 500.) The daily historical data used to train and test the system are collected for the period of October 15, 2013 to March 5, 2019. For each company, the system provides two separate predictions of the daily stock price by using (1) historical stock prices and (2) historical trends along with the historical daily net changes in stock price. The two predictions are mediated to select one as the final output of the hybrid system. For each company, the accuracy of the system was tested for the prediction of the most recent 98 consecutive days using the forecast accuracy measure of the Mean Squared Error (MSR). The results revealed that for every company the difference between the predicted and actual stock price is not statistically different from zero, which is the ideal (error-free) forecast.

Using Process Maps to Analyze Researchers' Productivity Behavior

Gilberto Ayala-Bastidas, Hector G. Ceballos, Francisco J. Cantu-Ortiz, Luciano Garcia-Banuelos
School of Engineering and Science, Tecnologico de Monterrey, Monterrey, Mexico

Abstract—The researchers' career has several factors that influence its development. Researchers continually have to face particular dilemmas: publishing in conferences or journals, choosing between a research topic or another, deciding whether to work individually or collaboratively and if so, if it will be an international collaboration or not; the publications made by the researcher over time evidence all these decisions. This work aims to model researchers' careers in terms of publication data (activities) to obtain insights about frequent paths researchers make along time. Our methodology involves generating an event log created by publication data extracted from a bibliometric database to discover a process structure representing researchers' long-term strategies. Our results show that it is possible to get insights about frequent paths in a process map. Keywords-Research

Utilizing Fuzzy Logic for Assessing "FAIRness" of a Digital Resource

Abdullah Alowairdhi, Xiaogang Ma
Department of Computer Science, University of Idaho, Moscow, Idaho, USA

Abstract—The goal of this research is to examine the possibility of utilizing fuzzy logic to evaluate the uncertainty of FAIRness level (findability, accessibility, interoperability, and reusability) of a digital resource. To date, there are no FAIRness evaluation studies based on fuzzy logic. To measure this uncertainty, we built a Fuzzy FAIR Assessment Framework (FFAF). FFAF is based on the three main steps of fuzzy logic: fuzzification, inferencing, and defuzzification. Applying the FFAF model on the FAIR data principles led to a specific FAIRness level result. Overall, this research shows that fuzzy logic is an efficient technique for evaluating the FAIRness level of a digital resource.

Extraction of Key Concept Relevance Graphs from Fourteen Decades of Psychoanalytic Journal Publications

Sheryl Brahnam, Rick Brattin, Andrew Crofford, Justin Freres
Information Technolgy & Cybersecurity, Missouri State University, Springfield, Missouri, USA

Abstract— Automatically tracing the history of concepts and terms by text mining diachronic corpora is a new multidisciplinary area of research. An essential step in the pipeline of such research is calculating the relevance of terms, with most work relying on simple frequency measures. In this study, we extract relevance graphs of key psychoanalytic concepts from a corpus built for this task, one that represents the entire history of psychoanalysis. Different measures of relevance are examined and shown to tell a different story; when combined, however, they provide a clearer picture of a term's significance over time.

A Low Cost LoRa-based IoT Big Data Capture and Analysis System for Indoor Air Quality Monitoring

Matthew Meli, Edward Gatt, Owen Casha, Ivan Grech, Joseph Micallef
Department of Microelectronics & Nanoelectronics, University of Malta, Msida, Malta

Abstract— This paper presents a low cost LoRa-based IoT big data capture and analysis system for indoor air quality monitoring. This system is presented as an alternative solution to expensive and bulky indoor air quality monitors. It enables multiple low cost nodes to be distributed within a building such that extensive location-based indoor air quality data is generated. This data is captured by a gateway and forwarded to a cloud-based LoRaWAN network which in turn publishes the received data via MQTT. A cloud-based data forwarding server is used to capture, format and store this big data on a cloud-based document-oriented database. Cloud-based services are used for data visualization and analysis. Periodic indoor air quality graphs along with air quality index and thermal comfort index heat maps are generated.

Energy Demand Forecasting and Error Correction with Decision Tree

Maria de Guadalupe Cota Ortiz, Pedro Flores Perez
Universidad de Sonora, Hermosillo, Sonora, Mexico

Abstract - Search solutions to optimize resources for energy demand is a complex problem. Factors such as the increase in energy consumption and environmental variation are basic to estimate the precision of the resource that will be generate. This article describes the procedure to correct errors on the results of energy demand forecasts, previously obtained with a library based on time series and the application of the 2G algorithm in the error correction stage. The experimental results indicate efficiency, however, for optimal results, it is necessary to have a larger dataset.

Intuitive Time-Series-Analysis-Toolbox for Inexperienced Data Scientists

Felix Pistorius, Daniel Baumann, Luca Seidel, Eric Sax
Institute for Information Processing Technologies (ITIV), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract—There are many different procedures to carry out a data mining project. Depending on the application, various methods have to be chosen, which is normally done by experts. For inexperienced users without machine learning experience, it is difficult to analyse their data without help. To simplify the access to data mining particularly for multivariate time series analysis, we propose an intuitive toolbox that guides step-by-step through data mining process. Furthermore, it supports inexperienced users with pre-suggested methods in every data mining process step. Therefore, specifications for such a toolbox are defined and a prototype for the realization of a toolbox is presented. The work steps which are based on the knowledge discovery process can be adapted and changed to suit different application scenarios.

A Grid Partition-Based Local Outlier Factor by Reachability Distance for Data Stream Processing

Raed Alsini, Omar Alghushairy, Xiaogang Ma, Terence Soule
Department of Computer Science, University of Idaho, Moscow, Idaho, USA
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
College of Computer Science and Engineering, University of Jeddah, Saudi Arabia

Abstract— In the field of big data, outlier detection is an important issue in many applications, such as communications, health, and network intrusion detection. Data streams present a challenge to the traditional outlier detection methods due to their unique characteristics, i.e., large volume and sequential structure. The Local Outlier Factor (LOF) is a popular algorithm used in anomaly detection to find outliers in a data stream. However, the LOF algorithm has a drawback. When a new point is added to the data stream, the algorithm needs to reprocess the measurement from the beginning of any change in the dataset. Additionally, LOF requires the whole dataset to be stored in memory. These issues were solved in the Grid Partition-based Local Outlier Factor (GP-LOF) algorithm. A new algorithm is introduced in this paper to improve the GP-LOF algorithm. It is based on the reachability distance measurement in LOF, and we name it Grid Partition-based Local Outlier Factor by Reachability distance (GP-LOFR) algorithm. GP-LOFR has slightly improved the accuracy rate of local outlier detection in our experiments with several real-world datasets.

A Literature Review of Data Mining Techniques Used in Collaborative Filtering Recommender Systems

Ashrf Althbiti, Rayan Alshamrani, Xiaogang Ma
Department of Computer Science, University of Idaho, Moscow, Idaho, USA

Abstract— As a result of the everincreasing number of available items in eservices, users are often overwhelmed. Therefore, it is essential to develop and apply algorithms to address the challenge of selection overload. Collaborative Filtering (CF) systems have been developed to help users to find what they might be interested in among a range of available selections. Moreover, CF systems have been widely discussed as an efficient approach to cope with the selection overload issue. This paper presents a literature review of the common CF techniques and data mining techniques used for CF.

Real-Time Data Visualization to Enhance Situational Awareness of COVID Pandemic

Sharad Sharma, Sri Teja Bodempudi, Aishwarya Reehl
Department of Computer Science, Bowie State University, Bowie, Maryland, USA

Abstract— Real-time data visualization can enhance decision making and empower teams with human-centric situational awareness insights. Decision making relies on data which comes in overwhelming velocity and volume, that one cannot comprehend it without some layer of abstraction. This research effort aims to demonstrate the data visualization of the COVID pandemic in real-time for the fifty states in the USA. Our proposed data visualization tool includes both conceptual and data-driven information. The data visualization includes stacked bar graphs, geographic representations of the data, and offers situational awareness of the COVID-19 pandemic. This paper describes the development and testing of the data visualization tool using the Unity gaming engine. Testing has been done with a real-time feed of the COVID -19 data set for immersive environment, non-immersive environment, and mobile environment.

A Novel Fuzzy Clustering Method based on GA, PSO and Subtractive Clustering

Thanh Le
School of Business Information Technology, University of Economics, Ho Chi Minh City, Vietnam

Abstract— Data clustering is a challenging problem in data science, requiring both accurate determination of the number of clusters and correct clustering of the data. While Fuzzy C-means (FCM) is a powerful algorithm that can cluster data into overlapping groups and converge quickly to local optima, it however depends on the choice of initial parameters and may not always reach a global optimum. In addition, FCM by itself cannot determine the correct number of clusters. In this study, we combine FCM with Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Subtractive Clustering (SC) algorithms for a novel algorithm, FZGPS, that both determines the correct number of clusters and efficiently constructs these clusters from a given dataset. We show that FZGPS performs effectively on both artificial and experimentally-derived gene expression datasets.

The Effect of COVID-19 on Various Demographics by Race in USA

Trisha Rayan, Adrian Brown, Andrei Carillo, Sharad Sharma
Thomas Jefferson High School for Science and Technology, Alexandria, Virginia, USA
Department of Computer Science, Bowie State University, Bowie, Maryland, USA

Abstract—There is growing concern that racial and ethnic minority communities around the United States are experiencing a disproportionate burden of infection rate and mortality from the coronavirus disease 2019 (Covid-19). While most research, media newspapers, websites, and television networks are providing statistical numbers of daily infection and death rate across US by state, these numbers fail to study the actual impact of COVID-19 to each race. Our approach has taken the top five races by population count in the US and has calculated the impact index by race for each state for COVID-19 infections and death rate. We also examine the rise in the utilization of hospitals as a result of the rise in cases of COVID-19 in the United states. We conclude that the African American race and Hispanic race is disproportionately impacted more than the white population for infection rate.

A Concise Review of Transfer Learning

Abolfazl Farahani, Behrouz Pourshojae, Khaled Rasheed, Hamid R. Arabnia
Department of Computer Science, University of Georgia, USA
Department of Information and Technology, Road and Urban Development Organization, Arak, Iran

Abstract—The availability of abundant labeled data in recent years led the researchers to introduce a methodology called transfer learning, which utilizes existing data in situations where there are difficulties in collecting new annotated data. Transfer learning aims to boost the performance of a target learner by applying another related source data. In contrast to the traditional machine learning and data mining techniques, which assume that the training and testing data lie from the same feature space and distribution, transfer learning can handle situations where there is a discrepancy between domains and distributions. These characteristics give the model the potential to utilize the available related source data and extend the underlying knowledge to the target task achieving better performance. This survey paper aims to give a concise review of traditional and current transfer learning settings, existing challenges, and related approaches.

CSCI-ISAI: ARTIFICIAL INTELLIGENCE

Stock Market Behaviour Prediction using Long Short-Term Memory Network and Gated Recurrent Unit

Samuel Olusegun Ojo, Pius Adewale Owolawi, Maredi Mphahlele, Chunling T. Du, Juliana Adeola Adisa
Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria, South Africa

Abstract—Stock Market behaviour prediction has been an area of interest within research for some time, and Recurrent Neural Networks have shown great promise in the way of solving time series based problems. In this study we analyse two Recurrent Neural Network based models. One makes use of Long Short Memory Networks and the other makes use of a variation of the Long Short Memory Networks called the Gated Recurrent Unit for the purpose of Stock Market behaviour prediction. A comparison is made between the two models based on training on the same stock market dataset. Results obtained show a greater accuracy for the Long Short Memory Network model in comparison to the Gated Recurrent Unit based model.

Identifying the Training Stop Point with Noisy Labeled Data

Sree Ram Kamabattula, Venkat Devarajan, Babak Namazi, Ganesh Sankaranarayanan
Electrical Engineering, University of Texas at Arlington, Texas, USA
Baylor Scott and White Health Research Institute, Dallas, Texas, USA
Department of Surgery, Baylor Scott and White Health Research Institute, Dallas, Texas, USA

Abstract - Finding an early stopping point at maximum obtainable test accuracy (MOTA) is a challenging problem when training deep neural networks (DNNs) with noisy labeled data. Recent studies assume either that i) a clean validation set is available or ii) the noise ratio is known, or, both. However, often a clean validation set is unavailable, and the noise estimation can be inaccurate. To overcome these issues, we provide a novel training solution, free of these conditions. We analyze the rate of change of the training accuracy under different conditions to identify a training stop region. We further develop a heuristic algorithm (AutoTSP) to find a training stop point (TSP) at or close to MOTA. We validate the robustness of AutoTSP through several experiments on various datasets, noise ratios and architectures.

A Conceptual Model for Real-Time Binaural-Room Impulse Responses Generation using ANNs in Virtual Environments: State of the Art

Daniel A. Sanaguano, Jose Lucio Naranjo, Roberto A. Tenenbaum
Escuela Politecnica Nacional, Quito, Ecuador
Universidade Federal de Santa Maria, Santa Maria, Brazil

Abstract—This work aims to give an overview of Artificial Neural Networks (ANN) approaches applied for BIRs generation published in the literature and to expose gaps in the academic research. The literature review shows that several successful studies are using ANNs approaches for BIRs generation with a reduction in the computational effort by up to 90% with respect to the Traditional Method. Nevertheless, these approaches are bounded by a fixed pair of a soundsource and binaural-receptor, meaning that they do not take into account dynamic variations in the position of the receptor. In this sense, this work also introduces a conceptual model for a real-time BIRs generator that considers a moving binauralreceptor using a set of Artificial Neural Networks.

A Summary Evaluation Method Combining Linguistic Quality and Semantic Similarity

Xingwen Wang, Bo Liu, Libin Shen, Yong Li, Rentao Gu, Guangzhi Qu
Beijing University of Technology, Beijing, P. R. China
Beijing University of Posts and Telecommunications, Beijing, P. R. China
Oakland University, Rochester, Michigan, USA

Abstract—Summary evaluation method is crucial to promote the development of text summarization technologies. However, most of the existing summary evaluation methods seldom consider the content integrity and readability of summaries simultaneously. This paper proposed a Linguistic Quality and Semantic Similarity Model (LQSSM) to evaluate generated summaries more comprehensively. Considering the readability of summaries, a linguistic quality evaluation network (LQEN) is proposed to evaluate summaries automatically. Meanwhile, a semantic similarity evaluation network (SSEN) is introduced to directly measure the informativeness between the summary and original text. Additionally, there is also a comprehensive evaluation using fusion indicators. The LQSSM and RecallOriented Understudy for Gisting Evaluation (ROUGE) methods were used to score the summaries generated by five models. The results showed that the proposed method gave further feedback from different angles on the basis of reaching the level of other methods.

Crossroad Accident Responsibility Prediction Based on a Multi-agent System

Helton Agbewonou Yawovi, Tadachika Ozono, Toramatsu Shintani
Department of Computer Science, Graduate School of Engineering - Nagoya Institute of Technology, Nagoya, Japan

Abstract—With the increasing number of motorized vehicles, road accidents are increasing year by year all over the world.. After an accident, the police investigate the circumstances of the incident and determine each actor's responsibilities. Our goal is to create a police support system. We focused on a multi-agent system that predicts each actor's responsibility in a road accident (especially crossroad accidents). The system uses the driving recorder video of a vehicle as the input data source, and it outputs the prediction of the responsibility of each actor in the accident. It consists of three agents: (1) Crash time detection and crash video split into images; (2) Traffic signs detection in the crash video; (3) Responsibility prediction using a knowledge system.

The Use of Video Captioning for Fostering Physical Activity

Soheyla Amirian, Abolfazl Farahani, Hamid R. Arabnia, Khaled Rasheed, Thiab Taha
Department of Computer Science, University of Georgia, USA

Abstract – To be provided soon.

A Proof of Sparseness, Optimality, and Convergence of an LP-SVR

Pablo Rivas, Korn Sooksatra
School of Engineering and Computer Science, Department of Computer Science, Baylor University, Waco, Texas, USA

Abstract—This article presents a proof of convergence and sparsity of a linear programming support vector machine for regression. First, the Support Vector Regression (SVR) problem is posed as a linear programming problem modeled on a primal and dual fashion leading to the definitions of optimality. Second, we describe a sequential optimization method based on variables decomposition, constraints decomposition, and primal-dual interior point methods for solving large-scale regression/classification problems. Third, based on the methodology, we present proof of convergence and optimality conditions of the sequential optimization and its ability to produce sparse solutions.

A Review of Machine Learning and Cryptography Applications

Korn Sooksatra, Pablo Rivas
Department of Computer Science, Baylor University, Waco, Texas, USA

Abstract—Adversarially robust neural cryptography deals with the training of a neural-based model using an adversary to leverage the learning process in favor of reliability and trustworthiness. The adversary can be a neural network or a strategy guided by a neural network. These mechanisms are proving successful in finding secure means of data protection. Similarly, machine learning benefits significantly from the cryptography area by protecting models from being accessible to malicious users. This paper is a literature review on the symbiotic relationship between machine learning and cryptography. We explain cryptographic algorithms that have been successfully applied in machine learning problems and, also, deep learning algorithms that have been used in cryptography. We pay special attention to the exciting and relatively new area of adversarial robustness.

Unsupervised Learning with Word Embeddings Captures Quiescent Knowledge from COVID-19 Drugs Literature

Tasnim Gharaibeh, Elise de Doncker
Department of Computer Science, Western Michigan University, Kalamazoo, Michigan, USA

Abstract— As COVID-19 patients flood hospitals worldwide, physicians are trying to search for effective antiviral therapies to save lives. However, there is currently a lack of proven effective medications against COVID-19. Multiple COVID-19 vaccine trials and treatments are underway, but yet need more time and testing. Furthermore, the SARS-CoV-2 virus that causes COVID-19 replicates poorly in multiple animals, including dogs, pigs, chickens, and ducks, which limits preclinical animal studies. We built an unsupervised deep learning model (CDVec) to produce word-embeddings using word2vec from a corpus of articles selectively focusing on COVID-19 candidate drugs that appeared in the literature to identify promising target drugs that could be used in COVID-19 treatment.

A Deep Learning based Customer Sentiment Analysis Model to Enhance Customer Retention and Loyalty in the Payment Industry

George Wanganga, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract— Both the industry and academia agree on the immense contribution of big data analytics and machine learning to competitive businesses. The payment industry would benefit from big data analytics and machine learning capabilities to harness their customers' opinions through sentiment analysis, thereby customizing their services and products to fit their customers' preferences. However, the challenge is implementing this competitive edge in small and medium-sized payment solution providers. This paper proposes a deep learning-based customer sentiment analysis model and a related (SaMS-PSP) algorithm that implements sentiment analysis within SaMS-PSP. Through experiments, we have demonstrated that our model has a super performance advantage over conventional machine learning methods and is more suited to handle "big data" applications such as customer sentiment analysis. This research has demonstrated that the sentiment analysis emotional polarity score can be used in a value-added customer orientation tool to promote customer retention and loyalty within SaMS-PSP.

Unified End-to-End Sentence Denoising

Zhantong Liang, Abdou Youssef
Department of Computer Science, The George Washington Univerisity, Washington, DC, USA

Abstract—It takes more than correct grammar to speak good English. In this paper, we describe the sentence denoising task that reduces the vagueness, redundancy, and irrationality of a grammatical sentence. We define a rich, linguistics-inspired noise taxonomy and establish the formal definition of the problem. A unified end-to-end model based on Transformer is proposed and an efficient algorithm for constructing the training data is given, together with a separate fine-tuning step to get the ideal model. Our method outperforms previous results and keeps good accuracy as the noise composition gets more complicated.

Performance Analysis of Tor Website Fingerprinting over Time using Tree Ensemble Models

Hyoungseok Oh, Donghoon Kim, Won-gyum Kim, Doosung Hwang
Department of Software Science, Dankook University, Yongin-si, South Korea
Department of Computer Science, Arkansas State University, Jonesboro, Arkansas, USA
AiDeep, Seoul, South Korea

Abstract—Tor (The Onion Router) ensures network anonymity by encrypting contents through multiple relay nodes. Recent studies on website fingerprinting (WF) showed that websites can be identified with high accuracy by analyzing traffic data. However, websites are changing over time by updating contents, which can significantly reduce the accuracy of WF attacks. This study analyzes the performance over time by using ensemble models with excellent WF attack performance. The experiment are conducted in two cases with the initial model. The not updated analyzes the accuracy of models made from initial data over time, whereas the updated adds data that has changed over time to update the model to analyzes the accuracy. The average accuracy of the initial ensemble models is over 90.0% and the Rotation Forest algorithm shows high performance of 93.5%. Comparing the models trained after 30 days with the initial model, the classification performance dropped in both cases; the not updated dropped by more than 30.0% and the updated dropped by about 10.0%. The experimental results suggest that WF using machine learning may require model learning on a regular basis.

Comparative Analysis of Machine Learning Models for Diabetes Mellitus Type 2 Prediction

Leila Ismail, Huned Materwala
Distributed Computing and Systems Research Laboratory, Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Al-Ain, Abu-Dhabi, United Arab Emirates, UAE

Abstract—Diabetes is one of the top 10 causes of death worldwide. Health professionals are aiming for machine learning models to support the prognosis of diabetes for better healthcare and to put in place an effective prevention plan. In this paper, we conduct a comparative analysis of the most used machine learning models in the literature to predict the prevalence of diabetes mellitus type 2. We evaluate the models in terms of accuracy, Fmeasure and execution time with and without feature selection using a real-life diabetes dataset. The detailed analysis is in the paper.

Detection of Hate Speech in Videos Using Machine Learning

Ching Seh Wu, Unnathi Bhandary
Department of Computer Science, San Jose State University, San Jose, California, USA

Abstract—With the progression of the Internet and social media, people are given multiple platforms to share their thoughts and opinions about various subject matters freely. However, this freedom of speech is misused to direct hate towards individuals or group of people due to their race, religion, gender etc. The rise of hate speech has led to conflicts and cases of cyber bullying,

causing many organizations to look for optimal solutions to solve this problem. Developments in the field of machine learning and deep learning have piqued the interest of researchers, leading them to research and implement solutions to solve the problem of hate speech. Currently, machine learning techniques are applied to textual data to detect hate speech. With the ample use of video sharing sites, there is a need to find a way to detect hate speech in videos. This research deals with classification of videos into normal or hateful categories based on the spoken content of the videos. The video dataset is built using a crawler to search and download videos based on offensive words that are specified as keywords. The audio is extracted from the videos and is converted into textual format using a Speech-to-Text converter to obtain a transcript of the videos. Experiments are conducted by training four models with three different feature sets extracted from the dataset. The models are evaluated by computing the specified evaluation metrics. The evaluated metrics indicate that Random Forrest Classifier model delivers the best results in classifying videos.

Radically Simplifying Game Engines: AI Emotions and Game Self-Evolution

John N. Carbone, James Crowder, Ryan A. Carbone
Forcepoint LLC, Richardson, Texas, USA
Colorado Engineering, Colorado Springs, Colorado, USA
University of Incarnate Word, San Antonio, Texas, USA

Abstract - Today, video games are a multi-billion-dollar industry, continuously evolving through the incorporation of new technologies and innovative design. However, current video game software content creation requires extensive and often-times ambiguous planning phases for developing aesthetics, online capabilities, and gameplay mechanics. Design elements can vary significantly relative to the expertise of artists, designers, budget, and overall game engine/software features and capabilities. Game development processes are often extensively long coding sessions, usually involving a highly iterative creative process, where user requirements are rarely provided. Therefore, we propose significantly simplifying game design and development with novel Artificial Cognition Architecture real-time scalability and dynamic emotion core. Rather than utilizing more static emotion state weighting emotion engines (e.g. ExAI), we leverage significant ACA research in successful implementation of analog neural learning bots with Maslowan objective function algorithms. We also leverage AIbased Artificial Psychology software which utilizes ACA's fine grained self-evolving emotion modeling in humanistic avatar patients for Psychologist training. An ACA common cognitive core provides the gaming industry with wider applications across video game genres. A modular, scalable, and cognitive emotion game architecture implements Non-Playable Character (NPC) learning and self-evolution. ACA models NPC's with fine grained emotions, providing interactive dynamic personality traits for a more realistic game environment and enables NPC self-evolution under the influence of both other NPC's and players. Furthermore, we explore current video game design engine architecture (e.g. Unity, Unreal Engine) and propose an ACA integration approach. We apply artificial cognition and emotion intelligence modeling to engender video games with more distinct, realistic consumer gaming experiences, while simultaneously minimizing software gaming development efforts and costs.

Machine Learning Techniques to Enhance Container Network Security

Abhinav Kommula, Yen-Hung (Frank) Hu, Mary Ann Hoppa, Samuel Olatunbosun
Monta Vista High School, Cupertino, California, USA
Norfolk State University, Norfolk, Virginia, USA

Abstract - Containers are designed as lightweight alternatives to Virtual Machines (VMs) with faster and more efficient deployment capabilities. As more applications are being run in the cloud, containers' role in deploying microservices is becoming increasingly important. Retrofitting new technology like containers into existing technology such as Linux introduces security vulnerabilities. In this paper, we analyze in detail several aspects of container security: shared resources like memory and network in the public cloud. Through experimentation, we programmatically proved that there are no threats in memory sharing between containers on Linux Ubuntu 20.04LTS. However, shared networking imposes security vulnerabilities. To mitigate these issues of shared networking, we propose a machine learning solution in which a stateful network bridge can learn critical information, such as Media Access Control (MAC) addresses of containers and Internet Protocol (IP) addresses and port numbers used by applications, through Layer 2 and Layer 3 processing of packets within a Docker-Compose cluster. This solution will not only help identify spoofing attacks, but it will also prevent network traffic snooping, thus securing application containers. We hope that our analysis will augment current efforts in implementing more secure containers and provide new avenues of research in container network security.

Data Poisoning on Deep Learning Models

Charles Hu, Yen-Hung (Frank) Hu
Woodside High School & Governor's School for Science and Technology, Newport News, Virginia, USA
Department of Computer Science, Norfolk State University, Norfolk, Virginia, USA

Abstract - Deep learning is a form of artificial intelligence (AI) that has seen rapid development and deployment in computer software as a means to implementing AI functionality with greater efficiency and ease as compared to other alternative AI solutions, with usage seen in systems varying from search and recommendation engines to autonomous vehicles. With the demand for deep learning algorithms that can perform increasingly complex tasks in a shorter time frame growing at an exponential pace, the developments in the efficiency and productivity of algorithms has far outpaced that of the security of such algorithms, drawing concerns over the many unaddressed vulnerabilities that may be exploited to compromise the integrity of these software. This study investigated the ability of poisoning attacks, a form of attack targeting the vulnerability of deep learning training data, to compromise the integrity of a deep learning model's classificational functionality. Experimentation involved the processing of training data sets with varying deep learning models and the incremental introduction of poisoned data sets to view the efficacy of a poisoning attack under multiple circumstances and correlate such with aspects of the model's design conditions. Analysis of results showed evidence of a decrease of classificational ability correlating with an increase of poison percentage in the training data sets, but the scale of which the decrease occurred varied with the specified parameters in the model design. Based on this, it was concluded that poisoning can provide varying levels of damage to deep learning classificational ability depending on the parameters utilized in the model design, and methods to countermeasure such were proposed, such as increasing epoch count, implementing mechanisms bolstering model fit, and integrating input level filtration systems.

Chinese Surgery Text ICD Coding Classification using Hybrid Machine Learning Strategy

Xiaoyuan Bao, Yunhaonan Yang, Kai Zhang
Medical Informatics Center, National Health Service Data Center, Peking University Health Science Center, Beijing, P. R. China
Peking University School of Public Health, Peking University Health Science Center, Beijing, P. R. China
Peking University People's Hosptial, Peking University Health Science Center, Beijing, P. R. China

Abstract—The automatic classification of clinical electronic medical records often uses ICD diagnostic codes to classify diseases, but there is a lack of research on the classification of surgical records using surgical codes. According to ICD-9 surgical codes and using the largest Chinese deidentified medical corpus, we evaluated the effects of 7 common word segmentation methods, 5 feature selection methods, 9 machine learning algorithms and BI-LSTM. Finally, we used three integrated strategies to test the effect of multi-classification on another data set. The result shows that 3-gram, LinerSVC and ADABOOST have the best general application effect, ADABOOST could be superior to BI-LSTM after optimization, and the effect of multi-classification could be improved by multimodel voting. Our study provides a reference for the classification of Chinese medical corpus in the future.

An Alert System: Using Fuzzy Logic for Controlling Crowd Movement by Detecting Critical Density Spots

Alaa Edris, Abdullah Alajlan, Frederick Sheldon, Terence Soule, Robert Heckendorn
Computer Science Department, University of Idaho, Moscow, Idaho, USA
Computer Science Department, University of Jeddah, Saudi Arabia
Computer Science Department, Technical and Vocational Training Corporation, Riyadh, Saudi Arabia

Abstract—When people form crowds without an orderly arrangement, we need to consider how to best control their movement to provide safety and maximize their experience. Controlling a large crowd is a complicated and costly operation, but it is important to prevent risky situations that can lead to trampling and crushing, and to provide general public safety. Crowd control forces must be able to organize people to provide successful crowd management. In this paper we examine two types of crowds: structured crowds, where people are heading towards the same goal and in generally the same direction, as happens in religious gatherings

like the Islamic Hajj or the Hindu Kumbh Mela; and unstructured crowds, where people walk in different directions, as occurs in train stations and in the centers of towns and cities. By identifying the locations of each individual in a crowd, many potential problems of controlling crowds can be detected and avoided, since we will able to identify the direction and speed of each individual and compare it to the surrounding crowd. We propose an alert system as a way of keeping crowds safe in risky situations. The alert system focuses on scrutinizing the status of individuals in order to inform the authorities in case of risk behavior. Fuzzy logic is proposed as the basis for the alert system for deciding if the locations of individuals are considered critical spots causing obstruction of crowd motion. The aim of this paper is to establish a system that is able to process and analyze the individuals' locations according to their current locations status. The system is using the fuzzy logic, as the machine starts to learn the critical density spots by pointing these locations.

