CSCI 2022 BOOK of ABSTRACTS

The 2022 International Conference on Computational
Science and Computational Intelligence (CSCI'22)

https://www.american-cse.org/csci2022/
December 14-16, 2022
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 2022 International Conference on Computational Science and Computational Intelligence (CSCI'22), December 14-16, 2022, Las Vegas, Nevada, USA.

The CSCI'22 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'21) is composed of the following Research Tracks: Computational Science (CSCI-RTCS); Computational Intelligence (CSCI-RTCI); Computational Biology (CSCI-RTCB); Cyber Warfare, Cyber Defense, & Cyber Security (CSCI-RTCW); Artificial Intelligence (CSCI-RTAI); Smart Cities and Smart Mobility (CSCI-RTSC); Big Data and Data Science (CSCI-RTBD); Education - CS & CE (CSCI-RTED); Health Informatics and Medical Systems (CSCI-RTHI); Mobile Computing, Wireless Networks, & Security (CSCI-RTMC); Software Engineering (CSCI-RTSE); Internet of Things & Internet of Everything (CSCI-RTOT); Social Network Analysis, Social Media, & Mining (CSCI-RTNA); Cloud Computing and Data Centers (CSCI-RTCC); Parallel & Distributed Computing (CSCI-RTPD); and Signal & Image Processing, Computer Vision & Pattern Recognition (CSCI-RTPC).

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 68 countries. About 59% 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 19%; and 24% 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'22. 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, De Gruyter, 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: Javier Gurrola, Patrick Kellenberger, Jennifer Carruth, and Tricia Yamaguchi.

We present the proceedings of CSCI'22.

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, School of Computing and Data Science, Wentworth Institute of Technology, Boston, 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

Co-Chairs, Steering Committee, CSCI'22

Conference Organization – CSCI 2022

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, School of Computing, The University of Georgia, Georgia, USA

Professor Leonidas Deligiannidis - Professor, School of Computing and Data Science, Wentworth Institute of Technology, Boston, 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

Members of International Program Committee (CSCI 2022)

- **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

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.

KEYNOTE LECTURES

Explainability and Accountability in Deep Few-Shot Learning: A Case Study in Medical Imaging Informatics

Dr. Ahmad P. Tafti
Faculty, Health Informatics, University of Pittsburgh, USA and
Director, Pitt Health + Explainable AI (Pitt HexAI)
Vice Chair, IEEE Computer Society at Pittsburgh

Abstract - Of late, deep learning computer vision algorithms have played as state-of-the-art strategies for a list of medical image analysis tasks, ranging from medical image segmentation and registration to image classification. However, there are several fundamental challenges that stop deep learning methods to obtain their full potential in healthcare domain. One can see that they often need a large column of labeled training data to achieve better accuracy and precision over traditional ML methods. This could lead to a list of problems in clinical settings, including but not limited to: 1) Annotating medical data/images requires significant medical knowledge which makes the annotation process expensive and very time consuming, 2) Providing a large amount of annotated/labeled data in a clinical setting is basically challenging due to high privacy-preserving standards in healthcare community, and 3) Medical images are often imperfect due to heterogeneity in medical imaging devices and equipment. In this talk, we will see how to build, train, test, and validate an explainable and accountable deep few-shot learning model to analyze medical images with only very few samples of manually annotated radiographs.

Measuring, Optimizing and Improving Collective Intelligence & Group Performance Using Artificial Intelligence

Massoud Alibakhsh
Founder & CEO, Xeba Technologies, Atlanta Georgia, USA
www.xeba.tech

Abstract - By examining a trained Neural Net for a given task, we can conclude that intelligence is coded in connections between the nodes. By way of analogy, an organization made up of humans organized to accomplish a defined task can succeed gloriously or fail miserably. What is the role of connectivity using various models and modes of communications systems amongst humans at work? This talk examines current group connectivity models and introduces a methodology to not only maximize group intelligence but transcend it to higher levels by fully and seamlessly integrating AI into workflow.

High-Level Development of Distributed Real-Time Applications Using Mobile Cloud and Software-Defined Networking

Prof. Sergei Gorlatch
Professor & Chair of Parallel and Distributed Computing,
University of Muenster, Germany

Abstract - We address the software development process for challenging class of Real-Time Online Interactive Applications (ROIA). ROIA are distributed applications connecting a potentially very high number of users who interact with the application and with each other in real time, i.e., a response to a user's action happens virtually immediately. Typical representatives of ROIA are massively multiplayer online computer games, advanced simulation-based e-learning and serious gaming. These applications are characterized by high performance and Quality-of-Service (QoS) requirements, such as: short response times to user inputs (about 0.1-1.5 sec); frequent state updates (up to 100 Hz); large and frequently changing numbers of users in a single application instance (up to tens of thousands simultaneous users). This talk will present our high-level development framework RTF and address two challenging issues in future Internet-based ROIA applications: a) using Mobile Cloud Computing for allowing high application performance when a ROIA application is accessed from multiple mobile devices, and b) managing dynamic QoS requirements of ROIA applications by employing the emerging technology of Software-Defined Networking (SDN).

CSCI-RTAI: ARTIFICIAL INTELLIGENCE

Characterization of Emerging AI Workloads: Neural Logic Machines and Graph Convolutional Networks

Cory Davis, Patrick Stockton, Eugene B. John, Zachary Susskind, Lizy K. John
The University of Texas at San Antonio, San Antonio, USA;
The University of Texas at Austin, Austin, Texas

Abstract - The present renaissance of artificial intelligence has created new domains of AI models. Some of these domains include Neuro-Symbolic AI (NSAI) and Graph Neural Networks (GNN). NSAI and GNN models have already demonstrated the capability to outperform state-of-the-art deep learning models in domains such as image and video reasoning, and network classification respectively. They have also been shown to obtain high accuracy with significantly less training data than traditional deep neural network models. Due to the recency of the fields' emergence and relative sparsity of published results, the performance characteristics of these models are not well understood. In this paper, we describe and analyze two recent models in these domains. We find that the symbolic model has less potential parallelism than traditional neural models due to complex control flow and low-operational-intensity operations and high cost of data movement. Additionally in the GNN model, we find an abundance of sparse matrix multiplication. Dense MM has a high potential for parallelism through usage of tensor cores, meanwhile new techniques for increasing parallelism in sparse matrix multiplication will be of extreme importance for GNN models.

Offline Reinforcement Learning using the Advanced Framework for Simulation, Integration, and Modelling Software

Althea Henslee, Haley Dozier, Abderahim Salhi, Indu Shukla, Jules White, Ian Dettwiller
Information Technology Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, Mississippi, USA;
Computer Science and Computer Engineering, Vanderbilt University, Nashville, Tennessee, USA

Abstract - Recent operational challenges have prompted a renewed interest in mission-engineering tools that can advance decision-making and innovation in defense. These tools have the ability to simulate high-fidelity combat engagements to analyze operational concepts, platforms, systems, and capabilities in multiple domains to develop optimal solutions to the intricate and difficult issues within the field of mission engineering. One such tool is the Advanced Framework for Simulation, Integration, and Modeling (AFSIM). AFSIM is a standard combat simulation tool that includes a variety of applications and plugins that can be used for systems engineering as well as engagement, mission, or campaign-level simulation. Due to software challenges, using AFSIM in combination with Reinforcement Learning (RL) poses particular challenges that must be overcome before RL will gain widespread use in the AFSIM community. This study proposes a novel way to utilize synthetic data created through AFSIM by using offline, or data-driven, RL.

CLPL: A Self-supervised Contrastive Learning Pseudo-Labeling Framework for Tabular Data

Abolfazl Farahani, Navid Hashemi Tonekaboni, Khaled Rasheed, Hamid R. Arabnia
School of Computing, University of Georgia, Athens, Georgia, USA;
Department of Computer Science, College of Charleston, Charleston, South Carolina, USA

Abstract - Semi-supervised and self-supervised learning methods have been recently applied to many domains and applications due to their promising solution in alleviating the need for large labeled datasets. One of the most recent trending semi/selfsupervised learning approaches that have achieved significant interest in the research community is pseudo-labeling. Pseudolabeling algorithms exploit a small set of labeled and an extensive collection of unlabeled data to learn the underlying representation of input data. In unstructured domains like image, text, and audio, pseudo-labeling can significantly benefit from contrastive learning to obtain a generalized model by learning abstract representations in an unsupervised fashion. However, contrastive learning is hardly applicable to the tabular domain as it requires data augmentation by applying a set of pre-defined transformations, while there are no such transformations that can be safely applied to the tabular data. Thus, we introduce a contrastive learning pseudo-labeling framework (CLPL) that reduces the generalization error by learning the underlying structure of class-specific representations. To the best of our knowledge, contrastive learning has not been used for pseudolabeling tasks in the tabular domain. Our experiments show the effectiveness of CLPL by comparing it with the known state-ofthe-art pseudo-labeling methods proposed for the tabular data.

High-Dimensional Probabilistic Time Series Prediction via Wavenet+t

Xiuxuan Sun, Jianhua Chen
Division of Computer Science & Engineering, Louisiana State University, Baton Rouge, Louisiana, USA

Abstract - High-dimensional time series inference plays a crucial role in various f ields (e.g., economic analysis, inventory analysis, electricity consumption, and stock market forecasting). However, classical time series models mostly deal with handling one dimension time series dataset with point estimate. In this work, we use a variant of WaveNet [9] in combination with a probability distribution (e.g., Student's t-distribution) for multivariate probabilistic time series prediction. WaveNet [9] and other related works [10, 11] used dilated causal convolutional neural networks (CNN) to extract the long/short term patterns from time series dataset. It also integrates residual network with the dilated causal CNN to solve the vanishing/exploding gradient problems and make models to be more expressive. Multi-step, probabilistic prediction for multivariate time series is generated by sampling from the conditional distribution (given input data) produced by the proposed WaveNet+t network. Our model demonstrates better or comparable performance on different realworld high dimensional time series dataset (e.g., Wikipedia with 9535 variables) when compared with the other state-of-the art multivariate probabilistic models.

Cloud-based Sepsis Prediction System with Neural Architecture Search Service

Yen-Han Chiang, Hsin-Yu Lee, Yi-Wei Huang, Hui-Shan Chen, Yi-Lun Pan
National Center for High-performance Computing, High Performance Computing, Hsinchu City, Taiwan;
National Center for High-performance Computing, High Performance Computing, Taichung City, Taiwan

Abstract - Sepsis is a common disease with prohibitive medical costs and life-threatening consequences. Early prediction of sepsis and initiation of antibiotics is widely recognized as essential determinant of patient survival. In this work, this study proposes a novel sepsis accuracy prediction method called Cloud NAS-based Sepsis prediction system, which utilizes cloud technology (Cloud NAS Framework) - Container technique and neural architecture search/optimization technology. Furthermore, this study fully integrates the hyperparameter tuning

mechanism and the matching neural network architecture search. The end user can use the model recommended by this study to land and make clinical sepsis and other related predictions to assist the first-line judgment. To verify the usability of this method, we also applied several public Datasets (e.g., MIMIC-IV) to conduct practical tests. The experimental data shows that this study not only searches for the model but also optimizes the hyperparameters in a lowcost way and mainly uses 2 different datasets for training and validation. Use the 2019 Pysionet dataset for model training and testing and MIMIC-IV for final verification. And the last model has an AUROC score of 0.95 within 12 hours of prediction and can predict the onset of sepsis eight hours earlier.

Reconstructing Word Representations from Pre-trained Subword Embeddings

Paula Lauren
Department of Mathematics and Computer Science, Lawrence Technological University, Southfield, Michigan, USA

Abstract - Subword embeddings are integral to transformers such as Generative Pre-trained Transformer (GPT) and Bidirectional Encoder Representations from Transformers (BERT), which are utilized in various natural language processing tasks. A subword is a subset of a word and a word can be decomposed into one or more subwords and characters. One challenge with subword tokenization is determining the optimal tokens for representing a word. This research proposes a refinement to the GPT tokenizer based on an analysis of subword tokens from GPT. The proposed approach reconstructs the word with the objective being minimization and refinement of subword tokens in the reconstruction of the word. The evaluation of the proposed approach is done using popular analogy datasets on semantic and syntactic tasks. This research shows the morphological emphasis on reconstructed word embeddings from pre-trained subword embeddings generated from GPT. The proposed approach reports an overall improvement of approximately ten percent for the refined GPT tokenizer on syntactic tasks.

Unsupervised Machine Learning Methods for Diagnosing Autism Spectrum Disorder Using Multimodal Data: A Survey

Antika Roy, Pablo Rivas, Mahee Noor Tayba

Rivas.AI Lab, USA;
School of Engineering and Computer Science, Department of Computer Science, Baylor University, Waco, Texas, USA;
Department of Computer Science, Baylor University, Waco, Texas, USA

Abstract - The neurodevelopmental disorder known as autism spectrum disorder (ASD) is becoming increasingly pervasive worldwide. It can be categorized when a person struggles to interact and communicate socially and also exhibits repetitious behaviors and interests. Although early intervention is crucial and may have long-term advantages in the lives of people with ASD, diagnosing ASD is difficult due to its heterogeneous characteristics and a large number of data from various aspects such as genetic, behavioral, electronic health records, and many other domains. This review paper offers an in-depth assessment of papers that use the most commonly used unsupervised machine learning techniques in ASD, such as “k-means clustering” and “Hierarchical clustering”. This research aims to identify and highlight the most recent unsupervised machine learning advances in the ASD literature while also attempting to highlight the significant contributions and limits of the selected works to provide insight for future researchers.

Experimental Study on Decision Fusion Parameters using Alpha Integration

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

Abstract - This paper presents an experimental study of the convergence and results of decision fusion parameters using alpha integration algorithms. The decision fusion methods are applied to the problem of detection of ultrasound pulses inserted in high-level background noise. The implemented alpha integration algorithms were compared with the following single classifiers: support vector machine, quadratic discriminant analysis, and random forest, and with the results of the fusion using the mean of the scores provided for those single classifiers. The convergence of the parameters of alpha integration-based methods is analyzed as well as the distribution of the scores (posterior probability) of all the implemented methods. The results show the capabilities of the alpha-integration methods to improve the classification performance over the other methods considered in comparisons.

Word Embedding Neural Networks to Advance Knee Osteoarthritis Research

Soheyla Amirian, Husam Ghazaleh, Mehdi Assefi, Hilal Maradit Kremers,
Hamid R. Arabnia, Johannes F. Plate, Ahmad P. Tafti
University of Georgia, USA; Quincy University, USA; Mayo Clinic, USA;
University of Pittsburgh, USA

Abstract - Osteoarthritis (OA) is the most prevalent chronic joint disease worldwide, where knee OA takes more than 80% of commonly affected joints. Knee OA is not a curable disease yet, and it affects large columns of patients, making it costly to patients and healthcare systems. Etiology, diagnosis, and treatment of knee OA might be argued by variability in its clinical and physical manifestations. Although knee OA carries a list of well-known terminology aiming to standardize the nomenclature of the diagnosis, prognosis, treatment, and clinical outcomes of the chronic joint disease, in practice there is a wide range of terminology associated with knee OA across different data sources, including but not limited to biomedical literature, clinical notes, healthcare literacy, and health-related social media. Among these data sources, the scientific articles published in the biomedical literature usually make a principled pipeline to study disease. Rapid yet, accurate text mining on large-scale scientific literature may discover novel knowledge and terminology to better understand knee OA and to improve the quality of knee OA diagnosis, prevention, and treatment. The present works aim to utilize artificial neural network strategies to automatically extract vocabularies associated with knee OA diseases. Our finding indicates the feasibility of developing word embedding neural networks for autonomous keyword extraction and abstraction of knee OA.

EXPANSE: A Continual and Progressive Learning System for Deep Transfer Learning

Mohammadreza Iman, John A. Miller, Khaled Rasheed, Robert M. Branch, Hamid R. Arabnia
School of Computing, University of Georgia, USA;
Institute for Artificial Intelligence, University of Georgia, USA;
Learning, Design, and Technology, College of Education, University of Georgia, USA

Abstract - Deep transfer learning (DTL) techniques attempt to tackle the limitations of deep learning, the dependency on extensive training data and the training costs by reusing obtained knowledge from source data for target data. However, the current DTL techniques suffer from either catastrophic forgetting dilemma (losing the

previously obtained knowledge) or overly biased pre-trained models (harder to adapt to target data) in fine-tuning pre-trained models or freezing a part of the pre-trained model, respectively. We propose a new continual/progressive learning approach for deep transfer learning to tackle these limitations. We extend the pre-trained model by expanding pre-trained layers (adding new nodes to each layer) in the model instead of only adding new layers. Hence the method is named EXPANSE. Our experimental results confirm that we can tackle distant source and target data using this technique. At the same time, the final model is still valid on the source data, achieving a promising deep continual learning approach. Moreover, we offer a new way of training deep learning models inspired by the human education system. We termed this two-step training: learning basics first, then adding complexities and uncertainties. The evaluation implies that the two-step training extracts more meaningful features and a finer basin on the error surface since it can achieve better accuracy compared to regular training.

Multi-Pollutant Ground-Level Air Pollution Prediction through Deep MeteoGCN-ConvLSTM

Pratyush Muthukumar, Shaurya Pathak, Kabir Nagrecha, Hirad Hosseini, Dawn Comer, Navid Amini, Jeanne Holm, Mohammad Pourhomayoun
Department of Computer Science, California State University Los Angeles, Los Angeles, California, USA; City of Los Angeles, Los Angeles, California, USA

Abstract - Air pollution is the fourth-largest threat to human health. The harmful effects of air pollutants have costed the global economy nearly $3 trillion. It is imperative that a solution for mitigating the harmful effects of the most pervasive groundlevel air pollutants . Carbon Monoxide (CO), Nitric Oxide (NO), Nitrogen Dioxide (NO2), Ozone (O3), and particulate matter 2.5 (PM2.5) . is implemented, especially in urban areas. Recent advances in deep learning such as the Convolutional Long Short Term Memory (ConvLSTM) architecture are capable of learning complex spatiotemporal patterns with multisource data. We propose a novel sequential encoder-decoder ConvLSTM architecture capable of predicting hourly CO, NO, NO2, O3, and PM2.5 spatially continuously over Los Angeles. Our model utilizes multisource satellite imagery collected from the ESA Tropospheric Monitoring Instrument (TROPOMI), remote-sensing data collected by the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) instrument onboard the NASA Terra+Aqua satellites, and site-monitoring sensor observations of atmospheric and ground-level air pollution, meteorological data, and wildfire data. Our results show that our MeteoGCNConvLSTM model is competitive with state-of-the-art approaches across all predicted air pollutants. Moreover, our results show the versatility of our model when provided with solely atmospheric satellite imagery and remote-sensing data as input.

A Tensor Decomposition in Multi-way Electroencephalogram (EEG) Data Analysis

Essence J. Stevenson, Roger B. Sidje
Paul W. Bryant High School, Tuscaloosa, Alabama, USA;
Department of Mathematics, The University of Alabama, Tuscaloosa, Alabama, USA

Abstract - Acquiring large datasets has advantages and disadvantages; an advantage being the bigger the datasets, the more information potentially captured; while a disadvantage being the amount of labor needed to process this information. To combat this disadvantage, researchers in a variety of fields increasingly rely on using tensors to represent high-dimensional data, and then using tensor decompositions to compress the data without losing significant information. One such field is electroencephalography (EEG), which is the study of electrograms that measure brain electrical activity. Having the ability to be continuously recorded for long periods of time, this could be hours, days, even weeks, EEG data can be massive. Here we discuss how tensor decomposition methods such as Parallel Factor (PARAFAC) analysis and Tucker decomposition can be executed on these large datasets.

Which Deep Learning Framework Should I Use: A Comparative Study for Deep Regression Modeling

Vishu Gupta, Wei-Keng Liao, Alok Choudhary, Ankit Agrawal
ECE Department, Northwestern University, Evanston, Illinois, USA

Abstract - The combined impact of deep learning techniques and computing resources with an increase in the availability of databases is transforming many research fields leading to technological advances to help solve real-life and research-related problems. In the recent years, various deep learning frameworks and libraries have been developed to implement these algorithms that can operate efficiently at large scale and heterogeneous environments. However, these implementations can vary depending on the framework leading to unexpected inconsistency in the results even for the same algorithm. This irregularity is seen more often in deep learning models trained using advanced parallel computing resources such as GPUs. In this study, we perform an investigation with three of the well-known deep learning frameworks: Tensorflow 1, Tensorflow 2 with Keras, and Pytorch for regression-based problems in physical sciences to analyze how the results vary depending on the framework. We implement different deep neural networks with varying complexity and perform accuracy, time, and computational based analysis to study the effect of the framework on model accuracy, training/testing times, reproducibility, and memory usage.

Machine Learning Methods for Evaluating Public Crisis: Meta-Analysis

Izunna Okpala, Shane Halse, Jess Kropczynski
University of Cincinnati, Cincinnati, USA

Abstract - This study examines machine learning methods used in crisis management. Analyzing detected patterns from a crisis involves the collection and evaluation of historical or near-realtime datasets through automated means. This paper utilized the meta-review method to analyze scientific literature that utilized machine learning techniques to evaluate human actions during crises. Selected studies were condensed into themes and emerging trends using a systematic literature evaluation of published works accessed from three scholarly databases. Results show that data from social media was prominent in the evaluated articles with 27% usage, followed by disaster management, health (COVID) and crisis informatics, amongst many other themes. Additionally, the supervised machine learning method, with an application of 69% across the board, was predominant. The classification technique stood out among other machine learning tasks with 41% usage. The algorithms that played major roles were the Support Vector Machine, Neural Networks, Naive Bayes, and Random Forest, with 23%, 16%, 15%, and 12% contributions, respectively.

On Unsupervised Reconstruction with Dressed Multilayered Variational Quantum Circuits

Pablo Rivas, Liang Zhao
School of Engineering and Computer Science, Department of Computer Science,
Baylor University, Waco, Texas, USA;
Marketing Department, St. Ambrose University, Davenport, Iowa, USA

Abstract - The advantages of unsupervised quantum machine learning are still under study and appear to be very promising. Trainable variational quantum circuits are one example of successful approaches to combining classic machine learning and quantum. However, there is no clear path toward a quantum advantage for different types of variational circuits. This paper furthers the research efforts in understanding the potential and applications of

hybrid quantum circuits. We study different circuits and see how similar they perform in an unsupervised learning task in an autoencoder configuration over a large multimodal dataset.

A Comparative Analysis of State-of-the-Art Time Series Forecasting Algorithms

Cathal Murray, Priyanka Chaurasia, Lynsey Hollywood, Damien Coyle
School of Computing, Engineering & Intelligent Systems, Ulster University, Londonderry, United Kingdom (UK);
Department of Hospitality & Tourism Management, Ulster University, Belfast, United Kingdom (UK)

Abstract - In recent years many new algorithms have been developed for applications in speech and image processing which can be repurposed for time series prediction. This paper presents a comprehensive comparative analysis of time series forecasting capabilities of eight such state-of-the-art algorithms, namely: Vanilla Long Short-Term Memory(V-LSTM) Gated Recurrent Unit (GRU), Bidirectional LSTM(BD-LSTM), Auto encoder (AELSTM), Convolutional Neural Network LSTM(CNN-LSTM), LSTM with convolutional encoder (ConvLSTM), Attention mechanism networks and the Transformer network. Model performances across five different benchmark datasets including fields of interests such as finance, weather and sales are evaluated. Whether direct or iterative prediction methods are optimal for forecasting is investigated. For efficient model optimization, the asynchronous successive halving algorithm (ASHA) is applied in the training folds in a 10 k-fold cross validation framework. Statistical tests are used to comprehensively compare algorithm performances within and across datasets. We show that whilst there are differences between all models, the differences are insignificant for the top performing models which include the Transformer, Attention, V-LSTM, CNN-LSTM and CV-LSTM. However, the transformer model consistently produces the lowest prediction error. We also show that the iterative multistep ahead prediction method is optimal for long range prediction.

Performance Reliability of Reinforcement Learning Algorithms in Obstacle Avoidance Game with Differing Reward Formulations

Brandon Hansen, Haley Dozier
Information Technology Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, M.S., USA

Abstract - When formulating environments for complex application areas, using analogies to games is beneficial as they provide convenient models to test algorithm performance in ways that are transferable to realistic environments. We propose a Frogger like grid based environment containing a simple action space, dynamic obstacles, and discrete game loop for testing Proximal Policy Optimization 2 and Deep Q-Network with comparisons to a heuristic and random agent. The environment contains four different reward function implementations, along with two different environment variations to explore adaptability. Seeing how these different parameters effect not just the average performance of the algorithm, but also the reliability of the performance is of concern as reliability determines the expectations of any single performance of an reinforcement learning agent. Experiments in these environments demonstrate common behaviors of reinforcement learning algorithms showing possible strengths and weakness of the approaches when applied to more complex decision-making scenarios. We explore these behaviors through evaluation techniques meant to measure the agents accumulation of reward in the game. Cross comparing these evaluation techniques elucidates the causation behind agent performance.

Effects of Selection Bias on Online Adversarial Aware SVM When Facing an Evasion Attack

Victor Barboza Morais, Pradeep Chowriappa
Program of Computer Science, Louisiana Tech University, Ruston, Louisiana, USA

Abstract - In evasion attacks, the success of machine learning (ML) depends on its ability to detect an attack in an adversarial setting. It is important that these ML models are trained regularly to keep abreast with evolving evasion attacks. We present an investigation of the effects of selection bias and class imbalance in training an Adversary-Aware Online SVM (AAOSVM). We show that incorrect samples can compromise the model's ability to detect attacks. Comparison results showed better True Positive Rates (TPR) performance than the Online Support Vector Machine (OSVM), and that AAOSVM was sensitive to selection bias but not to class imbalance.

Implementation Roadmap for Neural Networks in Array Databases

Otoniel Jose Campos Escobar, Peter Baumann
Jacobs University Bremen, Bremen, Germany

Abstract - Relational databases lack behind when handling array data and thus array databases were created to fill this gap. Array databases provide optimized storage, retrieval, and processing of multidimensional discrete data (MDD), also known as array data. Just like relational array databases, data processing in array databases is handled declaratively through an array query language that offers enough expressible power to define a myriad of operations. However, despite the advancements in array database technology, there is still a gap for describing machine learning (ML) algorithms and in particular neural networks which, in recent years, have been adopted for predicting phenomena in science and engineering. In this contribution we outline an implementation roadmap for defining neural networks in an array database. We identify the necessary linear algebra operators present in a feed forward neural network and use them to define the training and prediction operations of that network. We also cover other operators that, though they are not part of linear algebra, also need to be defined for a complete machine learning implementation.

Reinforcement Learning Agent Design and Optimization with Bandwidth Allocation Model

Rafael F. Reale, Joberto S. B. Martins
Program in Systems and Product Engineering - PPGESP, Federal Institute of Bahia - IFBA, Valenca, Brazil;
Computer Science Graduate Program - PPGCOMP, Salvador University - UNIFACS, Salvador, Brazil

Abstract - Reinforcement learning (RL) is currently used in various real-life applications. RL-based solutions have the potential to generically address problems, including the ones that are difficult to solve with heuristics and meta-heuristics and, in addition, the set of problems and issues where some intelligent or cognitive approach is required. However, reinforcement learning agents require a not straightforward design and have important design issues. RL agent design issues include the target problem modeling, state-space explosion, the training process, and agent efficiency. Research currently addresses these issues aiming to foster RL dissemination. A BAM model, in summary, allocates and shares resources with users. There are three basic BAM models and several hybrids that differ in how they allocate and share resources among users. This paper addresses the issue of an RL agent design and efficiency. The RL agent's objective is to allocate and share resources among users. The paper investigates how a BAM model can contribute to the RL agent design and efficiency. The AllocTC-Sharing (ATCS) model is analytically described and simulated to evaluate how it mimics the RL agent operation and how the ATCS can offload computational tasks from the RL agent. The essential argument researched is whether algorithms integrated with the RL agent design and operation have the potential to facilitate agent design and optimize its execution. The ATCS analytical model and simulation presented demonstrate that a BAM model offloads agent tasks and assists the agent's design and optimization.

Software-Based Mass Customization of Artificial Neural Networks and its Benefits

Dominic Rosch-Grace, Jeremy Straub
Department of Computer Science, North Dakota State University, Fargo, North Dakota, USA

Abstract - Artificial intelligence (AI) is a rapidly growing field of research that aims to create decision-making computing systems Machine learning (ML) is a subclass of AI that harnesses vast quantities of data to train computational models. ML agents can identify patterns in large datasets and render predictions based on the data they were trained with. Deep learning (DL) is a subclass of ML that develops artificial neural networks with multiple hidden layers of artificial neurons to solve data-intensive problems. ML/DL research requires specialized knowledge and systems, keeping many organizations from using it or achieving the full benefits it could provide. In this paper, a user-friendly approach to building AI/ML/DL models is presented which allows users to build models using diagramming software, thus, abstracting much of the complexity otherwise associated with building AI/ML/DL models, while also achieving benefits potentially only available via customization. The concept allows users to customize models at the neuron level, allowing them to specify attributes such as individual neurons. activation functions and connectivity. This work aims to broaden the availability of ML/DL technologies and unlock the benefits of low-level customization for users of all sizes.

Investigating Gender and Racial Bias in ELECTRA

Maryam Taeb, Yonathan Torres, Hongmei Chi, Shonda Bernadin
Department of Electrical & Computer Engineering, Florida A&M University, Tallahassee, Florida, USA;
Department of Computer and Information Sciences, Florida A&M University, Tallahassee, Florida, USA

Abstract - With the increased adaptation of natural language processing models in industrial applications such as hiring and recruitment, chatbots, social media monitoring, and targeted advertising, pre-trained language models (PTM) need fair and equal behavior across all ranges of demographic groups. ELECTRA has substantially outperformed BERT by predicting the original identities of the corrupted tokens over all input tokens rather than just the small subset that was masked out. Considering such enhancement and the ¼ less amount of computing required by ELECTRA, it can be one of the most suitable industrial applications. Therefore, it is crucial to understand its underlying architecture and tokenization protocol to identify any potential discrimination towards specific groups. This paper presents a fair operation from ELECTRAs' pre-trained network that shows the accurate classification of token replacements. This result is achieved via using a dataset with racially and gender-associated personal names, finetuning ELECTRA with the general language understanding evaluation (GLUE) benchmark, which analyzes the interactions of encoders and decoders using the Contextualized Embedding Association Test (CEAT) and sentiment association test. In addition, this paper will demonstrate that ELECTRA can achieve Biasaware Fair prediction with higher accuracy on downstream tasks after fully trained. This project is investigating the prediction of generator and discriminator on an initial word's token using the Named Entity Recognition (NER), and Part of Speech tagging (POS).

Evaluating the Performance Acceleration of Generalized Linear Solver using Normal Equation on Three Architectures for Tall Skinny Datasets

Tran Van Sang, Rie Shigetomi Yamaguchi, Ryosuke Kobayashi, Toshiyuki Nakata
The University of Tokyo, Tokyo, Japan

Abstract - In previous work, we effectively applied a Normal Equation method to solve the most draining task in the Generalized Linear Model training process on a tall-skinny real-world dataset. This paper generalizes this method by applying it to synthetic data in various sizes. Besides, we evaluated the method on a wider column of data to evaluate the scalability. In addition, we measured and made a comparison of the execution on three different architectures: Vector Machine, an up-to-date GPGPU, and x86 CPU, along with various compilers and BLAS implementations.

Comparisons of Machine Learning Methods for Human Activity Recognition Using Pseudo-Free-Living Data

Enas E. Alkhoshi, Khaled M. Rasheed, Hamid R. Arabnia, Frederick W. Maier, Jennifer L. Gay
School of Computing, University of Georgia, USA;
Shaqra University, Shaqra, Saudi Arabia;
Institute for Artificial Intelligence, University of Georgia, USA;
Department of Health Promotion & Behavior, University of Georgia, USA

Abstract - Recognition of human activities has become more critical in recent years. Practical approaches for classifying physical activities are becoming required in the scientific community, and such models are needed explicitly in health promotion and behavior. The data collected in real-world conditions differ from laboratory data, which has been the main focus of much research. As a result, in this work, we analyzed several machine-learning techniques and models for identifying human activity using a pseudo-free-living dataset obtained at the University of Georgia. We found that hierarchical metaclassifiers outperformed deep learning and classical models by 6% for classifying seven activities. Model personalization is promoted since it lowers the inter-subject variability of the dataset. We divided activities based on the Metabolic Equivalent of Task (MET), and we achieved 80% inter-subject accuracy and 87% accuracy by including 50% of the participant's data. Achieving high performance for machine learning models is challenging using real-world data.

Autonomous Sprinkler System with MAPE-K

Mira Kim, Jennifer Jin
IBM Automation, International Business Machines Corporation Costa Mesa (IBM), California, USA;
Computer Science and Engineering, California State University, San Bernandino, California, USA

Abstract - Sprinkler is a key component for farms and gardens. However, conventional sprinkler systems have limitations: underwatering and overwatering under evolving weather, not considering plant conditions, and burden of manual scheduling. We developed Autonomous Sprinkler System to remedy the limitations. We acquire environmental contexts with IoT sensors, cameras, and actuators, and apply MAPE-K for autonomic control. We implemented the Knowledge with a SVM classifier for inferring the situations and a Fast R-CNN for determining plant condition. Our experiments show the system yields a considerable savings of water consumption and promotes the plant health, all without human users. interventions.

A Framework of Soft Sensor Systems with Machine Learning

Woo Young Moon, Soo Dong Kim
Degital Platform Group, SK Inc., Seongnam-si, Gyeonggi-do, Korea;
School of Software, Soongsil University, Seoul, Korea

Abstract - A Soft-sensor is a means or a method to predict response variables that are difficult to predict by using the data of variables that can be easily obtained. There have been increasing technical demands on improving the accuracy of soft sensors and reducing development complexity since the soft sensor is more cost-effective and easier to collect data than hardware sensors. However, few systematical methods to select the optimal set of features for building soft-sensor models have been proposed, although feature selection is the most essential factor to improve the development quality of the soft sensors. Therefore, this thesis is to present a systematic method for generating soft-sensor models to enhance the model accuracy by measuring similarities of soft-sensor models and selecting the best feature set from the similarity analysis. The proposed method utilizes ML technologies to build soft-sensor models and presents an algorithm to build soft-sensor models by reusing existing and similar soft-sensor models to improve the accuracy of the softsensor models.

Service Chatbots and International Students: A Systematic Review

Sonipriya Paul, Lydia Ray
Department of Computer Science, Columbus State University, Columbus, Georgia, USA

Abstract - Mental health has predominantly been a social stigma where suffering individuals are in a constant combat with themselves. Youth and young adult students who would otherwise be the future of a nation are succumbing to depression or anxiety with the rise of socio-economic pressures. International students in the US are no exceptions and they fall prey to psychiatric and psychological issues. In such cases, they need easily accessible resources at their disposal. This paper surveys the existing research that studies various aspects of difficulties faced by international students in the United States and offers ideas of solutions based on computing and artificial intelligence through the effectiveness of chatbot systems.

A Method based on Deep Neural Network for Instance Segmentation of Retinal Lesions caused by Diabetic Retinopathy

Carlos Santos, Marilton Aguiar, Daniel Welfer, Marcelo Silva, Alejandro Pereira,
Marcelo Ribeiro, Bruno Belloni
Computer Center, Federal Institute of Education, Science and Technology Farroupilha, Alegrete, Brazil;
Postgraduate Program in Computing, Federal University of Pelotas, Pelotas, Brazil;
Postgraduate Program in Computer Science, Federal University of Santa Maria, Santa Maria, Brazil;
Federal University of Pelotas, Pelotas, Brazil;
Federal Institute of Education, Science and Technology Sul-Rio-Grandense, Passo Fundo, Brazil

Abstract - Diabetic Retinopathy is one of the main causes of vision loss and can be identified through ophthalmological exams that aim to locate the presence of retinal lesions such as microaneurysms, hemorrhages, soft exudates, and hard exudates. The development of computerized methods to perform the instance segmentation of lesions may support in the early diagnosis of the disease. However, the instance segmentation of retinal artifacts is a complex task due to factors such as the size of objects and their morphological characteristics. This article proposes a method based on a Mask R-CNN neural network architecture to perform instance segmentation of lesions associated with diabetic retinopathy. The proposed method was trained, adjusted, and tested using the public DDR and IDRiD Diabetic Retinopathy datasets, and implemented with the Detectron2 and OpenCV libraries. The proposed method reached in the DDR dataset, using the SGD optimizer, the mAP of 0.2660 for the limit of IoU of 0.5 in the validation step. The results obtained in the experiments demonstrate that the proposed method showed promising results in the instance segmentation of fundus lesions.

Enabling a Network AI Gym for Autonomous Cyber Agents

Li Li, Jean-Pierre S. El Rami, Adrian Taylor, James Hailing Rao, Thomas Kunz
Defence Research & Development Canada, Ottawa, Canada;
Department of Electrical and Computer Engineering, Queen's University, Kingston, Canada;
Department of Systems & Computer Engineering, Carleton University, Ottawa, Canada

Abstract - This work aims to enable autonomous agents for network cyber operations (CyOps) by applying reinforcement and deep reinforcement learning (RL/DRL). The required RL training environment is particularly challenging, as it must balance the need for high fidelity, best achieved through real network emulation, with the need for running large numbers of training episodes, best achieved using simulation. A unified training environment, namely the Cyber Gym for Intelligent Learning (CyGIL), is developed where an emulated CyGIL-E automatically generates a simulated CyGIL-S. From preliminary experimental results, CyGIL-S can train agents in minutes compared with the days required in CyGIL-E. The agents trained in CyGIL-S are transferrable directly to CyGIL-E, showing full decision proficiency in the emulated "real" network. Enabling offline RL, the CyGIL solution presents a promising direction towards simto-real in leveraging RL agents in real-world cyber networks.

An Evaluation of Machine Learning and Deep Learning Approach on Ekman Sentiment Classification

Abdou Aziz Ledoux Ngou Njikam, Elie Fute Tagne, Lionel Landry Sop Deffo
Nanjing University of Science and Technology, Nanjing, P. R. China;
University of Buea, Cameroon;
University of Dschang, Dschng, Cameroon; Telecom Paris, France

Abstract - The evolution of natural language processing during the last years has been phenomenal. Its applications are infinite and society as we know can no longer do without it. We have reached the point where works in natural language processing have been able to successfully collide with those in psychology, the outcomes of this are extraordinary. The one that focuses our interest is the basic emotion elaborated by the psychologist Ekman who classifies emotions into six basic states namely anger, disgust, fear, joy, sadness, and surprise. This study focuses on the evaluation of machine learning and deep learning model on the Ekman.s emotion classification. It first investigates four machine learning algorithms: random forest, support vector machine, naive Bayes, and logistic regression. Then two deep learning models: LSTM (Long Short-Term Memory networks) and BERT (bi-directional encoder representation). The study finishes with a comparative study of the different approaches. The dataset used is the Emotion Lines dataset.

Component-Wise Natural Gradient Descent on Deep Neural Networks

Tran Van Sang, Rie Shigetomi Yamaguchi, Mhd Irvan, Toshiyuki Nakata
University of Tokyo, Tokyo, Japan

Abstract - Natural Gradient Descent (NGD) is a well-known second-order network training method preconditioning the gradient descent with the inversion of the Fisher Information Matrix (FIM). Previously, we introduced an NGD-variant named ComponentWise NGD (CW-NGD) which approximates FIM as a block diagonal matrix. This paper extends CW-NGD to networks with Batch Normalization layers. For the first time, we evaluate it on a deep neural network. Additionally, to speed up the training, we formularize the update term in CW-NGD as the solution of a Linear Least Squares (LLS) equation. Correspondingly, we apply Cholesky LLS solver and reduce the training time by 2.50 times.

Various Patient Classification Model using ABR Data

Jun Ma, Seong Jun Choi, Jaehyon Kim, Min Hong
Department of Software Convergence, Soonchunhyang University, Asan, Republic of Korea;
Department of Otorhinolaryngology, Head and Neck Surgery, College of Medicine, Soonchunhyang University Cheonan Hospital, Cheonan, Republic of Korea;
Department Computer Education, Sungkyunkwan University, Seoul, Republic of Korea

Abstract - Auditory brain response (ABR) is an auditory test developed in the 1970s that enables objective testing regardless of age. However, ABR has the disadvantage that analysis and reading are essential because it is complicated to inspect and requires trained inspectors. In this paper, we present a deep learning Keras sequential model and a VGG16 model that distinguish between deaf patients and normal people using ABR data and compare the performance and accuracy of the model. It is expected that a deep learning model that automatically distinguishes deaf patients from normal people will be applied in the future using the proposed model.