Using EEG Data and NeuCube for the Study of Transfer of Learning

Mojgan Hafezi Fard, Krassie Petrova, Maryam Doborjeh, Nikola Kasabov
School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand
Ulster University, United Kingdom

Abstract— Deeper and long-lasting learning occurs through a critical review of prior knowledge in the light of the new context, and a transfer of the acquired knowledge to new settings. Attention to the task is one of factors that enable transfer of learning (TL). This study adopts a cognitive neuroscience approach to the study of TL. More specifically, it investigaes the relationship between attention to the task and prior knowledge. The study uses a brain like artificial intelligence (BLAI) architecture (NeuCube) which is based on spiking neural networks (SNN) to represent brain data during a series of cognitive tasks and interpret them in the context of the research question. The results of the experiment indicate that modelling and analysing spatio-temporal brain data (STBD) using the SNN environment of NeuCube suggested a better understanding of the process of TL, and the associated brain activity patterns and relationships. The outcomes of this study are used to inform the design of a follow-up study where SNN models will be built from STBD gathered from participants engaged in learning and in TL.

Satellite Image Atmospheric Air Pollution Prediction through Meteorological Graph Convolutional Network with Deep Convolutional LSTM

Pratyush Muthukumar, Emmanuel Cocom, Kabir Nagrecha, Jeanne Holm, Dawn Comer, Anthony Lyons, Irene Burga, Christa Hasenkopf, Mohammad Pourhomayoun
Department of Computer Science, California State University Los Angeles, Los Angeles, California, USA
City of Los Angeles, Los Angeles, California, USA
OpenAQ, Washington, D.C., USA

Abstract—Every five seconds, somebody around the world prematurely dies from the effects of air pollution. Air pollution is one of the world's leading risk factors for death. To mitigate the deadly effects of air pollution, it is imperative that we understand it, discover the patterns and sources, and predict it in advance. Air pollution prediction in real-time requires extremely powerful models that can solve this spatiotemporal problem in multiple dimensions. We used an advanced graph convolutional network coupled with a deep convolutional LSTM model to learn patterns over the spatial and temporal dimension in real-time. Our model employs a graph convolutional network that models meteorological features and extracts highlevel embeddings through unsupervised representation learning. We created a sequential encoder-decoder deep convolutional LSTM that allows for accurate and efficient satellite image based atmospheric Nitrogen Dioxide air pollution prediction over Los Angeles county 10 days into the future using data from 10 days in the past through the use of spatiotemporal satellite imagery and meteorological graph embedding inputs. Our results for predicting spatially continuous atmospheric Nitrogen Dioxide in Los Angeles over various time periods shows improvement in prediction over previous research done on this topic.

The Impact of Applying Recommendation Techniques on Traditional Shopping - A Survey

Manal Alghieth
Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

Abstract—Recommendation systems have been a very important tool because of the large growth of the data such as movies, books, music and shopping items. It offers to online shoppers a relative recommendation. Consequently, saving time and less complication. This study investigates the possibility and the impact of applying recommendation systems on (non-online) traditional shopping to offer the only related items to the client and make the products more personalized. A full analysis has been made based on a questioner and a quality measurement indicator and showed a good impact of recommendation systems (Collaborative filter.) on traditional shopping.

Collective Anomaly Detection for Multivariate Data using Generative Adversarial Networks

Chihiro Maru, Ichiro Kobayashi
Ochanomizu University, Japan

Abstract—Generative adversarial network (GAN) is used to model complex high-dimensional distributions of real-world scenarios. It has been applied to anomaly detection and making great achievements. However, most of the existing GAN-based anomaly detection methods cannot detect collective anomalies that change the behavior of multipoint data instances. Moreover, although many GAN-based methods for time-series anomaly detection have been proposed, there are few studies to handle collective anomalies in time-series data. Besides, there is still much room to improve the methods in terms of computational cost and the accuracy for detecting anomaly. We thus aim to propose a GANbased method to detect multi-dimensional collective anomalies with high accuracy. To correctly detect collective anomalies, we especially introduce an encoder into a GAN-based anomaly detection method to obtain the latent states of the real data. We furthermore adopt a sequence to sequence technique to both encoder and generator, recurrent neural network, and fully connected neural network for the discriminator. We conducted experiments using two types of datasets: artificial and natural, and verified the effectiveness of our GAN model.

Deep Learning Techniques for Stock Market Prediction in the European Union: A Systematic Review

Argyrios P. Ketsetsis, Christos Kourounis, Georgios Spanos, Konstantinos M. Giannoutakis, Pavlos Pavlidis, Dimitris Vazakidis, Theofanis Champeris, Dimitris Thomas, Dimitrios Tzovaras
Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece
Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece
Effect Computer Applications S.A., Athens, Greece

Abstract—Undoubtedly, stock market prediction constitutes one of the most popular and prominent problems that concerns a multidisciplinary audience. Indeed, various disciplines participate in this intriguing exercise including Economics, Statistics and Computer Science. Furthermore, the proliferation and high performance of machine learning methodologies and especially, deep learning techniques, have led the research community to adopt them in time-series forecasting such as stock prices. For this reason, a Systematic Literature Review (SLR) is conducted including primary studies that deal with the prediction of stock markets in the European Union (EU), using deep learning techniques. The proposed SLR involves 12 papers, indicating that there is not yet intense activity in this field, which thus appears open for further research.

Large Scale Data Mining for Banking Credit Risk Prediction

Suleiman Ali AlSaif
Computer Science Department, Deanship of Preparatory Year and Supporting Studies, Imam Abdulrahman Bin Faisal University, Saudi Arabia

Abstract—To become compliant with changing and stricter regulatory demands in the world, banks are focused on undertaking preventive measures for effective credit risk management. One of its drivers is to strengthen existing models to increase their predictive powers and help detect future defaults in advance. In this paper, we propose a new evaluation method to improve the overall performance of learning algorithms in the field of over-indebtedness prediction by using feature selection. Our objective is to find the most relevant subset of features related to the customer banking history for use in prediction model construction. Experimental results have shown that the proposed evaluation method is very useful to build a good predictor.

Artifical Intelligence Techniques Applied to Massive Collected Sensor Data in Subsurface Energy Applications

Mohammadreza Karbalaei Saleh, Fatemeh K. Saleh, Saeed Salehi
College of Engineering, IAUKB Univeristy, Alborz, Iran
College of Earth and Energy, University of Oklahoma, Norman, Oklahoma, USA

Abstract— Artificial Neural Network (ANN) has been used by the energy industry to identify critical well performance drivers since the 1990's. Present work deals with predicting the penetration rate of drill bit using back propagation neural network (BPNN). In this study, massive data collected by different sensor types were analyzed, trained, and used to present the prediction models. The optimum number of hidden layers and the number of neurons in the hidden layer were obtained using trial and error by calculating the mean square of error (MSE). The performance of the neural network is satisfactory while validated with field data.

LSTM Algorithm for Forecasting Events in Changing Electric Consumption

Ebrahim Najafi Kajabad, Sergey V. Ivanov, Ivan Khodnenko
Institute of Design & Urban, ITMO University, Saint Petersburg, Russia

Abstract— The industrial sections have always been thirsty to predict events and be aware of future sudden changes. Where using time-series forecasting based on deep learning techniques can have a significant effect on addressing this issue. In this paper, we present Stacked Long Short-Term Memory (LSTM) networks to predict significant electric consumption changes when a group of devices is activated or deactivated randomly. Three Power quadratic phases are used as a dataset. The goal is to predict multisteps in the future minutes. The supervised learning technique is used to transform series data to supervised data for training and prediction, and then the Walk-forward cross-validation is implemented to evaluate subsequent time steps prediction and output is a vector of prediction. Finally, the root-mean-square error (RMSE) method is employed as a metric for each lead time on an output vector. Furthermore, we compared our proposed model with the Vanilla and Bidirectional LSTMs architecture, in which the performance of our method shows that it was successful. Our experimental results support the view that enough data is crucial to event detection in time series data and suggest that Stacked LSTM is a good architecture with which to address it.

Transfer of Hierarchical Reinforcement Learning Structures for Robotic Manipulation Tasks

Christian Scheiderer, Malte Mosbach, Andres Felipe Posada-Moreno, Tobias Meisen
Institute of Technologies and Management of the Digital Transformation, University of Wuppertal, Germany
Institute of Information Management in Mechanical Engineering, RWTH Aachen University, Aachen, Germany

Abstract—While it is apparent that the transfer of knowledge between tasks is beneficial for training efficiency, the application of trained deep reinforcement learning agents to solve new tasks is not trivial. Especially when tasks are differently structured, retraining and fine tuning is not necessarily beneficial. Instead, it is often the most convenient approach to train a new agent from scratch. One potential solution for effectively reusing learned knowledge may be found in hierarchical reinforcement learning. In this paper we investigate the possibility of reusing low-level policies to improve training efficiency when learning manipulation tasks with an industrial robot. We consider four different scenarios and demonstrate for three of them an increased sample efficiency when training a high-level policy on top of pretrained low-level skills. In the fourth scenario we uncover the reason for a failed transfer to be an ambitious higher hierarchy level enforcing a relearning of the low-level skills.

Self Driving Cars: All You Need to Know

Elizabeth Spoehel, Shankar Banik
Department of Cyber & Computer Sciences, The Citadel, The Military College of South Carolina, Charleston, SC, USA

Abstract—Self-driving cars are coming closer and closer to being fact not fiction, but are we ready for them? In this research we analyze the current status in place for self-driving cars. We address the gaps that need to be filled, and we identify the questions that need to be answered before having selfdriving cars on the road becomes a reality. Towards this, we discuss four issues with self-driving: policies, safety, security, and psychological acceptability of users. Our research will help individuals understand different aspects of self-driving cars and their benefits and challenges. Our paper will educate policymakers on the areas that need to be addressed before we deploy self-driving cars on the roads in a larger scale.

Detection and Removal of Negative Requirements of Deadlock-type in Service-Oriented Architectures

Kenia Santos de Oliveira, Stephane Julia
Federal University of Uberlandia, Federal Institute of Brasilia, Brazil
Federal University of Uberlandia, Brazil

Abstract—This paper presents a method for the detection and removal of negative requirements of deadlock-type in ServiceOriented Architectures (SOA), which is represented by Interorganizational WorkFlow net (IOWF-net). An IOWF-net is formed through the composition of private WorkFlow nets (WFnets) that communicate through asynchronous communication mechanisms. New behaviors may then arise due to the interaction between distinct processes which may eventually lead the system to a deadlock situation. In order to identify the sequence of actions that lead to potential deadlock states, a backward and forward reasoning will be used. In particular, when it comes to the removing of the deadlock states caused by asynchronous message exchanges, synchronous rules are proposed. This approach will be effective in removing conflict situations encountered during the real-time execution of the services provided by the SOA model through immediate diagnostic actions.

Do We Know the Operating Principles of our Computers Better than those of our Brain?

Janos Vegh, Adam J. Berki
Kalimanos BT Debrecen, Hungary
University of Medicine, Pharmacy, Sciences and Technology of Targu Mures, Romania

Abstract—The increasing interest in understanding the behavior of biological neural networks, and the increasing utilization of artificial neural networks in different fields and scales, both require a thorough understanding of how technological computing works. However, von Neumann in his classic ”First Draft” warned that it would be unsound to use his suggested paradigm to model neural operation, furthermore that using ”too fast” processors vitiates his paradigm, which was intended only to describe (the timing relations of) vacuum tubes. Because of this, it is worth to scrutinize, how the present technology solutions can be used to mimic biology. Some electronic components bear a surprising resemblance to some biological structures. However, combining them with each other using different principles can result in systems with very poor efficacy. The paper discusses how the conventional computing principles, components and thinking about computing limit mimicking biological systems.

Weed Segmentation in Sugarcane Crops using Mask R-CNN through Aerial Images

Gabriel Alberto Mini, Daniel Oliva Sales, Maximillian Luppe
Department of Electrical and Computer Engineering (SEL-EESC-USP), University of Sao Paulo, Sao Carlos, Brazil
Institute of Mathematics and Computer Sciences (ICMC-USP), University of Sao Paulo, Sao Carlos, Brazil

Abstract—In this paper we describe an approach to detect weed regions in sugarcane crops using Mask Regions with Convolutional Neural Networks (Mask R-CNN) on aerial images. Experiments were held with twelve combinations using ResNet-50 and ResNet-101 as backbones. We used both “trained from scratch” and transfer-learning from pre-trained datasets combined with image augmentation techniques for training. The ResNet-101 model with pre-trained COCO achieved an Average Precision (AP50) of 65.5% and obtained values of 0.803, 0.707 and 0.752 for Precision, Recall and F1 score. These results suggest a potential use of Mask R-CNN for weed mapping using aerial images.

Black Political Participation in the Volunteer State

Antwain Leach, Sajid Hussain
Political Science, Fisk University, Nashville, Tennessee, USA
Computer Science, Fisk University, Nashville, Tennessee, USA

Abstract— There have been several explanations advanced to account for the unsteady levels of political participation among Blacks in recent years, but rather unfortunately, little research has been conducted into the role local issues may play. National personalities and elongated election seasons inundated with negative campaigning may have “turned-off” potential voters, but as the popular aphorism goes, “all politics are local,” and thus local conditions should be interrogated to help understand voting and other forms of political participation. To this end, this study seeks to probe more deeply into this phenomenon by examining local political participation in the American states. For reasons we outline below, we specifically focus on local political participation in the state of Tennessee (TN) Utilizing data from the 2016 Congressional Cooperative Election Study (CCE), we ask what local-level factors most likely influence political participation among Blacks in TN? In order to parse through the large 2016 CCES dataset more efficiently, we employ software programming tools to address our primary questions using computational thinking and research methods. This technology-driven style of political thinking demonstrates the utility of adopting computing tools and data-driven approaches to identify and address areas of concern across government and civic society. One implication of this work is that it highlights the need for political actors to think more locally when crafting strategies for outreach and engagement towards historically disadvantaged race groups.

Machine Understandable Contracts With Deep Learning

Rares Dolga, Philip Treleaven, Mendes Thame Denny
Department of Computer Science, University College London, London, United Kingdom, UK

Abstract—This research investigates the automatic translation of contracts to computer understandable rules trough Natural Language Processing. The most challenging aspect, which is studied throughout this paper, is to understand the meaning of the contract and express it into a structured format. This problem can be reduced to the Named Entity Recognition and Rule Extraction tasks, the latter handles the extraction of terms and conditions. These two problems are difficult, but deep learning models can tackle them. We think that this paper is the first work to approach Rule Extraction with deep learning. This method is data-hungry, so the research also introduces data sets for these two tasks. Additionally, it contributes to the literature by introducing Law-Bert, a model based on BERT which is pre-trained on unlabelled contracts. The results obtained on Named Entity Recognition and Rule Extraction show that pre-training on contracts has a positive effect on performance for the downstream tasks.

A Hybrid Artificial Intelligence, Machine Learning, and Control Algorithm Integration Framework for Embedded Systems using Semantic Web Technology

Jeffrey Wallace, Angelica Valdivia
Rocket Technology Systems, LLC, Washington, DC, USA

Abstract—A framework to integrate structurally different artificial intelligence, machine learning, and control algorithms is combined with an execution framework to create a powerful embedded system development platform. Control, decision, or algorithms providing an emulation of intelligent behavior in both declarative (interpreted) and imperative (compiled) paradigms can now be combined, for example Prolog and neural networks, respectively. This hybridization of algorithms provides more efficient overall control of systems in terms of resources such as compute cycles, network bandwidth and throughput, and memory speed and capacity. By providing an execution framework and control software that is native to embedded system and cloud architectures, and supports interactivity and time synchronization, the true utility of cloud computing and "big data systems" can be increased.

A Real-Time Traffic Surveillance and Security System using Transfer Learning and Edge Computing

Aaron Joseph Fernandez, Ajay K. S., Antony Jose, Austin Kuruvila M., Varun G. Menon, Vinod P., Xingwang Li, Mohammad R. Khosravi
Department of Computer Science and Engineering, SCMS School Of Engineering and Technology, Ernakulam, India
Department of Computer Applications, Cochin University of Science and Technology, Ernakulam, India
School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, P. R. China
Department of Computer Engineering, Persian Gulf University, Bushehr, Iran

Abstract—Two Wheeler's are one of the most commonly used vehicles worldwide, and the helmet is the most critical gear/safety equipment. There is no proper system that validates whether all the users are equipped with a helmet while they are riding. The existing traffic surveillance and security systems are designed only to identify the vehicles that over speed and jump over the red lights. These systems use interceptors at highways to capture the violators and consume a lot of manpower and vehicles for patrolling. An efficient traffic surveillance system to monitor whether the motorcyclist and the pillion rider is wearing a helmet is missing. This paper proposes an efficient and reliable traffic surveillance system using edge computing and transfer learning that continuously monitors the two-Wheeler's and sends alerts and information on the non-helmet riders to the nearest police vehicles. The edge computing is being applied at the input/entry-level (CCTV camera) and helps in reducing the delay of the entire system. The classification models are created by applying Transfer Learning on ResNet50, VGG16 architectures, and applied to real-time video footage from the CCTV camera to detect non-helmet riders. The proposed system gives an accuracy of 97.50%, reduces the manpower, and enables the police interceptors to work efficiently.

Design of Humanity by the Concept of Artificial Personalities

Taishi Nemoto, Takayuki Fujimoto
Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan

Abstract—Since the 2000s, the third artificial intelligence boom has occurred. Research on machine learning and deep learning is progressing, but challenges remain regarding realizing so-called 'human-like' general-purpose AI (Artificial Intelligence). In recent years, artificial intelligence research has been linked to cognitive science, and the question of what 'humanity' is and how to design 'humanity' has been raised as issues. The more robots resemble humans, the more the 'uncanny valley phenomenon' increases and the more people feel uncomfortable. Even though technology has advanced and realistic textures can be expressed, robots seem to be just 'artifacts.' The point of this study is to determine at what point humans feel 'human-like' and how to reproduce 'human-like' using a computer. In this study, in order to express human personality and characteristic gestures, we generate an 'artificial personality' (AP), and let people find the human touch that a real person possesses through that AP. For example, artificial reproduction of the intelligence of a deceased person is difficult with today's technology. However, AP enables to extract the characteristics of a person's gestures, routines, habits, and facial expressions in his or her lifetime, and to digitally recreate the person's personality based on the accumulation of 'flesh and blood' data. This study discusses the two elements and basic mechanisms that are necessary for AP research.

Forecasting Method based upon GRU-based Deep Learning Model

Ali Jaber Almalki, Pawel Wocjan
Department of Computer Science, University of Central Florida, Orlando, Florida, USA

Abstract—In this research, the world model has a modified RNN model carried out by a bi-directional gated recurrent unit (BGRU) as opposed to a traditional long short-term memory (LSTM) model. BGRU tends to use less memory while executing and training faster than an LSTM, as it uses fewer training parameters. However, the LSTM model provides greater accuracy with datasets using longer sequences. Based upon practical implementation, the BGRU model produced better performance results. In BGRU, the memory is combined with the network. There is no update gate and forget in the GRU. The forget and update gate are treated as one unit thus it is the primary reason of parameter reduction.

Machine Learning for Dense Crowd Direction Prediction Using Long Short-Term Memory

Abdullah Alajlan, Alaa Edris, Frederick Sheldon, Terence Soule
Computer Science Department, University of Idaho, Moscow, Idaho, USA
Computer Science Department, Technical and Vocational Training Corporation Riyadh, Saudi Arabia
Computer Science Department, University of Jeddah, Jeddah, Saudi Arabia

Abstract—The safety of a dense crowd is one of the most important matters for an event's organizers. Therefore, management of the crowd, and noticing any serious issues in advance becomes important. Developing a crowd simulation by using a social force model simulates the behavior of crowds in reality. The prediction of individual agents' behavior in the simulation and how the agents interact with each other can improve the safety of dense crowds. Depending on the success of Recurrent Neural Network(RNN) handling of sequential data, we propose a model that is based on Long Short Term Memory (LSTM) to predict individual agents' next locations. Our proposed approach will be tested two different densities of crowds, structured crowds, and unstructured crowds. In structured crowds, people generally move in one direction and head to the same destination, such as at the Islamic Hajj. In unstructured crowds people move in many different directions and head for different destinations, such as in public town squares.

Accuracy-aware Structured Filter Pruning for Deep Neural Networks

Marina Villalba Carballo, Byeong Kil Lee
Department of Electrical and Computer Engineering, University of Colorado, Colorado Springs, Colorado, USA

Abstract— Deep neural networks (DNNs) have several technical issues on computational complexity, redundancy, and the parameter size – especially when applied in embedded devices. Among those issues, lots of parameters require high memory capacity which causes migration problem to embedded devices. Many pruning techniques are proposed to reduce the network size in deep neural networks, but there are still various issues that exist for applying pruning techniques to DNNs. In this paper, we propose a simple-yet-efficient scheme, accuracy-aware structured pruning based on the characterization of each convolutional layer. We investigate the accuracy and compression rate of individual layer with a fixed pruning ratio and re-order the pruning priority depending on the accuracy of each layer. To achieve a further compression rate, we also add quantization to the linear layers. Our results show that the order of the layers pruned does affect the final accuracy of the deep neural network. Based on our experiments, the pruned AlexNet and VGG16 models' parameter size is compressed up to 47.28x and 35.21x with less than 1% accuracy drop with respect to the original model.

Artificial Intelligence for Computerized Adaptive Testing

Dena F. Mujtaba, Nihar R. Mahapatra
Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan, USA

Abstract—Artificial intelligence (AI) is increasingly used to provide customized and efficient e-learning, job search, and career development assistance to students and workers. Both students and jobseekers encounter assessments several times throughout their career and during job searches. Recently, organizations and academic institutions have employed computer adaptive testing (CAT), a computer-administered assessment that serves questions based upon the ability of the test taker. CAT aims to provide personalized assessments to test takers to accurately measure their proficiency (i.e., latent trait). There are several challenges in CAT, such as estimating the latent traits of an individual, generating questions, and question selection. Furthermore, these challenges become more complex as the number of latent trait dimensions being measured increases or if item responses are categorical rather than binary (e.g., using a 1 to 5 scale versus true or false). Traditional approaches employ psychometric and statistical models to make predictions. However, many approaches using machine learning, deep learning, and other AI techniques have emerged to provide better performance. In this paper, we provide a technique-oriented review of AI applications in CAT, and highlight the advantages, limitations, and future challenges in this problem area. We also reconcile different terms and notations used across psychometrics and AI to assist future research and development.

Mining and Analyzing Occupational Characteristics from Job Postings

Dena F. Mujtaba, Nihar R. Mahapatra
Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan, USA

Abstract—Hiring/recruitment is key to an organization's ability to position itself for success by attracting the right talent. Similarly, job search enables workers to connect to the right jobs in the right organizations. To assist in the hiring and job search processes, many technology solutions such as interest inventories, job recommendation models, job boards, and career pathway planning tools have been developed. However, solutions for preparing job postings are lacking. Job postings/ads play an essential role in hiring the right talent since they signal to the job seeker the knowledge, skills, abilities, and other occupationrelated characteristics (KSAOs) needed for a job. If the job ad does not convey the correct occupational characteristics, it is less likely that a well-qualified candidate will apply. Therefore, we present an interactive job ad visualization tool that analyzes the text in a job ad and matches phrases in it to a large occupational taxonomy of KSAOs. We combine O*NET, an occupational taxonomy, with natural language processing to perform semantic similarity matching between KSAOs for an occupation and ad text, and thereby assist job seekers in their search process and recruiters in preparing the job ad.

Sensor-Based Air Pollution Prediction Using Deep CNN-LSTM

Kabir Nagrecha, Pratyush Muthukumar, Emmanuel Cocom, Jeanne Holm, Dawn Comer, Irene Burga, Mohammad Pourhomayoun
Department of Computer Science, California State University Los Angeles, California, USA
City of Los Angeles, OpenAQ, California, USA

Abstract—The devastating impacts of air pollution have become more and more evident in recent years. As our measurement technologies improve, we gain better insight into the true impact of this deadly, yet often ignored, threat. The first step in reducing the damages caused by this problem is being able to analyze and predict its patterns. The problem of predicting air quality and the presence of particulate matter lies in the nature of the data needed to create an accurate system. The sheer number of factors affecting air quality mean that previously proposed approaches often utilize a great many sources of data, aiming to incorporate images, wind graphs, traffic information, and more. Yet in truth, most areas outside large metropolises lack ready access to high-quality data, preventing them from ever implementing an effective system. We propose a system utilizing a 1-D deep convolutional neural network to analyze past sensor readings and predict air pollutant concentrations up to a day in the future at a 3-hour resolution. We specifically developed this model for predicting PM2.5 values. The system receives PM2.5 sensor values and discovers temporal pattern in the data, which will be later used for prediction. By removing the dependency on complex data inputs, the system becomes accesible and easily implementable for any region. Despite this simplified approach, the results are comparable to — and often better than — any current state-of-the-art predictive systems in this domain.

Explanation Generation in a Kabuki Dance Stage Performing Structure Simulation System

Miku Kawai, Jumpei Ono, Takashi Ogata
Faculty of Software and Information Science, Iwate Prefectural University, Takizawa, Iwate, Japan
Faculty of Software and Information Technology, Aomori University, Edogawa City, Tokyo, Japan

Abstract—We are developing a stage performing structure simulation system. This paper introduces the generation of explanations into the system. First, we classified the explanations we collected. Thereafter, we prototyped a mechanism wherein the system automatically determines the content and method of an explanation based on arbitrary parameters. Through this trial, we will examine effective explanation methods for story generation.

Multi-Environmental Parameters Dashboard for Susquehanna River Basin using Machine Learning Techniques

Siamak Aram, Maria H. Rivero, Nikesh K. Pahuja, Roozbeh Sadeghian, Joshua L. R. Paulino, Michael Meyer, James Shallenberger
Harrisburg University of Science and Technology, Harrisburg, PA, USA
The Susquehanna River Basin Commission, Harrisburg, PA, USA

Abstract—Data-driven decision making is essential for environmental monitoring and protection of natural resources. The Susquehanna River Basin Commission (SRBC) collects water quality data to monitor the condition of the Susquehanna River Basin. The practice of water quality management can be evolved introducing a Geographic Information Systems dashboard that incorporates Artificial Intelligence models to predict environmental parameters in real-time. This study presents an operational dashboard that integrates Machine Learning model using the data collected from the SRBC's 62 different model sites from 2010 to 2019. Five daily time-series parameters were selected such as temperature, turbidity, pH, specific conductance, and dissolved oxygen. The results found the Random Forest as the best performance model to predict the specific conductance of water. These results outline an unprecedented tool for water quality management.

SYMPOSIUM ON ARTIFICIAL INTELLIGENCE BASED DEEP VIDEO DATA ANALYTICS

Chair: Dr. Charlie (Seungmin) Rho
Department Software, Sejong University, Gwangjin-gu, Seoul, Republic of Korea

An Activity Recognition Framework for Overlapping Activities using Transfer Learning

Muhammad Bilal, Muazzam Maqsood, Irfan Mehmood, Mubashir Javaid, Seungmin Rho
Sejong University, Gwangjin-gu, Seoul, Republic of Korea

Abstract— Activity recognition is gaining popularity with the increase in digital content. In video data, there is a lot of information hidden that needs to be explored. Human Activity Recognition (HAR) in video streams applies to many areas, such as video surveillance, patient health monitoring, and behavior analytics. Variation in the environment, view-point changes, occlusion, and illumination are some main challenges in HAR. Among other challenges, there is also a similar activity or overlapping activity issue that has not been explored much in past. Resolving overlapping activities classes issues can be a major contribution to overall Human Activity Recognition. Hand-crafted methods and traditional Machine Learning methods were extensively explored in past. Recently, many Deep Learning-based methods are achieved high accuracy. Convolutional Neural Network (CNN) and 3D CNN methods outperform other methods. In this paper, we proposed a Transfer Learning-based Human Activity Recognition (TLHAR) for video data streams. We used VGG16 and InceptionV3, two pre-trained CNN models, and utilized their prior training knowledge for efficient activity recognition. The purposed system outperformed existing activity recognition methods and showed state-of-the-art accuracy and less computational cost requirements than other techniques by taking the benefits of Transfer Learning.

An Efficient Liver Tumor Detection using Machine Learning

Anum Kalsoom, Anam Moin, Muazzam Maqsood, Irfan Mehmood, Seungmin Rho
Sejong University, Gwangjin-gu, Seoul, Republic of Korea

Abstract—Liver Cancer is among the most commonly diagnosed diseases in today's modern era. Liver tumor segmentation is a fundamental task to perform early diagnosis and recommend a treatment. Manual segmentation is the traditional approach to achieve the required results but it has always been a time-consuming process. There are some anomalies like ambiguous gray level color ranges similar to other neighboring organs, irregular tumor shapes, and various uneven tumor sizes which are overlooked. Due to these reasons, some semi-automated and even fully automated techniques have been put forward. However, the advancement of machine learning has been very accommodating for addressing this issue. In this paper, we propose an unsupervised machine learning technique combined with a supervised mechanism that accurately performs liver tumor segmentation. We perform clustering on our collected dataset and extract LBP features as well as HOG features from these clusters. Furthermore, we perform classification which is based on these extracted features using KNN. Furthermore, we have compared our results with two classifiers namely SVM and Ensemble to achieve a better understanding. Our proposed technique outperformed existing techniques and showed encouraging results when compared to other methods.

A Solution to Combined Economic Emission Dispatch (CEED) Problem using Grasshopper Optimization Algorithm (GOA)

Hina Bibi, Aftab Ahmad, Farhan Aadil, Irfan Mehmood, Mucheol Kim
School of Computer Science and Engineering, Chung-Ang University, Dongjack-gu, Seoul, Republic of Korea

Abstract— The Combined Economic Emission Dispatch (CEED) is a big issue of optimization of the power system service preparation. CEED 's working approach is to determine the best strategy for active generation of engaged units while taking into account both the objectives of reducing cost generation and emission content. This is highly complicated and multi constrained

problem. In this paper the solution of CEED are presents by using Grasshopper Optimization Algorithm (GOA). The efficacy of this algorithm was examined by its implementation on three separate standard IEE test system units, such as IEEE three-unit system, IEEE six-unit system and IEEE ten-unit system, the results of which are compared with other metaheuristic techniques. The distinguished feature of comfort zone operator of GOA assists in extracting stupendous simulations results of minimized fitness of multi objective functions illustrating the efficiency of GOA in term of highly optimal and feasible solution satisfying all the system equality and inequality operational constraint.

A Trust Assisted Matrix Factorization based Improved Product Recommender System

Hina Bibi, Asma Rahim, Muazzam Maqsood, Irfan Mehmood, Mucheol Kim
School of Computer Science and Engineering, Chung-Ang University, Dongjack-gu, Seoul, Republic of Korea

Abstract—Smart services is an efficient concept to provide services to the citizen in an efficient manner. The online shopping and recommender system play an important role in this scenario that provides efficient item recommendations to the citizens. Though, the majority of the latest recommender systems can't get effective and efficient prediction accuracy because of the sparsity of the item matrix against each user. Additionally, the recommendations are not reliable when tested upon larger datasets. To handle these problems, a trust-based technique is proposed, called trustasvd++, which fuses a user's trust data in the MF context. The offered strategy combines trust data and rating values to deal with the sparsity and cold start user's issues. Matrix Factorization (MF) has been recognized as a persuasive method for the formation of an effective Recommender System. Pearson correlation coefficient (PCC) is used as a similarity metric in the proposed technique. To assess the efficiency of the offered strategy, numerous datasets have been done on datasets including Epinions, Filmtrust, and Ciao.

Behavior-based Outlier Detection for Indoor Environment

Shinjin Kang, Soo Kyun Kim
Department of Computer Engineering, Jeju National University, Jeju-si, Republic of Korea
Jeju Special Self-Governing Province, Republic of Korea

Abstract - In this paper, we introduce a system that can detect the space outlier utilization of residents in indoor environment at low cost. Our system facilitates autonomous data collection from mobile app logs and the Google app server and generates a high-dimensional dataset required to detect outlier behaviors. For this, we used density-based clustering algorithm with t-distributed stochastic neighbor embedding (t-SNE). Our system enables easy acquisition of large volumes of data required for outlier detection. Our methodology can assist spatial analysis for indoor environments housing residents and help reduce the cost of space utilization feedback.

Artificial Intelligence with Wireless Sensor Network for Fire Detection

Faisal Saeed, Anand Paul, Seungmin Rho, Sangsoon Lim
Department of Computer Engineering, Sungkyul University, Anyang, Republic of Korea

Abstract— Artificial intelligence techniques such as intelligent search and intelligent agent is very appealing research for disaster surveillance such as Fire. A method for fire response is developed using the AI techniques. This is achieved by focusing wireless sensor networks as well as convolutional neural networks with Intelligent agents. The result analysis of this work is quite efficient.

Detection of Plant Diseases in the Images using Deep Neural Networks

Malik Urfa Gul, Anand Paul, Seungmin Rho, Sanghyun Seo
School of Computer Art, Chung-ang University, Anseong, Republic of Korea

Abstract— Agriculture suffers from crop diseases, and losses yield every year. Early detection of crop diseases can effectively decrease the loss. Leaves from crops are affected by the disease and can help farmers to detect any changes. Our study uses crops labelled dataset to train the Faster-RCNN model to identify if leaves are affected by any means. Our study shows more than 97% accuracy to detect disease in early stages that framers were unable to do in the past.

Blockchain based Healthcare System with Artificial Intelligence

Malik Junaid Jami Gul, Anand Paul, Seungmin Rho, Mucheol Kim
School of Computer Science and Engineering, Chung-Ang University, Dongjack-gu, Seoul, Republic of Korea

Abstract—Blockchain is not only for the financial domain anymore. It evolves to accommodate a broad range of domains and applications where trust and privacy are required for smooth transitions. As blockchain evolving, researchers are automating the blockchain tasks for better security and performance. The blockchain management system monitors specific tasks like transaction management, consensus, block security, and blockchain network security. Blockchain management system comprises of specialized engineers and blockchain software platform. This study explores the opportunity for Machine learning concepts to work with blockchain system management to automate tasks in the healthcare scenario. Reinforcement learning is used in this study to automate blockchain tasks with multiagents. Our study found that agents can be trained and perform tasks listed under the healthcare system's blockchain management system. Our study also suggests that storing and accessing data is efficient with machine learning concepts.

A 3D Real Object Recognition and Localization on SLAM based Augmented Reality Environment

Jongin Choe, Sanghyun Seo
School of Computer Art, College of Art & Technology, Chung-Ang University, Anseong, Republic of Korea

Abstract— this paper introduces a method for recognizing real world objects in AR (Augmented Reality) environment and visualizing virtual information based on the objects. Existing AR shows high dependence on markers. The use range of the virtual space is limited to narrow marker space and the placement and tracking of virtual objects in a 3D is also limited to the space centered on the marker. The method constructs a map of the space in the form of a point cloud using SLAM, and performs real-world object recognition of the constructed wide AR space in real time. As a result, the degree of freedom of space utilization increases, and it is helpful in applying AR such as information display of real-world objects by accurately recognizing and localizing the location of objects existing in the real-world space.

A Study on the Prediction of Emotion from Image by Time-flow using Color Analysis

Taemin Lee, Nahyuk Lee, Sanghyun Seo, Dongwann Kang
Department of CSE, Seoul National University of Science and Technology, Seoul, Republic of Korea

Abstract— When people are looking at a landscape, we feel different depending on time/season. This part is based on the colors of the landscape, even if you look at the same landscape due to the visual elements. We wanted to analyze human emotions, especially based on color, among these visual elements. For this purpose, machine learning models were established according to color and human emotion changes were analyzed quantitatively over time. Finally, we analyzed how much color affects human emotions.

ECG Signal Analysis for Patient with Metabolic Syndrome based on 1D-Convolution Neural Network

Chhayly Lim, Jung-Yeon Kim, Yunyoung Nam
Department of ICT Convergence Rehabilitation Engineering, Soonchunhyang University, Cheonan-si, Republic of Korea

Abstract— Metabolic syndrome (MetS) is a cluster of metabolic disorders associated with medical conditions: abdominal obesity, high blood pressure, insulin resistance, etc. People with MetS have a higher risk of cardiovascular diseases and type 2 diabetes mellitus. Hence, early detection of MetS can be useful in the field of healthcare. In this paper, we propose a 1D-Convolution Neural Network (1D-CNN) model for classifying the electrocardiogram (ECG) signals of the GBBANet online database into two classes: a group of people with the medical condition (MetS [n=15]) and a control group (CG [n=10]). The dataset consists of 5 ECG recordings per person. The proposed 1D-CNN model has achieved an overall accuracy of 88.32%.

Quality Evaluation of Fundus Images using Transfer Learning

Kuntha Pin, Jung-Yeon Kim, Yunyoung Nam
Department of ICT Convergence Rehabilitation Engineering, Soonchunhyang University, Cheonan-si, Republic of Korea

Abstract— Fundus images with poor quality may contain insufficient information to represent characteristics of eye diseases, such as diabetic retinopathy (DR), glaucoma, and agerelated macular degeneration (AMD). This negatively affects the performance of automated systems. In this paper, we present a method, which is based on the transfer learning technique with Inception-v3 model, to evaluate the quality of fundus images and eliminate fundus images that are classified as having poor quality. The original Inception-v3 model was modified by adding more layers at the classification layer of the model to adapt to quality fundus image evaluation. The experiment performed acceptable results of quality evaluation for fundus images into two categories: good and poor-quality images. Therefore, the proposed method can assess the quality of the image by identifying the contributing factors such as over dark and over bright.

A User-Centric Intelligent Context-Aware System for Realizing Internet of Things Environments

Kwanhee Kim, Sangoh Park
Department of Computer Science & Engineering, Chung-Ang University, Dongjak-gu, Seoul, Republic of Korea

Abstract— In the Internet of Things (IoT), platforms are being built that are capable of developing various services and delivering services to users through connection technology between various objects and sensors. However, the information provided on these platforms focuses on the state of objects and environments. It predicts the user's situation and does not provide intelligent services suitable for the current situation at the user's location. Future IoT needs a service that senses changes in various environments, controls each object, and recognizes a user-centered seamless situation. This paper proposes a dynamic mesh-type network configuration technique for analyzing the real-time situation of users and providing seamless services. We propose a mechanism to understand the user behavior and the surrounding environment through a network configured to recognize a user-centered situation. The parking lot system was simulated to evaluate the performance of the situational awareness system. The method applied to the parking lot system was compared with the best effort method and the situational awareness system proposed in this paper.

Dr. Answer AI Software for Prostate Cancer: Explainable Variable Importance of Predicting T Stage

Mi Jung Rho, Jihwan Park, Hyong Woo Moon, Jaewon Kim, Chanjung Lee, Choung-Soo Kim, Seong Soo Jeon, Minyong Kang, Ji Youl Lee
Department of Urology, Seoul St. Mary's Hospital, College of Medicine, Catholic University of Korea, Republic of Korea
LifeSemantics, Seoul, Republic of Korea
Department of Urology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
Department of Urology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Abstract— Dr. Answer AI software project for prostate cancer was conducted in South Korea. One of the things we need to develop an AI software model is creating a model that can be explained. We calculated the importance of variables of predicting T stage before developing the Dr. Answer AI software model and used it. We collected 7,128 cases clinical data of prostate cancer patients after the radical prostatectomy treatment: 1,723 were from 2008 to 2017. It is from hospital C, 2,751 from hospital S, and 2,654 from hospital A. We used the random forest algorithm to calculated the importance of variables. We choose the SMOTE+ENN to handle imbalance data sets. Accuracy with SMOTE+ENN is 90.7%. The top important variables in the T stage prediction were 1) initial PSA, 2) BMI, 3) max positive core count, 4) Gleason group, and 5) core ratio. We provide top important variables in T stage prediction to develop an AI software is creating a model that can be explained. Our research can provide guidelines in developing AI SW for prostate cancer.

CSCI-ISHI:
HEALTH INFORMATICS AND MEDICAL SYSTEMS

DRDr II: Detecting the Severity Level of Diabetic Retinopathy Using Mask RCNN and Transfer Learning

Farzan Shenavarmasouleh, Farid Ghareh Mohammadi, M. Hadi Amini, Hamid R. Arabnia
Department of Computer Science, University of Georgia, USA
School of Computing & Information Sciences, solid lab, Florida International University, Miami, Florida, USA

Abstract— DRDr II is a hybrid of machine learning and deep learning worlds. It builds on the successes of its antecedent, namely, DRDr, that was trained to detect, locate, and create segmentation masks for two types of lesions (exudates and microaneurysms) that can be found in the eyes of the Diabetic Retinopathy (DR) patients; and uses the entire model as a solid feature extractor in the core of its pipeline to detect the severity level of the DR cases. We employ a big dataset with over 35 thousand fundus images collected from around the globe and after 2 phases of preprocessing alongside feature extraction, we succeed in predicting the correct severity levels with over 92% accuracy.

A Fast and Accurate Myocardial Infarction Detector

Harold Martin, Walter Izquierdo, Ulyana Morar, Malek Adjouadi
Department of Electrical and Computer Engineering, Florida International University, Miami, Florida, USA

Abstract—We propose a novel pipeline for the real-time detection of myocardial infarction from a single heartbeat of a 12- lead electrocardiograms. We do so by merging a real-time R-spike detection algorithm with a deep learning Long-Short Term Memory (LSTM) network-based classifier. A comparative assessment of the classification performance of the resulting system is performed and provided. The proposed algorithm achieves an inter-patient classification accuracy of 95.76% (with a 95% Confidence Interval (CI) of ±2.4%), a recall of 96.67% (±2.4% 95% CI), specificity of 93.64% (±5.7% 95% CI), and the average J-Score is 90.31% (±6.2% 95% CI). These state-of-the-art myocardial infarction detection metrics are extremely promising and could pave the wave for the early detection of myocardial infarctions. This high accuracy is achieved with a processing time of 40 milliseconds, which is most appropriate for online classification as the time between fast heartbeats is around 300 milliseconds.

A Deep-Learning Approach for the Prediction of Mini-Mental State Examination Scores in a Multimodal Longitudinal Study

Ulyana Morar, Harold Martin, Walter Izquierdo, Parisa Forouzannezhad, Elaheh Zarafshan, Rosie E. Curiel, Monica Roselli, David Loewenstein, Ranjan Duara, Malek Adjouadi
Department of Electrical and Computer Engineering, Florida International University, Miami, Florida, USA
Department of Psychiatry & Behavioral Sciences, University of Miami, Florida, USA
Department of Psychology, Florida Atlantic University, Boca Raton, Florida, USA
Wien Center of Alzheimers Dis & Memory Disorders, Mt Sinai Med Center, Miami, Florida, USA

Abstract— This study introduces a new multimodal deep regression method to predict cognitive test score in a 5-year longitudinal study on Alzheimer's disease (AD). The proposed model takes advantage of multimodal data that includes cerebrospinal fluid (CSF) levels of tau and beta-amyloid, structural measures from magnetic resonance imaging (MRI), functional and metabolic measures from positron emission tomography (PET), and cognitive scores from neuropsychological tests (Cog), all with the aim of achieving highly accurate predictions of future Mini-Mental State Examination (MMSE) test scores up to five years after baseline biomarker collection. A novel data augmentation technique is leveraged to increase the numbers of training samples without relying on synthetic data. With the proposed method, the best and most encompassing regressor is shown to achieve better than state-of-the-art correlations of 85.07%(SD=1.59) for 6 months in the future, 87.39% (SD =1.48) for 12 months, 84.78% (SD=2.66) for 18 months, 85.13% (SD=2.19) for 24 months, 81.15% (SD=5.48) for 30 months, 81.17% (SD=4.44) for 36 months, 79.25% (SD=5.85) for 42 months, 78.98% (SD=5.79) for 48 months, 78.93%(SD=5.76) for 54 months, and 74.96% (SD=7.54) for 60 months.

An Edge Computing Based Situation Enabled Crowdsourcing Blacklisting System for Efficient Identification of Scammer Phone Numbers

Chen-Yeou Yu, Carl K. Chang, Wensheng Zhang
Department of Computer Science, Iowa State University, Ames, Iowa, USA

Abstract — The growth of telecommunication fraud has caused tremendous loss to end users. In particular, new technologies such as robocalling systems have been a new resource of harassment. Traditional approaches in detecting such activities simply rely on the construction of blacklisting number systems. However, criminals can easily masquerade their phone numbers simply by changing their numbers through VoIP (Voice over IP) or use virtual mobile numbers (VMN) with relatively low pricing, laxed ID checks and high-level API automation. In this paper, we present a novel situation-enabled approach to blacklist unwanted phone numbers while keeping high detection rate through distributed crowd sourcing. The system consists of two parts. First, we collect a user's daily schedule in time series as situational data and use the data to train Long Short Term Memory (LSTM) deep learning model. This model is used to predict the user's situation in the future. Then, we implement a semi-automatic tagging application to tag each incoming call by reading the call history against the predicted situation. An incoming phone number can be automatically tagged as malicious if it is in a wrong situation or could be benign otherwise. A user is also allowed to manually change the tagging afterwards if it is necessary. Second, a distributed crowdsourcing is used to aggregate highly ranked calling numbers from different devices in. When a higher-level blacklist has been built, it can be used to update local ones by propagating back to end user devices with edge local blacklist and edge foreign blacklist. A simple evaluation has been made against real incoming calls on Android phones. The results show that our system design can attain decent detection rates.

Empirical Analysis of Thermography Effectiveness for Health Diagnosis

Trasha Gupta, Rajni Jindal, S. Indu
Delhi Technological University, Delhi, India

Abstract - The lifestyle led by today's generation leads to the development of many chronic diseases. Such diseases rank amongst the first ten causes of death. To increase life's expectancy, a technology like thermography is required to monitor the health and identify the early signs of abnormality in the body. In this paper, we have studied the effectiveness of Thermography with Artificial Intelligence-based techniques for three different diseases and achieved an accuracy of 91.42%, 93.47%, 88.57% for DMR-IR, DB-FOOT-PLANTAR, TH-NOD-DB, respectively.

Time Series Forecasting of COVID-19 Infections in United Arab Emirates using ARIMA

Leila Ismail, Shaikhah Alhmoudi, Sumyah Alkatheri
Distributed Computing and Systems Research Laboratory, College of Information Technology, United Arab Emirates University, Al-Ain, Abu Dhabi, United Arab Emirates, UAE

Abstract— Machine learning time series models have been used to predict COVID-19 pandemic infections. Based on the public dataset from Johns Hopkins, we present a novel framework for forecasting COVID-19 infections. We implement our framework for the United Arab Emirates (UAE) and develop autoregressive integrated moving average (ARIMA) time series forecast model. To the best of our knowledge, this is the only study to forecast the infections in UAE using the time series model.

A Scoping Review of Clinical Unstructured Text Information Extraction

Xiaoyuan Bao, Shuangrong Fan, Kai Zhang
Medical Informatics Center National Health Service Data Center, Peking University Health Science Center, Beijing, P. R. China
School of Nursing, Peking University Health Science Center, Beijing, P. R. China
Peking University People's Hospital, Peking University Health Science Center, Beijing, P. R. China

Abstract—Extracting information from Electronic Health Records (EHRs) and applying it to clinical research and hospital management has been very effective. Compared with the capacity of Chinese EHRs data, there is still much work in research and application to be carried out on Chinese information extraction. Helping clinical professionals understand the progress of Information Extraction(IE) at home and abroad is intention of this scoping review. We obtain research articles from five databases including IEEE, MDELINE etc., and retain 189 articles screened by two authors. We analyze IE from five aspects, problem definition, data preprocessing, knowledgebase establishment, method training and testing, method practical application and upgrade iteration. Finally, we focus on summarizing and analyzing the state of art in Chinese information extraction works.

Healthcare Big Data Normalization Graph Theory Implementation

Atif Farid Mohammad, Peter Bearse, Intisar Rizwan I. Haque
Ontrak Inc., Santa Monica, California, USA

Abstract— This paper presents Healthcare Big Data Normalization using Computerized Provider Order Entry (CPOE) and application of Graph Theory. This is the process of entering physician orders directly into an electronic health record (EHR). CPOE replaces traditional pen and paper, email, fax, and telephone ordering methods. CPOE is an integral part of electronic medical records and a mandatory component for achieving Meaningful Use Stage 2 certification in health care. CPOE is vital because it helps reduce medical errors that can lead to morbidity and mortality and lowers health care costs. Relational databases are the most common type of database used in healthcare settings. The advantages of using a Relational Database Management System for CPOE are discussed, as well as the disadvantages. The Entity-Relationship diagram and schema for a medication CPOE system used in a small ambulatory medical clinic are provided. We also briefly discuss the potential use of a CPOE application and a NoSQL Open Source database, such as OrientDB, along with the benefits and challenges.

Linear Discriminant Analysis Applied to the Detection of Allergic Rhinitis in Patients

Gregory Stainhaouer, Stelios Bakamidis, Ioannis Dologlou
Institute for Language and Speech Processing ILSP/ R.C, "Athena", Athens, Greece

Abstract —This paper presents a system to detect symptoms of allergic rhinitis remotely by using uttered speech and by exploiting its specific spectral characteristics. Based on the principles of adaptive modeling and fundamental frequency variations (jitter) as well as speech analysis by means of acoustic models, the proposed technique achieves an efficient classification of patients from uttered speech using a Linear Discriminant Analysis (LDA) algorithm. A Singular Value Decomposition (SVD) based iterative approach is used for the accurate estimation of the jitter and Hidden Markov Models (HMM) are implemented to model the 32 phonemes. The final decision is derived by optimally combining the individual estimates to form vectors that are processed by an LDA based iterative algorithm providing better clustering of healthy and allergic subjects.

A Comparative Study on Machine Learning Algorithms for Predicting Breast Cancer Prognosis in Improving Clinical Trials

Neetu Sangari, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract— In recent years, machine learning algorithms have been more and more used in healthcare industry, especially in research areas involving human participants such as clinical trials and areas where the data is too expensive to collect. This research

project has conducted a comparative study on three well-known machine learning methods: Logistic Regression (LR), Support Vector Machine (SVM), and Naïve Bayes (NB) against the same dataset for predicting breast cancer prognosis in improving clinical trials. The experiment results have provided a comprehensive view of the patient's risk levels and risk factors to clinicians that benefit in effective and efficient treatment. This research has also demonstrated that different machine learning algorithms against the same dataset for breast cancer prognosis can have a difference in both performance and accuracy. Therefore, the comparative study on different machine learning algorithms can identify the most suitable machine learning algorithm to achieve cost-effective clinical trials.

Multiple Ways for Medical Data Visualization Using 3D Slicer

Ismail Mohammed Bahkali, Sudhanshu Kumar Semwal
Department of Computer Science, University of Colorado, Colorado Springs, Colorado, USA

Abstract— Computers can process large amounts of data. Medical practitioners can deliver better services and provide more accurate diagnoses and treatment regimens to patients. This document described how 3D Slicer allows Command Line Interface (CLI), Python, Jupyter, and MATLAB in software to process medial data. 3D Slicer has become useful software worldwide since 1997, especially in the medical field for preoperative visualization and analysis. Today, 3D Slicer is supported by The National Alliance for Medical Imaging Computing (NA-MIC), Neuroimaging Analysis Center (NAC), Biomedical Informatics Research Network (BIRN), The National Center for Image-Guided Therapy (NCIGT), The Harvard Clinical and Translational Science Center (CTSC), and the Slicer Community worldwide as a platform to develop new ideas. In this paper, we demonstrate our knowledge in using the 3D Slicer software.

Blockchain-based Enterprise Architecture for Comprehensive Healthcare Information Exchange (HIE) Data Management

Kofi Osei-Tutu, Shirin Hasavari, Yeong-Tae Song
Department of Computer and Information Sciences, Towson University, Towson, Maryland, USA

Abstract - Timely clinical data exchange has been the primary function of Health Information Exchanges (HIEs). However, there have been growing needs for sharing public health related data among government entities. Because the primary stakeholders for public health data are different from clinical data, data are stored over multiple systems, making the retrieval of related public health data items a hard process. In such situations, blockchain technology may be used to connect and securely exchange those scattered data items. In this paper, we proposed an enterprise architecture (EA) for HIEs to adopt blockchain technology so they could store and exchange both clinical and other health related data for efficient general public healthcare management. With the proposed solution, HIEs can not only exchange relevant data items, but also securely manage, record, and transfer complete health records. In our approach, we have incorporated a Hyperledger Fabric (HLF) Network blockchain technology that utilizes Blockchain as a Service (BaaS) to enterprise architecture so that HIEs can manage comprehensive health data.

Clustering County-Wise COVID-19 Dynamics in North Carolina

Man Sik Park, Seong-Tae Kim
Department of Statistics, Sungshin Women's University, Seoul, Korea
Department of Mathematics and Statistics, North Carolina A&T State University, Greensboro, North Carolina, USA

Abstract - The COVID-19 pandemic has caused an enormous number of confirmed cases and deaths in the U.S. This study aims to identify hidden clusters among counties in North Carolina using the COVID-19 data. Since individual states implement their own policies to cope with the COVID-19 pandemic, our study is limited to a single state, North Carolina. We incorporated two clustering techniques, dynamic time warping and deep learning autoeconder. These clustering techniques identified similar upper-level hierarchical clusters separating three metropolitan areas and other regions with slightly different sub-clusters in the countywise COVID-19 data. Our findings further understanding of county-wise COVID-19 dynamics and its implication.

A Comparative Study of N-gram and Skip-gram for Clinical Concepts Extraction

Susan Sabra, Vian Sabeeh
Department of Data Science, Eurecom, Biot, France
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA

Abstract—State-of-the-art technologies for clinical knowledge extraction are essential in a clinical decision support system (CDSS) to make a prediction of a diagnosis. Automatic analysis of a patient's health data is a requirement in such a process. The unstructured part of the data in electronic health records (EHR) is critical, as it may contain hidden risk factors. We present in this paper a comparative study of two well-known techniques N-gram and Skip-gram to enhance the extraction of risk factors concepts from the clinical narratives after applying initial natural language processing (NLP) techniques. We evaluate the use of both techniques using a case study dataset of patients' records with venous thromboembolism (VTE). Results of the techniques' comparative study yielded an advancement of N-gram precision while Skip-gram produced a better performance in terms of the recall measure.

COVID-19 Fuzzy Inference System

Fatema Alhammadi, Fatima Alkhanbashi, Maad Shatnawi
Department of Electrical Engineering Technology, Higher Colleges of Technology, Abu Dhabi, UAE

Abstract — The Coronavirus COVID-19 has been considered a pandemic due to its rapid spread increasing the number of affected cases and causing severe health issues and deaths all over the world. Meanwhile no particular treatment or vaccination has been identified for this disease, and therefore, the initial and early identification is crucial to control and break down the chain of COVID-19. In this research, a smart fuzzy inference system is proposed for initial identification of COVID-19 based on the patient symptoms and travel and contact history. The symptoms include cold, cough, fever, flu, breathing difficulties, throat pain and headache. Based on a particular patient data, the proposed system predicts the severity level of the disease that he/she has.

Workflow-based Anomaly Detection using Machine Learning on Electronic Health Records' Logs: A Comparative Study

Prosper K. Yeng, Muhammad Ali Fauzi, Bian Yang
Department of Information Security and Communication Technology, Norwegian University of Science and Technology, NTNU, Gjovik, Norway

Abstract—Timely access to patients' healthcare records is very essential. As a result, broad access to EHR is mostly provided to users in efforts towards complying with the availability trait of the CIA. However, this opens up the system for abuse and misuse. This paper, therefore, analyzed the workflows of healthcare staff's security practices in electronic health records (EHR) logs to determine anomalous security practices. Different classification types of machine learning algorithms were used. The EHR logs were simulated based on healthcare workflow scenarios. A number of machine learning algorithms were used to analyze the logs for deviations of accesses from the workflow. Based on the analysis results, all of the machine learning methods generally obtained a very good performance. The best performance on the non-normalized dataset is achieved by the Logistic Regression method with accuracy, precision, recall, and F1 value of 0.998, 0.849, 0.978, and0.909 respectively while Random Forest obtained the best result on Normalized data with accuracy, precision, recall, and F1 value of 0.998, 0.867, 0.836, and 0.851 respectively. It however remains challenging to detect malicious security practice if a malicious actor follows the workflow to access healthcare records with legitimate access right.

Integrated Health Care Delivery System with IoT Enabling Technology

Vivek Veeraiah, G. K. Ravikumar
Department of R&D, Computer Science, BGSIT, Adichunchanagiri University, Karnataka, India

Abstract - An integrated healthcare system, making use of Internet of things has immense benefits. A crucial factor for an efficient delivery of the health care system is the medical database of the patient on a real time basis. The mortality rate has deep rooted constrains in medical errors due to lack of medical information of the patient and unavailability of real time data. IoT enabled technology can bring about a drastic change in reducing the mortality rate. The value added is much beyond the money it can save people's life, with the outbreak of Covid-19 pandemic the benefits derived from IoT enabled technology has been remarkable. A striking feature of IoT enabled tools in the field of health care sector is that it can cater to a large population simultaneously and remotely. If IoT platform is implemented in healthcare services it can reduce the expenditure of health care services to the national GDP and can reduce the mortality rate. With digital transformation E-health apps are cost effective and less time consuming for individuals to monitor vitals at will. An emerging economy like India can be benefitted immensely from using IoT enabled tools in the field of health care services.

Toward Generating Synthetic CT Volumes using a 3D-Conditional Generative Adversarial Network

Jayalakshmi Mangalagiri, David Chapman, Aryya Gangopadhyay, Yaacov Yesha, Joshua Galita, Sumeet Menon, Yelena Yesha, Babak Saboury, Michael Morris, Phuong Nguyen
University of Maryland, Baltimore County, Baltimore, Maryland, USA
National Institutes of Health Clinical Center, Bethesda, Maryland, USA
Networking Health, Glen Burnie, Maryland, USA
OpenKneck Inc., Halethorpe, Maryland, USA

Abstract—We present a novel conditional Generative Adversarial Network (cGAN) architecture that is capable of generating 3D Computed Tomography scans in voxels from noisy and/or pixelated approximations and with the potential to generate full synthetic 3D scan volumes. We believe conditional cGAN to be a tractable approach to generate 3D CT volumes, even though the problem of generating full resolution deep fakes is presently impractical due to GPU memory limitations. We present results for autoencoder, denoising, and depixelating tasks which are trained and tested on two novel COVID19 CT datasets. Our evaluation metrics, Peak Signal to Noise ratio (PSNR) range from 12.53 - 46.46 dB, and the Structural Similarity index (SSIM) range from 0.89 to 1.

The Evaluation of Mobile Technology Adoption as a Employee Training Tool between Pre-COVID and COVID

Anastasia Tracy Biggs
School of Computer Science and Information Technology, Colorado Technical University, colorado, USA

Abstract--The purpose of this qualitative case study was to explore why and how corporate training managers can adopt mobile technology for employee training. The case study method explored the learning processes to determine if a learning model is appropriate for the use of mobile technology as a training tool (De Zan, De Toni, Fornasier, and Battistella, 2015, p. 341). This qualitative case study utilized interviews to explore how the use of mobile technology can be adopted to train employees. The use of interviews examined the degree of employee growth from mobile training (Alberghini, Cricelli, and Grimaldi, 2014, p. 260). Case study methodology answered how mobile technology through cause-effect relationships explored the lack of mobile technology adoption interventions between corporate managers and the organization (De Zan et al, 2015, p. 335) (Tsang, 2013).