Minimizing Computational Overhead While Retaining Gameplay Effectiveness in Starcraft Bots

David Kipnis, Wesley Deneke, Matthew Mitchell, Nils Rys-Recker
Computer Science Department, Western Washington University, Bellingham, Washington, USA

Abstract - Over the past decade, RTS games like Starcraft have emerged as testbeds for developing complex artificial intelligence. While several bots have been successful in their objective of beating opponents, no research has sought to minimize their computational resource usage. Reducing a models' computational complexity has the potential to expand the range of applications for such AI, which is the primary objective that VikingBot was designed to explore. By combining a heterogeneous agent model with a machine learning approach, VikingBot demonstrates that a Starcraft bot can minimize resource usage while retaining good gameplay effectiveness.

Predicting the Impact of Wildfire using Machine Learning Techniques to Assist Effective Deployment of Resources by the Fire and Forest Department

Nikesh Kumar Pahuja, Maria H. Rivero
Information Systems Engineering and Management, Harrisburg University of Science and Technology, Harrisburg, PA, USA

Abstract - Wildfire incidents occur every year due to humancaused factors, droughts, vegetation, and climate change. It is complex to predict severe wildfires and even more difficult to extinct them. Structural damage may occur during wildfires incidents. Therefore, understanding the intensity of these damages could help mitigate the risk for the region and local communities. This study builds a predictive model using various Machine Learning techniques as a real-time decision-support tool. The model predicts the severity of damage to the structures affected during the wildfires in the state of California, United States. The study used The California Department of Forestry and Fire Protection (CAL FIRE)'s historical data from 2013 until 2019. The study trained several supervised machine learning algorithms for classification, such as Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), Naïve Bayes (NB), and XGBoost (XGB), to predict the structural damage. Predictions were classified in the following categories of expected Fire damage: Inaccessible, Destroyed (50%), Major (26-50%), Minor (10-25%), Affected (1-9%), and No Damage. The best performing algorithm was the Decision Tree model. This study will allow fire and forest departments to forecast the expected fire damage when the wildfire is first reported. It will also enable the departments to prioritize actions and deploy their resources effectively to reduce the impact of the wildfires in the region.

Winter Wheat Crop Yield Prediction on Multiple Heterogeneous Datasets using Machine Learning

Yogesh Bansal, David Lillis, Tahar Kechadi
School of Computer Science, University College Dublin, Dublin, Ireland

Abstract - Winter wheat is one of the most important crops in the United Kingdom, and crop yield prediction is essential for the nation's food security. Several studies have employed Machine Learning (ML) techniques to predict crop yield on a county or farm-based level. The main objective of this study is to predict winter wheat crop yield using ML models on multiple heterogeneous datasets, i.e., soil and weather on a zonebased level. Experimental results demonstrated their impact when used alone and in combination. In addition, we employ numerous ML algorithms to emphasize the significance of data quality in any machine-learning strategy.

Epileptic Seizure Electroencephalography Signal Classification Using Machine Learning Algorithm

Zheng You Lim, Yong Li Neo
Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia;
Malacca Expartriate School, Melaka, Malaysia

Abstract - Epileptic seizure is a neurological disorder that will cause sudden disturbance in the brain. This will lead to the involuntary movement of the body, paralyze or even risk the life of the patient. This neurological disorder is commonly diagnosed based on electroencephalography (EEG) signals. As EEG signal is noisy and complicated, it is very time-consuming for the doctor to diagnose the patients. As technology advances, machine learning is now a common technology involved in our daily lives. Hence, we propose to conduct a series of experiments to find out which machine learning model is the best supervised learning model that can classify the EEG signal of a healthy person and epileptic patient. The results show that the Extra Tree classifier is the best machine learning model as compared to the other 8 machine learning models employed in this research. It achieves 99.2% in terms of AUC, 96.6% accuracy, and 95.3% recall rate in classifying the EEG signal of a healthy person and epileptic patient. The reliability of the Extra Tree classifier shows great potential that it can be applied in the epilepsy diagnosis tool to aid the doctor in diagnosing accurately in a shorter period of time.

Performance Analysis of a Micromodel-based Multinomial Classifier

Jawahar Y.V., Sanjay Nithin S, Varalakshmi M , Peer Mohideen
Vellore Institute of Technology (VIT), India

Abstract - Supervised learning methods demand large scores of labelled training data to achieve high predictive accuracy for classification problems. In contrast, the less accurate unsupervised learning models can work on raw data but fail to provide any useful insight. Micromodel-based multinomial classifier is a one-stop solution for increased performance with less data and less time. This paper investigates the effectiveness of micromodels for multiclass classification as against the existing supervised and unsupervised machine learning models and artificial neural networks. The model built using One-class Neural Network (OC-NN) to classify students' resumes into four different streams for admission into postgraduate programmes in top-ranked universities, outperforms some of the most powerful machine learning models.

Predicting Time Series Values with LSTM: Different Scenarios for Prediction

Alberto Arteta, Kristopher Allbright, Mehmet Sahinoglu, Luis Fernando de MIngo
Computer Science, Troy University, Troy, Alabama, USA;
Computer Science, Troy University, Montgomery, Alabama, USA;
Computer Science, UP, Madrid, Spain

Abstract - Stock price prediction is a key area of research within financial circles. With stock prices being notoriously volatile, predicting them has always been a subject of study in technical analysis and a known challenge. Even financial analysts struggle to identify and predict winning stocks. This is where pattern recognition modeling comes to play a major role. As trending in the stock market is widely targeted by Machine Learning, this paper shows a comparative study of different known approaches and introduces a new one by combining some of the known ones. Our hypothesis is tested using various forms of traditional Machine Learning algorithms as well as an LSTM (long short-term memory) deep learning model.

Autonomous Voice Recognition Wheelchair Control System

Mokhles M. Abdulghani, Wilbur L. Walters, Khalid H. Abed
Department of Electrical & Computer Engineering and Computer Science, Jackson State University (JSU), Jackson, Mississippi, USA

Abstract - Autonomous wheelchairs are essential for improving the mobility of individuals with disabilities or physical challenges. Advances in computer and wireless communication technologies have paved the way for the design of smart wheelchairs to match the needs of disabled people. This research paper introduces the design and implementation of an electric wheelchair controller using voice recognition. This model is based on a voice recognition algorithm for classifying the essential commands to drive the wheelchair. A pre-designed adaptive neuro-fuzzy controller has been used to generate the required control signals for activating the motors of the wheelchair. This controller alters data gathered from obstacle avoidance sensors and a voice recognition classifier. Deep learning algorithms have made many technological developments and influences modern daily lives. After testing our developed deep learning algorithm, the overall classification accuracy for distinguishing between eight voice commands was 98 %.

Hyperparameters and Features Impacts in Artificial Intelligence Applied to Stock Market

Victor Martins de Borba
Department of Data Analytics, Harrisburg University of Science and Technology, Harrisburg, USA

Abstract - Over the continuous utilization in real-world applications, Artificial Intelligence (AI) proved to be state-of-the-art technology, delivering more benefits at little incremental effort or cost to its utilization. This study proposes to apply a specific Machine Learning algorithm to a classic Financial Services application, stock price prediction, measuring its outcome according to different feature selections and hyperparameters tuning. Important to highlight those traditional methods (e.g., Fama-French), although primarily accepted and utilized in the market, are not fail-proof, and all have practical limitations. Using a Deep-Learning/Recurrent Neural Network algorithm, this Artificial Intelligence proof of concept reinforced the efficiency and applicability of the AI methodology. Moreover, the tested models were developed, allowing scalability at no cost and complexity for future uses. Finally, this paper discusses the impact of selected features and hyperparameters tunning in the final result and the efficiency of the tested AI models.

Application of Multi-Task Learning for Abnormal Diagnosis and Trip Variable Prediction in Nuclear Power Plants

Hye Seon Jo, Ji Hun Park, Man Gyun Na
Department of Nuclear Engineering, Chosun University, Dong-gu, Gwangju, Republic of Korea

Abstract - A nuclear power plant is a large facility composed of many components, and abnormal states occasionally occur in which components fail. In the event of abnormal states, if appropriate measures are not taken, the abnormal states can worsen and cause an unexpected reactor trip. Therefore, in order to provide the operator with key state information in case of abnormal states, abnormal diagnosis and trip variable prediction were performed based on multi-task learning (MTL). The MTL is a method of performing multiple tasks through a single model. Specifically, the progressive layered extraction method, one of the MTL structures, was applied. It efficiently transmits information between tasks through a gating network and progressive routing mechanism. The proposed model showed higher diagnosis accuracy and lower prediction error than the basic MTL model. If the key state information is provided to the operator through the proposed model, it will be able to contribute to reducing human error and preventing the aggravation of abnormal states.

Remaining Trip Time Prediction Using Light Gradient Boosting Machine in Nuclear Power Plants

Sang Won Oh, Sang Hyun Lee, Man Gyun Na
Department of Nuclear Engineering, Chosun University, Dong-gu, Gwangju, Republic of Korea

Abstract - Accidents can occur in nuclear power plants due to various causes. Operators perform diagnostics and mitigation tasks. In these tasks, operators must respond to rapidly changing situations. This may cause a human error. Therefore, many studies are conducted to support operators using artificial intelligence to reduce human errors. Among them, most of the researches in the prediction field predict the states of nuclear power plants and provide information to operators. This helps the operator to prepare ahead of time. Existing studies predict future trends in the parameters. It cannot predict the condition for more than a certain period of time, and there is a limit to identifying all the causes of the condition deterioration. Therefore, in this study, the remaining trip time prediction of nuclear power plants is performed. It also utilizes explainable artificial intelligence to determine the causes of predictive outcomes. The explainable artificial intelligence model shows the reason why the condition of nuclear power plants is deteriorating. Therefore, it is expected that this study will help operators to prepare emergency response and perform mitigation tasks.

A Drone Flight Control Using Brain Computer Interface and Artificial Intelligence

Mokhles M. Abdulghani, Arthur A. Harden, Khalid H. Abed
Jackson State University, Jackson, Mississippi, USA

Abstract - The human mind is a truly remarkable thing that does so much that we are not even aware of. Controlling machines using the concept of Brain-Computer Interface (BCI) is a practical method that opens the way to a fully synchronized method between human thoughts and controlled objects. Using BCI to control a drone will open the way toward smooth and high-response flight. Deep learning is a new-age skill that has made many breakthroughs and influenced modern technologies. It has made it possible to predict and identify even the most complex and abstract patterns that even we humans would be very challenged to catch ourselves. In this paper, a method of controlling a drone using BCI has been presented using an 8-channel Electroencephalogram (EEG) headset. Deep learning has been employed to process and classify human brain waves. After testing the resulting deep learning algorithm, the overall classification accuracy was 90% to distinguish between four different movements of the drone.

Few-shot Learning on Histopathology Image Classification

Ankit Kumar Titoriya, Maheshwari Prasad Singh
Department of Computer Science & Engineering, National Institute of Technology Patna, Patna, India

Abstract - Cancer detection in histopathology slides is not easy even today. CNN (Convolutional Neural Network) based Object identification and segmentation algorithms work very well. A large dataset of medical images is required for classification which may not be available especially for rare diseases. Therefore, deep learning and machine learning may not be effective for rare disease classification. If CNN architecture is trained on one dataset then it performs well but the same architecture may not achieve good accuracy on other datasets. So, generalization is one of main issues. This paper proposes FSL (Few-Shot Learning) to solve generalization and size of dataset. This paper uses Prototypical networks and MAML (Model Agnostic Meta Learning) simultaneously on four different datasets. Along with this, it has also been checked whether these two networks meet the concept of generalization or not. The paper also finds accuracy of both networks in 2-way, 3-way, and 5-way modes. Simulation results show that MAML achieves accuracy of 84.56% in the 2-way 2-shot 2 query mode. Further, simulation results show that Prototypical Network achieves accuracy of 74.575%, 61.9889% and 45.762% in 2-way 2-shot 2 query mode, 3-way 3-shot 3-query mode and 5- way 5-shot 5-query mode, respectively.

Cardiac Detection using YOLO-v5 with Data Preprocessing

Sungjin Lee, Jia Xibin, Ahyoung Lee, Hyo Wook Gil, Sungyeup Kim, Hong Min
Department of Software Convergence, Soonchunhyang University, Republic of Korea;
Department of Software Convergence, Faculty of Information Technology,
Beijng University of Technology, Beijing, P. R. China;
Department of Computer Science, Kennesaw State University, Georgia, USA;
Department of Internal Medicine, Soonchunhyang University, Cheonan Hospital, Republic of Korea;
Department of Research, All for One, Asan, Republic of Korea;
Department of Software Engineering, Soonchunhyang University, Republic of Korea

Abstract - Currently, interest in medical-related deep learning is dramatically increasing. Although this interest in deep learning is widely used in other fields, but it is very effective in medical image processing such as CT and MRI, which takes a lot of time for simple medical tests and analysis. In general, in deep learning using such image processing, it is possible to determine which algorithm is the most efficient by collecting data, preprocessing data, and using various models This paper conducted a research on cardiovascular CT images collected from Soonchunhyang University Hospital in Korea, and all of them used data collected for 3 years by professional medical staff. In the case of medical data, the number of data is very limited, so the results can vary greatly depending on how it is processed. Therefore, in this paper, research on an efficient deep learning method was conducted through image data preprocessing using Yolo.

What a Drag! Streamlining the UAV Design Process with Design Grammars and Drag Surrogates

Michael Sandborn, Carlos Olea, Anwar Said, Mudassir Shabbir, Peter Volgyesi,
Xenofon Koutsoukos, Jules White
Vanderbilt University, Nashville, Tennessee, USA

Abstract - Unmanned Aerial Vehicles (UAVs) continue to proliferate, revolutionizing tasks such as cargo transport, surveillance, and search and rescue operations. With the discovery of novel use cases or specialized tasks for aerial vehicles, there is an increased need for improved design space exploration and performance estimation techniques for candidate UAV designs. Typical pipelines for this design process rely on time-

consuming human efforts to identify productive design geometries or expensive computational approaches for performance analysis to reconcile aerodynamic, electrical, and physical interactions. In this work-in-progress paper, we propose the use of a design process that uses a design grammar for UAV design generation and a Graph Neural Network (GNN)-based drag surrogate trained on simulation data for accelerated UAV design space exploration. We formulate a UAV design grammar and provide preliminary performance results from the GNN drag surrogate for randomly generated designs. We expect our approach to accelerate the exploration of UAV design geometries using a learned surrogate drag model to circumvent resource-hungry Computer-aided design (CAD) and simulation routines.

Preliminary Results of Applying Transformers to Geoscience and Earth Science Data

Uko Ebreso, Justin Delgado, Yulia Kumar, Jenny Li, Patricia A Morreale
Kean University, Union, New Jersey, USA

Abstract - The transformers with Neural networks (NN)s become a dominant technology in AI/ML to achieve better accuracy in classification and prediction. We experimentally compare transformer based NN and convolutional NN's using various applications such as pollen detection and weather ITCZ prediction. Using Machine learning in Geoscience, Earth Sciences, and other natural sciences is not an entirely new concept but applying transformers to data from these environments creates many new opportunities. We apply Facebook's DETR transformer neural network, developed in 2020, to pollen and weather data to detect forty-four types of pollen and classify earth snapshots into two categories: having vs not having the phenomena of double Inter-Tropical Convergence Zones (ITCZs) in them. As we conduct our trials, we not only observe and document how the model performs on both datasets, but detect the biases present in each layer of the network and mitigate them as we go to tune the model and improve classification results even further.

Challenges of Self-Supervised Learning for Unified, Multi-Modal, Multi-Task Transformer Models

Graham Annett, Tim Andersen, Robert Annett
Computer Science, Boise State University, Boise, Idaho, USA;
Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA

Abstract - The recent success of multi-modal multi-task transformer models combined with their ability to learn in a scalable self-supervised fashion has presented evidence that omnipotent models trained with heterogeneous data and tasks are within the realms of possibility. This paper presents several research questions and impediments related towards the training of generalized transformer architectures.

To Shuffle or Not to Shuffle: Mini-Batch Shuffling Strategies for Multi-class Imbalanced Data Classification

Yuwei Mao, Vishu Gupta, Kewei Wang, Wei-Keng Liao, Alok Choudhary, Ankit Agrawal
Department of Electrical and Computer Engineering, Northwestern University, Evanston, Illinois, USA

Abstract - Mini-batch shuffling is important for the deep learning training process. Most people use the random shuffling method, which aims to produce a random permutation of the training dataset in every epoch. In this study, we explore minibatch shuffling for multi-class imbalanced data classification by investigating several shuffling strategies. We find that different order of input data can significantly affect the results of deep learning models. The results show that our proposed strategies can improve the accuracy by around 2%, demonstrating that higher diversity and lower imbalance ratio in each mini-batch can lead to better results.

Proposal of a Method for Inferring the Road Surface from Vibration Using Deep Learning on CanSat

Takuya Saito, Miho Akiyama
Faculty of Informatics, Tokyo University of Information Sciences, Chiba, Japan;
Graduate School of Electrical and Information Engineering, Shonan Institute of Technology, Fujisawa, Japan

Abstract - This paper describes a method for accurately classifying whether a CanSat is running on a hard surface or a soft, uneven surface based on acceleration data from vibration during running, using FFT and Deep Learning Image Classification. The CanSat cannot run smoothly on a hard surface, as in the case of ARLISS, and on a soft and uneven surface, as in the case of the Noshiro space event, unless the running speed and steering wheel angle settings are changed. Therefore, we considered the possibility of inferring whether the road surface is hard or soft from the vibration during running. This paper describes a method for converting acceleration data from running data into frequency-domain images by FFT analysis, and for classifying these images with high accuracy by using deep learning image classification. We found that this method can achieve a high classification accuracy of 100% for running data used for training, and more than 92% for other running data that was not used for training.

Machine Learning Algorithms for Natural Language Processing Tasks: A Case of COVID-19 Twitter data (Thailand)

Kunyanuth Kularbphettong, Rujijan Vichivanives, Pannawat Kanjanaprakarn, Kanyarat Bussaban, Jaruwan Chutrtong, Nareenart Ruksuntorn
Suan Sunandha Rajabhat University, Bangkok, Thailand

Abstract - This paper presents the use of natural language processing for the problem of information extraction and sentiment analysis. The dataset is from Twitter that has the information of people mentioning about COVID-19, this study has two tasks: (i) classification approach for information extraction task and (ii) deep learning approach for sentiment analysis task. In information extraction task, the data was gathered from twitter that related to COVID-19 information, and the sequence labelling method applied to classify text before giving it to classification algorithms (K-NN, Naïve Bayes, Decision Tree, Random Forest, and SVM). In sentiment analysis task, data was classified by onvert the word into index and using word embedding, then to process deep learning algorithm (Bi-directional GRU). The accuracy of two tasks are 98% and 79% respectively.

Qualitative Review of Art Generating AI Applications and their Relationship to NFTs

Emine Arikan, Siamak Aram
Harrisburg University of Science and Technology, USA

Abstract - The evolution of technology and science has brought groundbreaking developments to the current visual arts, and the application of digital technology has brought significant changes to the creation and aesthetic taste of traditional art. This review therefore investigates the current state-ofthe-art in artificial intelligence (AI) technologies and applications in generating visual art-while giving a brief history of the intersection of AI and art including the milestone advancements in neural networks. Fifteen interviews were conducted with technical artists who use art generating AI applications to gather insight into real life experiences. Based on the findings from these interviews, the state-of-the-art applications are reviewed and analyzed in six categories: Accessibility (cost), Barrier to entry (out of STEM), Novelty, Ethics and Morality, Control, non-fungible tokens (NFT) and Monetization which are widely discussed along with their success and limitations. The research concludes with

three main findings; (a) monetization of digital media through NFTs has a direct impact on the advancement of art generating AI applications, (b) there is a significant change in the traditional creative process with the integration of AI applications, AI is not just a tool but it's a creative agent that artists collaborate with (c) art generating AI applications can generate limitless possibilities within the same aesthetics as a result revolutionize the way humans create and interact with art.

A Study of Model Based and Model Free Offline Reinforcement Learning

Indu Shukla, Haley Dozier, Althea C. Henslee
Information Technology Laboratory, Engineer Research and Development Center (ERDC), U.S. Army Engineer Research and Development Center, Vicksburg, Mississippi, USA

Abstract - Reinforcement learning (RL) has received considerable attention for building autonomous systems with trained agents that interact with environments to learn optimal behavior. The RL requires a fundamentally online learning paradigm, one of the biggest obstacles for the widespread adoption of RL in scaling to many real-world scenarios. In this study we have explored the method of data collection and have applied offline model-free and model-based RL methods on classical OpenAIGym Cart-Pole environment. In this approach offline algorithms find a good policy from previously collected dataset.