Health Record Chain (HRC): Implementation of Mobile Healthcare System using Blockchain to Enhance Privacy of Electronic Health Record HER

Arij Alfaidi, Edward Chow
Department of Computer Science, University of Colorado at Colorado Springs, Colorado, USA

Abstract— Mobile health applications connect to wearable devices with inbuilt sensors to monitor critical human body parameters such as heart rate, pulse rate, body temperatures ,and others. This paper will introduce our system, Health Record Chain HRC, to investigate how to design and implement secure mobile health systems that leverage the smart mobile devices' new authentication and geo-location mechanism. HRC introduces an enhancing framework to ensure the users' security and privacy and conform to the related regulations (General Data Protection Regulation GDPR, Health Insurance Portability and Accountability Act HIPPA) using blockchain. The immutability of blockchain conflicts with GDPR' "Right to be forgotten" when the user can delete all of his/her data. This work implements a blockchain system with a mobile application and a web interface to address this conflict by hashing the Electronic health record EHR to Ethereum based blockchain.

An Open-source Application Built with R Programming Language for Clinical Laboratories to Innovate Process of Excellence and Overcome the Uncertain Outlook during the Global Healthcare Crisis

Maria Helena Rivero
Harrisburg University of Science and Technology, Harrisburg, PA, USA

Abstract— Clinical laboratories must thrive using supportive analytical tools to maintain the cost control and profitability during economic uncertainty, the shortages of laboratorians, and public healthcare emergencies like the COVID-19 pandemic. R Programming is a comprehensive open-source and platformindependent tool that can manage large datasets and display key performance indicators through real-time dashboards. This study seeks to innovate the performance excellence framework of smallsized clinical laboratories located in the United States, using R Programming's Shiny App dashboard. It will also encourage small-sized clinical laboratories to transition into a nextgeneration enterprise by digitizing their decision-making tools and improve performance using data analytics.

CSCI-ISCI: COMPUTATIONAL INTELLIGENCE

Bilateral Trade Flow Prediction Models Enhanced by Wavelet and Machine Learning Algorithms

Evdokia Maria Kottou, Tyler Andrew Grubelich, Xiaodi Wang
Department of Mathematics, Western Connecticut State University, Connecticut, USA
Farmington High School, Connecticut, USA

Abstract—International economy promotes a sense of global interdependence and offers mutual benefits to countries around the world through the system of imports and exports. In this study, we have utilized economic indicators as inputs in an algorithm scheme that is based on Machine Learning methods combined with Wavelet Transforms to predict the bilateral trade flow between pairs of trading countries. Utilizing this methodology will allow countries to strike more successful trade deals, and increase a nation's estimated Gross Domestic Product (GDP) per capita. This will ultimately help reduce poverty as governments reallocate funds to serve struggling populations and target the 2030 United Nations Agenda (UNA) [1].

An Auto Optimized Payment Service Requests Scheduling Algorithm via Data Analytics through Machine Learning

George Wanganga, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract— Traditional customer payment service scheduling approaches cannot cope with the modern demand for timely, high-quality service due to the disruption of big data within small and medium-sized payment solution providers (SaMS-PSP). While many customers have access to modern technologies to lodge their service requests easily and fast, SaMS-PSPs do not have equally automated big data-driven capabilities to handle the growing demands of these service requests. To effectively improve SaMS-PSP's customer payment service requests processing speeds, personnel optimization, throughput, and low latency scheduling, we have developed a new customer payment service request scheduling algorithm via matching request priority with the best personnel to handle the request based on data analytics through machine learning. Our experiments and testing have confirmed the merits of this new algorithm. We are also in the process of applying this new algorithm in real-world payment operations.

Classifying False Alarms in Camera Trap Images using Convolutional Neural Networks

Joseph Granados, Chris Halle, Gurman Gill
Department of Computer Science, Sonoma State University, Rohnert Park, California, USA
Center of Environmental Inquiry, Sonoma State University, Rohnert Park, California, USA

Abstract—Wildlife trapping cameras often capture false alarms when triggered by blowing vegetation or cloud shadows moving across the ground. Identifying these false alarms and distinguishing them from true capture events (images of actual animals, human, vehicle) requires a substantial amount of personnel time. Here we explore how convolutional neural networks can be used to develop an automated computer screening model for filtering out the false alarms. The models screening threshold can be varied to suit the requirements of the given camera network. For cameras used for real-time public education and outreach, a low screening threshold can be used. Based on using a screening threshold of 0.5 on a specific Tensorflow model, false alarms were classified with an average accuracy of 88.83 ± 4.29% and true capture events with 91.83±2.85% on a dataset of 23,930 images. A high screening threshold should be used for research purposes. By choosing a threshold of 0.97, only 0.37% of true capture events are misclassified and about 50% of false alarms are correctly classified, saving between 5.5 and 11 eight-hour workdays of personnel time. As part of this study, we also explore some of the ramifications of deploying the existing model to classify images from new camera networks.

Distance Correlation Sure Independence Screening for Accelerated Feature Selection in Parkinson's Disease Vocal Data

Dan Schellhas, Bishal Neupane, Deepak Thammineni, Bhargav Kanumuri, Robert C. Green II
Department of Computer Science, Bowling Green State University, Bowling Green, Ohio, USA

Abstract—With the abundance of machine learning methods and the temptation of using them all in an ensemble, having model-agnostic feature selection is incredibly alluring. Principal component analysis was developed in 1901 and has been a strong contender in this role since, but in the end is an unsupervised method. It offers no guarantee that the features selected have predictive power because it does not know what is being predicted. To this end, Peng et al. developed the minimum redundancy-maximum relevance (mRMR) method in 2005. It uses the mutual information not only between predictors but also with the response in its calculation. Estimating mutual information and entropy tend to be expensive and problematic endeavors, which leads to excessive processing times even for data that is approximately 750 by 750 in a Leave-One-SubjectOut jackknife situation. To remedy this, we use a method from 2012 called Distance Correlation Sure Independence Screening (DC-SIS) which uses the distance correlation measure of Szekely ´ et al. to select features that have the greatest dependence with the response. We show that this method produces statistically indistinguishable results to mRMR on Parkinson's Disease vocal diagnosis data 90 times faster.

Generative Truss Optimization for Support-Free Fused Filament Fabrication

Henrik Storm Forberg, Tonnes Frostad Nygaard, Mats Erling Hovin
Department of Informatics, University of Oslo, Oslo, Norway

Abstract—Fused Filament Fabrication is currently among the most commonly used Additive Manufacturing technologies but is highly reliant on temporary support structures during production. Implementing a generative optimization algorithm for support-free Fused Filament Fabrication could streamline the manufacturing process in terms of labor, time- and material use. Despite the current relevancy of Additive Manufacturing, there is a lack of research on structural optimization customized for support-free Fused Filament Fabrication. This research applies a generative optimization algorithm consisting of a multi-objective evolutionary algorithm and a local search algorithm to generate and optimize rigid-jointed 3D truss structures. The results show the capacity to generate and optimize support-free rigid-jointed truss structures with promising solutions to a multi-objective optimization task. This paper suggests that support-free structural optimization algorithms can impact how we design robotic bodies and parts in the future.

Recent Progress on Text Summarization Approaches

Suad Alhojely, Kalita Jugal
College of Engineering and Applied Science, The University of Colorado at Colorado Springs, Colorado, USA

Abstract—Text summarization produces a summary of a document by highlighting its most important content ideas. Researchers have been developing text summarization techniques since the 1950s. Most summarization deals with summaries of single documents, but recent summarization efforts have also produced summaries from clusters of documents. This paper reviews recent approaches in three categories: Extractive, Abstractive, and Hybrid text summarization along with specific methods within each category. Most authors focus on the Extractive approach, but the generated summaries are very different from human summaries regardless of techniques used.

Frequency Maps as Expert Instructions to lessen Data Dependency on Real-time Traffic Light Recognition

Thiago Almeida, Hendrik Macedo, Leonardo Matos, Bruno Prado, Kalil Bispo
Programa de Pos-Graduacao em Ciencia da Computacao (PROCC), Universidade Federal de Sergipe (UFS), Brazil
Departamento de Computacao (DCOMP), Universidade Federal de Sergipe (UFS), Sergipe, Brazil

Abstract—Research on Traffic Light Recognition (TLR) has grown in recent years, primarily driven by the growing interest in autonomous vehicles development. Machine Learning algorithms have been widely used to that purpose. Mainstream approaches, however, require large amount of data and a lot of computational resources. In this paper we propose the use of Expert Instruction (EI) to reduce the amount of data required to provide accurate ML models for TLR. Given an image of the exterior scene taken from the inside of a vehicle, we stand the hypothesis that the picture of a traffic light is more likely to appear in the central and upper regions of the image. Frequency Maps of traffic light locations were thus constructed to confirm this hypothesis. Results show increased accuracies for two different benchmarks, by at least 15%. The inclusion of EI in the PCANet achieved a precision of 83% and recall of 73% against 75.3% and 51.1% of its counterpart. We finally presents a prototype of a TLR Device with such EI model to assist drivers. To show the feasibility of the apparatus, a dataset was obtained in real time usage and tested in an AdaBSF and SVM algorithms to detect and recognize traffic lights. Results show precision of 100% and recall of 90.9%.

System Modeling by Ultra High Frequency Sigmoid and Sine Artificial Higher Order Neural Networks

Ming Zhang
Department of Physics, Computer Science and Engineering, Christopher Newport University, Newport News, Virginia, USA

Abstract — New open box and nonlinear system model of Ultra High Frequency Sigmoid and Sine Artificial Higher Order Neural Network (UGS-HONN) is presented in this paper. A new learning algorithm for UGS-HONN is also developed from this study. A time series data modeling system, UGS-HONN Simulator, is built based on the UGS-HONN system models too. Test results show that average error of UGS-HONN system models are closing to zero (10- 6). The average errors of Polynomial Higher Order Neural Network (PHONN), Trigonometric Higher Order Neural Network (THONN), and Sigmoid polynomial Higher Order Neural Network (SPHONN) models are from 2.8128% to 4.9076%. It means that UGS-HONN system models are 2.8128% to 4.9076% better than PHONN, THONN, and SPHONN models.

Evaluation of Machine Learning Based Regression Techniques for Prediction of Oil and Gas Pipelines Defect Length

Huda Aldosari, Raafat Elfouly, Reda Ammar
Computer Science and Engineering, University of Connecticut, Storrs, Connecticut, USA
Computer Science Department, Rhode Island College, Rhode Island, USA

Abstract—Magnetic flux leakage (MFL) signals allow the scale of metal failure defects present on a pipeline to be observed, located, and measured. An advanced and reliable intelligent pipeline monitoring system is needed to protect the local ecosystem from disastrous effects due to malfunctioning pipelines. MFL plays a vital role in gas pipeline examination; various studies have explored smart MFL based defect prediction systems. Machine learning-based defect prediction systems allow predicting the characteristics of the oil and gas defects in real-time. As fault detection control is applied to lowperformance embedded systems, prediction algorithms' reliability is of utmost importance. The intention was to extend the work of previous researchers and provide insight into the behavior of the selected classifiers with time as a robustness factor, an experimental design that constitutes the novelty of this study. For the said purpose, a prerecorded dataset comprised of MFL signals has been utilized. Prediction of defect length has been performed with several approaches, including linear regression (LR), linear regression with stochastic gradient descent (LRSGD), support vector machine (SVM), Gaussian process regression (GPR), boosting regression tree ensemble (BSTE), binary decision tree (BDT), stepwise (SW), and artificial neural network (ANN). The results indicated that ANN yields the best

prediction results for MFL based defect predictions, followed by GRP regression analysis. The presented results can be utilized to design and implement a defect length prediction model to be deployed underwater, providing realtime prediction results.

Optimal Artificial Neural Network Model for Prediction of Oil and Gas Pipelines Defect Length

Huda Aldosari, Raafat Elfouly, Reda Ammar
Computer Science and Engineering, University of Connecticut, Storrs, Connecticut, USA
Computer Science Department, Rhode Island College, Rhode Island, USA

Abstract—Magnetic flux leakage (MFL) signals are used for estimating not only the size, but also the shape of the faults causes by deteriorating metal that makes the oil and gas pipelines. Such defects, such as rust, which, if left undetected and poorly handled, may have devastating effects, both in terms of environmental degradation and loss of life, and also millions of dollars in repair costs to be incurred by the ownership firms. Algorithms focused on machine learning have demonstrated the ability to solve the issue by identifying and measuring the size and shape of such defects effectively. In particular, artificial neural networks (ANN) have shown great potential to generate high precision results. In this article, the optimization of ANN was carried out by using noisy and noiseless measurements of MFL signals. ANN optimization was conducted for the training function (12 separate training functions), the hidden neurons numbered between 1 and 100, and hidden layers numbered between 1 and 10. The output was calculated by root mean square (RMSE) error. It has been found that gradient descent momentum (GDM) and gradient descent (GD) exhibit bad performance outcomes than all other studied training functions, whereas all other studied training algorithms showed equal and comparable performance outcomes. The highest output outcomes have been found in the range of 1, 10 and 20 to 35 with regard to the number of hidden neurons. Network output deteriorates as the number of hidden neurons deviates from the optimal range observed. With regard to the optimal number of hidden layers, it has been observed that ANN yields better output results with 1,2,5 and 8 hidden layers for noiseless MFL signals and the best results with 1,2,5 and 7 hidden layers are observed for noisy datasets. To evaluate the oil and gas pipeline defects, the optimal inferred parameter range may be used to train the ANN model for improved performance outcomes.

ArtMatch: Classifying Famous Paintings and Matching Them with Children's Artwork

Ryan Z. Cheng
OHS Dayton, Ohio, USA

Abstract—This paper represents the progress of an ongoing effort to develop a mobile app called ArtMatch, which will help kids develop a passion for art. The proposed app will deepen children's appreciation for famous pieces of artwork while enhancing their creativity. By taking an image of artwork created by the user, ArtMatch will show the most similar painting done by a famous artist, using AI and deep learning technology. The app works through a program that clusters images based on their design and style. Throughout this process, the user will see the effectiveness of his/her painting through their match with well-known paintings. The program's present version can generate accurate matches for several works done by young children.

Simple Proofs of the Strong Perfect Graph Theorem Using Polyhedral Approaches and Proving P=NP as a Conclusion

Maher Hashem Heal
Baghdad University, Baghdad, Iraq

Abstract—The strong perfect graph theorem is the proof of the famous Berge's conjecture that the graph is perfect if and only if it is free of odd holes and odd anti-holes. The conjecture was settled after 40 years in 2002 by Maria Chudnovsky et. al. and the proof was published in 2006. However, that proof is lengthy and intricate and using a combinatorial approach. We provide simple short proofs of the strong perfect graph theorem using polyhedral methods. We first prove the weak perfect graph theorem by

polyhedral methods and use that to prove the strong perfect graph theorem. Our proofs emerge naturally from our work to calculate the capacity of multihop wireless networks. As a corollary of our proofs techniques we prove P=NP by proving there is an algorithm to find the maximum independent set or the independence number of any graph in polynomial time as a function of the number of the graph vertices.

Optimization of Sustainable Single-Machine Scheduling Problem

S. Mahdi Homayouni, Dalila B.M.M. Fontes
LIAAD, INESC TEC, Porto, Portugal
LIAAD, INESC TEC, & Faculdade de Economia da Universidade do Porto, Porto, Portugal

Abstract—This work considers sustainable scheduling of manufacturing operations and preventive maintenance activities in a single-machine environment where the machine works continuously in three eight-hour shifts per day. The jobs can be produced at different processing speeds, which reduces energy consumption and/or processing times. In a tri-objective mixed integer linear programming model, sustainability is attained through minimizing total weighted earliness/ tardiness - economic pillar, total energy consumption - environmental pillar, and number of undesired activities - social pillar. Moreover, a multi-objective genetic algorithm finds near optimal solutions in a timely manner. Numerical results will be presented at the conference.

Convolutional Neural Network Based License Plate Recognition Techniques: A Review

Oladapo Ibitoye, Temitayo Ejidokun, Joseph Dada
Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado Ekiti, Nigeria

Abstract— Identification of vehicle license plate has become more practical within the last decade in many applications such as; storage and retrieval of vehicular inflow records, automatic toll collection, parking fee payment, traffic monitoring, tracking of moving vehicles, recovery of stolen vehicles, etc. A lot of researches have been carried out on license plate recognition which has led to the development of new techniques and modification of existing techniques. This study provides a breakdown of some techniques used for the implementation of license plate recognition systems, with significant emphasis on Convolutional Neural Network (CNN). The strength and weaknesses of the highlighted techniques were discussed. Also, areas of further improvement of some of the selected CNNbased techniques were suggested.

Facial Expression Recognition for Hugging Type Vital Sign Measuring System

Wang Chao, Takehiko Ogawa, Ho Yihsin, Jun Hasegawa, Naoki Oshima
Mechanical and Electronic Systems Engineering Course, Takushoku University, Tokyo, Japan
Department of Electronics and Computer System, Takushoku University, Tokyo, Japan

Abstract - In recent years, artificial intelligence technology has been applied to medical doctors for medical treatment, disease analysis, health management, etc., and has been accelerated to improve the epidemic of new coronary pneumonia. In this regard, this research conceived a new type of hugging system, through deep learning to identify "surprise," "fear," "disgust," "anger," "happy," and "sadness" expressions to better grasp the patient's condition, and it helps alleviate the anxiety of patients. After experimental tests, facial expression cognition has certain feasibility in medical treatment.

VikingBot: Towards a Hybrid Artificial Intelligence for Starcraft

Wesley Deneke, Tyler Barger, Matthew Carter, Chris Lokken, Daniel Peterson
Computer Science Department, Western Washington University, Bellingham, Washington, USA

Abstract— The focus of this research is to develop a StarCraft bot that makes efficient use of computing resources. To this end, VikingBot employs a hybrid approach that pairs AI Planning and Machine Learning. The AI planning component handles high-level strategy management and the Machine Learning component handles individual control units in combat. In this paper, we describe our high-level approach and report initial results. This is an interim report on a work in progress.

An Efficient Local Outlier Factor for Data Stream Processing: A Case Study

Omar Alghushairy, Raed Alsini, Xiaogang Ma
Department of Computer Science, University of Idaho, Moscow, Idaho, USA
College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract— In the field of data mining and machine learning, outlier detection is considered a significant procedure for many applications, such as fraud detection in bank transactions and decision support systems. Data streams are a major player in the big data era. Currently, data streams are generated from various sources with huge amounts of data. This has led to difficulty when using older algorithms, which are designed for static data. The Local Outlier Factor (LOF) is one of these algorithms. The most challenging issue of the LOF is that it needs to preserve the whole dataset in computer memory. A new LOF that can deal with a data stream in limited memory is needed. This paper is a case study of several benchmark datasets for outlier detection that aim to increase the efficiency of the accuracy of local outlier detection in data streams.

CSCI-ISCC:
CLOUD COMPUTING AND DATA CENTERS

A Novel Cloud Authentication Framework

Latifa Khalid ALnwihel, Abdul Raouf Khan
Department of Computer Science, King Faisal University, Alahsa, Saudi Arabia

Abstract— Cloud computing is a medium that provides cost reduction with storage and fast computation environment. Various organization are marching to the cloud because of the several advantages provided such as remote access, cost efficiencies and simplified IT infrastructure and management. However, for any distributed or shared system, security is the main requirement. The research work discuss the numerous researchers work regarding cloud security issues, threats, known vulnerabilities and their solutions. Hence, the available threats in the cloud system are discussed to come up with an improved framework that provide data security through multiple techniques. The proposed research work combines cryptography, authorization and authentication mechanisms in order to ensure access control.

A Suggested Taxonomy for Governmental Clouds

Konstantinos Roungeris, Dimitris Zissis, George S. Androulakis
Department of Business Administration, University of Patras, Greece
Department of Product and Systems Design Engineering, University of Aegean, Greece

Abstract— Cloud Computing has undoubtedly been the most disruptive IT paradigm of the last two decades, spanning a wide area of technologies and implementations. Its emergence is already transforming the public sector with relevant eGovernment use cases. The adoption of various cloud-related initiatives as part of national strategies is driving the need to use a common framework to better describe, analyse and compare those implementations, known as "Governmental Clouds". In addition, for any interoperability to be effective in the long run, a standard framework and common language to describe the relevant ontologies should exist. This paper aims to introduce and discuss a taxonomy of three different models based on the existing and emerging Cloud Computing governmental best practices and available use cases, addressing the lack of relevant concrete definitions.

Kubernetes-based Workload Allocation Optimizer for Minimizing Power Consumption of Computing System with Neural Network

Ryuki Douhara, Ying-Feng Hsu, Tomoki Yoshihisa, Kazuhiro Matsuda, Morito Matsuoka
Graduate School of Information Science and Technology, Osaka University, Osaka, Japan
Cybermedia Center, Osaka University, Osaka, Japan

Abstract—Edge computing has been attracting attention due to the spread of the Internet of Things. For edge computing, containerized applications are deployed on multiple machines, and Kubernetes is an essential platform for container orchestration. In this paper, we introduce a Kubernetes based power consumption centric workload allocation optimizer (WAO), including scheduler and load balancer. By using WAO built with power consumption and response time models for actual edge computing system, 9.9% power consumption was reduced compared to original Kubernetes load balancer. This result indicates that the WAO developed in this study exhibits promising potential for task allocation modules as a micro service platform.

Intrusion Detection System: The Use of Neural Network Packet Classification

Nery Ruiz, Bryan Tavera, Abdel-Shakour Abuzneid
Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, Connecticut, USA

Abstract-- Despite recent advances in cloud processing power and network connectivity to handle massive network traffic, networks are still vulnerable to Distributed Denial of Service (DDoS) attacks. With the recent proliferation of the Internet of Things (IoT), unsecured devices are fueling the ever-growing botnets, which allow creating larger malicious networks. Current mitigation techniques need to adapt to a new growing size of zero-day attacks to protect network services to consumers and block malicious connections. Deep learning enables machines to find the solution to many complex problems. This paper evaluates the performance of the Simple Neural Networks, Convolutional Neural Networks, and Recurrent Neural Networks in detecting DDoS attacks when trained with the CSE-CIC-IDS2018 Dataset. This research will discuss the presented datasets and the efficiency of the proposed networks. The trained data was obtained from a realistic dataset that holds different forms of intrinsic volume, protocol, and web-based attacks.

Secure Cloud Storage Migration

Edward Eisenberger, Sai Teja Kutalam, Vamsi Varma Datla, Abdel-Shakour Abuzneid
Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, Connecticut, USA

Abstract— Cloud Storage is a cost-effective and agile platform for both users and businesses. Transferring data to and between Cloud Storage systems presents a vulnerability for the data. This paper proposes a method for securely migrating between Cloud Storage systems using Public Key Exchange and Cipher Block Chaining (CBC).

Information Flow Control to Secure Data in the Cloud

Fahad Alqahtani, Salahaldeen Duraibi, Predrag T. Tosic, Frederick T. Sheldon
Department of Computer Science, University of Idaho, Moscow, Idaho, USA
Department of Computer Science, Prince Sattam Bin Abdulaziz University, Al- Kharj, Saudi Arabia, KSA
Computer Science Department, Jazan University, Jazan, Saudi Arabia, KSA
Department of CSEE, Eastern Washington University, Spokane, Washington, USA
Department of Mathematics and Statistics, Washington State University, Pullman, Washington, USA

Abstract— Data security remains a major concern for organizations considering the use of cloud services to store their confidential, business-critical data. In this paper, we investigate how information flow control can be used in the cloud to enhance the confidence of enterprises, so they can safely and securely adopt cloud solutions for their data storage needs. We discuss how different techniques can be used with the CloudMonitor tool to guarantee the protection of data in the cloud. We then give an overview of how centralized and decentralized information flow control systems operate, and the comparative advantages and disadvantages of each approach. Our analysis suggests that CloudMonitor can achieve better data security with the use of decentralized information flow control. We then discuss different decentralized information flow tracking tools applied to monitoring data in the cloud. CloudMonitor enables the consumers and the providers of cloud services to agree on acceptable security policies as well as their implementation, to ensure secure data storage in the cloud.

Evidence for Monitoring the User and Computing the User's Trust

Maryam Alruwaythi, Kendall Nygard
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia
Department of Computer Science, North Dakota State University, Fargo, North Dakota, USA

Abstract— Cloud computing helps organizations to dynamically increase the resource needs as and when needed, without the need to purchase them. Security is a basic concern in cloud computing, and threats can occur both internally and externally. The basic security protection such as traditional access control (TAC), virus detection, and intrusion detection are unable to manage

variety of malicious and network attacks. The number of users might get hacked because of limitation in basic security protection. To implement a secure, reliable, and safe cloud-computing environment, we need to consider the trust issue. A trusted cloud is guaranteed to be safe from user terminals; combined with the concept of a trusted network , it evaluates, forecasts, monitors, and manages the user's behavior to eliminate malicious datacenter attacks which are performed by unwanted cloud users and hackers; as a result, there is improved cloud security. In this paper, we are addressing trust issue in the cloud computing based on user behavior and what kind of evidence must be considered to implement sufficient model to calculate user' creditability.

Density-Based Server Placement for Collaborative Virtual Services

Sakir Yucel
NetApp, Wexford, PA, USA

Abstract— The boundaries on different domains are blurring as corporations and infrastructure providers are collaborating to offer end-to-end services over the networks and infrastructures of various network, service, cloud, content delivery and other infrastructure providers as well as the customer premises. To capture opportunities and shine in competitive service market, infrastructure and service providers need to excel in addressing the changing customer requirements and in the operations and management of the resources. They should welcome effective collaboration with other network and infrastructure providers in delivering quality services to customers. Effective sharing of the infrastructure resources is essential in meeting the customer demands and reducing the cost. Server placement problem for end-to-end virtual services becomes a crucial optimization challenge for providers in such collaborative environments. We formulate the collaborative virtual server placement problem and suggest density-based clustering algorithms to address this problem.

Biometrics Based Access Framework for Secure Cloud Computing

Ashokkumar R. Patel
Florida Polytechnic University, Lakeland, Florida, USA

Abstract— This paper is focused on the topic of the use of biometrics framework and strategy for secure access identity management of cloud computing services. This paper present's a description of cloud computing security issues and explored a review of previous works that represented various ideas for a cloud access framework. This paper discusses threats like a malicious insider, data breaches, and describes ways to protect them. It describes an innovative way portrayed a framework that fingerprint access-based authentication to protect Cloud services from unauthorized access and DOS, DDoS attacks. This biometrics-based framework as an extra layer of protection, added then it can be robust to prevent unauthorized access to cloud services.

CSCI-ISED:
EDUCATION - STEM, COMPUTER SCIENCE AND COMPUTER ENGINEERING

The System's Holding me Back - Challenges of Teaching Computing in Further Education

Jordan Allison
Technical and Applied Computing, University of Gloucestershire, Gloucester, United Kingdom, UK

Abstract—The UK further education (FE) sector has been plagued by continual policy reforms and has been described as the poor relation of the UK education sector. Meanwhile, it has been reported that there is a 'digital skills' gap, placing more importance on the quality and delivery of computing education. By conducting twelve semi-structured interviews from those who teach computing in FE colleges, seven challenges were identified that influence their teaching practice. Through the identification of these challenges, more targeted measures can now be taken to mitigate these challenges. Further, this paper contributes to the relatively short supply of research that investigates computing education within FE colleges.

A Framework for Effective Continuing Professional Development: The Case of Computer Science Teachers within Further Education Colleges

Jordan Allison
Technical and Applied Computing, University of Gloucestershire, Gloucester, United Kingdom, UK

Abstract—With rapid changes in technology, computer science teachers need to undertake continuing professional development (CPD) to ensure that their teaching is relevant and leads to positive pupil outcomes. Unfortunately, existing research has typically neglected what is required for effective CPD for computer science teachers who work in general further education colleges. This paper presents the findings from thirteen semi-structured interviews from those who work in this setting, and outlines a framework detailing four main areas of activity required for effective computer science teacher CPD. This framework should help inform CPD providers in creating more effective computer science teacher CPD.

Designing a Parallel Programming Course for Lower-Division Students

Xuguang Chen
Computer Science Department, Saint Martin's University, Lacey, Washington, USA

Abstract -- This paper covers the design of a course introducing parallel programming. The design is partly based on the feedback collected by the author when teaching other parallel computing courses offered for senior undergraduates and graduates, plus the conversations with lower-division students outside of class. It targets the students typically having completed CS1 and CS2 courses and puts more emphasis on gaining parallel programming skills through hands-on exercises. Its ideal lecture schedule is that each week, one meeting introduces the concepts and theories of parallel computing in a traditional classroom and the other discusses parallel programming skills in a computer lab. Unlike the courses designed for senior undergraduates or graduates, its instructional approaches are more inclined to student-centered and teamwork approaches. Its coursework will include term projects, the topics of which are selected from the instructor' research area and tailored for this class. The course design has been completed and will be implemented in the spring of 2021 as a pilot.

Evaluation of Group Projects in an Undergraduate Data Structure Course

Anurag Dasgupta
Computer Science, Valdosta State University, Georgia, USA

Abstract— Exposure to group projects allows students to explore and extend the concepts they learn in the classroom. The author experimented with semester long projects besides the regular topics of a typical data structure course in an undergraduate setting. The students were encouraged to select different data structures as project topic to solve different practical problems. The goal of this paper is to find out what type of group projects work the best in a data structure course. The paper focuses on how the students fared with the different data structure projects.