Machine Learning Opportunities for Improving Logistics within the Army

Joseph E. Jabour, Arnold L. Thomas, Indu Shukla, Brandon E. Hansen, Haley R. Dozier, Ian D. Detwiller, Althea C. Henslee
US Army Engineer Research and Development Center, Vicksburg, Mississippi, USA

Abstract - This poster provides a brief insight to the workings of logistics within the Army. While the Army is a massive network of various different branches, each with their own uses of logistics efforts, this work effort aims to provide the concept, samples of research, and current use cases of machine learning to reinforce the strategy and efficiency of logistics work in the Army.

Designing of Prompts for Hate Speech Recognition with In-Context Learning

Lawrence Han, Hao Tang
Ridge High School, Basking Ridge, New Jersey, USA;
Computer Information Systems, The City University of New York, New York City, New York, USA

Abstract - In-context learning is a recent paradigm in natural language understanding, where a pretrained large language model (LLM) directly performs a new task without any update to its parameters by taking a test instance, new task description and a few training examples (e.g. input-label pairs) as its input. However, performance has been shown to strongly depend on the task description and selected training examples (both together termed as prompts here). In this paper, we use GPT-3 as the LLM, hate speech recognition as the new task, and we investigate how to design effective prompts for better performance. Our preliminary experimental results show that: (1) substantial number of input-labels pairs are necessary for good performance (2) informative task descriptions can further boost performance by ingesting our prior knowledge as inference guidance.

Artificial Intelligence use in Congestive Heart Failure Mid-Term Findings

Atif Farid Mohammad, Rajan Vanjani, Jagadisan Suresh
Apexon, Plano, Texas, USA

Abstract - The Heart Rate Variability represents the flexibility of the heart to the outer and inner stimuli. Decreased HRV may be a predictor of adverse cardiovascular outcomes. In view of the nonlinear, non-stationary, and extremely complex elements of the cardiovascular framework control instrument, the ability to analyze the hidden elements accurately has been limited by direct HRV steps. In this review, by separating multimodal highlights, a computerized system for breaking down HRV flags. To test the position execution, powerful AI procedures are used, for example, Support vector machine (SVM), Decision tree (DT), k-nearest neighbor (KNN), and Random Forest classifiers. The impressive presentation of the technique of the random forest method shows that it plays an enormous role in the identification of congestive heart failure (CHF) and can be essential in the transmission of knowledge useful for medication. In this paper, we review the results of the mid-term research about Congestive Heart Failure (CHF) using different approaches to get an understanding of the problem and then address the models of machine learning that can be applied to operate on the available data on this subject. Some results are obtained and established in this paper based on the research carried out on the data given by platforms regarding these diseases.

Intelligent Processing of Judicial Documents Based on Deep Learning

Jingzi Wang, Chao-Lin Liu
Department of Computer Science, National Chengchi University, Taipei, Taiwan

Abstract - Legal judgement prediction is a critical application in the law and AI field, and has received great attention in recent years. In this paper, we focus on the recommendations of applicable law articles for a given criminal case. This is a multilabel problem in nature, since it is common for a sequence of criminal activities to violate the law in many aspects. We compare two different strategies for identifying the articles. The first one employs only the Lawformer, which is a Longformer-based model designed specifically for the legal domain, and the second one will combine the applications of the Lawformer with a traditional technique for keyword selection. The experimental results indicate that integrating the deep learning and the traditional methods provide a better result.

Artificial Intelligence Role in Chronic Condition Decision Extrapolation

Atif Farid Mohammad, Mike Boese, Amruta Gandhe, Ishdeep Kaur, Ramyaa Ganesh
Data Science & AI/ML, Apexon Inc., Texas, USA

Abstract - Chronic Conditions are long lasting diseases that usually last a year or more and require continuous medical treatment and attention throughout. To prevent them from rapidly increasing and being inherited, it is essential to control and manage them with support systems and programs. The Chronic Condition Decision Support System is intended to train nurses and physicians to understand chronic conditions and make decisions for the patient.s wellbeing. It has been predicted that the chronic condition management market is going to boom in the coming years due to the rapidly increasing chronic cases among population as well as the advancement of medical technologies. To build an effective chronic management solution it is also important to understand the process, its capabilities and performance. Currently, the chronic management solution is widely being used all around the globe and it is going to be a major player in the upcoming time.

CSCI-RTCI:
RESEARCH TRACK ON COMPUTATIONAL INTELLIGENCE

Stacked Differential Evolution for Data Envelopment Analysis

Pavel Kromer, Maryam Olfati
Department of Computer Science, VSB-Technical University of Ostrava, Ostrava, Czech Republic

Abstract - Data envelopment analysis (DEA) is a popular method for the complex evaluation of the efficiency of entities that perform similar operations with different ratios of inputs and outputs. In this work, we use a modern version of the differential evolution algorithm, j2020, to analyze DEA instances. It is used in the context of a novel DEA solution strategy in which the high-level task of DEA instance analysis is transformed from a set of low-dimensional optimization problems into a lower number of higher-dimensional ones. It allows the evaluation of the efficiency of multiple DMUs in a single run, decreases the number of individual optimization problems that need to be solved to assess complete DEA instances, and as a result provides an opportunity for the use of optimization algorithms that can efficiently operate in high-dimensional search spaces.

Spoken Digits Classification based on Spiking Neural Networks with Memristor-based STDP

Danila Vlasov, Yury Davydov, Alexey Serenko, Roman Rybka, Alexander Sboev
Complex of NBICS Technologies, NRC "Kurchatov Institute", Moscow, Russian Federation;
Department of Automation Technology, National Taipei University of Technology, Taipei, Taiwan

Abstract - Spiking neural networks are commonly attributed to the third generation of neural networks. They mimic biological neurons more closely by processing information in the form of impulses (spikes) and are characterized by low power consumption and ease of hardware implementation. This paper shows two approaches to the task of classifying audio data represented by the spoken digits dataset using spiking neural networks with memristive plasticity. It is shown that both supervised and unsupervised learning methods based on local plasticity can be successfully used for audio classification. The networks achieve accuracies ranging from 80% to 94% depending on the network topology, plasticity type and the way of decoding output neuronal activity. The results obtained in the paper can be a step towards creating devices capable of recognizing audio signals on the fly.

Sampling Strategies for Exploratory Landscape Analysis of Bi-Objective Problems

Pavel Kromer, Vojtech Uher, Andrejaana Andova, Tea Tusar, Bogdan Filipic
Department of Computer Science, VSB - Technical University of Ostrava, Czech Republic;
Department of Intelligent Systems, Jozef Stefan Institute and
Jozef Stefan International Postgraduate School, Ljubljana, Slovenia

Abstract - Exploratory landscape analysis (ELA) is a popular method for the understanding of complex, often black-box optimization problems. It tries to approximate and describe the surfaces formed by the fitness and other characteristic values associated with problem solutions on top of the multidimensional solution spaces. Sampling is the initial step of the ELA pipeline. It is a strategy for selecting a limited number of solutions, i.e., points in the multi-dimensional solution space, for which the fitness function(s) are evaluated. Consequently, the fitness landscape is approximated and its properties are drawn from these fitness values. In this work, the properties and

the impact of various sampling strategies on the analysis of the fitness landscape are studied in the context of bi-objective optimization. Extensive computational experiments show that the use of different sampling strategies affects both the value of high-level landscape features and their usability for problem classification. The results also demonstrate that the magnitude and significance of the impact depend on problem dimension and sample size.

A Novel Hybrid Multi-resolution Image Registration Algorithm with Curvelet Transform and Artificial Neural Networks

Hyunjong Choi, Xiao-Hua Yu
Department of Electrical Engineering, California Polytechnic State University, San Luis Obispo, California, USA

Abstract - The purpose of image registration is to transform multiple images of the same subject taken from different points of view, times, depths, or sensors into one coordinate system. In this study, a novel hybrid image registration approach is developed based on curvet transform, discrete cosine transform, and artificial neural networks. Curvelet transform is an emerging multi-resolution analysis method that can effectively represent objects with highly anisotropic elements such as lines and curves. The proposed algorithm combines the orientation selective property of curvelet transform and the "energy compaction" property of discrete cosine transform together to provide a more efficient way to extract features of curvilinear structures from images. Besides, the learning ability and nonlinear mapping ability of artificial neural network provide a flexible and intelligent tool for data fusion on feature matching and parameter estimation. The performances of the proposed approaches are studied and compared with other methods on medical magnetic resonance images (MRI) via computer simulations.

Is There a Limit to the Utility of Analogy? A Case Study in Knowledge Graph Triple Representation

Alexander Kalinowski, Yuan An
Department of Information Science, Drexel University, Philadelphia, USA

Abstract - Knowledge graph embedding methods are known to be highly dependent on the locally closed world assumption (LCWA). This assumption has practicality in training neural networks for the link prediction task, but is ill-posed for representing knowledge graph triples as first class objects. In this paper, we explore an alternate sampling paradigm, namely pairwise triple similarity scoring (PTSS), and detail the impact of the sampling parameter on downstream predicate prediction tasks. We specifically seek to find the limit at which more negative samples do not provide a statistically significant lift in performance, giving insights into how to efficiently sample and train such models. Our main finding indicates that there is a point of diminishing return on the number of analogous, pairwise samples selected; this point can be found prior to experimental hyperparameter sweeps. Additionally, our experiments show that there are two classes of models: those that benefit from additional sampling, and those that are less impacted. The root cause for these differences is driven by changes in the distributions of similarity scores, depending on the seed knowledge graph embeddings selected. Our work demonstrates the importance of selecting the correct seed embedding method, largely dependent on the topology of the underlying knowledge graph.

Accuracy Analysis of Supervised and Unsupervised Techniques on Breast Cancer Datasets

Elham Jahanpeikar, Mahdi Ebrahimi
Department of Computer Science, California State University, Northridge, Los Angeles, California, USA

Abstract - Modern medicine benefits extensively from technology as a valuable assistant to diagnose diseases at early stages and cure them with less complication and more success. In the health sector, cancer diseases undeniably, possess the most variety and complexity of all diseases. Therefore, this paper focuses on breast cancer and applies supervised, unsupervised, and deep learning as one candidate from each major machine learning field to evaluate different models and analyze their outcomes. We aim to answer the following questions: Which dataset produces the best result and improves its metrics and performance using machine learning optimization techniques? Will there be any definite decision on which machine learning models or algorithms have superiority over the others? The original and diagnostic datasets are trained by Logistic Regression, K-means clustering, and multi-layer perceptron or artificial neural network model (ANN). Then, unsupervised techniques, heatmaps, and Principal Component Analysis (PCA) are used to reduce dimensionality and concise the dataset for any probable improvements. The original dataset produced better results for the machine learning models, and ANN obtained the best accuracy score. The comprehensive and systematic calculation of the metrics and indexes of the breast cancer datasets and the thorough optimization by unsupervised technics is the novelty of this research. The comparison between these two datasets has not been approached before. The clustering by K-means creates novel visualization of the datasets, which could give the experts in the field ideas of the cancerous mass's characteristics.

Knowledge Discovery of Bovine Tuberculosis in the Eurasian Badger using Machine Learning Techniques

Kyle McBride, Aleksandar Novakovic, Adele H. Marshall, Emily Courcier
School of Mathematics and Physics, Queen's University Belfast, Belfast, United Kingdom (UK);
Joint Research Centre in AI for Health and Welness, Ontario Tech University, Oshawa, Canada;
Veterinary Epidemiology Unit, DAERA, Belfast, United Kingdom (UK)

Abstract - Bovine tuberculosis (Mycobacterium bovis) is a disease of cattle with severe consequences for agriculture in the British Isles. The Eurasian badger (Meles meles) is implicated in the spread and maintenance of bovine tuberculosis in the cattle population and various measures have been trialed in badgers to control infection. A five-year pilot Test, Vaccinate and Remove investigation (TVR) was carried out in a 100km2 area of Northern Ireland that tested, vaccinated, and removed infected badgers. This study used machine learning techniques in order to predict whether a badger has bovine tuberculosis using data collected from the TVR study. Several machine learning models - Decision Trees, Random Forests, Logistic Regression, XGBoost - were created and attempted in order to classify the data with the highest accuracy. Synthetic Minority Oversampling Technique (SMOTE) was also carried out due to imbalance in the data. The C5.0 decision tree model was chosen as the final model. This model was the most appropriate choice as it achieved a very high AUC score with a value of 0.974 in training and 0. 962 in testing. It also had the benefit of being a white-box model. Almost all of the variables were found to be significant, including the visual diagnostic tests used in the study, thus supporting their importance. The final model gives confidence in current diagnostic tests to accurately identify infected badgers and helps to inform future diagnostic test regimes. This study represents one of the first applications of machine learning in wildlife disease control.

Noise Suppression Using Gated Recurrent Units and Nearest Neighbor Filtering

Arghavan Asad, Rupinder Kaur, Farah Mohammadi
Department of Electrical and Computer Engineering, Toronto Metropolitan University, Toronto, Canada

Abstract - A technique to enhance noisy speech through machine learning and digital signal processing methods is proposed in this paper. In the first step of enhancement, Mel-frequency cepstral coefficients are extracted from the noisy speech and fed to a gated recurrent unit (GRU) network which estimates a sequential gain vector used to improve the signal-to-noise ratio (SNR) of the noisy speech. In the second step of enhancement, nearest neighbor filtering is applied to generate an estimate of the isolated noise spectrogram in the noisy speech. This estimate is used to compute a soft mask which is multiplied with the frequency spectrum of the enhanced noisy speech from the first step. This two-step process achieves good results in SNR conditions of greater than 5 db. Under this threshold, the output speech can be distorted. Noisy speech is artificially generated through two datasets consisting of speech, and noise files to create a training and testing dataset for the GRU network.

Contextual Fusion of Classifiers under Variable Atmospheric Conditions for Coastal Surveillance

Oceane Tran, Anne-Laure Jousselme, Gilles Tedeschi, Lionel Gardenal
CS Group, La Garde, France;
University of Toulon, France

Abstract - Monitoring maritime areas is a challenging task which generally benefits from sensing systems such as radars or cameras, with different electromagnetic or optical capabilities. Infrared (IR) cameras in particular allow night and day surveillance, but their image is affected by maritime atmospheric conditions, e.g., the aerosol concentration. Further classification process thus inherits this image degradation resulting in possibly poor target classification results. To overcome this issue, we propose in this paper a contextual classifiers fusion system where two neural networks are trained into two environmental contexts and further combined with Bayesian reasoning. Individual classifier's reliability is considered, enabling to balance between the two classifiers depending on the uncertain context of use. Additionally, we apply an imprecise decision rule for a greater flexibility allowing a compromise between two criteria of accuracy and specificity. Results are obtained on simulated classifiers. outputs as well as on a synthetic dataset of IR images degraded with atmospheric context. It is shown that the proposed approach allows the possibility to increase accuracy while foregoing specificity and adds explainability.

Stress Detection Using Physiological Signals Based On Machine Learning

Ashmita Hota, Sung-Won Park
Department of Electrical Engineering and Computer Science, Texas A&M University-Kingsville, Texas, USA

Abstract - Stress can be defined as the body's attempt to control itself in response to changes in the environment. Due to stress work performance may suffer and the risk of neurological issues such as hypertension and psychological illnesses such as anxiety disorder may rise. In today's world, an increasing number of people are experiencing some form of stress. Comprehension of stress cognition is required along with the capacity to build systems with stress cognition characteristics. A methodology of stress detection using physiological signals based on machine learning is presented in this paper. Physiological signals such as respiration, sweat gland activity on the skin of hands, heart rate, and electromyogram were recorded while driving from multiple healthy participants in various situations and locations. The signal is then segmented for various time intervals such as 100, 200, and 300 seconds, depending on the levels of stress. Statistical features were retrieved and made available to the classifiers namely Support Vector Machine (SVM) and k-Nearest Neighbor (KNN) algorithm. We achieved the highest accuracy of 96% with 100 and 200-second long signal, and 98% with 300-second long signal.

Automatic Music Mastering using Deep Learning

Jared Smit, Byeong Kil Lee
Department of Electrical and Computer Engineering, The University of Colorado, Colorado Springs (UCCS), Colorado, USA

Abstract - The process of mastering music has long been a tedious component of music production. Since the early ages of recorded music, it has presented a hurdle for many musicians, requiring vast technical knowledge that many musicians do not have. As technology has improved and innovations in the industry have been made, the mastering process has become simpler and has allowed many more hobbyist musicians to produce their music, without needing the help of an expert. While it has become easier, the mastering process is still a deterrent for many would-be musicians. It is both time-consuming and expensive if outsourced to a professional producer, and it also vitally impacts the quality of the song. Deep learning can excel in taking in large amounts of data and performing tasks. It is also excelling in creative applications such as art and literature, with deep learning models creating paintings and stories. The same technology has not been widely explored for music mastering. With the use of deep learning, the music mastering process can be simplified much further, even automating it completely. This research explores this idea and some of the methods that could be used for such a task. It provides evidence that this concept is viable, results of various experiments with many deep learning model parameters, as well as topics for further exploration.

Fault Detection and Mitigation for Revenue Loss Minimization

Sepiribo L. Braide
Department of Electrical Engineering, Rivers State University, Port Harcourt, Nigeria

Abstract - Transmission line fault detection is an important aspect of monitoring the health of a power plant since it indicates when suspected faults could lead to catastrophic equipment failure. This research looks at how to detect generator and transmission line failures early and investigates fault detection methods using Artificial Neural Network approaches. Monitoring generator voltages and currents, as well as transmission line performance metrics, is a key monitoring criterion in big power systems. Failures result in system downtime, equipment damage, and a high danger to the power system's integrity, as well as a negative impact on the network's operability and dependability. As a result, from a simulation standpoint, this study looks at fault detection on the Trans Amadi Industrial Layout lines. In the proposed approach, one end's three phase currents and voltages are used as inputs. For the examination of each of the three stages involved in the process, a feed forward neural network with a back propagation algorithm has been used for defect detection and classification. To validate the neural network selection, a detailed analysis with varied numbers of hidden layers was carried out. Between transmission lines and power customers, electrical breakdowns have always been a source of contention. This dissertation discusses the use of Artificial Neural Networks to detect defects in transmission lines. The ANN is used to model and anticipate the occurrence of transmission line faults, as well as classify them based on their transient characteristics. The results revealed that, with proper issue setup and training, the ANN can properly discover and classify defects. The method's adaptability is tested by simulating various defects with various parameters. The proposed method can be applied to the power system's transmission and distribution networks. The MATLAB environment is used for numerous simulations and signal analysis. The study's main contribution is the use of artificial neural networks to detect transmission line faults.

EEG Classifier Using Wavelet Scattering Transform-Based Features and Deep Learning for Wheelchair Steering

Mokhles M. Abdulghani, Wilbur L. Walters, Khalid H. Abed
Department of Electrical & Computer Engineering and Computer Science,
Jackson State University (JSU), Jackson, MS, USA

Abstract - A Brain-controlled wheelchair functions via a brain-computer interface (BCI), an electroencephalogram (EEG) worn on the patient's forehead to detect neural activity. Classifying human thoughts still needs a significant effort to achieve a reliable BCI. This paper presents the design and testing of a mind-based control of a wheelchair for physically challenged users. The design of this system is focused on the acquisition of the EEG brain signals to create control of the wheelchair. All the laboratory measurements of human brain activities have been recorded using human control commands of wheelchair navigation. The received data are used to design a human brain-based control mechanism using wavelet scattering and deep learning techniques. After testing the designed deep learning algorithm, the overall classification accuracy was 95% to distinguish between four commands of the wheelchair (Forward, Stop, Left, and Right).

Automatic Detection of Pathological Myopia using Smartphone App

Sarvat Ali, Shital A. Raut
Department of Computer Science, Visvesvaraya National Institute of Technology, Nagpur, Maharashtra, India

Abstract - Pathological Myopia (PM) is an excessive axial elongation associated with myopia that leads to structural changes in the posterior segment of the eye (including posterior staphyloma, myopic maculopathy, and high myopia-associated optic neuropathy) and that can lead to loss of vision. To show the significance and real-time device utility of Artificial Intelligence in the field of ophthalmology, we have attempted to design a mobile-based application for Android users to detect PM automatically. This app attempted to bring detection of PM at the fingertip of general people serving as their personal screening tool. We have used smartphone compatible, pretrained MobileNetV3 small model with a custom layer at the end to train the model on the dataset of 400 images provided by the PALM challenge. The model reported an Area under Curve (AUC) of 0.9927 on the validation data of 400 fundus images. The optimized model version is imported to the Android Studio to develop the smartphone application, which on the input of the fundus image can report it as PM or Non Pathological Myopia (NPM). This app can be used by ophthalmologists as a second opinion or by general physicians in rural areas to report severe cases to specialists.

Survey and Comparison of Nature Inspired Optimization Algorithms

Jam Jahanzeb Khan Behan, Sajjad Haider
CellCarta, Antwerp, Belgium;
Faculty of Computer Science, Institute of Business Administration, Karachi, Pakistan

Abstract - Nature Inspired Meta-heuristics algorithms have made a profound contribution to the field of optimization ever since they have been initiated in the late 1980s and have gained immense popularity in the last decade. The aim of this research survey is to study and compare the performance of different types of nature inspired meta-heuristics algorithms by the means of evaluating them over several benchmark functions and compare the best so far value obtained by the algorithms. Every algorithm is run multiple times to ensure diversity in results and are compared using four different measures.

Toward CORDIC-based Hyperbolic Function Implementation for Neural Engineering Hardware

Ardavan Vakil, Miad Faezipour
School of Engineering Technology, Electrical & Computer Engineering Technology,
Purdue University, West Lafayette, Indiana, USA

Abstract - Researchers have proposed various models pertaining to the biological and mathematical aspects of neuronal spiking patterns. However, there is not much effort seen in running hyperbolic neuronal spiking pattern models in real-time hardware applications. Since hyperbolic functions are not synthesizable for computer/digital hardware, the CORDIC algorithm can be used to implement efficient hardware running the hyperbolic models of neuronal spiking patterns. Initially, CORDIC was specifically designed for real time computations in areal applications. In this research-work-in-progress, the development of the CORDIC algorithm with Verilog hardware description language is presented that can be used toward designing neural engineering hardware trigonometric and hyperbolic functions.

Towards Practical Onboard PHM System for Aircraft Engines

Jian Cao, Jinyi Xu, Jiaqi Huang, Weidong Yang
School of Computer Science, Fudan University, Shanghai, China;
COMAC Shanghai Aircraft Design and Research Institute, Shanghai, China;
Zhuhai Fudan Innovation Institute, School of Computer Science, Fudan University, Shanghai, China

Abstract - Aircraft engines are essential for modern airplanes and monitoring their status is paramount for keeping our journey safe. Hence, various PHM systems for aircraft engines have been designed to supervise the engine operation and provide maintenance advice for engineers. The first step of those systems is to identify whether there exist faults from substantial amount of sensor data. Recent studies have shown the dominant capability of deep learning methods in discovering latent abnormal patterns, but the inherently high computational cost of deep learning models has made them impractical when deploying on embedded systems where onboard PHM systems reside. We proposed a more practical approach that only uses simple SVM model and achieved 86.7% overall accuracy for detecting gas path components degradation and 0.826 F1-Score for detecting rotor collision and abrasion on our engine dataset.

CSCI-RTCS:
RESEARCH TRACK ON COMPUTATIONAL SCIENCE

Inteplato: Generating Mappings of Heterogeneous Relational Schemas using Unsupervised Learning

Leonard Traeger, Andreas Behrend, George Karabatis
Information Systems, University of Maryland, Baltimore County, Maryland, USA;
Communications Engineering, Technical University of Cologne, Cologne, Germany

Abstract - The growth of unstructured, unclean, and incomplete data poses quite a problem on attempts to integrate information from various disparate repositories. This paper describes Inteplato, a scalable web-based framework for domainindependent schema mappings between numerous heterogeneous databases. Starting with schema XML exports, the local schemas are visualized in a knowledge graph. A novel algorithm generates clusters based on similarity of local concepts by utilizing fuzzy string, synonym intersection, data type, and constraint similarities. Using these clusters, another algorithm generates mappings between global and local level concepts. The autonomy of each participating database system is maintained, and the mappings enable code automation for global-to-local query propagation. Accuracy was used as a metric to evaluate the generated mappings between local and global concepts.

Connected Hyperbolic Complex Networks

Rolland Vida, Jozsef Biro
Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary

Abstract - Generative models play important roles in analyzing and predicting relevant characteristics of real-world complex networks. The predictive power of these models lies in the possibility that large amount of synthetic networks similar to real ones at some extent can be produced and by analyzing them statistically significant inferences can be performed. Synthetic networks based on the hyperbolic geometry turned out to be good generative models of real-world networks, they reproduce several macroscopic behaviour of real networks and can help in assessing the scalability of new network functions. The original hyperbolic generative model by Krioukov et al. [1] lacks an important property, namely it can not provide with 100% probability that the resulted synthetic network is connected. In other words for some parameter regions, the networks fall into fragments resulting disconnected regions of subgraphs. Inevitable, from the viewpoint of modeling real networks the connectedness is an expected property. In this paper it is shown that with a slight extension of the original generation rule of the hyperbolic networks the connectedness can always be ensured. We also show analytical and numerical results on that there will be no significant changes in the macroscopic properties of networks like average degree, degree distribution and clustering.

Web-based 3D Smog Visualization for Air Pollution Analysis

Jordan Limperis, Felix Hamza-Lup, Weitian Tong
Department of Computer Science, Georgia Southern University, Statesboro, Georgia, USA

Abstract - The increase in the transaction of data and availability of augmented reality devices (virtual, augmented, and mixed reality) have led to the formation of a suitable environment for the real-time communication of complex air pollution data to a diverse population of individuals. In recent years, air pollution has become a global concern and thus it is important to constantly monitor air pollution and communicate it to citizens. In this study we design a web-based, 3-Dimensional virtual reality platform to portray air pollution risk through a realistic smog effect over the Joshua Tree Landscape. Through user natural language processing evaluation, this platform is shown to increase user risk perception of air pollution and its relation to environmental landscapes. Our model promotes the user's exploration of an air pollution landscape and provides tools to portray air pollution data to citizens.

Comprehensive Review of the Informative Path Planning, Autonomous Exploration and Route Planning using UAV in Environment Monitoring

Marcela Aparecida Aniceto dos Santos, Kelen Cristiane Teixeira Vivaldini
Department of Computer Science, Federal University of Sao Carlos, Brazil

Abstract - Unmanned Aerial Vehicles (UAVs) have been used in several applications for monitoring environments and mapping. To carry out the mapping of these environments, the UAV needs to decide which path to follow to collect as much information about the environment to maximize the search area. In the literature, these issues are being addressed within the area of Informative Path Planning (IPP), Route Planning (PR) and Autonomous Exploration of environments. As a way to clarify these problems and their objectives in robotics, this article aims to present a comprehensive review on these areas highlighting their approaches. For this, a comprehensive review of the main existing methods to solve them was carried out and this study serves as a starting point and a guide for everyone interested in exploring the monitoring area for data acquisition in unknown environments.

A Modeling Strategy using Bayesian Optimization with POMDP for Exploration and Informative Path Planning of UAVs in Monitoring of Forest

Marcela Aparecida Aniceto Dos Santos, Kelen Cristiane Teixeira Vivaldini
Department of Computer Science, Federal University of Sao Carlos, Brazil

Abstract - Unmanned Aerial Vehicles (UAVs) have been used for several applications in monitoring complex and unknown environments. The challenge is to plan missions for the UAV in situations where the vehicle needs to visit and explore an area and analyze it in real-time to define the route to be followed. This visit occurs with the search area's maximization from the definition of the trajectories, making it possible to collect information to acquire knowledge about the environment and provide a map. This type of problem is known as Informative Path Planning (IPP) and Autonomous Exploration (AE). In this context, Bayesian Optimization (BO) has been adopted. Moreover, there is a need to define the planning for decision-making based on information about the environment considering specific restrictions. Partially Observable Markov Decision Processes (POMDP) can be used to define the planner responsible for the decision-making. Therefore, considering these methods and the strategy of exploring the environment in a continuous 3D space, this paper proposes a development of a modeling strategy that is exploration and informative path planning using Sequential Bayesian Optimization with POMDP in forest monitoring with Canopy gap. A sequential decision-maker was developed under uncertainty based on the Sequential Bayesian Optimization approach with Partially Observable Markov Decision Process (BO-POMDP).

N-Body Performance with a kD-Tree: Comparing Rust to Other Languages

Jonathan Rotter, Mark C. Lewis
Department of Computer Science, Trinity University, San Antonio, Texas, USA

Abstract - This paper presents the results of a more representative N-body benchmark utilizing a kD-tree implemented in multiple languages. We find that while Rust is slightly slower than C or C++ for smaller simulations, it is the fastest language for simulations at the scale we use in actual research. On the other hand, Go is constantly 1.5-2x slower than Rust. The JVM is competitive with Go for intermediate-size simulations but struggles when we reach one million particles. As expected, scripting languages are the slowest, though it is rather remarkable how much slower Python is than Node.js. Also surprising is that our attempts to speed up the Python implementation using NumPy made it significantly slower.

Finite Difference Methods with Improved Properties and their Application to Solving Some Model Problems

V. R. Ibrahimov, M. N. Imanova
Institute of Control System: A. Huseynov, Baku, Azerbaijan;
Computational Mathematics, Baku State University, Baku, Azerbaijan;
Science Development Foundation under the President of the Republic of Azerbaijan, Baku, Azerbaijan

Abstract - As is known finite difference methods or multistep methods can been applied to solve scientific and technical tasks as the numerical methods with different properties. For this, here have considering application of the finite difference methods to solve initial-value problem for both ODEs and Volterra integro-differential equation by using some similarities between these tasks. It is known that initial value problem (ivp) for ODE can been received from the ivp for the Volterra integro-differential equation as the partial case. By using these relations of the named problems, here the methods, which are used in solving of the ODEs, have modified and applied to solve ivp for the Volterra integro-differential equations. For this purpose have used the multistep methods with constant coefficients. And have recommended some ways for application of above mentioned methods to calculation of definite integrals. And also have given some recommendation to select appropriate methods. By using some model problems have illustrated the advantages of the investigated here methods which have recommended to solve above mentioned problems. To increase the reliability of the receiving results obtained by using some numerical methods, here have considered the possibility of using bilateral methods.

An Improved Sudoku Solver

Adiba Afif Suha Binta Wadud, M. Abdullah-Al-Wadud
Bangladesh International School, English Section, Riyadh, Saudi Arabia;
Department of Software Engineering, College of Computer and Information Sciences,
King Saud University, Riyadh, Saudi Arabia

Abstract - Sudoku is a popular puzzle game played by people of all ages. Consequently, many methods are developed to solve these puzzles on computer. However, these methods have deficiencies, especially in terms of the time taken to solve the puzzles. In this paper, we propose a method which is more efficient than the methods used before. This uses a few additional data structures and updated algorithms to make the method faster at solving various Sudoku puzzles.

The 2-MAXSAT Problem Can Be Solved in Polynomial Time

Yangjun Chen
Department of Applied Computer Science, The University of Winnipeg, Winnipeg, MB, Canada

Abstract - By the MAXSAT problem, we are given a set V of m variables and a collection C of n clauses over V. We will seek a truth assignment to maximize the number of satisfied clauses. This problem is NP-hard even for its restricted version, the 2-MAXSAT problem by which every clause contains at most 2 literals. In this paper, we discuss a polynomial time algorithm to solve this problem. Its time complexity is bounded by O(n^2 M^3). Hence, we provide a proof of P = NP.

Refining a Parameter Setting Evolutionary Approach for Fire Spreading Models Based on Cellular Automata

Maria Eugenia A. Ferreira, Danielli A. Lima, Luiz G. A. Martins, Gina M. B. Oliveira
Federal University of Uberlandia (UFU), Campus Santa Monica, Computer Faculty (FACOM), Uberlandia, MG, Brazil;
Federal Institute of Triangulo Mineiro (IFTM), Campus Patrocinio, Information Technology Department (DTI), Laboratory of Intelligence Computational and Robotics (LICRo), Patrocinio, MG, Brazil

Abstract - Forest fires have increased significantly due to climate change affecting diverse biomes. Fire propagation modeling is essential in preventing and controlling the damage caused by this phenomenon. Cellular automata were demonstrated to be effective when constructing such models. However, adjusting the many parameters involved in these models is a complex task. Recently, an evolutionary approach to parameter adjustments of a fire simulation model based on CA has been proposed. This paper aims to continue this study by refining the method. Different experiments were carried out to analyze the sensitivity of the evolutionary approach to parameter adjustment, including the generation of bases from other models and the inclusion of heterogeneous vegetation.

An Open and Fully Decentralised Platform for Safe Food Traceability

Eric Tutu Tchao, Elton Gyabeng, Andrews Tang, Joseph Barnes Nana Benyin, Eliel Keelson, Jerry John Kponyo
Distributed IoT Platforms, Privacy and Edge-Intelligence Research Lab,
Kwame Nkrumah University of Science and Technology, Ghana

Abstract - Concerns about food safety have grown across society in recent years. Building a trustworthy traceability system is essential for effectively identifying and preventing food safety issues as well as tracing the responsible parties. The entire food supply chain, which includes the stages of production, processing, warehousing, transportation, and sale, must be precisely recorded, shared, and traced. Traditional traceability systems suffer from problems like data invisibility, tampering, and the leakage of sensitive information. This paper proposes an open platform for a food safety traceability system that indefinitely and incessantly stores and records all transactions, events, and activities on the blockchain's immutable ledger linked with IPFS - a peer-to-peer decentralised file system - for storing and providing maximum transparency and traceability. The platform leverages the blockchain's characteristics such as immutability, transparency, smart contracts, and consensus algorithms to make it ideal for food safety traceability systems. But more importantly, it mirrors the food supply chain making it a pluggable toolbox for all stakeholders across the food chain to adapt to their system irrespective of the food products they deal with since it is a multi-asset system as well. It could be even adapted for non-food products that have a supply chain similar to the typical food supply chain. Simulation results show that there is the complete success of all the blockchain transactions on our platform with real-time responsiveness.

New Algorithms for the Computation of Inverse Transforms

Jun Zhang, Ruzong Fan, Fangyang Shen
Department of Computer Science and Engineering Technology,
University of Maryland Eastern Shore, Maryland, USA;
Department of Biostatistics, Bioinformatics, and Biomathematics,
Georgetown University, Washington, DC, USA;
Department of Computer Systems Technology, New York City College of Technology,
City University of New York, Brooklyn, New York, USA

Abstract - The Laplace Transform has been widely used for about two centuries in problem solving in mathematics, engineering and sciences. The Sumudu Transform is very recent, it is as powerful as the Laplace Transform and has many nice features. Traditionally, when these transforms are used, the calculations of the inverse transforms are necessary; unfortunately, the calculations of Inverse Laplace Transform and Inverse Sumudu Transform are problematic and challenging. The authors studied Sumudu Transform in the computational approaches. We shall introduce a few novel algorithms on the computations of Inverse Sumudu Transform and Inverse Laplace Transform in this paper. The algorithms introduced here are straightforward to understand and powerful for problem solving, and some demo versions were implemented in the Maple Computer Algebra Systems.

Modeling of Electrical Distribution System Reliability with Local and Interregional Bidirectional Centralized Electric Vehicle Charging Stations

Janine C. G. Darato, Jordan R. Perlas, Taufik Taufik, Ali Dehghan Banadaki, Majid Poshtan
Electrical Engineering Department, Cal Poly State University, San Luis Obispo, California, USA

Abstract - This paper developed a method to evaluate the impact of bidirectional electric vehicle (EV) charging on power system reliability using Synergi Electric software. Load profiles, EV availability, and EV state-of-charge (SOC) were important factors considered in this study. The analysis in this study is based on local and interregional vehicle-to-grid (V2G) implementation at different load points in the system. In general, local V2G implementation improves system reliability over interregional V2G power flow. System Average Interruption Duration Index (SAIDI) and System Average Interruption Frequency Index (SAIFI) improvements increased with increasing SOC but were less progressive between higher SOCs. Based on the simulation results, the method proves to be sufficient to calculate SAIDI and SAIFI reliability indices as the simulation results corroborate with the theory.

Incentives Effect on the Size of Mining Pool for Power-of-Work Blockchain Systems

Yean-Fu Wen, Chun-Yu Huang
Graduate Institute of Information Management, National Taipei University, Taiwan (R.O.C)

Abstract - Blockchain mining is a trending topic because bitcoin prices and mined incentives are attractive. The trend of blockchain mining is joining a mining pool, which also allows recording of a block. Pool selection is a major problem for the proof-of-work consensus protocol because larger mining pools provide higher winning probability; however, numerous miners share the reward. We modeled a mining process as a draw ball problem to simulate various sizes of mining pools according to the existing mining mechanism used for bitcoin blockchains. An average winning rate and individual reward were set as metrics to help miners select the pool. By utilizing the ball drawing computational simulation model, we analyzed and compared the mining winning rate, reward, and margin reward with several reward mechanisms. The simulation results provided managerial insights for individual miners. Moreover, they support advanced reward mechanism selection for mining pool operators.

A Review of Counterfeit Product Identification Models

Selallo M. Rahlagane, Topside E. Mathonsi, Tshimangadzo M. Tshilongamulenzhe
Tshwane University of Technology, Pretoria, South Africa

Abstract - In recent years, the introduction of blockchain technology has drawn considerable interest in a wide array of fields of study. However, this technology has mainly been utilized to track financial and health issues. The public, on the other hand, must contend with counterfeit goods. Most people put their trust in manufacturers and brands. In most cases, it is not easy to determine whether the information on the product labels is genuine or not. Almost every product has a fake copy of itself. In South Africa, most foreign owned tuck-shops sell similar products at lower prices by compromising the original quality of well-known brands. Globally, the major issue is that even original product experts cannot distinguish fakes from original products. In this paper, we propose a model that uses blockchain technology to identify fake products using Quick Response (QR) codes that manufacturers print on their original products during manufacturing. The advantage of this technology is that data is recorded in a way that makes it hard to alter, hack, or cheat the system. QR codes will be used to identify each product on a blockchain. Each blockchain will include the details of the product and a QR code. Customers can scan the QR code using their smartphones to determine whether the product is real or fake. This will curb the issue of tuck-shops that sell counterfeit products in the country. The result of this proposal will be a mobile application developed to differentiate between counterfeit and original products.

Formulations and Algorithms to Find Maximal and Maximum Independent Sets of Graphs

Maher Heal, Kia Dashtipour, Mandar Gogate
Baghdad University, Baghdad, Iraq;
School of Computing, Edinburgh Napier University, Merchiston Capus,
Edinburgh, United Kingdom (UK)

Abstract - We propose four algorithms to find maximal and maximum independent sets of graphs. Two of the algorithms are non-polynomial in time, mainly binary programming and non-convex multi-variable polynomial programming algorithms. Two other algorithms run in polynomial time seek to find a maximum independent set. The algorithms depend on our earlier work. The main advantage and the difference of the new algorithms is that we do not need to enumerate the maximal cliques of the graphs. We applied the algorithms to some graphs from DIMACS and other graphs and their performance was seen to be adequate.

Intelligent Computational Approach to Constructing Adequate Statistical Decisions under Parametric Uncertainty of Applied Stochastic Models

Nicholas A. Nechval, Gundars Berzins, Konstantin N. Nechval
BVEF Research Institute, University of Latvia, Riga, Latvia;
Aviation Department, Transport and Telecommunication Institute Riga, Latvia

Abstract - The technique used here emphasizes pivotal quantities and ancillary statistics relevant for optimization or obtaining prediction limits (or intervals) for anticipated outcomes under parametric uncertainty and is applicable whenever the statistical problem is invariant under a group of transformations that acts transitively on the parameter space. It does not require the construction of any tables and is applicable whether the experimental data are complete or Type II censored. The exact prediction limits on order statistics associated with sampling from underlying distributions can be found easily and quickly making tables, simulation, Monte-Carlo estimated

percentiles, special computer programs, and approximation unnecessary. The proposed technique is based on a probability transformation and pivotal quantity averaging. It is conceptually simple and easy to use. The discussion is restricted to one-sided prediction limits. Finally, we give practical numerical examples, where the proposed analytical methodology is illustrated in terms of the one-parameter exponential distribution. Applications to other log-location-scale distributions could follow directly.

Computation in Markoff-Hurwitz Equations III

Shanzhen Gao
Department of Computer Information Systems, Reginald F Lewis College of Business,
Virginia State University, Petersburg, Virginia, USA

Abstract - In the more than 100 years since Markoff-Hurwitz Equations, they play a decisive role, have turned up in an astounding variety of different settings, from number theory to combinatorics, from classical groups and geometry to the world of graphs and words, from discrete mathematics to scientific computation. This paper is the continuation of our previous papers [1] (we refer to as Part I), and [2]. In this paper we provide a new proof to K.Guy and R.Nowakowski's problem [22]. We challenge the readers to prove or disprove our conjectures. We list many theorems that we have already proved. But we do not provide proofs since we would like to challenge the readers to prove them independently and hopefully get better proofs.