Remote Collaboration Potential in STEM Education using Bare Machine Computing Research

Nirmala Soundararajan, Joel Weymouth, Ramesh Karne, Alexander L. Wijesinha, Navid Ordouie
Computer and Information Sciences, Towson University, Towson, Maryland, USA

Abstract— Bare Machine Computing (BMC) is a paradigm used in programming applications to run on a Bare Machine. It consists of a CPU, Memory, and I/O without an Operating System and resident mass storage. It utilizes an Application Development model software that interfaces directly with the hardware. It also has potential as a powerful teaching and research tool in undergraduate STEM programs, notably Computer Engineering and Computer Science. Before the COVID-19 Pandemic, researchers used the University lab environment, and Bare Systems connected via LAN/WAN. The Pandemic restrictions put members of the research team in different locations. Hence, the researchers adopted a Remote Collaboration model of Concurrent Participation, Communication, and Development for Bare Machine to continue this research. This paper discusses the advantages and disadvantages of moving the BMC research entirely online and measures needed to address the Configuration of BMC Development Environment, Communication, Collaboration, and Diverse Skill Set. The paper covers online technical research because most literature concentrates on only online survey research's benefits and drawbacks. This paper also addresses BMC application in presenting STEM subjects to students, mainly undergraduate students. Finally, the paper includes remote research that fulfills the Department of Education (DOE) requirements of student access.

Observable Learning Outcomes Among Tertiary Mathematics Students in a Newly Implemented Blended Learning Environment

Mary Ruth Freislich, Alan Bowen-James
School of Mathematics and Statistics, University of New South Wales, Sydney, Australia
Le Cordon Bleu Business School, Sydney, Australia

Abstract— The study compared learning outcomes for two cohorts of mathematics students in an Australian university, studying in adjacent years, in two sequential core mathematics subjects. The only change in delivery of teaching was the replacement of one of two face-to-face tutorials with an online tutorial. The online tutorial tasks were designed to lead the student through each topic using steps of increasing difficulty as scaffolding. Learning outcomes were assessed by a method based on the SOLO taxonomy, which provided a common scale for scoring learning outcomes observable in examination answers. In the students' first mathematics subject, scores were significantly higher for the group doing online tutorials, but differences were not significant in the second. The conclusion was that the online tutorials helped students beginning university study but was less important later.

Unsupervised Functional Analysis of Graphical Programs for Physical Computing

Tom Neutens, Francis Wyffels
IDLab, Department of Information Technology, Ghent University - IMEC, Belgium

Abstract—Assessing students' code is a challenging and time intensive task for teachers. Facilitating this process is essential to stimulate teachers to teach programming in their classroom. In this paper we describe a new technique to autonomously assess program functionality for different programs in a physical computing context. Using a qualitative analysis on different scenarios, we show that our method can be used to group programs with similar functionality and distinguish programs with different functionality. Our approach will facilitate code assessment for teachers and is a first step toward a more intelligent way to evaluate how children learn programming.

A Mobile Application as Didactic Material to Improve Learning on Distributed Architectures

Guadalupe Ortiz, Alfonso Garcia-de-Prado, Juan Boubeta-Puig, Halina Cwierz
Department of Computer Science and Engineering, University of Cadiz, Spain
Department of Computer Science and Engineering, University of Extremadura, Spain
Department of Computer Architecture and Technology, University of Cadiz, Spain

Abstract—The university has been applying a teaching methodology based on continuous assessment over the last few years. The main advantage of this methodology is that students acquire the knowledge gradually, so that they consolidate it better in their memory; however, they run the risk of forgetting what was studied in the first parts of the course when they finish it. To motivate students to review what they learnt, we have implemented an augmented reality mobile application and methodology with which students review the contents of the subject through the serious game. Moreover, the game focuses on the specific difficulties of each student by reinforcing those questions that are most difficult for them. In this paper, we also present how the game has been customized for various courses from the Degree in Computer Science and Engineering at the University of Cadiz and which results were obtained.

A Data Mining based Optimization of Selecting Learning Material in an Intelligent Tutoring System for Advancing STEM Education

Olanrewaju Ogunkunle, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract— Subsequent to the data deluge of the internet era and the recent advancement in big data technologies, it is easy to affirm the continuous application of such technological innovation to tackling a wide array of students' educational needs. The field of artificial intelligence and machine learning have improved education learning outcomes. However, the problem of generalized traditional supportive collaboration scripts for all students irrespective of the student's learning traits and position on the learning spectrum leads to less than optimum result in their educational pursuits. This paper presents a novel approach that uses data mining algorithm to optimize the selection of educational resources for students based on their learning traits and the six factors that cofound instructional content and delivery with a focus on students with learning disabilities for STEM subjects.

Zoom Sandwich: An Adaptable Model for Distance Learning

Fadi Muheidat, Lo'ai Tawalbeh
School of Computer Science and Engineering, California State University, San Bernardino, California, USA
Department of Computing and Cybersecurity, Texas A&M University, San Antonio, Texas, USA

Abstract— Distance learning is the education of students who may not often be physically present. Internet technology has made many possible ways of distance learning by open educational tools and services, such as e-learning and MOOCs. The distance learning technologies are divided into two types of delivery: synchronous learning and asynchronous learning. The two methods can be combined to create a hybrid learning model. In this work, we will discuss an adaptable hybrid teaching model we call it Zoom Sandwich. The model utilizes the benefits of both methods and integrates active learning techniques where students learn the concepts and topics online and then come to the classroom for application through interactive groups and class level discussions utilizing technology breakout rooms and chat platforms. We also consider the best practices to humanize online learning and increase students' and instructors' social presence. Covid-19 Pandemic changed the way institutions conduct business and provide services. To fight the spread of coronavirus, schools across the country have temporarily closed and transitioned to distance learning. This provided an opportunity for cybercriminals to hack and expose human vulnerabilities. We are proposing a set of security best practices and policies to countermeasures the cyberattacks.

Achieving ABET Accreditation: An Outcome Assessment Case Study

John Carelli
Computer Science and Information Technology Department, Kutztown University, Kutztown, Pennsylvania, USA

Abstract— ABET is an organization offering accreditation of degree programs for institutions of higher education in science, engineering, and computing. Institutions seek ABET accreditation as a demonstration of the quality of their programs. One important component of the accreditation process involves assessment of student performance against ABET defined student learning outcomes. This, in turn, requires the establishment of an on-going assessment process, together with a culture of continuous improvement. A case study is presented herein detailing a process of assessment developed at one institution which recently achieved ABET accreditation. It is hoped that the information provided here will serve as a useful guide to other institutions seeking to improve their assessment processes and/or receive ABET accreditation

Analyzing Coding Behaviour of Novice Programmers in Different Instructional Settings: Creating vs. Debugging

Tom Neutens, Francis Wyffels
IDLab, Department of Information Technology, Ghent University, Ghent, Belgium

Abstract—Many methods exist for teaching programming in a physical computing context. However, it is not clear what the advantages and disadvantages of these methods are in practice. Our research compares two methods for integrating programming in a primary robotics workshop. Both methods use a custom-designed programming environment based on Google Blockly. Moreover, one method lets learners create their programs from scratch while the second method requires the learners to fix faulty programs. We compared the differences between the integration methods by analyzing programming environment logging data and linked it to the results on a programming knowledge test. Our results show that the learners in the create group spend more time solving the programming assignment and are more often distracted by code blocks they don't need. The learners in the debug group require less time. However, they perform a disproportionate amount of code changes and apply the "tinkering" strategy more often. Nevertheless, the drawbacks of the create method, the learners in this group did show significantly higher scores on our programming knowledge test.

MannaTeam: A Case of Interinstitutional Collaborative Learning

Daniela Eloise Flor, Eduardo Henrique Molina da Cruz, Ayslan Trevisan Possebom, Carlos Roberto Beleti Junior, Rodrigo Hubner, Linnyer Beatriz Ruiz Aylon
Federal Institute of Paranavai (IFPR), Brazil
Federal University of Technology Parana (UTFPR), Brazil
State University of Maringa, (UEM), Brazil

Abstract—This case addresses the challenge of developing a novel educational strategy that promotes hard and soft skills using different types of emerging technologies. We achieved this with a horizontal collaborative learning strategy, between universities, and vertical collaborative learning, between universities and schools, organized by the MannaTeam network. In addition to sharing experiences, knowledge, laboratories and materials among network partners, we popularized our vision of education for the 21st century – Education 5.0, and invested in a project of female empowerment, showing society the reasons and impacts of the gender gap in technological areas. The success of our contemporary vision of education and the efforts of MannaTeam motivate us to continue. There is much to be done to stimulate innovation in education and the broad understanding that ability has no gender.

Cultivating Positive ICT Perceptions: Application of the MST-tree Model to the 'Guyanese Girls Code' Initiative

Alicia Layne, Penelope Defreitas, Juanelle Marks, Rayad Lackhan
Department of Computer Science, University of Guyana, Georgetown, Guyana

Abstract—Globally, women account for a small percentage of students and professionals in ICT-related fields. In order to increase female participation in this area, governments worldwide have been facilitating targeted interventions. One such initiative is the 'Guyanese Girls Code' (GGC) training program. This program was launched in 2018 with the aim of using Scratch and the BBC micro:bit to introduce girls at the Grade 7 to 9 levels to ICT education. In this paper, we propose a novel teaching model, the 'Motivation, Support and Teaching Components' tree model (MST-tree model). This model was critical to the design of GGC's curriculum and teaching strategies. An evaluation was conducted to gather information on the program participants' ICT perceptions and their experience of the training. The key findings revealed that the participants had a significant interest in learning about ICT even though they perceived the area to be difficult and local training opportunities to build their knowledge and skills were insufficient. Further, the girls portrayed high levels of selfefficacy, since they were interested in the program modules that they perceived to be most challenging. These findings suggest that the teaching model may address negative ICT perceptions when teaching females at the Grade 7 to 9 levels. As such, the MST-tree model has the potential to inform the design and execution of similar training programs and interventions.

Educational Approach for a BIM Collaboration

Anabelle Rahhal, Samia Ben Rajeb, Pierre Leclercq
LUCID Lab, University of Liege, Liege, Belgium
BAtir ULB, Polytechnic School of Brussels, Brussels, Belgium

Abstract— This article provides feedback on the methodology put in place for learning BIM in order to prepare future designers to integrate these new methods and tools into their practice. The objective of this course is to encourage and make learners aware of the challenges of cooperation for the success of architectural projects which mobilize complementary skills and expertise divided between distributed responsibilities. Based on recent advances in collaborative design, the course anticipates professional practices: it presents emerging BIM methods and tools for cooperation and implements them in an original technological context, dedicated to collaborative work, between the design and the execution phase of an architecural project. At the end of this article, we present to you the contributions, the limits and the perspectives envisaged in favor of a better implementation of BIM towards a Collaboration 4.0 in the architecture and construction sector.

MAESTRO: a semi-autoMAted Evaluation SysTem for PROgramming Assignments

Alessandro Bertagnon, Marco Gavanelli
Department of Engineering, Ferrara University, Ferrara, Italy

Abstract—Many works in the literature highlight the importance of formative assessment to improve the learning process. Formative assessment means providing assignments to the students and giving them feedback during the course. It is not very used in Italian Universities because it is highly time-consuming for the professors, that typically do not have help in the process of homework grading. In this work, we focus on programming exercises in computer science subjects, and propose a tool to semi-automatically grade and give feedback to the students. The tool was used in a computer language course on functional programming in a M.Sc. degree; the students evaluation of the course show a steep increase in the students appreciation. The tool is currently used in a course at undergraduate level on C programming. Index Terms—Distance learning methods and technologies; Computer and web-based software for instruction; Debugging tools and learning; software engineering programming issues and laboratory practice.

New Trends in Pedagogical Agents in Education

Luis Alfaro, Claudia Rivera, Jorge Luna-Urquizo, Elisa Castaneda, Jesus Zuniga-Cueva, Maria Rivera-Chavez
Dept. Academico de Ingenieria de Sistemas e Informatica. Universidad Nacional de San Agustin, Peru
Dept. Academico de Administracion, Universidad Nacional de San Agustin, Peru
Dept. Academico de Ingenieria Industrial, Universidad Nacional de San Agustin, Peru
Facultad de Ciencias e Ingenierias, Universidad Catolica de Santa Maria, Arequipa, Peru

Abstract— A pedagogical agent is an anthropomorphic virtual character used in an online learning environment for educational purposes. Pedagogical Agents fulfill a variety of educational purposes, for which they can have a high degree of adaptability, be very versatile, enable highly realistic simulations, focus on socio-cultural requirements of students, as well as having the ability to foster commitment, motivation and responsibility as well as improve learning and performance of cognitive subjects. In this work, we review the literature on Pedagogical Agents in order to identify and analyze them, as well as to evaluate the impact of the implementation of these systems in the field of pedagogy, considering also the different elements that they incorporate, such as: text, voice, 2D and 3D images, and human figures. The factors that drive the demands of new characteristics, functionalities and socio-cultural environments for the development of contemporary pedagogical agents are analyzed.

4 Year Comparison Of Undergraduate Students Provided Personal Laptop Computers: Initial Research

Gary Cantrell, Hussain Aljafer
Computer Science and Information Systems, Southern Utah University, Cedar City, Utah, USA

Abstract - This study explores the effectiveness of Southern Utah University's undergraduate laptop cohort program where students are given a laptop and assigned to a faculty mentor for the duration of their time at SUU. The authors seek to determine if the laptops are used, if they are used for what was intended, and if the inclusion in this program has a beneficial effect on the student's grade point average (GPA). The program is available to both computer science and information system majors. Both groups are examined together and separate. Data was collected with a student survey and from historical records of average GPA for seniors fall and spring semesters from 2016 to 2020. Students were grouped by cohort and non-cohort students.

Fitting a Four Year Computer Science BS Degree into Three Years: A Case Study

Nathan Barker, Laurie Harris
Computer Science & Information Systems Department, Southern Utah University, Cedar City, Utah, USA

Abstract—The typical advertised time to complete a bachelor's degree in the USA is four years. Although many students take much longer than this, some students want to complete their degrees in as short a time as possible. These students could be motivated by getting into the job market sooner, reducing the costs associated with getting their degrees, pressure from parents or family, or pressure from themselves to excel. Beyond student expectations, there are other reasons a University may want to create a path for students to get their degree sooner than usual, such as political pressure. This study describes the challenges of creating a pathway for students to complete a four-year Computer Science degree in 3 years faced by a comprehensive regional university.

Learning and Teaching Undergraduate Introductory Programming Courses in Java - The Use of an IDE vs Command Line

Hussain A. Aljafer, Gary Cantrell
Computer Science and Information Systems, Southern Utah University, Cedar City, Utah, USA

Abstract—A topic of debate among Java teachers is whether to use an Integrated Development Environment (IDE) or a text editor paired with a command line compiler to teach introductory Java programming courses. Is it really in the student's favor to start their programming journey using an IDE or is it better to just use any text editor to write the code and then compile and run the code using the command line? Which approach will help the students really understand the programming concepts and be able to debug their code when they have errors? In this paper we discuss the advantages and disadvantages of both approaches and we compare the performance of two groups of students in which the first group used a text editor with command line commands and the other group used Eclipse IDE.

More on Computer Architecture Simulators for Different Instruction Formats

Xuejun Liang
Department of Computer Science, California State University - Stanislaus, Turlock, California, USA

Abstract—Several simple computer architecture simulators are developed and implemented for four different instruction formats, including stack-based, accumulator-based, two-address, and threeaddress machines. The simulators for the first two machines have been reported in [1]. This paper will present details for the remaining two instruction formats. These simulators can be used to assemble and run assembly language programs on these architectures. Several examples are given to illustrate how to develop assembly language programs to deal with arrays, loops, subroutines, and recursions on these different computer architectures. Sizes and performances of these assembly language programs are discussed. Students will have a better understanding of computer architectures by using these simulators on their assembly language programming assignments. In addition, students can also modify these simulators to add more instructions or pseudo-instructions.

A Lightweight Visual Programming Tool for Machine Learning and Data Manipulation

Ivan Khodnenko, Sergey Ivanov, Kirill Prokofiev
Institute of Design and Urban Studies, ITMO University, Saint Petersburg, Russia

Abstract—In recent decades, visual programming tools have focused on scientific workflow tasks. However, the current focus has shifted to machine learning tasks, which are most often amenable to unification. The existing solutions in this area demonstrate many alternatives with the possibility of using models from various machine learning packages. The proposed solution is based on the concept of minimal graphical notation, where all typical operations on data and machine learning models are possible.

Another distinguishing feature of the proposed solution is the explicit separation of data and models, making the proposed notation clearer for perception. In the proposed engine, computational graph traversal does not have a rigorous sequence. The call of the sequence of calculations depends on the specific model that the user currently needs.

Discovery of Research Trends in Computer Science Education on Ethics Using Topic Modeling

Sarah Parsons, Natalia Khuri
Department of Computer Science, Wake Forest University, Winston-Salem, North Carolina, USA

Abstract—To evaluate how ethics is addressed in computer science education, unsupervised text mining was performed to identify six salient topics and their trends in educational research publications. The majority of articles focused on the integration of ethics into introductory computer science courses, in particular those targeting non-majors. Models derived from a larger data set of research abstracts on ethics in computing, revealed additional topics in human-computer interaction and artificial intelligence, currently underrepresented in computer science education research. The agreement between automatic and manual topic assignments was 89%. Therefore, our approach provides a time-efficient and reproducible evaluation system to support curricular decision-making.

Crowdsourcing Exams to Increase Student Engagement in an Online Information Technology Class: An Experience Report

Lisa L. Lacher, Cody M. Gibson
College of Science & Engineering, Department of CS, University of Houston - Clear Lake, Houston, Texas, USA

Abstract – To achieve better success rates in online courses, we need to cultivate student engagement. Students learn more when they are actively engaged. However, most instructors still use traditional lecture methods. Because instructors have not embraced active learning in their face-to-face classes, many are finding it challenging to incorporate active learning into their courses as they move online. Prior research has found that although students learn more from active learning activities, they do not feel like they are learning more, which can have a detrimental impact to their satisfaction with the activity and the course. Studies have found that with more training and promotion, students' attitudes can change; however, this takes additional time and effort which instructors and students can find discouraging. Our research reports on the experience of using a crowdsourcing activity to create exam questions. Without any promotion from the instructor, students reported that this activity helped them feel engaged with the course materials and helped them learn. We discuss the contributing factors that added to student positivity, as well as challenges in adopting this active learning method.

Unofficial API and Browser Extension Development for Augmenting Student Resources

Trey Stone, Asma Ahmed, Travis Vensel, Yanyan Li
Department of Computer Science & Information Systems, California State University San Marcos, California, USA

Abstract— Official university resources for students have been grossly outpaced by the technological expectations of incoming generations, often being supplemented with third-party services to provide desired functionality. Enrollment services are a common example of this, typically lacking any means to view course feedback and evaluations. In response, students use external review sites such as RateMyProfessors (RMP) for this information, manually searching for and assessing each professor's feedback for every course they consider. Integration of these reviews into the student portal would be a substantial optimization, but the closed-source nature of the sites prevents direct innovation from outside developers. To solve this, we created a browser extension for the California State University San Marcos student site capable of injecting RMP rating data directly alongside existing course information. The resulting product allows for inline availability of the requested content without it needing to manually fetched for each professor, greatly improving convenience of access and enabling more informed decision-making.

Operation Results of the First Year of IoT Making Things Program at University

Takuya Saito
Faculty of Engineering, Shonan Institute of Technology, Fujisawa, Japan

Abstract— This paper presents the background of the 'IoT: Making Things' program established at Aichi University of Technology in Japan, as well as its educational mechanism and results after first year of offering. 'IoT: Making Things' is a crossdisciplinary special program offered to a class of 15 students selected from the three departments of Aichi University of Technology. Students wishing to enroll in the 'IoT: Making Things' program are selected through application form submission and evaluation of GPA scores and interviews. In 2017, the first year of offering, there were many more applicants than anticipated. The results of this study confirmed that students and society have high expectations and demands for IoT-related programs.

CSCI-ISPC:
SIGNAL & IMAGE PROCESSING, COMPUTER VISION & PATTERN RECOGNITION

A Review of Convolutional Neural Networks and Gabor Filters in Object Recognition

Mehang Rai, Pablo Rivas
Department of Computer Science, Baylor University, Waco, Texas, USA

Abstract—Convolutional neural networks (CNNs) have become a classic approach to solving challenging computer vision problems. Much of its success is due to its ability to discover optimal filters that capture non-trivial spatial relationships in data. Other vital components include advances in optimization, regularization, and overfitting prevention strategies. However, recently, researchers have observed closely the connections between what CNNs learn in the layers that capture low-level features and filter-banks such as Gabor filters. Gabor filters have been used in computer vision tasks long before CNNs were popularized with good performance. This paper presents a review of the literature concerning approaches that involve both Gabor filters and CNNs. We pay close attention to successes and opportunities for future research in the intersection of these two computer vision tools.

Comparisons of Full Body and Facial Dog Identification

Ridnarong Promya, Somying Thainimit, Chalermpol Charnsripinyo, Yasuharu Koike
Department of Electrical Engineering, Kasetsart University, Bangkok, Thailand
Internet Innovation Research Team, National Electronics and Computer Technology Center, Bangkok, Thailand
Department of Information and Communication Engineering, Tokyo Institute of Technology, Tokyo, Japan

Abstract—Free-roaming dog population survey is very important in the veterinary research area. Modern technologies can promote the automation process of dog surveys. This paper investigates two networks, which are Facenet and Dlib for dog identification. These two networks predict embedded vectors which are a reduced dimension vector of images containing distance information that is aligned with dog identities. The full body and face images of dogs are fed into the networks. Our experiments show that facial images provide more discriminative embedded vector than full body images. The average accuracy of both networks with face and full body images are 91.43% and 87.32%, respectively. In overall, both networks provide good accuracy and promising results for dog recognition applications.

Complexity-Based Convolutional Neural Network for Malware Classification

Kenneth Brezinski, Ken Ferens
Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada

Abstract—Malware classification remains at the forefront of ongoing research as the prevalence of metamorphic malware introduces new challenges to anti-virus vendors and firms alike. One approach to malware classification is Static Analysis - a form of analysis which does not require malware to be executed before classification can be performed. For this reason, a lightweight classifier based on the features of a malware binary is preferred, with relatively low computational overhead. In this work a modified convolutional neural network (CNN) architecture was deployed which integrated a complexity-based evaluation based on box-counting. This was implemented by setting up max-pooling layers in parallel, and then extracting the fractal dimension using a polyscalar relationship based on the resolution of the measurement scale and the number of elements of a malware image covered in the measurement under consideration. To test the robustness and efficacy of our approach we trained and tested on over 9300 malware binaries from 25 unique malware families. This work was compared to other award-winning image recognition models, and results showed categorical accuracy in excess of 96.54%.

Multi-Class Weather Classification Using ResNet-18 CNN for Autonomous IoT and CPS Applications

Qasem Abu Al-Haija, Mahmoud A. Smadi, Saleh Zein-Sabatto
Department of Electrical and Computer Engineering, Tennessee State University, Nashville, Tennessee, USA
Department of Electrical Engineering, The Hashemite University (HU), Zarqa, Jordan

Abstract — Severe circumstances of outdoor weather might have a significant influence on the road traffic. However, the early weather condition warning and detection can provide a significant chance for correct control and survival. Therefore, the auto-recognition models of weather situations with high level of confidence are essentially needed for several autonomous IoT systems, self-driving vehicles and transport control systems. In this work, we propose an accurate and precise self-reliant framework for weather recognition using ResNet-18 convolutional neural network to provide multiclass weather classification. The proposed model employs transfer learning technique of the powerful ResNet-18 CNN pretrained on ImageNet to train and classify weather recognition images dataset into four classes including: sunrise, shine, rain, and cloudy. The simulation results showed that our proposed model achieves remarkable classification accuracy of 98.22% outperforming other compared models trained on the same dataset.

Scene Text Recognition With Linear Constrained Rectification

Gang Wang, Hua-Ping Zhang, Jian-Yun Shang
School of Computer Science, Beijing Institute of Technology, Beijing, P. R. China
School of Software, Beijing Institute of Technology, Beijing, P. R. China

Abstract—Scene Text Recognition remains a challenging problem because of various text styles and image distortions. This paper proposed an end-to-end trainable model with a rectification module network. The rectification module adopts a polynomial based spatial transform network to rectify the distorted input image, the feature representation between the rectification and encoding step is shared. The model can be trained with the scene text images and the corresponding word labels. With the flexible rectifying and feature sharing, this model outperforms previous works through the extensive evaluation results on the standard benchmarks, especially on irregular datasets, 80.2% on IC15 and 85.4% on CUTE, more specifically.

Defect Detection in PV Arrays Using Image Processing

Akshat V. Patel, Lifford McLauchlan, Mehrube Mehrubeoglu
Department of Electrical Engineering and Computer Science, Texas A&M University - Kingsville, Texas, USA
Department of Engineering, Texas A&M University - Corpus Christi, Texas, USA

Abstract—Renewable energy continues to be an important part of the energy field. New wind, photovoltaic (PV) and hydro plants continue to be installed. Many PV farms have been in operation for years and are exhibiting defects and less efficient operation due to age as well as being constantly exposed to weather conditions, such as rain, ice, cold, heat and hail. In this research image processing operations are applied to PV panels to determine defects or damaged areas/panels. The proposed method can be utilized in real-time to determine the damaged areas and count the number of damaged panels.

Image Classification of High-Performance Liquid Chromatography Chromatograms with Neural Networks

Akhil Mandalapu, Sebastian Calzadilla, Sean Mondesire
School of Science, St. Thomas University, Miami, Florida, USA

Abstract—High-Performance Mass Spectrometry Liquid Chromatography (MS-HPLC) data analysis has led to insights within the fields of metabolomics and guides researchers to potential genes of interest for future cancer research. Current MS-HPLC classification techniques are time-intensive and require human supervision. Statistical and automated approaches, such as denoising and feature detection, have been explored with moderate success. An approach that has yet to be adequately evaluated for MS-HPLC is feature classification through artificial neural networks. To effectively implement a neural network method for feature classification, the optimal network structure must be determined. Discovering the optimal structure could lead to potential automation of MS-HPLC feature classification and could result in saving time for researchers. Therefore, this research determines the optimal neural network structure for MS-HPLC image classification. Results indicate high accuracy of 96.8% and could lead to the automation of HPLC-MS data analysis.

Quantifying Plastic Bottle Debris in Waterways Using Image Processing

Katlin Walden, Mehrube Mehrubeoglu
Department of Engineering, Texas A&M University - Corpus Christi, Texas, USA

Abstract— Waterway debris can impact many aspects of marine environments including marine navigation and recreational fishing in bodies of water. Waterway plastic debris pollution is occurring at such a large scale globally that it is onerous to track and quantify the sheer quantity of misplaced plastics. Digital image processing is one effective way to monitor waterway debris, but the associated complex processes have inherent challenges. The variety of debris size and shape renders all plastics detection impractical without expert review. In this project, image processing algorithms are utilized to detect debris pieces, and estimate the total weight and volume of all plastic bottles in a scene based on known weight and volume of plastic water bottles. The typical weight and volume of plastic water bottles has been used as a reference to estimate the total weight and volume of plastic bottle debris as quantitative measures in digital waterway debris images. The results of estimated plastic bottle debris weight and volume from a variety of debris scenes are presented in this research.

Correction of Gain Mismatch for Time Interleaved Analog to Digital Converter System

Rohan K. Balar, Sung-won Park
Department of Elcectrical Engineering and Computer Science, Texas A&M University-Kingsville, Kingsville, Texas, USA

Abstract— For digital signal processing, an analog signal should be sampled at least at the Nyquist rate. In many applications such as radar systems and software defined radios that deal with super high frequency, required sampling rates may exceed currently available analog-todigital converters (ADCs). To increase the sampling rate time interleaved ADC (TI-ADC) systems are often used. In TI-ADC system several ADCs are interleaved to increase the sampling rate. Ideally all ADCs in the system will have the same gain. Due to imperfection of ADCs there is a gain mismatch. The gain mismatch results in spurious peaks in the spectrum and reduces the dynamic range of the TI-ADC system. In this paper we will first explain how the gain mismatch affect to the signal spectrum. We will also devise a method to estimate the gain mismatch and correct it.

Object Detection in Haze Enhanced by Preprocessing Image Dataset with Synthetic Haze

Binghan Li, Yindong Hua, Mi Lu
Electrical and Computer Engineering, Texas A&M University, Texas, USA
Electrical and Computer Engineering, Stony Brook University, New York, USA

Abstract - Object detection in hazy environment has always been a difficult task in the autonomous driving field. Huge breakthrough is hard to achieve due to the lack of large-scale hazy image dataset with detailed labels. In this work, we present a simple and flexible algorithm to generate synthetic haze to MS COCO training dataset, which aims to enhance the performance of object detection in haze when taking the new synthesized hazy images as training dataset. Our algorithm is inspired by the Multiple Linear Regression Dark Channel Prior (MLDCP), and we obtain a general model that can add synthetic haze to haze-free images by implementing Stochastic Gradient Descent (SGD) to the reversed MLDCP model. We further evaluate the mean average precision (mAP) of Mask R-CNN when we train the network with the Hazy-COCO training dataset and preprocessing test hazy dataset with existing single image dehazing algorithms.

A Fast Histogram Equalization and KDE to aid a Supervised Algorithm to Count Eucalyptus Seedlings

Guilherme Pereira Jorge Franze, Emanuel Rocha Woiski
South America Zone Analytics Department, Ambev SA, Jaguariuna, Brazil
Mechanical Engineering Department, Sao Paulo State University, Ilha Solteira, Brazil

Abstract—In a previous work, we developed two supervised approaches for automatic seedling counting in a UAV high-definition image from an Eucalyptus plantation stand of approximately 62 acres. We observed that the deployment of accessible and faster algorithms for automatic seedling/tree detection and counting suitable for field laptops are scarce. Moreover, researchers have reported troubles with the variable quality of the acquired aerial imagery, which plays a fundamental role in the algorithms degrading their discriminative power. Analyzing the bimodal BNDVI pixel distribution histograms, we noticed that, regarding the illumination quality, the distance between the two modes for any distribution seems to remain almost invariant. Therefore, the main core of the present work was to devise a fast BNDVI histogram equalization and to deploy a KDE in order to apply to a one-dimensional supervised counting method. From the resulting PDF, we defined a hard threshold to structure the solution as a binarized classification problem. We used the F1- score to predict the amount of seedlings within properly defined testing tiles taken from the large image. For our case of study, since the precision and recall resulted larger than 95%, the FP and FN are very small in comparison to the true number of seedlings.