Entry and Exit Decision Under Uncertainty Where Prices Include Jump Diffusion Processes

Kangrong Tan, Shozo Tokinaga
Faculty of Economics, Kurume University, Japan;
Department of Economics, Kyushu University, Japan

Abstract - Firms make decisions of investment under uncertainty to avoid additional cost. This paper deals with entry and exit decision under uncertainty where prices include jump diffusion processes. Conventional works assume the price follows an ordinary Brownian motion. But in real market, cases in which the prices evolve in the presence of sudden rise and fall must be taken into account. Recently, works are shown for modeling time series based upon jump diffusion processes, and this approach has been successfully applied to the fields such as optimization of production and stock exchange. For entry and exit problem, we use the model in which jumps are included when the values of time series vary relatively larger, so that the corresponding probabilities of jumps become larger. We then obtain the value functions for the cases in which prices are modeled with jump diffusion processes in simple forms. As a result, we can estimate the optimal prices (thresholds) for entry and exit decisions which are seemingly deviated from thresholds under the prices following only ordinary Brownian motion. Moreover, by setting the recovery time after jumps (required for returning to the previous level of prices) longer than the original assumption of rapid returning, we yield the deviation of time for entry and exit problem based upon jumps. We find our approach is applicable to cope with a wider range of real business data.

Effectiveness of PMV-based Controls for Commercial Buildings in the USA

Maya Arida, Ahmed Megri, Rand Talib
North Carolina A&T State University, Greensboro, North Carolina, USA;
University of Cincinnati, Cincinati, Ohio, USA

Abstract - Using an accurate energy model, this paper simulated existing commercial buildings to compare PMVbased comfort control with conventional thermostatic control and conduct a sensitivity analysis using the Design of Experiments method with monitored energy consumption data. This analysis was applied considering the impact of building envelope characteristics including thermal insulation level of exterior walls and air leakage rate on the ability of both control options to maintain indoor thermal comfort while minimizing cooling energy consumption. The analysis findings indicated that PMV-based comfort control is viable for striking a middle ground between thermal comfort and heating energy consumption with a PPD level at 8% for any combination of walls. R-value and air infiltration rate while temperature-based controls result in an unacceptable indoor thermal comfort performance especially for low R-values regardless of the air infiltration rate with PPD level reaching over 76%. Also, the results show that in a comfort-controlled space, the average radioactive temperature and occupant-related features like metabolic rate and clothing level have a far bigger effect on energy use than other parameters like relative humidity.

Fibonacci Numbers in Memory of Richard K. Guy

Weizheng Gao, Jianning Su, Olumide Malomo, Somasheker Akkaladevi,
Ephrem Eyob, Adeyemi A. Adekoya, Shanzhen Gao, Julian D. Allagan
Department of Mathematics and Computer Science, West Virginia State University, West Virginia, USA;
Department of Mathematics and Sciences, Midway University, Midway, Kentucky, USA;
Department of Computer Information Systems, Reginald F Lewis College of Business,
Virginia State University, Virginia, USA;
Department of Mathematics, Computer Science and Engineering Technology,
Elizabeth City State University, Elizabeth City, North Carolina, USA

Abstract - Mathematicians and Computer Scientists love Fibonacci numbers because they seem to be very easy, but they are related to many challenging topics. There are around 7300 items on Fibonacci numbers in the On-Line Encyclopedia of Integer Sequences. We will discuss some properties, algorithms and Python programs used to generate Fibonacci numbers, conjectures, and applications of Fibonacci numbers. We would like to offer an award to the first person who proves or disproves one of our conjectures in this paper. We would like to use this paper in memory of our dear friend, Richard, the Man Who Loved Problems [7].

Processing Time Comparison and Applications of Multiplying 2 by 2 Matrices using Strassen Algorithm

Shanzhen Gao, Olumide Malomo, Ephrem Eyob
Department of Computer Information Systems, Reginald F Lewis College of Business,
Virginia State University, Petersburg, Virginia, USA;
Lifetime Fellow of the Institute for Combinatorics and its Applications

Abstract - Multiplying matrices can be very challenging although it seems straightforward. Many researchers have studied the multiplication of two 2 x 2 matrices by using the Strassen Algorithm in the past 50 years. They focused on the complexity from the mathematical and algorithmic points of view. We will discuss the running time comparison of two algorithms from a practical point of view. Several open problems are then posted to challenge our readers.

Automatic Plant Disease Detection Using Deep Learning

Saeka Rahman, Sudip Vhaduri, Miad Faezipour
School of Engineering Technology, Electrical and Computer Engineering Technology,
Purdue University, West Lafayette, Indiana, USA;
Department of Computer and Information Technology,
Purdue University, West Lafayette, Indiana, USA

Abstract - Plants get affected by different types of diseases. Each year a significant amount of food loss occurs due to various diseases globally. To ensure food security worldwide, the minimization of food loss due to plant diseases is essential. Diagnosing the diseases at the right time is crucial for effectively treating plants. Usually, the farmers or plant scientists diagnose the diseases. They perform the diagnosis by visually inspecting the leaves or different parts of the plants. This task is timeconsuming and requires experience. Sometimes this type of manual detection process becomes more difficult due to the number of plants and the similarity of the symptoms. To alleviate this process, an automated plant disease detection model is developed using the Convolutional Neural Network (CNN) architecture of deep learning. The proposed model's computational expense and run-time are reasonably low. The model is validated using 2-D color images of potato leaves, including healthy and infected, collected from the PlantVillage dataset of the Kaggle public website. Test accuracy, the required number of parameters, and model run-time are considered to evaluate the model. The result is comparable to the state-of-the-art models to detect plant diseases with fewer resource requirements.

Fault Identification in Wind Turbines: A Data-Centric Machine Learning Approach

Danielle Pinna, Rodrigo Toso, Rafelli Coutinho, Ana Isabel Pereira, Diego Brandao
PPCIC - CEFET/RJ, Rio de Janeiro, Brazil;
Research - Microsoft, Boston, Massachussets, USA;
CeDRI - IPB, Braganca, Portugal

Abstract - The last few years have been marked by the transition of the world energy matrix, predominantly with wind and solar sources considered clean energies. Wind turbines, responsible for the energy conversion process, are complex and expensive equipment susceptible to several failures due to multiple factors. Monitoring turbine components can assist in detecting failures before they occur, reducing equipment maintenance costs. This work compares machine learning techniques in a data-centric approach to wind turbine failure detection. Preliminary results demonstrate the importance of feature selection in this problem.

Hibernation Based Hybrid Booting for Baseboard Management Controller

Ajung Kim, Jaehoon Ahn, Younghwan Kim, Jiman Hong
School of Computer Science and Engineering, Soongsil University, Seoul, South Korea;
Intelligent IDC Project Office, Korea Electronics Technology Institute, Seongnam, South Korea

Abstract - In the event of server failure, the host system can quickly diagnose the problem using the Baseboard Management Controller. For faster booting Baseboard Management Controller, the hibernation technique can be used. However, a fast booting by using the hibernation technique takes significantly less time but booting by using the hibernation technique may take longer than a cold booting, depending on the size of the hibernation image. In this paper, we propose a hybrid booting technique that dynamically performs cold booting or hibernation-based booting by predicting the size of the hibernation image based on memory usage. The proposed technique derives a relational expression between the size of the hibernation image based on memory usage through regression analysis and predicts the size of the hibernation image and it can dynamically determine the faster booting between cold booting and hibernation-based booting.

CSCI-RTCB:
RESEARCH TRACK ON COMPUTATIONAL BIOLOGY

Collagen a1(XI) Structure Prediction by Alphafold 2

Abu Sayeed Chowdhury, Julia Thom Oxford
Biomolecular Sciences Graduate Program, Center of Biomedical Research Excellence,
Boise State University, Boise, Idaho, USA;
Department of Biological Sciences, Center of Biomedical Research Excellence,
Boise State University, Boise, Idaho, USA

Abstract - Collagen alpha1(XI) is a minor fibrillar collagen involved in the critical regulation of collagen fibrils such as nucleation, assembly, and regulation of fibril diameter. The amino propeptide domain of the collagen alpha1(XI) is retained on the surface of the collagen fibril for an extended period of time and may play a crucial role in the interaction with extracellular matrix glycosaminoglycans and other proteins during the process of fibrillogenesis. Understanding the mechanism of action of this protein will ultimately help us understand the organization and assembly of the extracellular matrix that underlies the structural integrity of connective tissues.

A Flexible Variant of the Neighbor-Joining for Building Phylogenetic Trees

A. Chastel Lima, Eloi Araujo, Marco A. Stefanes, L. C. S. Rozante
Federal University of Mato Grosso do Sul, Campo Grande, MS, Brazil;
Federal University of ABC, Santo Andre, SP, Brazil

Abstract - An important problem in Computational Biology is to rebuild a phylogenetic tree from a set of species S and the respective evolutionary distances between each pair of specie in S. The most used methods for accomplishing this task are those distance-based due to their efficiency and accuracy. Find a phylogenetic tree can be performed by Neighbor-Joining (NJ) method, a high accuracy distance-based method, but whose running can be impractical when the amount of species is large, since it is a cubic time algorithm in the worst case. The Unweighted Pair Group Method Arithmetic Mean (UPGMA) is another distancebased method faster than Neighbor-Joining whose complexity is quadratic whether good data structures are used, however its accuracy is not so good when compared to Neighbor-Joining. In this article we present a flexible variant of the NJ algorithm that is faster than the NJ and can construct phylogenetic trees with good accuracy. Through experimental tests we demonstrate that with a good calibration our Flexible Neighbor-Joining reach solutions with accuracy better than UPGMA and with time performance better than both NJ and UPGMA methods.

Three Genome-scale Approaches Support that Lungfish is the Closest Living Relative of Land Vertebrate, But Not Coelacanth

Yunfeng Shan, Youjun Zhou, Guo Li, Xiu-Qing Li, Robin Gras
College of Mathematics and Computer Sciences, Guangxi Science and Technology Normal University, P. R. China;
College of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin, P. R. China;
Fredericton Research and Development Centre, Agriculture and Agri-Food Canada, Fredericton, New Brunswick, Canada;
School of Computer Science, University of Windsor, Windsor, Ontario, Canada;
Department of Natural History, Royal Ontario Museum, Toronto, Ontario, Canada

Abstract - The origin of tetrapod has been one of intense debating open questions for decades between coelacanth (Latimeria chalumnae) and lungfish (Protopterus annectens). For resolving this incongruence in phylogenies, a genome-wide data mining approach is used to retrieve 43 shared genes of seven taxa from GenBank and further 1001 orthologous genes of ten taxa from the Ensembl and NCBI. We used the maximum gene-support tree approach and the majority-rule branch approach to analyze 43 nuclear genes encoding amino acid residues and compared these results to those inferred with the concatenation approach. Our results successfully provide strong evidence in favor of the lungfish-tetrapod hypothesis, but rejecting the coelacanth-tetrapod hypothesis based on significantly fewer gene supports and lower taxon jackknife probabilities for the coelacanth-tetrapod clade than the lungfish-tetrapod one with the maximum gene-support tree approach and the jackknife method for taxon subsampling. When more and more genomic data become available in recent years, sequence data of 1001 shared genes was mined. We used the maximum gene-support approach with this larger dataset successfully to infer that lungfish is the closest relative of land vertebrates with a significant difference at $p < 0.01$ (Chi-Square test) in gene support values between a maximum gene-support tree and the second most gene support tree with ML methods. The second most support to the maximum (SM ratio), a relative value, is a better support index than a single absolute value of support to show the insight of the phylogenetic support. Our results also show increasing the number of shared genes is much more effective than increasing the number of taxa.

Screening Results by Colorectal5 of 169 Microarrays

Shuichi Shinmura
Professor Emeritus of Seikei University, Japan

Abstract - The discriminant theory has four serious problems. The author studies a new discriminant theory (Theory1) and develops a Revised Optimal-LDF (RIP) by integer programming (IP), finding a minimum NM (MNM) and solving four serious problems of Theory1. Every data has a unique MNM, and .MNM=0. defines LSD. If a two-group is LSD, other LDFs are useless because ER of RIP always becomes 0 (Problem1, 2, 3). We find two new facts. Fact1 (LDF.s coefficient and NM relationship) shows that the combinatorial theory is more proper for Theory1. Fact2 is the monotonous decrease of MNM (MNMk>=MNM(k+1)). Although the discriminant theory is not inferential statistics (Problem4), 10-fold cross-validation (Method1) solves it. Method1 evaluated the discriminant result by the average ERs of 10 validation samples (M2). Since 2015, RIP has found six firstgeneration, and 163 second-generation Microarrays were LSD. Fact2 can explain that LSD has a Matryoshka/Nested data structure containing small LSDs (Small Matryoshkas, SMs) to the smallest LSD (Basic Gene Set, BGS). The Matryoshka feature selection method (Method2) can split 169 arrays into many SMs with less than n genes. MNMs of SMs are zero, which separates cancer and normal subjects. In addition to SM, we found RIP.s BGS and logistic regression's Degree of Freedom (DF) decompositions. We find the vital Fact3 that every LSD has four universal data structures of discrimination data. Fact3 is the gospel for high-dimensional gene data analysis of every living and opens the new discriminant world of big data. This paper proposes the screening method for cancer gene diagnosis (Theory2) to find vital BGSs with "M2=0 and less than five genes". If physicians study the relationship between vital BGSs and legacy oncogenes found by medical research, they can reveal the character of vital BGS as multivariate oncogenes.

Investigating Protein Structure Populations from Simulation Data using Unsupervised Learning

Gideon K. Gogovi, Joshua Kiddy K. Asamoah, Gabriel Obed Fosu
Department of Mathematics and Statistics, University of Houston-Downtown, Houston, Texas, USA;
Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Abstract - Data obtained from molecular dynamics simulation provides important intuition into the dynamical interactions of biological molecules. The chronicles of sequential time-dependent atomic motions of configurations obtained from simulation and the derived properties estimated from molecule's trajectory is specified by this sequence. Therefore, knowing how to efficiently extract representative structures from simulation data is important because often, we will want to identify changes in conformation of a protein structure when simulation is performed. We use unsupervised machine learning techniques to cluster such data and investigated a few of protein structural properties. The algorithms implemented in this paper presents clusters of the simulation data that tends to group frames from an adjacent block of time together, even when sampling at 10 ps intervals. We found that sampling of conformational space for a shorter run simulation may not be able to completely visit all structures that belong to a specific cluster. But for the sufficiently long simulation, the systems revisit previous clusters repeatedly. Cluster populations change rapidly at the initial stage of the simulations, but became steady before each got to their terminal values, indicating equilibrium attainment. Investigation of protein structure properties also attest the correspondence between clusters of protein structures obtained from the clustering algorithms.

An Overview of Machine Learning in Biology

Abigayle Hoppe, Ehyo Marie Degny, Callie Siverson, Frank Djarnie, Mohammed Mahmoud
Department of Computer Science & Information Systems,
Minnesota State University, Moorhead, Minnesota, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - There is a long history between Machine Learning and its application in Biology. Machine Learning has allowed the field of Biology to have many advancements. Such include studies for prediction and discovery and different Machine Learning techniques for specified types of biological data. The versatility of techniques and frameworks has helped improve Machine Learning every day, and hopefully, this will help us improve and become more efficient on new things we discover, and grow our database on biological data. In this paper, we will discuss the different applications of Machine Learning in Biology, i.e., Synthetic Gene Circuits, Convolutional Neural Networks, Recurrent Neural Networks and Generative Adversarial Networks. Since the biological and medical fields have been quickly growing into a data-rich environment, Machine Learning has become a vital tool to sort through all of this data coming out. The application of Machine Learning helps in two main facets. The first way is to help classify and predict tasks that a machine can quickly do. Secondly, it does not allow much human input, which will help minimize human bias or performance issues. Neural Networks go hand-in-hand with Deep Learning as they are one of its techniques. By utilizing Deep Learning and Neural Networks, we can expand our learning in Biology far more than ever before.

CSCI-RTBD:
RESEARCH TRACK ON BIG DATA AND DATA SCIENCE

Word2vec Embeddings for Categorical Values in Synthetic Tabular Generation

Waldemar Hahn, Martin Sedlmayr, Markus Wolfien
Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, Technische Universit, Dresden, Germany

Abstract - Although more and more generative models for synthetic tabular data exist, not all of them handle numerical and categorical data equally well. Some of the approaches are solely limited to numerical data. To extend the application of these methods beyond numerical data, we propose a Word2Vecinspired approach for converting categorical values into numerical values. We demonstrate on the Census Income dataset that the proposed embeddings are capable of learning semantic relationships for ordinal variables. In general, we observed that with larger embedding sizes the quality of the learned embeddings increases. We trained state-of-art CTGAN models on this data and compared it to CTGAN's built-in method for learning categorical data. Our proposed method achieved comparable results. We, therefore, suggest our proposed method as a versatile algorithm that can improve on the synthetic tabular data generation without the need to change existing architectures.

Sparrow Tracer: Scalable Real Time Metrics from Event Log Pipelines at Twitter

Riya Chakraborty, Lohit Vijayarenu, Zhenzhao Wang, Praveen Killamsett
Twitter Inc, USA;
Columbia University, USA

Abstract - Streaming event pipelines are one of the core components of Twitter's Data Infrastructure. Twitter Sparrow is a project responsible for aggregating, processing and delivering user action generated events from microservices to data warehouses and data lakes in real time. User action generated events are converted into datasets used for data processing and data analytics use cases. This project is built using different on premise and cloud services. One of the important requirements of such a streaming event pipeline is the ability to measure important metrics such as latency, event count, event drop vs. success rate, and more. These metrics are responsible for defining the health of the streaming pipeline as well as providing valuable insights to users of the events. In this paper we introduce Sparrow Tracer which is a novel way to capture the event metrics using the concept of tracer events.

CDAP: A Cultural Algorithm for Data Placement in Big Data Workflows

Mahdi Ebrahimi, Seyed Ziae Mousavi Mojab, Najmeh Najmi
Department of Computer Science, California State University, Northridge, Los Angeles, California, USA;
College of Business & IT, Lawrence Tech. University, Southfield, Michigan, USA

Abstract - The performance of executing big data workflows in the cloud highly depends on the placement of the workflow original datasets. An optimal data placement reduces the data movement among virtual machines, and as a result, it significantly reduces the workflow makespan. Data placement is an NP-hard problem, therefore in this paper, we propose CDAP (Cultural algorithm Data Placement), a novel metaheuristic data placement strategy

based on Cultural Algorithms (CA) to improve the performance of a workflow by minimizing the data movement among the virtual machines during workflow executions. The effectiveness of CDAP is demonstrated through extensive experiments where we evaluated our proposed method against a set of well-known data placement strategies.

An Agri-Food Data Platform for Food Safety and Certification

Nikos Papageorgopoulos, Danai Vergeti, Elena Politi, Dimitris Ntalaperas, Xanthi S. Papageorgiou, Manos Karvounis, Giannis Stoitsis, Braminir Rakic, Milos Kotlar, Simeon Petrov, Antoniy Kunchev
Robotics & Cognitive Systems Unit, UBITECH Athens, Greece;
AGRROKNOW Athens, Greece; PROSPEH Belgrade, Serbia;
SIRMA AI Sofia, Bulgaria

Abstract - Food safety is undergoing through tremendous challenges over the last years, with food scandals and contamination issues putting constant pressure to global markets, while consumers demands for high quality of products are increasing. This raises the need for increasing stakeholders. knowledge of the food production process and adopting data sharing practices in the product and supply chain management. Data sharing platforms can undertake the role of creating high value from data while facilitating secure and mutually beneficial multi-partner data sharing. Our proposed system aims to deliver an industrial data platform that will facilitate the exchange and connection of data between different food safety actors, who are interested in sharing information critical to certification, while boosting the way that food certification takes place in Europe.

SPECTS: Price Elasticity Computation Model using Thompson Sampling

Srividhya Sethuraman, Uma Maheswari G, Sharadha Ramanan
TCS Research and Innovation Labs, Tata Consultancy Services Ltd, Chennai, India

Abstract - A large international Kids' apparel chain retailer, operating in 800 stores and e-commerce, wanted to quantify the relationship of price with sales for all their seasonal products. The price elasticity model had to capture the behavior of seasonal items including limited price changes, highly volatile sales and need to adapt to changes in the environment. We developed a model based on Sequential Price Elasticity Computation using Thompson Sampling (SPECTS) to compute price elasticity values. SPECTS consists of four different Thompson Sampling based algorithms, namely, Distribution approximation, XGBoost regressor, GAN and MCMC sampling methods, for determining parameters of price elasticity distribution. The output of SPECTS is the average price elasticity value from these four methods. We applied the model to the Clothing collection for the Fall season. We employed a distributed computing and parallel execution framework in the cloud. Our solution was later scaled and implemented for remaining 1532 styles and 7300 products.

Application of an Oversampling Method for Improving Road Surface Classification

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

Abstract - This paper presents the application of oversampling methods to improve the road surface classification performance. A large number of multimodal experiments were implemented with the following configuration: three accelerometer vibration signals from acceleration in the X-, Y-, and Z-axis; three microphone acoustic signals; two speed channels; and the torque and position of the handwheel. Data acquisitions were made under

different settings: three worm-gear interface configuration; hands on or off the wheel; vehicle speed (constant speed of 10, 15, 20, 30 km/h, or accelerating from 0 to 30 km/h); and road surface (smooth flat asphalt, stripes or cobblestones). The objective was to identify the different kinds of road surfaces where the car travels during the experiments. Testing of the system considered small training datasets augmented using two oversampling methods: Synthetic Minority Oversampling Technique (SMOTE) and Generative Adversarial Network Synthesis for Oversampling (GANSO). It is demonstrated the capabilities of data augmentation to reduce the classification requirements for the size of training datasets in the proposed application.

A Recurrent Neural Network for Prediction of the Economic and Financial Indicators in Context of COVID-19 Pandemic

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

Abstract - A recurrent neural network (RNN) was developed to explore the impact of the COVID-19 pandemic on the stochastic macroeconomic and financial indicators of the yield spread (Spread), crude oil prices (Oil), recession (USrec), and two stock market indices of the volatility index (VIX) and Wilshire 5000 total market index (Wil5000). A time-series dataset was obtained from the Federal Reserve Bank of St. Louis (for Jan/2/1990 - Feb/2/2022). For each indicator, separately, the dataset was partitioned into "before" and "during" the pandemic using the dataset.s breakpoint (transition block) established based on the indicator behavior. The results revealed: (a) VIX was explained by Wil5000, Spread, Oil, and USrec more accurately before the pandemic, indicating that other observed and unobserved factors arising from the COVID-19 pandemic would affect the VIX more than the macroeconomic and financial indicators, (b) USrec is predicted less accurately compared to other indicators during the pandemic, which shows the sensitivity of this indicator to health and geopolitical challenges, and (c) effects of the indicators on the bond market diminished during the pandemic. The sensitivity analysis suggests that the results remain highly robust to the changes in the number of records chosen for the different test sets.

New Version of Fast Automatic Determination of Cluster Numbers for High Dimensional Big Data

Zohreh Safari
Department of Computer Science, Elizabeth City State University, Elizabeth City, North Carolina, USA

Abstract - The main issue in clustering algorithms is how to efficiently define number of clusters automatically. Considering both the quality of clustering and efficiency of clustering algorithm during determination of number of clusters can be a trade off that was our main purpose to overcome with. Successfully, In our approach the best number of clusters for a large data set of high dimensional data automatically would be determined with respect to clustering quality and efficiency. We carried out experimental studies on our five previous data sets [43] and four new larger ones by which we found that our procedure has the flexibility of choosing different criteria to determine the optimal K under each of them. The procedure takes the advantages of Bisecting K-Means algorithm, and indicated higher efficiency compared with Ray & Turi method when it comes to find the best number of clusters.

Leveraging K-hop based Graph for the Staffing Recommender System with Parametric Geolocation

Yan Wang, Amar Henni, Azade Fotouhi, Leopold Mvondo Ze
Sogeti Innovation Labs, Capgemini, Meudon, France

Abstract - The graph-based methods in the staffing Recommender Systems (RecSys) can make a more accurate recommendation by leveraging the relationships. In order to discover implicit relationships, a Knowledge Graph (KG) which correlates Curriculum Vitae (CV), competence, jobs can be enriched with a large amount of semantics. However, the previous study is strictly based on the relationships of competences in both CVs and jobs. It is interesting to propose an approach to predict the jobs in which the competencies are not explicitly described in the CVs. To achieve this approach, it is necessary to estimate the distances between different competences. Moreover, the estimation of the transportation time has a key role in providing better recommended jobs as it can improve the quality of candidates working conditions. To alleviate the above issues, at first, a pipeline is reused to extract the competence keywords from CVs and jobs, and a normalized system is developed to enrich their semantics. We propose two transformer models - Sentence Bidirectional Encoder Representations from Transformers (SBERT) and Bidirectional and Auto-Regressive Transformers (BART) for the similarity measurements between the semantics of competences. And then we propose a k-hop based causality method to predict the potential jobs and a method of parametric geolocation to transform the transportation time into proximity index. At last, we apply the weight average method to calculate the final matching score for the recommended jobs. The experiment is based on Normalized Discounted Cumulative Gain (NDCG).

A Study on the Performance Comparison of Five Popular Machine Learning Models Applied for Loan Risk Prediction

Ihsan Said, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract - The application of machine learning algorithms resulted in the development of predictive tools trained on data routinely obtained from consumers' financial behavior and used to predict credit card defaults. Logistic Regression, Support Vector Machine, Gradient Boosting, AdaBoost, and Random Forest were used to predict credit card payment defaults to improve and reduce financial institutions' risks. All five models were applied to the same dataset. The original dataset has been supplemented in order to increase the accuracy of the model as well as the metrics used for model performance evaluation. Aside from being useful in general research, the predictions made by machine learning algorithms are particularly useful in research involving human participants, such as banking tools, and in fields where data collection is too expensive. The study established a solid foundation for bridging the gap between insufficient computer technology-based tools and the use of machine learning algorithms to improve the performance and accuracy of credit card default prediction while also saving time and money in traditional banking tools, which was previously lacking.

Creating a Dataset used for Applying Machine Learning Technique to Accurately Forecast the Energy Cost in Home-Based Small Businesses

Edwin Arrey Agbor, Yanzhen Qu
Colorado Technical University, Colorado Springs, Colorado, USA

Abstract - A systematic approach to creating datasets for applying machine learning techniques to forecast the energy cost in home-based small businesses accurately remains a primary requirement in this practice. However, a ready-to-use dataset remains a significant challenge when performing energy cost forecasting in home-based

small businesses. Unfortunately, much research has not been conducted to address this issue directly despite its recurrence in real-world applications. The CRMDV approach proposed in this study is designed to provide a framework for creating a ready-to-use time series dataset that can be leveraged via machine learning techniques to forecast the energy cost in home-based small businesses accurately.

Moving RDBMS to NoSQL Paradigms

Hiwa Ali Faraj
Information Technology, University of Sulaimani, Sulaymaniyah, Iraq

Abstract - Current business requirements and application infrastructure paved the way toward not only SQL (NoSQL) paradigms. The business requirements impose on information technology that brings a new atmosphere for the apps that require handling massive amounts of heterogeneous data types in the limited time and provide Highly Available Transaction (HAT) requirements. Each of Relational Database Management System (RDBMS) and NoSQL models support a group of features that bring certain functionalities to utilize business requirements. The key features to determine a specific data model is technology capabilities to achieve business requirements. Due to the shortages of Atomicity, Consistency, Isolation, and Durability (ACID) transaction, relational databases are unable to manage the atmosphere's requirements, then, Brewer's theorem, data consistency, data availability and partition-tolerance (CAP) emerge an alternative to the modern business requirements. Traditional relational databases still support the applications that require to enforce relational transactions protocol, restrictions, and constraints. In parallel with both models, polyglot persistence has come out, coexisting with both models in the single web server or an enterprise server. This paper concentrates on the roles of various factors of moving toward NoSQL, such as handling big data, heterogeneous data types, Internet of Things (IoT), highly available transactions requirements, eschew ACID constraints, high performance, software-as-a-service business, dynamic modeling schema, and data scaling. Moreover, an overview of nature and modeling of both relational and non-relational databases, limitation of relational database, and reasoning to move to NoSQL is presented.

Generating Random XML Files Using DGL

Peter M. Maurer, John Carbone
Department of Computer Science, Baylor University, Waco, Texas, USA

Abstract - DGL is a language for specifying and generating random data for testing and simulation. Output can take many forms and can be placed in files, mysql databases and C++ internal variables. In many cases, data files containing raw input data are created. The XML feature of DGL simplifies the generation of XML output. The XML data can be stored in an XML file or can be streamed into another program. The XML feature can generate XML files, DTD files or both simultaneously. Several new production types have been implemented. These new production types assist the user in creating syntactically correct XML data.

Statistical Relationship between Internet Data Indicators and the Brazilian Stock Exchange

Kethlyn Campos Silva, Lucas Jose de Faria, Deborah Silva Alves Fernandes,
Marcio Giovane Cunha Fernandes, Afonso Ueslei Fonseca, Fabrizzio Soares Alves de Melo Nunes Soares
EMC, Universidade Federal de Goias, Goiania, Brazil;
Informatica, Instituto Federal Goiania, Goiania, Brazil;
INF, Universidade Federal de Goias, Goiania, Brazil;
CCET, Universidade Estadual de Goias, Goiania, Brazil

Abstract - This work presents a statistical analysis between indicators obtained from internet data in Portuguese - news sentiment and Google Trends - and data on the Brazilian stock market through Spearman's rank correlation coefficient. The methodology used to collect, pre-process and obtain each indicator is detailed. Data from the years 2019 to 2021 were obtained. For the sentiment analysis of the news, a CNN model (Convolutional Neural Network) was adopted, which obtained an F1-score of 96%. As a result, some interesting correlations were obtained, among which, an inverse correlation characterized as "moderate" (according to the Cohen scale) between news sentiment and adjusted closing price in 2019; between search volume and closing price, a negative and "very large" and a positive and "large" correlation between trade volume and search volume. In both 2020 and 2021, negative coefficients defined as "large" were found, taking into account the closing price and trading volume.

Exploring Attribute Selection and Classification Methods for Predicting Heart Disease

Richard Jacob Stupka, Trinadh Chowdary Chigurupati, Devin Jackson, Demontae Moore
College of Computing & Software Engineering, Kennesaw State University, Georgia, USA

Abstract - Machine learning affords the opportunity to analyze enormous data sets and extract patterns and correlations that can.t be seen with manual examination. These classifiers are either well suited to certain data set types or more appropriate for other types. In our paper we will explore the importance of individual attributes in a large numeric data set on the accuracy of various classifiers, as well as ensemble methods applied of those various classifiers.

Multi-user Searchable Attribute Based Encryption for Outsourced Big Data

Kyle Astudillo, Mahdi Ebrahimi, Adam Kaplan
Computer Science Department, California State University, Northridge, California, USA

Abstract - Petabytes of data are generated every minute, and this creates a corresponding demand for servers which ingest, process, store, and maintain this data. The scale of data has become sufficiently large that normal file access and processing present different challenges than in prior decades. In the Web 2.0 era, data owners commonly stored and shared data using on-premise servers. In the past decade, cloud computing has quickly become the dominant hosting paradigm, as it offers data owners the benefits of cost-savings, flexible scaling, and ease of administration. However, cloud computing presents a myriad of security challenges. Attribute based encryption has been key to tackling these challenges. We present a method to allow users to maintain their own secure data on a shared system, while eliminating the time penalty and single pointof-failure associated with traditional authorization techniques. This is implemented as a Representational State Transfer API, removing the burden of managing auxiliary files or encryption libraries when uploading documents. This system also supports data encrypted at rest, removing concerns of administrator access to unencrypted contents in storage. We leverage

key policy attribute-based encryption (KP-ABE) to store this data, search its contents, and receive/decrypt this data. We further demonstrate that KP-ABE can be implemented as part of a distributed or centralized authorization method. We compare KP-ABE against AES cryptography, and present system metrics from its execution on representative data.

Using Reinforcement Learning with Transfer Learning to Overcome Smart Building Cold Start

Mikhail Genkin, Jenn J. McArthur
Department of Architectural Science, Toronto Metropolitan University, Toronto, Canada

Abstract - Smart buildings aim to optimize energy consumption by applying artificial intelligent algorithms. When a smart building is commissioned there is no historical data that could be used to train these algorithms. On-line Reinforcement Learning (RL) algorithms have shown significant promise, but their deployment carries a significant risk, because as the RL agent initially explores it's action space it could cause significant discomfort to the building residents. In this paper we present ReLBOT, a new technique that uses transfer learning in conjunction with deep RL to transfer knowledge from an existing, optimized smart building, to the newly commissioning building, to reduce the adverse impact of the reinforcement learning agent's warm-up period. We demonstrate improvements of up to 2.5 times in the duration, and up to 250 times in prediction variance for the reinforcement learning agent's warm-up period.

Cyber-Physical System Based Data Mining and Processing Toward Autonomous Agricultural

Abdulrahman M. Abdulghani, Mokhles M. Abdulghani, Wilbur L. Walters, Khalid H. Abed
Department of Electrical & Computer Engineering and Computer Science,
Jackson State University, Jackson, MS, USA

Abstract - Recent advancements in the internet and the proliferation of sensing devices have made it possible to deploy and communicate heterogeneous and bridge data in a variety of systems that link physical objects to the real world. This breakthrough also provides a lot of benefits for the farming industry, including improved resource leadership and human workforce. With objective evidence gleaned by sensors with the goal of increasing production and durability, principal benefits emerge. This type of automated and data-driven farm management relies on data to boost efficiency while reducing resource waste and environmental contamination. Smart farming, along with automated alternatives that use Artificial Intelligence (AI) approaches, lays the foundation for future food production. This paper proposed an Autonomous Agricultural Cyber-Physical System (AA-CPS) as a framework to predict the precise crop that fits the farm based on the soil and weather data. The data is collected by sensory technology, associated data mining techniques, and autonomous tractors in the field leveraging the best recommendation about what crop to be grown on any farm around the world.

Dynamic Text Modeling and Categorization Framework based on Semantics Extraction and Similarity Checking

Gaukhar Dauzhan, Alisher Iskakov, Miras Kenzhegaliyev, Dinh-Mao Bui
Computer Science Department, School of Engineering and Digital Sciences,
Nazarbayev University, Astana, Kazakhstan

Abstract - Semantics extraction is necessary to obtain relevant knowledge from general data. Recently, many researchers have developed numerous techniques to retrieve the semantics and use this information for categorization purposes. Primarily, the methodologies can be grouped into two main approaches: knowledge-based and corpus-based approaches. In this work, we studied how to achieve the same purpose using semantics extraction and matching methods. In order to prove the effectiveness of our approach, we utilized the problem of news article categorization as a use case. In fact, we developed a dynamic tool that adopts the proposed technique to crawl and classify articles from many well-known newspapers. Furthermore, the topic of articles can be grouped by their meta-information or by user-defined categories flawlessly. For generalization, the Universal Sentence Encoder model has been used for distributed semantic text representation and trained on a large processed Wikipedia corpus to achieve better accuracy. Besides, there is a possibility to enhance the results further using multi-label categorization or optimizing the meta-categories from the training dataset shortly.

Structuring Context Entities for Knowledge Discovery

Nour Matta, Nada Matta, Agata Marcante, Nicolas Declercq
LIST3N, University of Technology of Troyes, France;
Namkin, Troyes, France

Abstract - Understanding and representing the meaning of words requires a specific presentation of their context. In previous work, we introduced a context awareness approach that aims to extract context entities to represent the context of a text and provide a reference for knowledge discovery from textual data. This paper will provide a semantic network to structure the different relationships between the identified context entities.

Understanding the Variables Affecting Decision-making using Semantic Study

Imen Gmach, Nada Chaabane, Amal Abdulrazzaq AlFryyan, Nahla O.A.Mustafa, Nabila Alomari, Aya Hellal,
Abeer Rafi Abdullah, Reem Matahen, Sajeda Qasim Aldiabat, Enas Kanan
Saudi Arabia

Abstract - This article assesses the current condition of strategic decision-making. It's a matter of researching and comprehending the factors that influence decision-making. Our empirical study was based on qualitative research using TROPES software for semantic analysis. The findings describe the various factors that influence the decision-making process, with a focus on the impact of the recommendation on strategic decision-making. We have already mentioned that strategic decisions are made based on a collection of data and consultation. The recommendation is crucial in this decision.

Hey There, You Are Behaving Like a Troll: Using K-Means Algorithm to Cluster Trolling Traits

Sariah Lopez-Fierro, Carlos Chiriboga-Calderon, Jose Lopez-Fierro, Ruben Pacheco-Villamar
Departamento de Tecnologia, Soluciones Wandarina S. A., Guayaquil, Ecuador;
Universidad E-Espiritu Santo, Guayaquil, Ecuador

Abstract - This work presents the results obtained after targeting down and following the Ecuadorian Presidential 2021 Election trolls. To achieve this, new tweets after the Elections were downloaded, and "trolling behavior" was spotted and measured, as an attempt to increase the accuracy of a K-Means Algorithm, to cluster Twitter trolls.

Proposal of a Multi-purpose Marketing Data Analysis Program Design Methodology

Donggyu Choi, Jongwook Jang
Institute of Smart IT, Dong-eui University, Busan, Republic of Korea;
Department of Computer Engineering, Dong-eui University, Busan, Republic of Korea

Abstract - Marketing is a strategy for the sale of goods, and shopping malls and vendors consider it important. Companies set up strategies through marketing departments or professional companies and collect related data directly, but data analysis eventually requires collecting data directly, even if some tools are used, and most of the automatically collected data is only normalized data. In this paper, we present the necessity and method of performing a program that can be used fundamentally by automatically performing the functions of existing data analysis tools. As a method, crawling and deep learning techniques are utilized, and the results of simple analysis for automatic data collection and evaluation are provided. Start-ups and small businesses can reduce the need for primary preparation and basic costs and efforts to use the data analysis tools used in marketing.

CSCI-RTNA:
RESEARCH TRACK ON SOCIAL NETWORK ANALYSIS, SOCIAL MEDIA, & MINING

Social Network Analysis of Biomedical Research Growth at a Primarily Undergraduate Institution

Amrina Ferdous, Diane Smith, Tracy Yarnell, Julia Thom Oxford
Center of Biomedical Research Excellence, Boise State University, Boise, Idaho, USA;
Computing Program, Center of Biomedical Research Excellence, Boise State University, Boise, Idaho, USA

Abstract - The National Institutes of Health Institutional Development Award Programs support the establishment and growth of biomedical research infrastructure in states that receive a low level of federal funding for biomedical research. The purpose of this investigation was to analyze the growth in research productivity over time. This program fostered an environment in which a biomedical research program could be developed and allowed to grow at Boise State University, a primarily undergraduate institution. The growth of the biomedical research community can be visualized through social network analysis.

Knowledge-Infused Dynamic Embedding for Predicting the Severity of Suicidal Ideation in Social Media

David M. Lee, Hamidreza Moradi
The University of Mississippi Medical Center, Mississippi, USA

Abstract - Depression accounts for an increased number of suicides in the United States. To identify those at risk and notify the authorities, social media can play an invaluable role. Posts made on social media can be collected to train predictive models capable of identifying individuals at risk. However, studies with an unsupervised method of data labeling are susceptible to misidentify individuals presenting signs of depression from those who provide support. Moreover, these studies tend to classify depression as a binary outcome, not considering its severity levels. This necessitates the use of accurate and efficient modeling techniques for the existing scarce but professionally annotated datasets. While pre-trained embeddings proved to be an efficient method of text representation, not all may fully encode emotions and sentiment polarities present in mental health related posts. In this work, we propose the use of a dynamic embedding infused with emotion and polarity knowledge for a more accurate representation of depression severity and suicidal ideation. The emotion, polarity, and context-aware generated embeddings are then utilized by bidirectional RNN (GRU and LSTM) model to provide the most accurate predictions. The results show that knowledge-infused dynamic embeddings with bidirectional LSTMs leads to an 8% increase in the overall AUC with improved prediction accuracy of suicidal ideations at its highest level.