StimulEye: A Computer Vision Based Concussion Detector

Nicolas Elia, David Monge, Charles Kinzel, Fadi Muheidat
Computer Science and Engineering, California State University, San Bernardino, California, USA

Abstract— Traumatic brain injuries (TBI) are highly prevalent in contact sports, affecting the physical safety of athletes of all ages. Despite this, there is currently no affordable and objective way to test for concussions on the sidelines. Current sideline concussion tests are either trivial and subjective or overly complex and financially unfeasible. This paper presents a method for concussion detection using a smartphone. The proposed smartphone TBI test uses the camera to analyze the reactions of players' pupils when subjected to the external stimulus provided by the smartphone's LED. The results of this research show that sufficient data and accuracy are obtained by the application, StimulEye, to efficiently and accurately detect a concussion.

Exploring Generalization Capability for Video Forgery and Detection based on Generative Adversarial Network

Ying Lin, Yanzhen Qu, Yuanpei Li, Zhishen Nie
School of Software, Yunnan University, Kunming, P. R. China
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA
School of Software, Zhengzhou University, Zhengzhou, P. R. China

Abstract-With the development of digital image processing technology based on deep learning, the potential risk of using related technologies to threaten the security of multimedia information is increasing. Because the generated human face effect largely depends on the completeness of the input sample set, most of the current deep forgery models have the problem of human side-face collapse. This paper has studied the deep forgery technology of Deepfacelab and Faceswap, and adjusts the original auto-encoder-based model architecture to a generative adversarial network. By using the harmonic mean of cross entropy and mean square error as the loss function, the improved model can reduce the probability of some frames being discarded during training. Meanwhile, by adjusting key characteristics and the weights of features in different frames, it further optimizes the cross-dataset detection performance. Experimental results have shown that the improved model can keep more facial details while still maintain high human face clarity. The detection performance is improved and the cross-dataset average error rate of the deep detection model is about 35%.

Improved Image Semantic Segmentation Based on Cascade Data Augmentation

Khwaja Monib Sediqi, Hyo Jong Lee
Department of Computer Science and Engineering, Jeonbuk National University, Jeonju, South Korea

Abstract— A large portion of the research for the task of image semantic segmentation focuses on optimizing deep learning networks to improve the state-of-the-art results. In this work, however, we take a different approach in order to enhance the performance of state-of-the-art methods for image semantic segmentation. Benefiting from the internal learning process of deep learning models we propose a Cascade Data Augmentation (CDA) technique that can effectively increases the size of data based on the characteristics of the dataset in hand. CDA utilizes basic image transformation techniques such as geometric and photometric transformations to augment the size of data. In particular, CDA blends these transformations by assigning two adjustable parameters that are 1) the probability of applying the transformation operation and 2) the magnitude of the operation. We evaluate the effectiveness of our method on the validation set of Pascal Context benchmark dataset for image semantic segmentation. Experimental results prove that incorporating CDA with baseline deep learning models can increase their performance. In particular we attain a average gain of ≈ 4.72 pixel accuracy and ≈ 1.9 mean intersection over union PascalContext dataset. Moreover, CDA is a generic method that can be integrated with any deep learning model and attains improved performance while adding trivial extra training time of the network.

Study on Parachute Entanglement Prevention Method Using Image Recognition in CanSat

Miho Akiyama, Takuya Saito
Graduate School of Electrical and Information Engineering, Shonan Institute of Technology, Fujisawa, Japan
Faculty of Engineering, Shonan Institute of Technology, Fujisawa, Japan

Abstract—In this paper, we propose a method to prevent CanSat from getting entangled in the parachute using image recognition, and describe the results of evaluating its effectiveness in actual CanSat competition. There is a simulated small planetary exploration rover called CanSat, and CanSat competitions are held in various regions as an experiment similar to the development of the most feasible planetary exploration rover on the ground. In the CanSat competition, assuming the development of an actual planetary exploration rover, it is required to have a structure that can withstand impacts such as fully autonomous control and rocket launch similar to the planetary exploration rover, release in the sky, and landing on the ground. Therefore, landing is performed using a deceleration mechanism such as a parachute, but there are many cases where CanSat gets entangled with the parachute and becomes unable to run after the start of running. To solve this problem, we propose a parachute avoidance operation method for CanSat that uses a parachute structure and image recognition that are difficult to get entangled with CanSat. In addition, by actually confirming the operation at the CanSat competition and using the proposed method, it was confirmed in a demonstration evaluation experiment that CanSat was not entangled in the parachute.

On the Discriminative Properties of Principal Component Analysis based on L1-Norm

Ruben Martin-Clemente, Vincent Zarzoso, J. L. Camargo Olivares
Department of Signal Processing and Communications, University of Seville, Seville, Spain
CNRS, I3S Laboratory, University of Cote d'Azur, Sophia-Antipolis, France

Abstract—Principal Component Analysis (PCA) is one the most widely-used techniques for the analysis of multivariate data. Unfortunately, PCA is extremely sensitive to the presence of large outliers in the data. To overcome this drawback, a robust variant of standard PCA, based on the L1-norm, has been proposed in recent years. This variant, called L1-PCA behaves like traditional PCA, while offering robustness against the presence of large outliers in the data. This paper shows that, combined with a whitening pre-processing, L1-PCA is also endowed with discriminative properties, allowing it to solve binary classification problems in an unsupervised way, thus sparing the need for training data.

Securing Three Dimensional Regions with Stereo Vision

Leonard D. Litvak, Leonidas Deligiannidis
Department of Computer Science & Networking, Wentworth Institute of Technology, Boston, MA, USA

Abstract—Although there exist many different solutions for securing volumes of space with various kinds of security cameras such as conventional monocular or thermal cameras, using stereo vision has not been attempted as an alternative method. By combining motion detection via image subtraction and range-finding via stereoscopic vision, a single stereo camera can secure a large three-dimensional volume within its field of view. With this, many individual areas can be secured by a single stereo camera placed further away that would otherwise have to be secured by many conventional or thermal cameras placed near areas of interest. Our solution works well for depth-ranges of up to 20 feet. At ranges beyond 20 feet, the disparity map becomes too noisy. Within the 20 feet range, the depth detection is highly accurate with only 0.18% error.

Determining the Number of Endmembers of hyperspectral Images using Clustering

Jose Prades, Addisson Salazar, Gonzalo Safont, Luis Vergara
Institute of Telecommunications and Multimedia Applications, Universitat Politecnica de Valencia, Valencia, Spain

Abstract— Some applications require knowing how many materials are present in the scene represented by a hyperspectral image. In a previous paper, we presented an algorithm that estimated the number of materials in the scene using clustering principles. The proposed algorithm obtains a hierarchy of image partitions and selects a partition using a validation index; the estimated number of materials is set to the number of clusters of the selected partition. In this algorithm, the user must provide the image and the maximum number of materials that can be estimated (P). In this paper, we have extended our algorithm so that it does not require P as input parameter. The proposed method iteratively performs the estimation for several increasing values of P and stops the process when a certain condition is met. The results obtained with five hyperspectral images show that our algorithm approximately estimates the number of materials in that images.

Multi-Stage CNN-Based Monocular 3D Localization and Pose Estimation

Ali Babolhavaeji, Mohammad Fanaei
Department of Electrical and Computer Engineering and Computer Science, University of Detroit Mercy, Michigan, USA

Abstract—This paper aims to design a 3D object detection model from 2D images taken by monocular cameras by combining the estimated bird's-eye view elevation map and the deep representation of object features. The proposed model has a pretrained ResNet-50 network as its backend network and three more branches. The model first builds a bird's-eye view elevation map to estimate the depth of the object in the scene and by using that estimates the object's 3D bounding boxes. We have trained and evaluate it on two major datasets: a syntactic dataset and the KIITI dataset.

Multiple Attention Mechanism Neural Network in Garment Image Segmentation

Xu Yingheng, Zhong Yueqi
College of Textile, Donghua University, Shanghai, P. R. China
Key Laboratory of Textile & Technology, Ministry of Education, Donghua University, Shanghai, P. R. China

Abstract - Instance segmentation of clothing images is a task that fashion analysts are paying more and more attention to in recent years. The segmentation of different clothing components allows designers to better design new fashion items and allows consumers to better understand design concepts. Most of the current methods are based on deep convolutional neural networks (DCNN). However, most of the current instance segmentation neural networks are limited by the size of the receptive field and cannot capture the global dependence. For clothing images, the use of contextual information between different clothing and collocations can obtain fine-grained and higher clothing segmentation images. Previous studies have shown that attention-based methods can obtain non-local dependencies of the whole image and are mostly used for panoramic segmentation of aerial images. For instance segmentation of clothing images, we propose a new dual-branch attention module based on the Non-local attention mechanism, called Multiple Attention MaskRCNN (HAMaskRCNN). Specifically, for the attention module, we use two branches: position attention and channel attention. After feature fusion, the FPN module and the attention module are connected in parallel to form a multiple attention module. We use the Imaterialist-fashion (2019) dataset to conduct experiments and compare with the benchmark to prove the effectiveness of our HAMaskRCNN.

A Chest X-ray Image Retrieval System for COVID-19 Detection using Deep Transfer Learning and Denoising Auto Encoder

O. F. Layode, Md Mahmudur Rahman
Computer Science Department, Morgan State University, Baltimore, Maryland, USA

Abstract— The COVID-19 pandemic is the defining global health crisis of our time which is currently challenging families, communities, health care systems, and government all over the world. It is critical to detect and isolate the positive cases as early as possible for timely treatment to prevent the further spread of the virus. It was found in few early studies that patients present abnormalities in chest radiography images that are characteristic of those infected with COVID-19. In the current context, a rapid, accessible and automated screening tool based on image processing of chest X-rays (CXRs) would be much needed as a quick alternative to PCR testing, especially with commonly available X-ray machines and without the dedicated test kits in labs and hospitals. Several classifications based approaches have been proposed recently with encouraging results to detect pneumonia based on CXRs using supervised deep transfer learning techniques based on Convolutional Neural Networks (CNNs). These black box approaches are mainly non-interactive in nature and their prediction represents just a cue to the radiologist. This work focuses on issues related to the development of such an automated system for CXRs by performing discriminative feature learning using deep neural networks with a purely data driven approach and retrieving images based on an unknown query image and performing retrieval evaluation on currently available benchmark datasets towards the goal of realistic comparison and real clinical integration. The system is trained and tested on an image collection of 1700 CXRs obtained from two different resources with encouraging results based on precision and recall measures in individual deep feature spaces. It is hoped that the proposed system as diagnostic aid would reduce the visual observation error of human operators and enhance sensitivity in testing for Covid-19 detection.

A Fast and Efficient Method for Detection of Seizure in Elettroencephalogram using Log-energy Entropy and Support Vector Machine

Luigi Pavone, Jaime F. Delgado Saa, Slavianka Moyanova
IRCCS NEUROMED, Pozzilli, Italy
Department of Electrical and Electronics Engineering, Universidad del Norte, Barranquilla, Colombia

Abstract— We present a method for automatic detection of seizures in EEG that might help clinicians by speeding up the process of seizure detection. The method consists of extraction of Log-Energy Entropy from band-passed EEG and use of a Support Vector Machine (SVM) classifier. Furthermore, using multiple regression analysis, we evaluated the effect of some characteristics of the patients on the performance of the method. We found that the type of epilepsy is the major factor, which influenced the performance of the method. The high performance of the method makes it feasible also for real-time applications.

Object Detection in Degraded Visual Environments using Compressive Sensing

Mohammed Abuhussein, Aaron Robinson
Department of Electrical and Computer Engineering, The University of Memphis, Tennessee, USA

Abstract—Compressive Sensing (CS) has proven its ability to reduce the number of measurements required to reproduce images with similar quality to those reconstructed by observing the Shannon-Nyquist sampling criteria. By exploiting spatial redundancies, it was shown that CS can be used in target recovery and object detection. In this paper we propose a method that incorporates an effcient use of CS to locate a specific object in zero-visibility environments. We show that with the use of an over-complete dictionary of the target our technique can perceive the location of the target from hidden information in the scene. This paper will compare previously implemented algorithms with our, list the shortcomings evident in their outputs, explain our setups, detail the differences in dictionary structures, and present quantified results to support its efficacy in the results section.

New Applications of an Oversampling Method based on Generative Adversarial Networks

Addisson Salazar, Luis Vergara, Gonzalo Safont
Institute of Telecommunications and Multimedia Applications, Universitat Politecnica de Valencia, Valencia, Spain

Abstract—This paper presents new applications of a novel oversampling method based on Generative Adversarial Networks (GAN). Two challenging applications are approached: classification of stages of a neuropsychological test (Barcelona test) using electroencephalographic (EEG) data from epileptic patients; and classification of sleep stages using electrocardiographic (ECG) data from apnea patients. The results of the GAN-based method are compared with the commonly used interpolation-based method synthetic minority oversampling technique (SMOTE). The results show the superiority of the GAN method over SMOTE to improve the probability of error of the two-class classification proposed applications.

Detecting Keypoints for Automated Annotation of Bounding Boxes using Keypoint Extraction

Kaito Ishizaki, Kasuki Saruta, Hiroshi Uehara
Department of Management Science and Engineering, Akita Prefecural University, Akita, Japan
Department of Information and Computer Science, Akita Prefecural University, Akita, Japan

Abstract—Object detection requires an enormous amount of training data annotated by bounding boxes. All bounding boxes are manually drawn, which leads to highly expensive labor costs. Therefore, this study proposes automatic bounding box annotation of training data for object detection. The keypoints to identify object regions in pictures are extracted, which can then be used for drawing bounding boxes automatically, thus, reducing manual labor requirements. When our proposed method is used for pictures of road signs, keypoints that identify road sign regions in the pictures are detected; these keypoints are found to be highly accurate for drawing bounding boxes.

A Computer Vision Framework for Quantification of Feather Growth Patterns

Tyler Thompson, Anna Vickrey, Mike Shapiro, Edward Hsu
Department of Biomedical Engineering, University of Utah, Salt Lake City, Utah, USA
Department of Biology, University of Utah, Salt Lake City, Utah, USA

Abstract—This study aims to examine phenotypic variations in pigeon feather growth patterns by rendering computed tomography (CT) scans as point clouds, and developing machine learning based, feature extraction techniques to isolate the feathers, and map the growth patterns on the skin.

Performance Analysis of Network Pruning for Deep Learning based Age-Gender Estimation

Autumn Knight, Byeong Kil Lee
Department of Electrical and Computer Engineering, University of Colorado, Colorado Springs, Colorado, USA

Abstract—With the advance of visual AI technology, agegender estimation plays a fundamental role in identifying individuals. As deep learning technologies are emerging, identification schemes show significant progress and can handle many challenges of unconstrained imaging conditions. Research on age-gender estimation has begun applying deep convolutional neural networks (CNN) as a framework. However, due to large memory footprints and computational workloads, deep neural networks are hard to apply to on-device training and inference for embedded devices which have limited hardware resources. To solve this issue, network model pruning has been proposed as an efficient approach to reduce the model redundancy without significant degradation of the performance. In this paper, we modeled and characterized several pre-training models with architecture variations on

baseline age-gender estimation before applying pruning schemes. For each of the models, three types of pruning comprised of weight, layer and filter pruning are applied and the pruning results are analyzed in terms of complexity and accuracy to find optimal pruning conditions. Combined schemes of pre-training models and network pruning techniques are discussed, and their results are compared with the original model's. Based on our experiments, the actual size of a fully trained prediction model can be reduced by as much as 90% with an accuracy loss of 2%~9%.

Development of Image Pre-processing System for GEO-KOMPSAT-2 GOCI-II

Jinhyung Park, Hyun-su Lim, Jun-Yeong Bok
Image Data System Development Division, Korea Aerospace Research Institute, Daejeon, Korea

Abstract— Korea Aerospace Research Institute launched GEO-KOMPSAT-2B, the second satellite of its Geostationary Observation Satellite series, in February 2020. The GOCI-II is a payload embedded on GEO-KOMPSAT-2B continuing GOCI's geostationary ocean monitoring for Korean Peninsular as well as Earth fulldisk. It is expected that more various marine information production with its 4-time improved performance in spatial resolution. In this paper, we introduce development of DPS to pre-process image data of GOCI-II. The GOCI-II DPS is designed to process in real-time and reliable without operator's interference. Completing in-orbit tests, the GOCI-II DPS currently starts its nominal operations during the next 10 years of GEO-KOMPSAT-2B lifetime to provide useful ocean image data.

Next Generation of Gallery Sharing in VR

Wyatt Phillips, Leonidas Deligiannidis
Department of Computer Science & Networking, Wentworth Institute of Technology, Boston, MA, USA

Abstract—The use of digital photo albums, social media posts, and embedded videos connect people closer to their memories. We believe the next generation of visual albums will immerse individuals through modern virtual reality technologies. This paper examines new methods for enhancing photo albums and visual content; achieved in highly interactive and realistic environments where users are presented with interactable frames, 360 imagery, videos, and dynamic exhibits.

Image-Based Determination of the Growth or Shrinkage of Wounds at the Dermal Layer

Berra Z. Barkana, Duha A. Barkana, Miad Faezipour
Masuk High School, Monroe, CT, USA
Departments of Computer Science & Engineering and Biomedical Engineering, University of Bridgeport, Bridgeport, CT, USA

Abstract—Acute and chronic wounds affect people of all ages around the world. Monitoring the size of open wounds has been proven to be effective for the diagnosis and prognosis of several medical complication such as ulcers, chronic wounds, and a variety of infections. This work presents an image-based approach to calculate the percentage of growth or shrinkage of wounds at the dermal layer. Image segmentation is applied to determine the region of interest in the wound image. Changes in the wound size (area) are computed to quantify the healing percentage.

Human Temperature Scanning from a Distance

Leonidas Deligiannidis
Department of Computer Science and Networking, Wentworth Institute of Technology, Boston, MA, USA

Abstract—In this work we present an inexpensive, yet accurate, solution of measuring human temperature from a distance. The need for such solution is essential during pandemics. During COVID-19, one of the most common symptoms, for those who develop symptoms, is fever. We believe a tool that measures multiple peoples' temperature from a safe distance can be valuable.

As people enter buildings, airports, hospitals, etc. they can be scanned automatically from a safe distance. The system can alert the authorities for further assessment. Even though such a tool does not prevent the spread of a virus by itself, it can help contain the virus following additional measures such as wearing a face mask, frequent hand washing, and social distancing.

CSCI-ISMC:
MOBILE COMPUTING, WIRELESS NETWORKS, & SECURITY

Traffic-based Congestion Management Algorithm for Wireless Sensor Networks

Tshimangadzo M. Tshilongamulenzhe, Topside E. Mathonsi, Maredi I. Mphahlele, Dp. Du Plessis
Department of Information Technology, Tshwane University of Technology, Pretoria, South Africa

Abstract— In this research paper, Wireless Sensor Networks (WSNs) is described as a self-organized network system formed by the collection of nodes that are capable of sensing, monitoring, processing, and communication in an autonomous manner. This WSNs uses real-time applications for monitoring each event that happened during the packet transmissions. Sensor nodes (SNs) are deployed randomly in a specific environment. Each sensor node uses a short-range transmitter, to route the packets from source node towards the destination node. This research paper demonstrated the network congestion problem that happens during packet transmission from one node to another node in WSNs. Moreover, network congestion normally happens when packets distribution in a network overload both transmission channel and nodes which leads to link congestion in WSNs. This further decrement the packet delivery ratio, network throughput, and high end-to-end delay. Therefore, the Traffic-based Congestion Management (TCM) algorithm is proposed to manage network congestion in WSNs. The TCM algorithm was developed by integrating two algorithms namely: Routing Congestion Control (RCC) algorithm and Traffic Rate Adjustment (TRA) algorithm. Network Simulator 2 (NS-2) was used as a simulation tool. Simulation results showed that the proposed TCM algorithm performed better than RCC and TRA algorithms with the improvement of packet delivery ratio by 68.7%, network throughput increased by 98.9 %, and end-to-end delay reduced by 34.2%.

Controller Area Network Security Requirements

Vinayak Tanksale
Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA

Abstract—Controller Area Network (CAN) is the dominant communication standard for intra-vehicle communications in automobiles. The CAN protocol is designed to be light-weight in order to increase speed and efficiency. However, this severely limits the protocol's ability to support any kind of security countermeasures. Some of the devices that utilize CAN lack the resources to perform the required cryptographic computations. The rapid increase in automobile communications has exacerbated the need for efficient security solutions. This paper addresses requirements for such CAN security solutions.

Lightweight Multi-factor Authentication for Underwater Wireless Sensor Networks

Ahmed Al Guqhaiman, Oluwatobi Akanbi, Amer Aljaedi, C. Edward Chow
Department of Computer Science, University of Colorado Colorado Springs, Colorado, USA
Department of Computer Networks and Communications, King Faisal University, Hofuf, Saudi Arabia
College of Computing and Information Technology, University of Tabuk, Tabuk, Saudi Arabia

Abstract—Underwater Wireless Sensor Networks (UWSNs) are liable to malicious attacks due to limited bandwidth, limited power, high propagation delay, path loss, and variable speed. The major differences between UWSNs and Terrestrial Wireless Sensor Networks (TWSNs) necessitate a new mechanism to secure UWSNs. The existing Media Access Control (MAC) and routing protocols have addressed the network performance of UWSNs, but are vulnerable to several attacks. The secure MAC and routing protocols must exist to detect Sybil, Blackhole, Wormhole, Hello Flooding, Acknowledgment Spoofing, Selective Forwarding, Sinkhole, and Exhaustion attacks. These attacks can disrupt or disable the network connection. Hence, these attacks can degrade the network performance and total loss can be catastrophic in some applications, like monitoring oil/gas spills. Several researchers have studied the security of UWSNs, but most of the works detect malicious attacks solely based on a certain predefined threshold. It is not optimal to detect malicious attacks after the threshold value is met. In this paper, we propose a multi-factor authentication model that is based on zero-knowledge proof to detect malicious activities and secure UWSNs from several attacks.

Development of an Effective and Secure Communication System in a Quarantine Situation

Madara Pratt, Sarma Cakula, Ginta Majore, Egons Buss
Vidzeme University of Applied Sciences, Valmiera, Latvia
Latvijas Mobilais Telefons Sia (LMT), Riga, Latvia

Abstract— Most of the human interaction is non-verbal, and communication using technology-based means is not as effective and face-to-face meetings. There is a new set of complex challenges in organizations when face-to-face communication is transferred to technology-based communication. Not all people have the natural ability for effective communication. It is especially critical within online communication. The global pandemic in 2020 emphasized the importance of technology in our work communication. The new digitalized work setting creates new challenges and problems. Technology-based communication system developers have to consider many constructs. The aim is to investigate the impact of communication technologies on employee motivation and satisfaction, recognize the challenges and develop 4EM (For Enterprise Modeling) model for effective communication system in an quarantine situation for secure remote workplace. Recommendations for organization leaders were presented and the model of requirements using 4EM approach has been created.

Intelligent Energy Efficiency Algorithm for the 5G Dense Heterogeneous Cellular Networks

Topside E. Mathonsi, Tshimangadzo M. Tshilongamulenzhe
Department of Information Technology, Tshwane University of Technology, Pretoria, South Africa

Abstract – Over the past years, telecom network operators have seen an endlessly growing demand for ubiquitous highspeed wireless access and an exceptional increase in connected wireless devices. As a result, we have seen a high growth in traffic volumes. The fifth-generation (5G) heterogeneous cellular networks (HetNets) have been developed by telecom network operators to meet the growing mass data capacity. The 5G dense HetNets is made up of multiple radio base stations (RBSs)/small base stations (SBSs) to increase the coverage and system capacity which led to a high number of network elements. Hence, significantly increases power consumption and lessening the energy efficiency for the telecom network operators. Energy efficiency algorithms have been developed for the 5G dense HetNets, however, these existing energy efficiency algorithms do not satisfy the throughput QoS requirements such as minimalized packet loss, longer battery lifetime, reliability, and high data rates. In addition, real-time traffic types such as voice and video requires high computational load at the terminal side, which have an undesirable impact on energy/battery lifetime which further affects the throughput QoS performance. As a result, this paper proposed an Intelligent Energy Efficiency (IEE) algorithm for throughput QoS and energy efficiency enhancement in 5G dense HetNets. In the proposed

IEE algorithm, a deep neural network (DNN) was used to determine the cell capacity ratio for the SBSs. Hence, the SBSs cell capacity ratio was employed as decision criteria to put the SBSs into a sleep state. In addition, transferable payoff coalitional game theory was used in order to ensure real-time applications have a higher priority over non-real time applications. Numerous computer simulation results illustrated that the proposed IEE algorithm experienced an average packet loss of 2.6%, the energy consumption of 3.4 joules, and produced a network throughput of 97.4%.

Replay Spoof Attack Detection using Deep Neural Networks for Classification

Salahaldeen Duraibi, Wasim Alhamdani, Frederick T. Sheldon
Department of Computer Science, University of Idaho, Moscow, Idaho, USA
Department of Computer Science, University of Jazan, Jazan, Saudi Arabia
Department of Computer and Information Sciences, University of the Cumberlands, Williamsburg, USA

Abstract— In this paper, we explore the use of the deep learning approach for replay spoof detection in speaker verification systems. Automatic speaker verifications (ASVs) can be easily spoofed by previously recorded genuine speech. In order to counter the issues of spoofing, detecting spoofing attacks play an important role. Hence, we consider the detection of replay attack spoofing that is the most easily accomplished spoofing attack. In this light, we propose a deep neural network-based (DNN) classifier using a hybrid feature from Mel-frequency cepstral coefficient (MFCC) and constant Q cepstral coefficient (CQCC). Several experiments were conducted on the latest version of ASVspoof 2017 dataset. The results are compared with a base line system that uses the Gaussian mixture model (GMM) classifier with different features that include MFCC, CQCC, and the hybrid feature of the two. The experiment results reveal that the DNN classifier outperforms the conventional GMM classifier. It was found that the hybrid-based features are superior to single features, such as CQCC and MFCC in terms of equal error rate (ERR). In addition, like many previous researchers have found, it turned out that high-frequency regions of speech utterance convey much more discriminative information for replay attack detection.

A Framework for Mobile Malware Forensics

Abdullah Mujawib Alashjaee, Michael Haney
Computer Science Department, University of Idaho, Moscow, Idaho, USA
Computer Science Department, Northern Borders University, Arar, Saudi Arabia

Abstract—Mobile malware is a serious threat to mobile devices. Effective forensic investigations are needed to identify malware authors and thwart this growing threat. This paper investigates existing conventional digital forensics tools against mobile malware. In addition, the paper proposes a mobile malware forensics framework that combines readiness, live, and reactive forensic measures that provide detection, acquisition, and the steps taken to analyze acquired data in labs. The paper focuses on the investigations involving Android mobile devices but the results can be tested against other mobile operating systems. The result obtained in the collection and analysis of Android malware using conventional tools suggest that such tools are less effective at investigating mobile malware. The proposed framework highlighted in this paper can significantly assist digital forensics investigators in the process of mobile malware forensics investigations.

Network Intrusion Detection with XGBoost and Deep Learning Algorithms: An Evaluation Study

Amr Attia, Miad Faezipour, Abdelshakour Abuzneid
Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT, USA

Abstract— This paper introduces an effective Network Intrusion Detection Systems (NIDS) framework that deploys incremental statistical damping features of the packets along with state-ofthe-art machine/deep learning algorithms to detect malicious patterns. A comprehensive evaluation study is conducted between eXtreme Gradient Boosting (XGBoost) and Artificial Neural Networks (ANN) where feature selection and/or feature dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are also integrated into the models to decrease the system complexity for achieving fast

responses. Several experimental runs confirm how powerful machine/deep learning algorithms are for intrusion detection on known attacks when combined with the appropriate features extracted. To investigate unknown attacks, the models were trained on a subset of the attack datasets, while a different set (with a different attack type) was kept aside for testing. The decent results achieved further support the belief that through supervised learning, the model could additionally detect unknown attacks.

An Efficient Localization for Indoor Environment using Classification Algorithms

Wyatt Towne, Aos Mulahuwaish, Kayhan Zrar Ghafoor, Halgurd S. Maghdid
Department of Computer Science and Information Systems, Saginaw Valley State University, Michigan, USA
Department of Software Engineering and Informatics, Salahaddin University-Erbil, Iraq
School of Mathematics and Computer Science, University of Wolverhampton, Wolverhampton, UK
Department of Software Engineering, Faculty of Engineering, Koya University, Kurdistan Region, F. R. Iraq

Abstract— The positioning accuracy of various kinds of devices is often calculated by using Wi-Fi fingerprints as a data source of Received Signal Strength (RSS). Through different matching algorithms, these RSS values are related to corresponding positioning coordinates for positioning estimation. In this paper, we compare two well known Machine Learning (ML) structures for calculating positioning accuracy: Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM) and a Decision Tree (DT). Both matching algorithms run through a series of datasets in two parts. In the first part, the datasets contain raw, non-normalized RSS values. In the second part, the RSS values are all normalized to remove noise in the datasets. After extensive testing and simulations, the tests showed that as long as RSS values are first normalized, a Decision Tree offers better average positioning accuracy.

Hybrid Physical Layer Security for Passive RFID Communication

A. Gouissem, K. Abualsaud, E. Yaacoub, T. Khattab, M. Guizani
Computer Science and Engineering, Qatar University, Doha, Qatar
Electrical Engineering, Qatar University, Doha, Qatar

Abstract—Thanks to its low cost, small weight and energy efficiency, passive radio frequency identification (RFID) backscatter communications systems have attracted a lot of attention in several application fields. However, such devices have limited computational capabilities and resources which makes them unable to incorporate traditional security protocols and are therefore vulnerable to several types of attacks including cloning and counterfeiting. Therefore, in this paper, a novel hybrid RFID tags identification and malicious devices detection system is proposed by exploiting the estimated tags locations and manufacturing imperfections. In particular, an iterative approach is proposed to estimate the minimum power response at each frequency of the tag in addition to its location. The conducted simulation results show the efficiency of this technique in detecting all the malicious tags and classify the legitimate ones under different network configurations.

Comparison between Automatic Repeat Request (ARQ) Protocols and Solving the Buffer Problem

Mohammed Rajhi, Hatim Madkhali
University of California Santa Cruz, California, USA

Abstract— Today is the era of fast and dynamic internet and communication technologies which make communication convenient compared to the past. Communication technology, through different channels, such as mobile phones and computers, has become pervasive. In almost every means of telecommunication, retransmission protocols have played a significant role in the protocol stack layers. The primary aim of this study is to compare Automatic Repeat Request (ARQ) protocols and solve the Buffer problem. The first section compares ARQ protocols (Go-Back-N, selective repeat, stop and wait ARQ), using TCP and UDP. The study compares these protocols' efficiency and changes the behaviors of stop and wait protocols to selective repeat. The second section includes solving the buffer problems. The paper contains the theoretical part as well.

Safe Selfie

Matthew Blair, Davis Jeffords, Eric Lilling, S. Banik
Department of Cyber & Computer Sciences, The Citadel, The Military College of South Carolina, Charleston, SC, USA

Abstract—In this paper, we discuss the importance and proof of concept for a picture-taking app that will remind the user of their surroundings. The background to this issue is that several people die a year because they are in a situation unsafe for taking pictures, but are too preoccupied with their phone to realise the danger. This application will take many factors into consideration, including user velocity, local emergency contact information, and geological hazards to warn the user of safety issues. The goal of this application is to reduce the amount of accidental injuries or deaths related to taking pictures in unsafe areas.

Development of Vehicle Management System using Location Data Collected by 920MHz LoRa

Daiki Nobayashi, Yasufumi Niwa, Kazuya Tsukamoto, Takeshi Ikenaga
Department of Electrical and Electronic Engineering, Faculty of Engineering, Kyushu Institute of Technology, Fukuoka, Japan

Abstract—The MaaS (Mobility as a Service), which collects, analyzes, and utilizes various data obtained various types of vehicles, is expected to solve various problems about traffic jams, autonomous driving, social community, etc. In this paper, we propose the vehicle management system at a low cost and effecivery using LoRa and Wi-Fi communication. Furthermore, we propose the application to estimate the arrival time assuming the use by a rental car company. The effectiveness of the proposed method is evaluated through demonstration experiments.

CSCI-ISSE: SOFTWARE ENGINEERING

Automated Testing of Mobile Applications Using a Robotic Arm

Demian Frister, Aleksandar Goranov, Andreas Oberweis
Karlsruhe Institute of Technology (KIT), AIFB, Karlsruhe, Germany
FZI Research Center for Information Technology, Karlsruhe, Germany

Abstract - Recent developments in mobile testing have raised the importance of black-box testing and the usage of automated test procedures. In order to secure a flawless user experience, developers are required to develop error free applications. The realization of mobile tests with the support of robotic equipment ensures new possibilities to run tests without further human interaction. In this paper we discuss different approaches for testing mobile applications with robotic arms and additionally share our insights on a prototype suited to automatically test mobile applications. The implemented prototype can perform black-box tests by utilizing an algorithmic approach based on tree-search.

Large-Scale Agile Implementation in Large Financial Institutions: A Systematic Literature Review

Chris H. Hoeseb, Maureen Tanner
Department of Information Systems, University of Cape Town, Cape Town, South Africa

Abstract— Organizations must constantly evolve and adapt to meet customer demands in an ever-changing business environment. The most notable differences have been in the way fintech companies use agile practices to speed up software development and their time to market for new products. This has impacted the way how large financial institutions operate. Large Scale Agile Development (LSAD) is popular, and many studies are being conducted in that domain. However, there is a need to understand the research landscape and the level of theorization on all the studies that where conducted, so that future research opportunities and needs can be identified in this domain. This study aims to synthesize and analyze the research trends and provide a synopsis on the levels of theorization in LSAD research while at the same time also recommending areas for future research in the field of LSAD in the financial sector. Using a systematic literature review protocol, we identified 39 primary studies that focused on LSAD in the financial services sector between 2015 and 2020. The studies were limited to English language only and were sourced from the Google Scholar, IEEE Xplore and ScienceDirect electronic databases. Of all papers reviewed, 14 produced Analysis theories, 9 produced Explanation theories, 3 produced Explanation & Prediction theories, and 7 produced Design & Action theories. Interestingly, none produced Prediction theories. 6 of the papers were systematic literature reviews. The findings indicate that LSAD research in the financial sector has not yet matured in the explanation & prediction theories, especially in the prediction-type theories. This provides an avenue for future research to provide the building blocks for theory development. The focus of future studies should be on building a predictive theory around LSAD Teams in the financial Sector.

Impacts of the Space Technology Evolution in the V&V of Embedded Software-Intensive Systems

Carlos L. G. Batista, Tania Basso, Fatima Mattiello-Franciso, Regina Moraes
National Institute for Space Research (INPE), Sao Jose dos Campos, Brazil
University of Campinas (UNICAMP), Limeira, Brazil
University of Coimbra (UC), Coimbra, Portugal

Abstract—CubeSat-based nanosatellites are composed of COTS components and rely on its structure and standardized interfaces. A challenge in the nanosatellites context is to adapt the V&V (Verification and Validation) process to answer to the increase importance of the embedded software, to reduce the artefacts to be delivered aiming at cutting cost and time and still complying with international standards. This work presents an analysis of the strategy adopted in a real nanosatellite for the development of

the OBDH software embedded in NanosatC-BR2 mission. The goal is to discuss the impact that the standardization of the structure and interfaces of the CubeSat impose on the V&V process of the SiS and to highlight the challenges of "New Space Age" for the use of existing V&V techniques and methods. Index Terms—CubeSat, nanosatellites, V&V techniques

Robustness Testing of Safety-critical Systems: A Portable Insulin Pump Application

Aiman Gannous, Anneliese Andrews, Lamees Alhazzaa
Department of Computer Science, University of Denver, Colorado, USA

Abstract—As a consequence of software becoming an indispensable component of any safety-critical system, verification of such systems will not be sufficient without showing that the software will not create or contribute to a hazardous situation that might cause a serious injury, or even loss of human or animal life. In this paper we show an application of testing framework designed for testing safety-critical systems to a Portable Insulin Pump System, and investigate the framework applicability and its contribution to the certification process in the medical domain. The testing framework design is motivated by providing sufficient evidence for an efficient safety certification process; thus it integrates modeling, safety analysis, and combinatorial testing techniques to provide broadened testing activity outputs that generate diverse types of evidence.

Detecting Software Security Vulnerability during an Agile Development by Testing the Changes to the Security Posture of Software Systems

Benjamin Arnold, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract—The purpose of this quantitative quasiexperimental study is to identify the possible correlation between software changes and the likelihood that software releases developed using an agile methodology like DevOps will introduce vulnerabilities into the software application when integrated. There are several scholarly articles that provide details on how Agile development methodologies like scrum and DevOps rely on automated testing for security. The majority of literature on the subject recommend manual security and penetration testing, but there is currently no objective measure to determine when this manual testing should take place. In Agile scrum and in DevOps, manual security testing is usually conducted after a large feature is completed and integrated into production. If a correlation can be found between aspects of software changes and their propensity to introduce vulnerabilities into a software application, then that data can be used to build an objective process for measuring when manual security testing should be performed in Agile development.

Automated Estimation of the Rate of Equivalent Mutants

Amani Ayad, Ali Mili
SUNY, Farmingdale State College, New York, USA
New Jersey Institute of Technology, NJIT, Newark, New Jersey, USA

Abstract - Whereas most past research on mutation equivalence has focused on analyzing a base program and a mutant to determine whether they are equivalent, we favor a totally orthogonal approach: We argue that in addition to being difficult and error-prone, the determination of whether two programs are equivalent is also often unnecessary. For most practical applications, it is not necessary to identify equivalent mutants individually; it is sufficient to estimate their number. In this paper, we discuss an automated tool that does so by static analysis of the base program.

Neural Network Model for Use in Performing Pitch Correction in a Voice-Driven Musical Instrument

John Carelli
Computer Science and Information Technology Department, Kutztown University, Kutztown, Pennsylvania, USA

Abstract— A neural network model is presented for use in musical note recognition and pitch correction using only recently sung notes captured in real time. The goal is to improve, by correcting inaccuracies in singing execution, the performance of a voice driven musical instrument that translates sung pitch into notes played by a separate virtual instrument. This is accomplished without knowledge of musical key and, in order to enable real-time response, with a minimal number of recently sung notes. Model development, training, and testing in the voicedriven instrument are described. Overall, the study provides a unique and potentially useful investigation of a human/computer interaction with application in live musical performance.

Methods of Implementation, Maturity Models and Definition of Roles in DevOps Frameworks: A Systematic Method

Luciano de Aguiar Monteiro, Ioram Schechtman Sette, Domingos Savio M. Pessoa Monteiro, Washington Henrique Carvalho Almeida, Anderson Cavalcanti de Lima
Center of Advanced Studies and Systems of Recife - CESAR, Brazil

Abstract— The software development industry has been evolving and has produced new development standards and models for providing services. Agile methodologies have reached their plateau with DevOps, which has increased software quality and hastened its delivery. However, a gap in terms of its formalization and unresolved issues concerning its adoption have become relevant. This study seeks, by using a Systematic Mapping, to elucidate existing gaps in this area such as methods for implementing DevOps, the use of maturity models and defining the roles of those who participate in this process. This Systematic Method was applied to five databases (ACM Digital Library, IEEE Xplore, Science Direct, Springer Link and Scopus), during which automatic searches are conducted with a view to answering the research questions. 32 studies returned answers to these. 11 maturity models were found; however, only two of them had been validated by companies, one of which only applied to IBM software.

Coding Overhead of Mobile Apps

Yoonsik Cheon
Department of Computer Science, The University of Texas at El Paso, El Paso, Texas, USA

Abstract — A mobile app runs on small devices such as smartphones and tablets. Perhaps, because of this, there is a common misconception that writing a mobile app is simpler than a desktop application. In this paper, we show that this is indeed a misconception, and it's the other way around. We perform a small experiment to measure the source code sizes of a desktop application and an equivalent mobile app written in the same language. We found that the mobile version is 19% bigger than the desktop version in terms of the source lines of code, and the mobile code is a lot more involved and complicated with code tangling and scattering. This coding overhead of the mobile version is mostly due to the additional requirements and constraints specific to mobile platforms, such as diversity and mobility.

Incremental Contract-based Verification of Software Updates for Safety-Critical Cyber-Physical Systems

Yosab Bebawy, Houssem Guissouma, Sebastian Vander Maelen, Janis Kroger, Georg Hake, Ingo Stieran, Martin Franzle, Eric Sax, Axel Hahn
R&D Transportation Division, OFFIS e.V., Oldenburg, Germany
Institute for Information Processing Technologies, Karlsruhe Institute of Technology, Karlsruhe, Germany
Department of Informatik, University of Oldenburg, Oldenburg, Germany

Abstract—Software updates are indispensable for the continuous development of Cyber Physical Systems (CPS): They allow for low-cost bug-fixing, fast adaptation to new or changing environments, or adding new functionality throughout the CPS's life-cycle. Due to the urgent need for some safety-critical updates, their verification and validation may need to happen as fast as possible without loss of quality. For this reason, incremental checks targeting specifically the introduced changes and their impact on the system are essential as they speed up the validation process. In this paper, we introduce a concept for such an incremental verification for different types of updates by using contract-based design and verifying the integration of the introduced changes by checking their compliance with the contractually agreed assumptions and guarantees. We demonstrate our approach by applying two update types to an Adaptive Cruise Control (ACC) system and verifying the impact of the changes within the environment of the changed module(s).

An Open Web-based Platform for Enhancing the Visibility of Brazilian Researches

Alvaro Magri Nogueira da Cruz, Adriana Barbosa Santos, Rogeria Cristiane Gratao de Souza, Thiago Luiz Parolin
Department of Computer Science and Statistics, Sao Paulo State University, Sao Jose do Rio Preto, Brazil

Abstract—The scientific research developed in Brazil is prominent in the global scene. Despite this, it still lacks visibility outside the scientific sphere and better recognition for Brazilian researchers. This paper presents a web-based platform proposed to reinforce the open access initiatives to expand scientific knowledge. It is focused on publishing research articles by audiovisual format. The audiovisual resources can introduce another dynamic for scientific communication in Brazil. The characteristics and subcharacteristics of the ISO/IEC25010 quality model were adapted to define the platform's design, its architecture and user interface.

Validation Support Tool to Cross-check the Behavioral Flows on a Requirements Analysis Model using the State Transition Model

Hikaru Morita, Saeko Matsuura
Graduate School of Engineering and Science, Shibaura Institute of Technology, Saitama, Japan

Abstract—We propose a method to evaluate and improve the validity of the required specifications by comparing models from different viewpoints. Inconsistencies are automatically extracted from the model in which the analyst defines the service procedure based on the initial requirement, and automatically compared with the state transition model created by an evaluator different from the analyst from the same initial requirement. The identified inconsistencies are fed back to the analyst to improve the required specifications. We developed a tool for extraction and comparison, and discuss the effectiveness by applying the method to an example requirements specification.

2D Animation of Recursive Backtracking Maze Solution Using JavaFX Versus AWT and Swing

Anil L. Pereira
School of Science and Technology, Georgia Gwinnett College, Lawrenceville, Georgia, USA

Abstract—This paper describes software developed by the author, that constructs a two-dimensional graphical maze and uses animation to visualize the recursive backtracking maze solution. The software is implemented using three Java graphics technologies, Java Abstract Window Toolkit, Java Swing and JavaFX. The performance of these technologies is evaluated with respect to their graphics rendering frame rate and heap memory utilization. The scalability of their rendering performance is compared with respect to maze size and resolution of the displayed maze. A substantial performance gap between the technologies is reported, and the underlying issues that cause it are discussed, such as memory management by the Java virtual machine, and hardware accelerated graphics rendering.

Spikes in Agile Software Development: An Empirical Study

Hussein Al Hashimi, Andy Gravell
Electronics and Computer Science, University of Southampton, Southampton, UK
King Saud University, Riyadh, Saudi Arabia

Abstract— Spikes can be an essential component of the agile development cycle, because they assist the teams, for both technical and functional issues, to identify any uncertainty in a user story, leading to a more efficient solution to the problem. The use of spikes in agile software development (ASD) can enable organizations to produce quality software by employing the required technical expertise, planning the entire development cycle and ensuring that the client's requirements are adhered to. This study aims to examine the use of spikes in ASD. It explores the role, efficiency and efficacy of spikes in various software development domains through the different agile methods. An exploratory research design is adopted to achieve this purpose, whereby mixed methods are used to collect concurrently both qualitative and quantitative data from the experts recruited to the study. Through the survey, it establishes that the primary role of spikes is risk management through investigations to understand user stories and reveal any uncertainty. Conclusively, the study findings imply that spikes have become an essential tool for most agile teams in ASD. The efficiency and effectiveness that are reported show that the majority of experts in software development have realized the value of using spikes in their processes.

Recommending Attack Patterns for Software Requirements Document

Mounika Vanamala, Jairen Gilmore, Xiaohong Yuan, Kaushik Roy
North Carolina Agricultural and Technical State University, North Carolina, USA

Abstract—To develop secure software, software developers need to know the potential threats to the software. Knowledge captured in the Common Attack Pattern Enumeration and Classification (CAPEC) database can help software developers to understand how attackers target application weaknesses. In this paper, we present a method of recommending CAPEC attack patterns based on software requirement specification (SRS) documents. The method uses topic modelling to extract topics from each attack pattern and to extract topics from the software system description, user classes, use cases, and function requirements within the SRS documents. Attack patterns are recommended by calculating the distance measure of each attack pattern topic distribution and each SRS topic distribution using cosine similarity. Attack patterns are then ranked from maximum to minimum. The top attack patterns are then recommended to the software developers as the most relevant to the software system under development.

An Analysis on Scrum Methodology in Global Software Development – GSD

Areeba Ahmad, Joan Lu, Steve Wade, Richard Hill, Lee McCluskey
University of Huddersfield, Huddersfield, United Kingdom, UK

Abstract— The development of applications in collaborative production environments poses problems relating to teamwork, control and communication. Global software development (GSD) has increasingly identified agile techniques that require regular contact and self-organization between remote sites to mitigate challenges. The main aim of this paper is to present an analysis of the challenges in GSD using Scrum practices. This study discusses the Scrum practices in GSD from existing database libraries from 2009 to 2020. It is found that there is relatively quite wide activity in the area of GSD and it may be difficult to apply Scrum in GSD but it provides solutions to the issues people face in this area.

Developing Software Using Agile and Design Thinking Framework

Mahrukh Mirza, Soma Datta
University of Houston-Clear Lake, Houston, Texas, USA

Abstract—There are several software companies that develop several products. Are all of them used? Even after software companies following Agile framework methodology to develop their software's, many software's still fail to meet the requirements and eventually customers end up not using them. This is usually because the software teams fail to completely understand the user requirements. Agile framework and Design Thinking framework can be used together to gain customer empathy and to overcome this problem. This study discusses how Agile framework and Design Thinking supports the same principles and can be used to reduce software development failures. A Flashcard application was developed to validate Agile framework Design Thinking framework. The results are discussed in the validation and result section.

Reasoning Heuristics for the Theorem-Proving Platform Rodin/Event-B

Jacobus Gideon Ackermann, John Andrew van der Poll
School of Computing, College of Science, Engineering and Technology, University of South Africa, South Africa
Graduate School of Business Leadership (SBL), University of South Africa, South Africa

Abstract—Developments in formal- and mathematical logic; and computing the past couple of decades have paved the way for the automation of deductive reasoning. However, despite theoretical and technological advances in computing, the rapid growth in the search space for complex proofs where the reasoner explores the consequences of irrelevant information, remains problematic. The challenge of a combinatorial explosion of the search space can in many cases be addressed by heuristics. Consequently, in this paper we investigate the extent to which heuristics may usefully be applied in discharging complex settheoretic proof obligations using the hybrid reasoning environment, Rodin/Event-B. On the strength of our experiments, we develop a set of heuristics to aid the theorem-proving environment in finding proofs for set-theoretic problems which could not be obtained using the default settings. A brief exposition of related work in this area is presented towards the end of the paper.

Numerical Expression Treatment for Pseudo Natural Programming Language

Toshiyuki Maeda, Masumi Yajima, Akiyoshi Wakatani
Department of Management Information, Hannan University, Japan
Department of Economics, Meikai University, Japan
Faculty of Intelligence and Informatics, Konan University, Japan

Abstract—This paper addresses a multi-lingual pseudo natural language programming environment, so as to introduce basics of computer science and support to develop IoT application. In this paper, we at first explain the background of this research. After that, we describe the implementation of our system, expecially the treatment of numerical expressions.

CSCI-ISSC:
SMART CITIES AND SMART MOBILITY

Bus Pass Time Estimation based on Optimal Data Gathering from a Slow Mobility Server

Carlos Garcia-Maurino, Pedro J. Zufiria, Alejandro Jarabo-Penas
Depto. Matematica Aplicada a las TIC, ETSI Telecomunicaccion, Universidad Politecnica de Madrid, Madrid, Spain
Information Processing and Telecommunications Center, Universidad Politecnica de Madrid, Madrid, Spain

Abstract—Statistical description and prediction of bus arrival times is relevant for public transport users since it allows more timewise efficient journeys. This work is focused on characterizing the real behavior of buses based on past arrival estimation data. The main goal is to estimate real bus pass times by optimally collecting data from an intercity bus arrival time estimation system which is limited in petition handling capacity. This requires to model the server behavior prior to the design of the data collection system. In addition, it also requires the design of an algorithm to estimate the bus real passing time considering that only the provided estimated time of arrival is available. This information can be useful for designing alternative online arrival time estimators based on supervised learning which could potentially improve the estimator efficiency.

Headway Estimation in Urban Buses based on Available Arrival Time Estimators

Alejandro Jarabo-Penas, Pedro J. Zufiria, Carlos Garcia-Maurino
Depto. Matematica Aplicada a las TIC, ETSI Telecomunicaccion, Universidad Politecnica de Madrid, Madrid, Spain
Information Processing and Telecommunications Center, Universidad Politecnica de Madrid, Madrid, Spain

Abstract— Nowadays, public transport is an indispensable urban service, helping to improve urban mobility by making it accessible to everyone, as well as reducing the pollution inside big cities. For this reason, any efforts for characterizing and improving urban public transportation systems are welcome. In the present paper we perform an analysis of the data corresponding to the arrival time estimators of the urban buses provided by the EMT (municipal transport company) of Madrid, with the aim of improving their performance and reliability. To achieve these objectives, we have developed several algorithms that, using the EMT provided data, estimate the position, arrival times and running times between stops. Finally, we estimate the headways between the buses, which can be used to monitor the performance of the service in each line.

Generating Indoor Navigation Routes using Beacons

Alicia Martinez-Rebollar, Hugo Estrada-Esquivel, Luis Lopez-Garcia, Leon Torres-Restrepo, Javier Ortiz-Hernandez
Department of Computer Science, National Technology of Mexico, CENIDET, Cuernavaca, Mexico

Abstract—The Internet of Things is promoting the generation of smart buildings. These buildings have as main requirement the navigation of interiors. However, GPS technology, which is used to carry out positioning, cannot be used within a building because satellite signals do not travel well through solid materials. In this paper, we present an indoor navigation proposal, which uses beacons technology and smartphones. Our software application obtains information from the context and generates the best route to reach the destination within an intelligent building. The Dijkstra algorithm was used to process all the information. Hence, our proposal aims to combine different technologies, and adapt developed algorithms to indoor navigation. The results obtained are encouraging and show that it is possible to obtain good results using this type of technology.

A Lightweight Framework for IoT Smart Solutions

Javier Ortiz-Hernandez, Juan Antonio Miguel-Ruiz, Manuel Erazo-Valadez, Leon Torres-Restrepo, Alicia Martinez-Rebollar, Manuel Mejia-Lavalle
Department of Computer Science, National Technology of Mexico, CENIDET, Cuernavaca, Mexico

Abstract— Successive technological improvements have made possible the development of IoT with a high capacity for communication and data collection, providing numerous opportunities for application. As a result, attention from researchers in both academic and industrial environments has been drawn to the subject matter. Despite all the enhancements till date, there are still several open issues that represent the main challenges for IoT. One of these concerns resides in the management of the devices responsible for collecting environmental data. The functionality and reliability of the services it provides depends on the correct management of the sensor devices. In this article we propose a framework for the development of a lightweight IoT smart solution using the FIWARE Open-Source Platform. This solution scheme is intended for an indoor localization system using Wireless Fidelity (WiFi) and Bluetooth Low Energy (BLE) signals.

Energy Usage of Deep Learning in Smart Cities

Supadchaya Puangpontip, Rattikorn Hewett
Department of Computer Science, Texas Tech University, Lubbock, Texas, USA

Abstract— Deep learning has increasingly become an essential component of many Smart City functions including smart city lightings, emergency rescues, smart drainage, and smart parking. These functions operate continuously in real-time throughout the day. Thus, excessive energy usage of deep learning computation can negatively impact economic benefits and efficiency of smart cities. The situation can escalate when dealing with resourceconstrained large-scale smart cities of huge Internet-of-Things and networks with large numbers and varieties of sensors. To effectively sustain, manage and protect smart cities from failures due to energy overload, the awareness of energy consumption by deep learning computation is unavoidably necessary. Most recent research in smart cities focuses on using deep learning to perform certain tasks but does not address energy issues. This paper presents a formal approach to estimating energy consumption of deep learning and illustrates its use in smart cities. In particular, we develop a fine-grained mathematical model that extends an existing model to include the quantification of MAC (multiplyand-accumulate) operations as well as data access from a memory hierarchy. This paper focuses on deep and convolutional neural networks. We describe the proposed approach and validate the results obtained from our model by comparing them against those of existing work. The proposed approach is applied to three (deep) neural systems in smart cities, namely smart drainage, smart transportation and smart parking systems, all of which yield promising results.

Creating a Real-Time Geocoding System: Implications of Open Source for the Public Safety

Murat Ozer, Michael Zidar, Rustu Deryol, Said Varlioglu, Ibrahim Sevki Eldivan, Halil Akbas
School of Information Technology, University of Cincinnati, Cincinnati, Ohio, USA
Department of Criminology, University of South Florida, Sarasota, Florida, USA
Lodi, New Jersey, USA
Antroplogy Sociology Criminology, Troy University, Troy, Alabama, USA

Abstract—Today, maps are important components of public safety. Collecting geographic information data enables any public safety agency to smartly allocate their limited resources to those places where the public safety needs are the greatest. However, the current applications of address geocoding are generally static requiring specifically assigned personnel to geocode the addresses using a geographic information system (GIS) software. Realtime geocoding services like Google Maps are an alternative to these static systems by offering real-time geocoding capabilities; however, these closed systems come usually with a price that many agencies cannot afford. Given this context, we developed a public safety geocoding system using available open-source data and systems to enable any public safety agency to continuously enjoy the power of geographic information systems for their smart decision-making process. The proposed model is not limited to public safety and can be used for any purpose where people are used to using Google Maps Geocoding API or other commercial geocoding services.

Black Ice Detection using CNN for the Prevention of Accidents in Automated Vehicle

Hojun Lee, MinHee Kang, JaeIn Song, Keeyeon Hwang
Department of Urban Design & Planning, Hongik University Seoul, Korea
Department of Smartcity, Hongik University, Seoul, Korea
Department of Urban Planning, Hongik University, Seoul, Korea

Abstract— Black ice is recognized as the main cause of major accidents in winter because it has characteristics that are difficult to identify with the naked eye. This is expected to be a potential cause of accidents in the era of automated vehicles as well. Accordingly, this study presents a CNN-based black ice detection plan to prevent traffic accidents caused by black ice. Due to the characteristic of black ice that is formed only in a certain environment, the data was augmented and the image of road environment in various environments was learned. Test results show that the proposed CNN model detected black ice with 96% accuracy and reproducibility(recall).

Dynamically Adjustable PID for Adaptive Motion Control: PID++ Algorithm Introduction and Applications

Thomas F. Arciuolo, Miad Faezipour
Department of Computer Science & Engineering, University of Bridgeport, Bridgeport, CT, USA
Department of Biomedical Engineering, University of Bridgeport, Bridgeport, CT, USA

Abstract—The Holy Grail of Robotic Motion Control is the implementation of an algorithm which is humanlike in operation and low in computational cost. Today, most robotic motion control approaches are highly sophisticated and by implication, computationally costly. A proposed new algorithm, with the Proportional-Integral-Derivative (PID) at its heart, called the PID++, will achieve humanlike control while being computationally lightweight. This relatively simple algorithm will have the speed and precision of a machine and have characteristics that are unmistakably human. Unlike current robotic motion control algorithms that are tedious at best to implement, lacking in automatic situational adaptability, and tending to be static in nature, the PID++ algorithm will be just the opposite. The PID++ algorithm will be a radically new, simple, and computationally lightweight approach to humanoid motion control, using periodic, minor adjustments and basic arithmetic, based on the real-time encoder position input, to achieve a stable, precise, controlled, dynamic, adaptive control system, for linear motion control, in any direction regardless of load. This algorithm will be complete in its implementation, and truly dynamic and adaptive by design. Engineers will be able to use this algorithm in commercial, industrial, biomedical, and space applications alike. Motion control will be easy to implement on even the smallest microcontrollers (MCU) using a single command interface and without the need of reprogramming or reconfiguration. This paper discusses potential PID++ applications.

Collaborative Mobile Surveillance System for Smart Cities

Laisa C. C. De Biase, Samira Afzal, Pablo Calcina-Ccori, Geovane Fedrescheski, Marcelo K. Zuffo
Department of Electronic Systems Engineering, Polytechnic School, University of Sao Paulo (USP), Brazil
Institute of Mathematics and Statistics, University of Sao Paulo (USP), Brazil

Abstract—In this poster, we design the Collaborative Mobile Surveillance System (CMSS) integrating vehicles and Internet of Things (IoT), for an intelligent and near real-time safety management. The system, consisting of Artificial Intelligence (AI) and network communication methods, aims to improve the sense of safety for drivers in smart cities. The CMSS will detect and notify emergency events, such as crime and lifethreatening situations, using sensors, AI-based audio processing and computer vision, and send notifications to emergency services (e.g., police, ambulance, and firefighting) enabling immediate action.

CSCI-ISPD: PARALLEL AND DISTRIBUTED COMPUTING

Data Optimization for Large Batch Distributed Training of Deep Neural Networks

Shubhankar Gahlot, Junqi Yin, Mallikarjun (Arjun) Shankar
Oak Ridge National Lab, Oak Ridge, Tennessee, USA

Abstract—Distributed training in deep learning (DL) is common practice as data and models grow. The current practice for distributed training of deep neural networks faces the challenges of communication bottlenecks when operating at scale, and model accuracy deterioration with an increase in global batch size. Present solutions focus on improving message exchange efficiency as well as implementing techniques to tweak batch sizes and models in the training process. The loss of training accuracy typically happens because the loss function gets trapped in a local minima. We observe that the loss landscape minimization is shaped by both the model and training data and propose a data optimization approach that utilizes machine learning to implicitly smooth out the loss landscape resulting in fewer local minima. Our approach filters out data points which are less important to feature learning, enabling us to speed up the training of models on larger batch sizes to improved accuracy.

Distributed Ray Tracing of Large Scenes using Actors

Elizabeth M. Ruetschle, Kurt D. Hardee, Mark C. Lewis
Department of Computer Science, Trinity University, San Antonio, Texas, USA

Abstract—We present work on implementing an actorbased ray-tracing system that can work with distributed geometry so that scenes can exceed the memory requirements of a single machine. The goal of this work is large-scale visualizations of planetary rings to compare with Cassini observations. Going through several iterations, we get to a configuration that is flexible and allows different options for distributing work across the geometry. We present performance results for a number of variations and different views of the scene.

Distributed Ray Tracing of Large Scenes using Spark

Erica Cater, Connor Weisenberger, Mark C. Lewis
Department of Computer Science, Trinity University, San Antonio, Texas, USA

Abstract—We present work on implementing a ray tracing system that can work with distributed geometry so that scenes can exceed the memory requirements of a single machine using the Apache Spark big-data system. We look at a progression of implementations beginning with implementations that simply distribute the work and moving to implementations that allow the geometry itself to be distributed. We find that Spark makes it rather simple to distribute work across a cluster. While distributing the geometry across the cluster is more challenging, the implementation is still rather short given that it is capable of rendering scenes that are larger than any of the individual machines in the cluster could handle.

Recursive MaxSquare: Cache-friendly, Parallel, Scalable in situ Rectangular Matrix Transposition

Claudio A. Parra, Travis Yu, Kyu Seon Yum, Arturo Garza, Isaac D. Scherson
Department of Computer Science - Systems, University of California, Irvine, California, USA

Abstract—An in situ rectangular matrix transposition algorithm is presented based on recursively partitioning an original rectangular matrix into a maximum size square matrix and a remaining rectangular sub-matrix. To transpose the maximum size

square sub-matrix, a novel cache-friendly, parallel (multithreaded) and scalable in-place square matrix transposition procedure is proposed: it requires a total of Θ(n 2 /2) simple memory swaps, a single element temporary storage per thread, and does not make use of complex index arithmetic in the main transposition loop. Recursion is used to transpose the remaining rectangular sub-matrix. Dubbed Recursive MaxSquare, the novel proposed rectangular matrix in-place transposition algorithm uses a generalization of the perfect shuffle/unshuffle data permutation to stitch together the recursively transposed square matrices. The shuffle/unshuffle permutations are shown to be efficiently decomposed using basic vector/segment swaps, exchanges and/or cyclic shifts (rotations). A balanced parallel cycles-based transposition is also proposed for comparison.

Reconfigurable Many-Core Embedded Computing Platform with Geometrical Bus Interconnection

Tirumale Ramesh, Khalid Abed
Department of Electrical & Computer Engineering and Computer Science, Jackson State University, Jackson, Mississippi, USA

Abstract—There is an increasing interest in many-core embedded system of the scale of hundreds designed specifically with high degree of parallelism. These systems require high performance interconnection with minimum cost. In this paper, we propose a reconfigurable architecture based on a Geometrical Reduced Bus System (GRBS) interconnection and present interconnect characterization in terms of cost, fault tolerance and effective memory bandwidth. GRBS has b number of reduced buses given by b= $\mathbf{min}(\boldsymbol{n},\boldsymbol{m})/\boldsymbol{k}$ where n is the number of processor cores and m is number of memory modules and k is the bus reduction factor for number of buses. We present geometrical connections such as rhombic, trapezoidal, cyclic, balanced, staircase and pyramid to connect to either processor cores or memory modules. Our results show that by using reduced bus connections, we can achieve comparable memory bandwidth and 1.5x bandwidth per cost. GRBS provides sustainable memory bandwidth for single bus failure.

Energy-Efficient Heterogeneous Computing Of Parallel Applications Via Power Capping

Kishwar Ahmed, Samia Tasnim, Kazutomo Yoshii
University of South Carolina Beaufort, South Carolina, USA
Florida A&M University, Tallahassee, Florida, USA
Argonne National Laboratory, Lemont, Illinois, USA

Abstract—High-performance computing (HPC) systems require significant energy to run their operations. A typical HPC system contains various heterogeneous hardware components, including CPUs and GPUs. Power capping is a widely-used feature in processors to achieve an upper limit of power allocation. In this paper, we exploit power capping capability in modern processors to achieve energy-efficiency in heterogeneous computing system. We first develop an optimal power cap allocation model considering heterogeneous computing platform. Next, we develop a simulator based on a parallel discrete-event simulation engine to simulate our proposed power capping allocation model. Finally, we perform trace-based simulation experiments to demonstrate effectiveness of our model. Experiments demonstrate that our proposed model is capable of achieving various level of energy-tosolution reduction for different parallel applications considering heterogeneous computing platform.

Coupling Storage Systems and Self-Describing Data Formats for Global Metadata Management

Michael Kuhn, Kira Duwe
Otto von Guericke University Magdeburg, Magdeburg, Germany

Abstract—Traditional I/O stacks feature a strict separation of layers, which provides portability benefits but makes it impossible for storage systems to understand the structure of data. Coupling storage systems with self-describing data formats can offer benefits by making the storage system responsible for managing file metadata and allowing it to use structural information for selecting appropriate storage technologies. Our proposed storage architecture enables novel data management approaches and has

the potential to provide significant performance improvements in the long term. By making use of established self-describing data formats, no modifications are necessary to run existing applications, which helps preserve past investments in software development. Specifically, we have designed and implemented an HDF5 VOL plugin to map file data and metadata to object and key-value stores, respectively. Evaluations show that our coupled storage system offers competitive performance when compared with the native HDF5 data format. In some cases, performance could even be improved by up to a factor of 100.

Proposed Design for Effectively Expand Adaptive-ticks Feature in the Linux Kernel to Full Tickless Function

Abdullah Aljuhni, Shaji Yusuf, C. Edward Chow, Oluwatobi Akanbi, Amer Aljaedi
Department of Computer Science, University of Colorado, Colorado Springs, Colorado, USA
Intel Corporation, India Dev Center, India

Abstract— Operating systems use timers available in the processor to trigger timer interrupts that help the kernel manage several tasks, ranging from time management of system software to scheduling and task switching. Traditionally, operating systems program the system timer to tick at regular intervals ranging from 1 ms to 100 ms, depending on the required system response and type of load. The periodic tick timer is the simplest and most common way of managing the system's management activities. However, advancements in microprocessors and operating system scheduler allows for a more efficient and better-performing solution by eliminating regular timer interrupts. In this research, we illustrate the implementation of Adaptive-ticks feature and analyze the performance of Adaptive-ticks scheduler in Linux kernel version 4.0.9. We also propose a new full Adaptive-ticks design to support multiple tasks in the ready queue. This new proposed design extends the current Adaptiveticks feature to improve performance and power efficiency by getting rid of unwanted interrupts.

Parallel Computation of Standard Competition Rankings over a Sorted Array

Jingyuan Liang, Jonathan Bisnett, Alan Hylton, Janche Sang, Chansu Yu
Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, Ohio, USA
Space Communications and Navigation, NASA Glenn Research Center, Cleveland, Ohio, USA

Abstract—The Standard Competition Ranking (SCR) is a commonly adopted ranking strategy and has been used in a wide range of applications, such as statistics, text mining, image processing, and so on. Though the sequential implementation of the SCR can be executed in linear time, it is not straightforward to design parallel algorithms for the SCR. In this paper, our focus is on the novel use of the parallel prefix computation method for calculating the SCR on a many-core Graphics Processing Unit (GPU). We also design a pthreads-based algorithm on a multicore CPU which adopts a modified binary search to find the first item's rank in each partitioned segment. By integrating the modified binary search with the prefix computation, we later design and implement a more efficient hybrid algorithm on the GPU. The experimental results show that, as compared with the sequential execution on the CPU, our pthreads-based algorithm on a 12-core CPU can be roughly 8 times faster, while the hybrid algorithm on the GPU can achieve more than two orders of magnitude speedup.

von Neumann's Missing "Second Draft": what it should contain

Janos Vegh
Kalimanos BT Debrecen, Hungary

Abstract—The computing science is based on the classic computing paradigm created for vacuum tubes only, and using it for today's technological conditions was told to be unsound by its inventor. The reason is that, unlike three-quarters century ago, the transmission time is not negligible, aside from the processing time. The paper points out that the theory is perfect, but now it is used outside of its range of validity, given that the neglection it makes is not valid anymore. For today, the computing paradigm became the major obstacle for creating more effective computing systems: the correct science background must be included.

Feature Selection for Learning to Predict Outcomes of Compute Cluster Jobs with Application to Decision Support

Adedolapo Okanlawon, Huichen Yang, Avishek Bose, William Hsu, Dan Andresen, Mohammed Tanash
Department of Computer Science, Kansas State University, Manhattan, Kansas, USA

Abstract—We present a machine learning framework and a new test bed for data mining from the Slurm Workload Manager for high-performance computing (HPC) clusters. The focus was to find a method for selecting features to support decisions: helping users decide whether to resubmit failed jobs with boosted CPU and memory allocations or migrate them to a computing cloud. This task was cast as both supervised classification and regression learning, specifically, sequential problem solving suitable for reinforcement learning. Selecting relevant features can improve training accuracy, reduce training time, and produce a more comprehensible model, with an intelligent system that can explain predictions and inferences. We present a supervised learning model trained on a Simple Linux Utility for Resource Management (Slurm) data set of HPC jobs using three different techniques for selecting features: linear regression, lasso, and ridge regression. Our data set represented both HPC jobs that failed and those that succeeded, so our model was reliable, less likely to overfit, and generalizable. Our model achieved an R2 of 95% with 99% accuracy. We identified five predictors for both CPU and memory properties.

Parallel Data Indexing and Storage on a Cots Cluster

Anil L. Pereira
School of Science and Technology, Georgia Gwinnett College, Lawrenceville, Georgia, USA

Abstract—In this paper, a data transfer, indexing and storage system on a commodity-off-the-shelf cluster using parallel processing, asynchronous file Input/Output, direct memory access and asynchronous User Datagram Protocol sockets is proposed. Also, a performance evaluation framework for the system is described. There are two main considerations in developing the system. First, as data communication networks support increased data rates due to fiber optical cables and more efficient network devices, better data transfer and storage methods are required to exploit the speed of the networks. Second, applications in particle physics, climate modeling and weapon systems simulation generate petabytes of data from a single experiment. The challenge is to index and store the data as soon as it is produced and preprocessed by several instruments.

Scalable Distributed Checkpointing Algorithm

Jinho Ahn
School of Computer Science and Engineering, Kyonggi University, Suwon Gyeonggi, Republic of Korea

Abstract—A communication-induced checkpointing algorithm, named HMNR, was introduced to effectively use control information of every other process piggybacked on each sent message for minimizing the number of forced checkpoints. Then, an improved algorithm, called Lazy-HMNR, was presented to lower the possibility of taking forced checkpoints incurred by the asymmetry between checkpointing frequencies of processes. Despite these two different minimization techniques, if the high message interaction traffic occurs, Lazy-HMNR may considerably lower the probability of detecting Z-cycle free patterns due to its shortcoming. Also, there is no prior research attempt to design the algorithms considering network topologies for making the number of forced checkpoints as few as possible with control information piggybacked on each message. This paper introduces a new Lazy-HMNR algorithm for group communication-based distributed systems to synergistically decrease the number of forced checkpoints in a more effective manner.

CSCI-ISCS: COMPUTATIONAL SCIENCE

The Impact of Big Data on AI

Souad Demigha
CRI, University of Paris 1 Sorbonne, Paris, France

Abstract—Big Data refers to data that can't be processed with traditional applications due the challenge of capturing, storing, transferring, querying, fast processing and updating data in such large amounts. The Big Data concept often uses analytics involving Artificial Intelligence (AI), Machine Learning and Deep Learning. The paper investigates the impact of Big Data in the use of AI methods and techniques.

A Novel Naive Bayesian Approach to Inference with Applications to the MNIST Handwritten Digit Classification

Kai Wang, Hong Zhang
Department of Computer Science, Georgia Southern University, Statesboro, Georgia, USA

Abstract—Naive Bayesian approach is an effective method for many data analysis problems such as pattern classification and machine learning. However, it often suffers from the underflow problem when the input data has a high dimension. Such a problem is often addressed by taking logarithms and working in the transformed domain. In this paper we propose a novel approach to this problem based on geometric means and apply it to the classical MNIST handwritten digit classification problem. The results show that it not only achieves satisfactory accuracy but also demonstrates its power of presenting "second best" guesses that are meaningful and useful in the pattern classification domain.

Efficient Seed Volume Measurement Framework

Chendi Cao, Mitchell Neilsen
Department of Computer Science, Kansas State University, Manhattan, Kansas, USA

Abstract—Modern seed breeding programs require the ability to analyze seeds efficiently to be useful. Even simple measures such as volume and density can be challenging to compute efficiently with modest equipment. Accurately measuring seed volume becomes a highly under-constrained problem. Multiple images from different perspectives are required. This paper presents an efficient and affordable 3D single seed volume measurement system to extract image contours and compute volumes using a modified volume carving method in a controlled lab environment. The framework is constructed with a turntable, a stepper motor controlled by an Arduino microcontroller, three orthogonal cameras, and camera control via a modest computer used for data acquisition and processing. For testing, images are captured using only a side camera from different angles by rotating the turntable. Then, the framework processes the multiple images in parallel and reconstructs 3D seed objects to calculate the volume based on the voxel numbers. The proposed framework: (1) generates single seed 3D geometries for visualization, (2) calculates precise seed volumes within seconds, and (3) achieves less than a 3% error rate on a reference ceramic sphere.

A Class of Generic Approximate Sparse Pseudoinverse Matrix Techniques based on Incomplete QR Factorization

A. D. Lipitakis, C. K. Filelis-Papadopoulos, G. A. Gravvanis, D. Anagnostopoulos
Department of Informatics and Telematics, Harokopio University of Athens, Greece
University College Cork, Cork, Ireland
Department of Electrical and Computer Engineering, School of Engineering, Democritus University of Thrace, University Campus, Xanthi, Greece

Abstract—A class of Generic Approximate Sparse Pseudoinverse matrices based on modified incomplete QR factorizations with Givens rotations is presented. The filtering criteria for the Incomplete QR factorization have been modified to improve the quality of the incomplete upper triangular factor R, by removing the restriction on the number of nonzero elements per row. Moreover, the Givens rotation matrices are dropped during the incomplete factorization process in order to reduce memory requirements. The proposed approximate pseudoinverse preconditioning scheme is computed, using a "restricted" solution process, based on approximate pseudoinverse sparsity patterns, derived from powers of the sparsified incomplete upper triangular matrix R. The proposed preconditioning scheme is used in conjunction with the Explicit Preconditioned Conjugate Gradient Least Squares method. Numerical results indicating the applicability and efficiency of the proposed algorithmic schemes for solving various model problems are given.

Simulation of TCP-100 Facility System Level Model for Operation Training Purposes

Luis J. Yebra, Francisco M. Marquez, Pedro J. Zufiria
Plataforma Solar de Almeria, CIEMAT, Tabernas, Spain
Dpto. Matematica aplicada a las TIC, ETSI Telecomunicacion, Universidad, Politecnica de Madrid (UPM), Madrid, Spain
Dpto. de Automatica, Universidad de Alcala (UAH), Alcala de Henares, Spain
Dpto. Matematica aplicada a las TIC, Information Processing and Telecommunications Center, ETSI Telecomunicacion, Universidad Politecnica de Madrid (UPM), Madrid, Spain

Abstract—The TCP-100 parabolic trough collectors research facility at Plataforma Solar de Almer´ıa (CIEMAT) has replaced the 32 years old ACUREX research plant, widely used in Automatic Control research. The new TCP-100 structure poses new challenges in the design of more sophisticated operation modes. This paper presents a hybrid first principles system level model of the TCP-100 facility, which is based on two connected parts: a previously published non-linear continuous dynamic model, and a discrete part modelling the behaviour of an operator. This model is able to cover the different operation modes of the plant and proves its applicability for operation training activities. The discrete part of the model characterizes the operation procedure using both the StateGraph formalism in the Modelica object oriented modelling language, and the Dymola tool. We illustrate the validity of our hybrid model by simulating two different operation procedures applied to the TCP-100 plant model in a typical operation day in which the system passes through several operating modes.

Design and Application of the Prevention Model based on the Examination of Academic Plagiarism

Ziran Fan, Takayuki Fujimoto
Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan

Abstract—Currently, the plagiarism is continuously seen in the academic fields. The one of the reasons is thought that the accurate application of plagiarism results in obscurity. To solve this issue related to the plagiarism, this study examines the typical examples of the plagiarism in recent years and surveys the relevant references of plagiarism, and proposesthe general definition of the plagiarism from the aspect of "uniqueness." Moreover, this study considers the prevention of the plagiarism along the idea that attempts to prevent the plagiarism at the stage before plagiarism occurring. It is different from the ordinary idea that focuses on how to detect the plagiarism after it happened. We conceive the prevention model of the plagiarism based on that idea, design the application utilizing the model and show its mechanism.

Hyperbolic Trees in Complex Networks

Zalan Heszberger, Andras Gulyas, Andras Biro, Laszlo Balazs, Jozsef Biro
MTA-BME Information Systems Modelling Research Group, Department of Telecommunications and Media Informatics, Faculty of EE and CS, Budapest University of Technology and Economics, Budapest, Hungary

Abstract—The two-dimensional hyperbolic space turned out to be an efficient geometry for generative models of complex networks. The networks generated with this hyperbolic metric space share their basic structural properties (like small diameter or scale-free degree distribution) with several real networks. In this paper, we present a new model for generating trees in the two-dimensional hyperbolic plane. The generative model is not based on known hyperbolic network models: the trees are not inferred from the existing links of any network; instead, the hyperbolic tree is generated from scratch purely based on the hyperbolic coordinates of nodes. We show that these hyperbolic trees have scale-free degree distributions and are present to a large extent both in synthetic hyperbolic complex networks and real ones (Internet autonomous system topology, US flight network) embedded in the hyperbolic plane.

Solving Cryptarithmetic Puzzles by Logic Programming

Feng-Jen Yang
Department of Computer Science, Florida Polytechnic University, Lakeland, Florida, USA

Abstract—As a personal interest of study, I tried a logic programming approach towards the problem solving of cryptarithmetic puzzles that are commonly discussed as a subcategory of constraint satisfaction problems in the literature of artificial intelligence. While there are possibly several methods capable of solving constraint satisfaction problems, I took into consideration the efficiency as well as the completeness that will identify all possible solutions under the specified constraints and exclude trivial and useless solutions from the perspective of real-life practice. In this paper, I demonstrated an approach that can be adapted to solve most of the constraint satisfaction problems especially within the context of cryptarithmatic puzzles. This method will also perform forward checking to have early backtracking and prevent searching the entire search tree exhaustively.

Preventing Drowning in Information: A Topic Model Approach to Relating Information on Strategic Scanning

Alexis Miranda Carrillo, Edison Loza-Aguirre, Carlos Montenegro
Escuela Politecnica Nacional, Facultad de Ingenieria en Sistemas, Quito, Ecuador
Escuela Politecnica Nacional, Departamento de Informatica y Ciencias de la Computacion Quito, Ecuador

Abstract— Information overload lead managers to not adequately use the relevant information they collect in Strategic Scanning. Such information comes in small pieces of text that are dispersed in terms of time, language, and sources. These characteristics of information from Strategic Scanning usually prevent to identify connections with the previously collected information. In this paper, we propose an alternate tool for dealing with this issue by using topic analysis techniques. The tool provides a quick reading interface, in which the proximity relationship between various texts can be easily visualized. We compare our tool with two mechanisms for clustering and proximity measurement. Our tool excelled in terms of execution time and of the pertinence of results.

Environmental-Economic Dispatch with Renewable Sources Forecasting and Energy Storage

Elisa Espinosa-Juarez, Jorge Luis Solano-Gallegos, Fernando Ornelas-Tellez
Faculty of Electrical Engineering, Universidad Michoacana de San Nicolas de Hidalgo, Morelia, Mich., Mexico

Abstract— This paper presents the problem of multiobjective economic dispatch for an electrical system including renewable energy penetration and energy storage. The wind power and photovoltaic power forecast, as well as the power demand forecast is considered in the formulation of the onehour ahead environment-economic dispatch. Case studies are presented, which show the effectiveness of the methodology implemented to analyze the impact of renewable energy sources and energy storage in order to minimize the generation costs and the pollution emissions produced by the conventional generation units, throughout of a period of time.

Is Entropy Enough for Measuring Privacy?

Sevgi Arca, Rattikorn Hewett
Department of Computer Science, Texas Tech University, Texas, USA

Abstract—Anonymization is critical to privacy. It helps protect the identity and sensitive information of individuals from their profile data. Knowing the degree of anonymity attained is an important step to advance privacy and anonymization techniques. However, little research has focused on articulating a measure to quantify the quality of anonymization. On the other hand, many have used popular Shannon's entropy, a wellestablished measure from information theory, as a way to measure anonymity. In this paper, we take a closer look at the meaning, the distinction and the relationship between anonymity and entropy with respect to privacy. We argue that, even though information entropy is used amply as a metric for anonymity, it is not a befitting measure. Furthermore, although parts of the entropy's information theory are relevant, they alone are not adequate to be a proper measure for anonymity. This paper presents a simple, intuitive, and theoretically grounded measure for anonymity. We provide a comparison analysis between our measure with other entropy-based metrics along with experiments to show the effectiveness of our proposed measure.

Proximity in the Brain

Zalan Heszberger, Andras Gulyas, Andras Biro, Laszlo Balazs, Szabolcs Mezei, Jozsef Biro
MTA-BME Information Systems Modelling Research Group, Department of Telecommunications and Media Informatics, Faculty of EE & CS, Budapest University of Technology and Economics, Budapest, Hungary

Abstract—The structural navigability of complex networks is an important question in the function-structure perspective of complex network analysis. This may provide hints on the underlying mechanisms that have been forming the structure of networks for a desirable level of navigation. It has been already discovered that greedy navigational cores as minimalistic networks with 100% greedy navigability considerably present in many real networks, including the structural networks of the human brain. Because the greedy navigational core is not unique, the connection between the level of its presence in a network and the structural navigability of that network is far from clear. In this paper, we deal with a special subset of the greedy navigational core, the so-called greedy proximity links (GPL), whose presence is necessary for 100% greedy navigability of a network. We show that the greedy proximity links are highly present in the brain networks, and the presence is consistent throughout the the individual subjects.

All Nearest Neighbors Query Including Scores Road Network

Hyo-Kyun Kim, Tae-Sun Chung
Department of Software, Ajou University, Suwon, Gyeonggi-do, South Korea

Abstract— This paper introduces an improved ANN (All Nearest Neighbor) algorithm using the SCL (Standard Clustered Loop) algorithm to reduce the consumption of computing resources that can occur when searching for the data object nearest to the query object in the process of executing the algorithm. Additionally, a method to improve ANN algorithm is proposed. When the

algorithm is executed, it is a situation in which the user finds a data object adjacent to the user. In this case, our technique applies the criteria set provided by users.

Designing a Composite Platform for Operational Efficiency

Wangjie Xu, Takayuki Fujimoto, Ziran Fan
Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan

Abstract—This paper focuses on the problem of in-house split information systems and the resulting decline in operational efficiency. According to the result of the authors' investigation for the business environment of small and medium-sized enterprises, the simplification of information and communication among employees has become an issue to be solved in many enterprises today in order to improve operational efficiency. Based on this, in this paper, we design a composite platform that integrates information tools commonly used in companies. By covering all information-related operations with one platform, we support improvement of companies' operational efficiency. We also show the effectiveness of the proposed platform by comparing it with existing information systems. We mainly examine the design concept of the composite platform and its construction requirements.

Using Serde to Serialize and Deserialize DIS PDUs

Noah W. Scott, Douglas D. Hodson, Richard Dill, Michael R. Grimaila
Air Force Institute of Technology, WPAFB, Ohio, USA

Abstract—Serialization is the process of translating a data structure into a format that can be stored and/or transmitted, and then subsequently reconstructed at a later time to create an identical clone of the original. The use of data serialization assures data objects can be transmitted, stored, and reliably reconstructed across differing computer architectures, even with different data type sizes or endianness, with no additional effort. Serializing the data in an architecture-independent format prevents the problems of byte ordering, memory layout, or representing data structures in different programming languages. This is especially important in the context of live, virtual, and constructive (LVC) simulation environments where multiple geographically separated computers, each with many independent threads, are connected and must communicate with as little latency as possible to remain near "real-time" like in terms of responsiveness. In this paper, we demonstrate the use of Serde, a Rustbased systems programming language crate, to serialize and deserialize IEEE standard Distribute Interactive Simulation (DIS) Protocol Data Units (PDUs) to support DIS-based network interoperability. The results show that Serde is an efficient mechanism for serialization/deserialization when using the inherently safe Rust programming language.

A Framework for Modeling a Real-Time Radar System

Brennen T. Garland, Douglas D. Hodson, Scott L. Nykl, Richard Dill, Michael R. Grimaila
Air Force Institute of Technology, Wright Patterson AFB, Ohio, USA

Abstract—Models are useful mathematical abstractions of the real-world systems. These models are parameterized, to the detail necessary, to enable the study of systems dynamics under a wide variety of different conditions that would be difficult, if not impossible, to create otherwise. In this paper, we present the Mixed Reality Simulation Platform (MIXR), a framework for modeling radar systems with sufficient detail needed to support typical real-time interactive military training exercises and experiments. MIXR has been in development and use since the late 1980s, with the current incarnation being written in the object-oriented C++ programming language. The usefulness of the framework to help simulation engineers and software developers rapidly prototype both standalone and distributed applications has proven itself over the years, as many different simulation products have been created and used within the military domain. In this seminal paper, we focus exclusively on how radar systems are modeled within the MIXR code-base.

Computational Methods and Techniques for Case-Based Reasoning (CBR)

Souad Demigha
CRI, University of Paris 1 Sorbonne, Paris, France

Abstract—Case-based reasoning (CBR) is a computational model of problem solving. It uses a representation of specific episodes of problem solving to learn to solve a new problem. CBR uses the experience of past problem solving when solving a new problem. CBR systems store past experience as individual problem solving episodes. These two parts of case-based reasoning are modeling as a computational model. In this context, we propose to present a methodology allowing the performance of the CBR system. The methodology consists of designing experiments to control the variability in the behavior of the CBR system and obtaining empirical performance data.

CDMI: A Clockwise-Displacement Algorithm to Compute Multiplicative Inverse

Hashim Abu-Gellban, Long Nguyen
Department of Computer Science, Texas Tech University, Lubbock, Texas, USA

Abstract—A multiplicative inverse (MI) algorithm is used in several fields, like cryptography algorithms. There are many MIs, however, these algorithms suffer from using several multiplication and division operations, which take more execution time than additions and subtractions. We created a new algorithm called Clockwise-Displacement (CDMI) by using addition and subtraction operations in the iterative steps instead of multiplications and divisions. Additionally, numerous MIs face the undecidable problem because of the floating-point issue. Whereas, CDMI tackles this issue by converting the domain from the floatingpoint space to integer space. Therefore, CDMI declines the time consuming to calculate the multiplicative inverse by applying fewer divisions and multiplications (expensive operations) and addresses the rounding error issue in some MI Algorithms.

CSCI-ISCB: COMPUTATIONAL BIOLOGY

First Success of Cancer Gene Data Analysis of 169 Microarrays for Medical Diagnosis

Shuichi Shinmura
Emeritus Professor, Seikei University, Kashiwa, Japan

Abstract - I succeed in the first high-dimensional gene data analysis of 169 microarrays. I completed the new theory of discriminant analysis in 2015 and developed Revised IP-OLDF (RIP) based on a minimum number of misclassifications (Minimum NM, MNM). As an applied research theme, RIP discriminated six microarrays (data) and showed six MNMs were zero. The Matryoshka feature selection method (Method2) divided data into many linearly separable gene sets (Small Matryoshkas, SMs). After 2019, RIP and Method2 confirm these facts by 163 data. Although many researchers studied this theme from 1995, they did not know LSD-discrimination and failed in this theme.

Characterizing Focal and Generalized Epileptic Networks Using Interictal Functional Connectivity

Elaheh Zarafshan, Hoda Rajaei, Parisa Forouzannezhad, Ulyana Morar, Mercedes Cabrerizo, Malek Adjouadi
Department of Electrical and Computer Engineering, Florida International University, Florida, USA
Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Massachusetts, USA

Abstract —Using electroencephalography (EEG) data from epileptic patients1 , we investigated and compared functional connectivity networks of three various types of epileptiform discharges (ED; single, complex & repetitive spikes) in 4 regions of the brain. Our results showed different connectivity patterns among three ED types within-and between-brain regions. The one-way ANOVA test indicated significant differences between the mean of the average connectivity matrices (ACMs) of the single spike, which characterize focal epilepsy, and the other two ED types (complex & repetitive) which characterize generalized epilepsy. The interictal EEG segments, through the connectivity patterns they yield, could be considered as one of the key indicators for the diagnosis of focal or generalized epilepsy.

Development of Multistage RFE-SVR Model to Predict Radiation Sensitivity

Sharmin Nahar Mithy, Grisselle Centeno, Ibrahim Khalilullah
Industrial and Management Systems Engineering, University of South Florida, Tampa, Florida, USA
Department of Data Science and Business Analytics, Florida Polytechnic University, Lakeland, Florida, USA
Department of Computer Science, Lamar University, Beaumont, Texas, USA

Abstract— As radiation therapy (RT) has been in usage as a universal part of cancer patient treatment, the correlation between RT and patient profiles have become of interest to the research scholars and clinicians. There have been many studies which suggest a strong relation between radiation therapy and the genomic expression profiles of the patients. The analysis of the gene expression profiles poses a huge challenge due to the high dimensionality and the class imbalance problem. Identifying the best useful genes and eliminating the redundant ones is one of the key factors when analyzing genomic data. In our study, we have established a prediction model which identifies the radiation sensitivity index for the patients. For our study we have used SF2 (surviving fraction following irradiation with 2 GY) as the radiation sensitivity index. This multistage RFE-SVR model identifies the most useful features to predicts the SF2 value. We eliminate features step-by-step from the original set of features with a combination of filter and embedded method to avoid the loss of function.