Sentiment Analysis and Topic Modeling on COVID-19 Vaccines using Twitter Data

Daniel Ojeda, Eric Landaverde, Daehan Kwak
Department of Computer Science and Technology, Kean University, Union, New Jersey, USA

Abstract - Ever since the public began quarantining due to Covid-19 in 2020, people have been waiting for life to go back to normal. Until April 2021, "normal" activities were only allowed by being socially distant or wearing a mask. However, in early 2021, CDC announced that a vaccine would soon be released to the public. This announcement seemed to be good news for some, but for others, this was another obstacle on the way to normalcy. People have shown pushback against the Covid-19 vaccine due to the uncertainty of its effectiveness coupled with its potential side effects. Vaccine hesitancy has a negative impact on society and poses a real threat to public health. This is an important issue worldwide, and questions arise about how the general public feels about getting vaccinated. Therefore, the purpose of this study is to analyze the sentiment of Twitter users towards vaccination, specifically the Covid-19 vaccine. This study collects data through Twitter IDs to pick up on hashtags and keywords relating to the Covid-19 vaccine via the Twitter API and Tweepy. Tweets are put through a sentiment analysis tool to get a general idea of the sentiment. Furthermore, topic modeling is used to understand the topics discussed when mentioning the Covid-19 vaccine. By analyzing the sentiment towards the Covid-19 vaccine, we hope to provide the first step towards mitigating the risk associated with vaccine hesitancy.

A k-Way Partitioning Framework for Compression on Social Networks

Chandra N Sekharan, Sridhar Radhakrishnan, Michael R. Nelson
Department of Computing Sciences, Texas A&M University Corpus Christi, Texas, USA;
School of Computer Science, University of Oklahoma, Norman, Oklahoma, USA

Abstract - Social networks have been studied in the context of matrices or graph theory for efficient representation, and manipulation. Knowledge gleaned from these graphs are useful to better coordinate events, advertise, and in recommendation of friends or games. Many social networks represented as graphs or matrices are large containing 100's of millions or even billions of entries. Hence it has become important to store them efficiently using data compression. The authors have used different kinds of compressed data structures such as quad-tree, array of binary trees (ABT), and differential, compressed binary trees (DCBT) in previous research for representing massive social networks and performing various queries [15]-[17]. In this paper, we provide a k-way partitioning framework that generalizes the array of binary trees structure by providing a combinatorial basis that is shown to lead to substantial improvements in compressing social network graphs.

Parenthood Mining Using Hashtag Social Network Mining Approach

Xiaowen Wang, Mourad Oussalah, Mika Niemela, Tiina Ristikari
Center of Computer Vision and Signals Analysis, University of Oulu, Finland;
Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Finland;
Center for Life-long Health Research Unit, University of Oulu, Finland;
Itla Children's Foundation, ITLA, Helsinki, Finland

Abstract - Low birth phenomenon is a global trend with an important economic and societal consequences that stress global and holistic initiatives in western countries. This paper attempts to contribute to this issue by developing an original approach to mine parenthood patterns from Twitter social media platform. Relevant hashtags are employed to develop a social network-based approach to distinguish relevant communities. The three relevant online communities were identified. We conducted a topical based analysis to comprehend the discussion trend and sentiment analysis to further monitor issues of parenthood concerns. Examples of concerns raised by the community analysis include anxiety in autism parenting, stress in single parenting, college loans, and debts in pandemic time.

An Analysis of Public Sentiment around Face Mask During COVID-19

Eric Landaverde, Daniel Ojeda, Daehan Kwak
Department of Computer Science and Technology, Kean University, Union, New Jersey, USA

Abstract - The impact of the covid-19 pandemic continues to be felt as we face the aftermath the pandemic has caused globally in terms of loss of life, economic damage, civil unrest, and political instability. Mixed messaging and partisanship from leaders worldwide have made proposed solutions a divisive political stance. The polarization of covid-19 and the sentiments around proposed solutions warrant investigation; thus, this study's objective is to examine the sentiments and topics in global public discourse around one of the proposed measures to contain the spread of covid-19; mask usage. This study looks to examine the sentiments around the global discourse of Twitter users on mask usage to investigate the relationship between public sentiment, fluctuations in covid-19 cases, and topics of discussion and influential events through sentiment analysis and topic modeling. Sentiment analysis and Topic Modeling are conducted on the dataset after it is preprocessed to analyze the abstract topics discussed over the time period. The results are then visualized to highlight key topics present in the discussion around mask usage and the accompanying sentiment that is then examined in relation to global events and case numbers during the same time. The goal is to understand how the nature of global discourse can impact the perceptions of mask usage and how that relates to behavior that continues the spread of the pandemic.

PSR: Probabilistic Serendipitous Recommendations

Hyeseong Park, Kyung Whan Oh, Uran Oh
Department of Computer Engineering, Sogang University, Seoul, South Korea;
Department of Computer Engineering, Ewha Womans University, Seoul, South Korea

Abstract - Recent recommender systems studies have started focusing on accelerating serendipity to offer users diverse perspectives. However, recommending serendipitous items is known to decrease users. satisfaction due to low recommendation accuracy. In this paper, we propose a probabilistic serendipitous recommendations (PSR) model that aims to discover serendipitous items through the fully-connected combinations of unrelated user-provided information. The proposed model provides an unexpected result, while user's answers are maintained not to decrease recommendation accuracy. An online user study with 40 participants revealed that the proposed PSR model provides serendipitous items by using the intersection of usefulness and unexpectedness from the recommended results.

A Smart Trust Based Approach for Sybil Node Detection in Online Social Networks

Ranjan Kumar, Yashwanth Dasari, Prabhat Kumar, Akash Sinha
Department of Computer Science and Engineering, National Institute of Technology Patna, India

Abstract - Forged identities (Sybil) have the potential to damage the trust of the platform in the long run by doing malicious works. Owing to the severity of the threats posed by these Sybil nodes, various graph based mechanisms have been proposed to identify them. However, these approaches are more suitable for the classical scenario which considers that it is difficult to establish the linkage between sybil and honest nodes. Contrary to this assumption, the modern scenario assumes that such linkage can be easily established using simple strategies. The work proposed in this paper makes use of a trust based approach to detect sybil nodes in an online social network. The proposed approach is more suitable for the modern scenarios since the strategies to link sybil and honest nodes are based upon the concept of the trustworthiness of the nodes. The proposed approach calculates aggregate trust value of the nodes which is further compared with a benchmark threshold value for deciding whether the node is Sybil or Benign. The proposed model has been theoretically analyzed and the obtained results clearly exhibit the efficiency of the proposed work over the existing approaches.

A Study on Public Sentiment of Comments from COVID-19-related Videos

Hualong Yu, Chao Xu, Guang Yu
School of Management, Harbin Institute of Technology, Harbin, China

Abstract - To provide decision-making references for the optimization of media release strategies and the correct guidance of public sentiment in the era of COVID-19, this paper selects a series of videos created by "Lin Chen Hearing", a Media practitioner who attracted significant attention during the lockdown of Wuhan, China. We establish a sample dataset for video communication and public comments and use the natural language processing technology to build a sentiment classification model. We analyze the impact of video information on public sentiment, hoping to provide s reference for the formulation of media release strategies, agenda setting and public opinion control in the era of COVID-19.

A Study on Personal Identifiable Information Exposure on the Internet

Ningning Wu, Robinson Tamilselvan, Talha Tayyab
Department of Information Science, University of Arkansas at Little Rock, Little Rock, Arkansas, USA

Abstract - Personal Identifiable Information (PII) is any information that permits the identity of an individual to be directly or indirectly inferred. It should be protected against random access. This paper studies the extent of PII exposure on the Internet. It is hoped that the results of this study can help raise the Internet users. awareness on privacy protection.

Predicting the Trends of Stock Price by CNN-BiLSTM-AM and Fuzzy Twin Support Vector Machine from Social Networks

Pei-Yi Hao
Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan, ROC

Abstract - Because stock prices are affected by so many factors, it is difficult to predict with a simple model. Deep learning has excellent feature learning ability. Support vector machine has excellent generalization ability. This research will combine the advantages of both models. This study will propose a hybrid deep model, which is composed of convolutional neural network (CNN), bidirectional long short-term memory (BiLSTM) and attention mechanism (AM), to automatically learn important features. CNN is used to extract the position-invariant features of the input data, BiLSTM is used to extract long-term dependent features, and AM is used to capture the influence of different time feature states on the stock closing price in the past to improve the accuracy of prediction. Next, this study feeds the acquired features to the fuzzy twin support vector machine to establish the best stock price prediction model. Compared with the classic stock price prediction model, the prediction accuracy rate of the proposed method is significantly better than the advanced stock price prediction models.

Analyzing Climate Change Discussions on Reddit

Mohammad S. Parsa, Haoqi Shi, Yihao Xu, Aaron Yim, Yaolun Yin, Lukasz Golab
University of Waterloo, Waterloo, Canada

Abstract - Climate action is one of the United Nations Sustainable Development Goals. We contribute to this effort by analyzing climate change topics on the Reddit social curation platform, which contains over 100,000 discussion communities. We leverage the community-oriented structure of Reddit to answer two questions: 1) Which online communities discuss climate change? and 2) What do they discuss? In addition to the presence of groups of activists and skeptics, we find climate change discussions on question-answering and opinion-sharing communities, as well as in a community for teenagers. We then run a topic modeling algorithm on the five largest climate change communities. We find concerns about skeptics, political policies doubting the severity of climate change and the consequences of climate change, as well as discussions of climate actions that can be taken by individuals. Finally, an analysis of the URLs mentioned in discussions highlights influential content creators and underscores the need to avoid misinformation on these sources.

Security and Privacy Challenges in Voice-based Social Networks

Youna Jung, Dylan Palmer, Mark Shelton
Computer and Information Sciences, Virginia Military Institute, Lexington, Virginia, USA

Abstract - Voice-based social networks (VBSNs) like Clubhouse have explosively grown since they were introduced. During the pandemic, VBSNs allowed people to communicate with others via audio while many were isolated at home. Their services successfully met the need for social interactions, and they quickly established themselves as one of the major social network service providers alongside Facebook and Twitter. The unique features of VBSNs open the possibilities for diverse uses of them. As one of the application areas, we focus on emergency and disaster management. Using VBSNs, we can address the limitations of existing disaster response using online social networks. For the practical use of VBSNs during disasters, we, however, must resolve the security and privacy issues. To this end, in this paper, we define the characteristics of VBSNs and identify the security and privacy issues to be considered. The result of the user survey and threat scenarios show us that there is an urgent need for research on this topic. Then, we present potential solutions and challenges.

An Experiment on Dynamic Extraction of the Human-Relationships on the Web Using Closeness Centrality

Ayaka Ichikawa, Ryo Onuma, Hiroki Nakayama, Hiroaki Kaminaga, Youzou Miyadera, Shoichi Nakamura
Fukushima University, Japan;
Tsuda University, Japan;
Yamagata University, Japan;
Tokyo Gakugei University, Japan

Abstract - Opportunities to conduct advanced research and analysis using the Web have been increasing. In these work, accurate understanding of the human-relationships involved with target of analysis contributes to its results. However, scope in which human relationships can be grasped is limited due to bloating the Web. In our project, we have aimed to develop methods that enable to observe human relationships expressed by descriptions on the Web from various perspectives. In this paper, we mainly describe methods for dynamically extracting the human relationships on the Web using closeness centricity. Moreover, we describe an experiment that dynamically generates the humanrelationship networks in accordance with people who the observer focuses and discuss the effectiveness of our methods based on its results.

Using Topological Analysis to Investigate Ture and False Information Diffusion

Amin Riazi, Yingfeng Wang
Department of Management, University of Tennessee at Chattanooga, Tennessee, USA;
Department of Computer Science and Engineering, University of Tennessee at Chattanooga, Tennessee, USA

Abstract - Social media provides an online platform for users to spread information including true and false stories. Both human users and social bots play critical roles in diffusing true and false information. This study investigates the user behavior patterns in Twitter information diffusion by analyzing the topological properties of retweet networks. Our results show humans and bots have different behavior patterns in true and false information diffusion. Our finding suggests the roles of human and bot users in spreading true and false stories are worth further investigation.

CSCI-RTCW:
RESEARCH TRACK ON CYBER WARFARE, CYBER DEFENSE, & CYBER SECURITY

Nano-Resolution Visual Identifiers Enable Secure Monitoring in Next-Generation Cyber-Physical Systems

Hao Wang, Xiwen Chen, Abolfazl Razi, Michael Kozicki, Rahul Amin, Mark Manfredo
School of Computing, Clemson University, South Carolina, USA;
School of Electrical, Computer and Energy Engineering, Arizona State University, Arizona, USA;
MIT Lincoln Laboratory, Massachusetts Institute of Technology, Massachusetts, USA;
Morrison School of Agribusiness, Arizona State University, Arizona, USA

Abstract – Today's supply chains heavily rely on cyber-physical systems such as intelligent transportation, online shopping, and E-commerce. It is advantageous to track goods in real-time by web-based registration and authentication of products after any substantial change or relocation. Despite recent advantages in technology-based tracking systems, most supply chains still rely on plainly printed tags such as barcodes and Quick Response (QR) codes for tracking purposes. Although affordable and efficient, these tags convey no security against counterfeit and cloning attacks, raising privacy concerns. It is a critical matter since a few security breaches in merchandise databases in recent years has caused crucial social and economic impacts such as identity loss, social panic, and loss of trust in the community. This paper considers an end-to-end system using dendrites as nano-resolution visual identifiers to secure supply chains. Dendrites are formed by generating fractal metallic patterns on transparent substrates through an electrochemical process, which can be used as secure identifiers due to their natural randomness, high entropy, and unclonable features. The proposed framework compromises the back-end program for identification and authentication, a web-based application for mobile devices, and a cloud database. We review architectural design, dendrite operational phases (personalization, registration, inspection), a lightweight identification method based on 2D graph-matching, and a deep 3D image authentication method based on Digital Holography (DH). A two-step search is proposed to make the system scalable by limiting the search space to samples with high similarity scores in a lower-dimensional space. We conclude by presenting our solution to make dendrites secure against adversarial attacks.

SREP+SAST: A Comparison of Tools for Reverse Engineering Machine Code to Detect Cybersecurity Vulnerabilities in Binary Executables

Thomas Ryan Devine, Maximillian Campbell, Mallory Anderson, Dale Dzielski
Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, West Virginia, USA

Abstract - Cybersecurity of mission components is vital for safety-conscious industries. This paper examines the effectiveness of using existing software reverse engineering products (SREPs) to first reverse engineer binary executables and then detect cybersecurity vulnerabilities through static application security testing (SAST) on the output (SREP+SAST). We analyzed 2.3 million lines of code from test suites and open source software. Results showed that SREP+SAST revealed 48% of the 20,129 vulnerabilities detected by SAST on the source code, including 35% of high-risk vulnerabilities (HRVs). We introduce Smaug, a novel, open source, customized SAST ruleset optimized for HRV detection. Smaug boosted our our best-performing combo by 20% and increased the HRV detection rate to 55%. For further validation, we traced vulnerabilities in source code linked to specific CVEs and found that Smaug improved the detection rate for CVE-specific vulnerabilities by 67% for our bestperforming combo. Our methods and code are publicly available and we believe that further improving SREP+SAST could lead to enhanced security for systems that depend on COTS technologies.

Classifying RDP Remote Attacks on User Interfaces to Industrial Control Systems

Ryan R. Ramirez, Thuy D. Nguyen, Neil C. Rowe, Joseph T. Meier
Department of Computer Science, Naval Postgraduate School, Monterey, California, USA

Abstract - The Microsoft Remote Desktop Protocol (RDP) is popular for remote access, but its use for industrial control systems (ICSs) is risky because of their many vulnerabilities. Recognizing RDP attacks is also difficult because most RDP traffic is encrypted, and ICS traffic has many differences from non-ICS traffic. Our experiments obtained data from a hardened powergrid honeypot to characterize real RDP attacks on ICSs by malicious signatures, Windows event logs, and traffic metadata. Severity of malicious traffic varied widely and require novel labeling methods. This work can provide early warning to defenders about RDP attacks against ICS systems.

FDIA Detection Methods on a Navy Smart Grid AMI Data Set using Autoencoder Neural Networks: A Case Study

Preston Musgrave, Preetha Thulasiraman
Naval Postgraduate School, Monterey, California, USA

Abstract - In 2019, the Naval Facilities Engineering Command (NAVFAC) deployed its first smart grid infrastructure in Norfolk, VA, enabling shore commands to meet energy goals set by the Secretary of the Navy. However, with increased functionality and control comes increased vulnerability to malicious cyber activity. In this paper we aim to address anomaly detection in the Navy smart grid using autoencoder neural networks. Our specific focus is on anomalies in the sensor data originating within the advanced metering infrastructure (AMI) of the Navy smart grid. An efficient autoencoder model was developed through benchmarking experiments with open source Modbus data sets. We then used NAVFAC provided AMI data to train the autoencoder model to detect 14 different false data injection attacks (FDIA). Twenty six AMI data features were manipulated to simulate FDIAs. Accuracy, precision and recall scores were used to quantify model performance. We compare the performance of our autoencoder with that of a deeper neural network model to show that a smaller, leaner model can be equally effective in classifying and identifying FDIAs. Our experiments showed that our autoencoder model achieved an average 90%-95% on precision, accuracy and recall scores over multiple combinations of optimizers and activation functions. This work is a case study on the use of unsupervised machine learning methods for data anomaly detection on the Naval smart grid.

Enhancing Security in Communications between Unmanned Aerial Vehicles (UAVs)

Ilias Panagiotopoulos, George Dimitrakopoulos, Elena Politi
Department of Telematics and Informatics, Harokopio University of Athens, Athens, Greece

Abstract - Unmanned aerial vehicles (UAVs) are widely anticipated to fundamentally change the air transport including, among others, air-taxi operations, cargo transport, delivery of goods, delivery of medical equipment, inspection of infrastructures, police surveillance, and emergency services support. Intelligent transportation systems (ITS) and communication interactions are important components for the successful operation of UAVs. In the field of connectivity UAV networks it is very important for the aviation industry to manage communications among nearby UAVs and between UAVs and ground infrastructure units. The above cooperative communication systems bring the promise of improved air safety and optimized air traffic. For the successful deployment of UAV communications it is essential to make sure that "life-critical safety" information cannot be modified by external

or internal within the network attackers. In this basis, lack of security in UAV networks could be one of the key hindrances to the wide spread implementations of such vehicles. Moreover specific operational parameters (highly moving UAVs, frequently and fast changed connectivity) make the problem very novel and challenging. To address this need, we propose a new variant of the Diffie-Hellman key-exchange algorithm for secure and authenticated UAV communication interactions. Our analysis demonstrates the secure-efficiency of our scheme towards the deployment of UAV networks in the air transport area.

Development of an Autonomous Retesting Penetration Testing Paradigm

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

Abstract - This paper presents a paradigm for automating penetration testing for the purpose of retesting systems to ensure that previously detected issues have not been reintroduced. This approach, which is patterned off of path-based attack techniques, is described and an implementation of the proposed paradigm, using the Blackboard Architecture, is presented. The efficiencies and potential pathway to more automated testing provided by the proposed paradigm are also discussed.

Mapping Zero-Click Attack Behavior into MITRE ATT&CK Mobile: A Systematic Process

Awad Mussa, Zachary Daher, Brendan Martin, Connor Morgan
School of Computing and Analytics, Northern Kentucky University, Highland Heights, Kentucky, USA

Abstract - A zero-click attack is a sophisticated class of attacks where a smartphone device can be compromised remotely without any user interaction. Recently, there have been traces of successful zero-click attacks on iPhones by Pegasus, hacking spyware that can infect billions of phones running either iOS or Android operating systems. However, there is a lack of behavioral profile for a zero-click attack, which is critical for a response and prevention plan and security awareness. To that end, we aim to devise and apply a systematic process that leverages CISA-MITRE-ATT&CK-Mapping-Guide and MITRE ATT&CK framework for mapping zero-click cyber threat intelligence reports into the MITRE ATT&CK matrix. Based on the forensic data, we have created a zero-click attack behavioral profile that consists of 65 techniques and 13 different tactics. To validate our mapping results, we compared our mapping results with a MITRE ATT&CK analyst expert from the cybersecurity industry and we have found many similarities between the two mappings. The proposed process was not difficult to follow and apply by novice analysts. The created profile has the potential to provide knowledge for detection and protection. Also, the device process has the potential to assist novice team members to map cyber threat intelligence into the MITRE ATT&CK framework.

A Multiagent CyberBattleSim for RL Cyber Operation Agents

Thomas Kunz, Christian Fisher, James La Novara-Gsell, Christopher Nguyen, Li Li
Systems and Computer Engineering, Carleton University, Ottawa, Canada;
Defence Research and Development Canada, Ottawa, Canada

Abstract - Hardening cyber physical assets is both crucial and labor-intensive. Recently, Machine Learning (ML) in general and Reinforcement Learning RL) more specifically has shown great promise to automate tasks that otherwise would require significant human insight/intelligence. The development of autonomous RL agents requires a suitable training environment that allows us to quickly evaluate various alternatives, in particular how

to arrange training scenarios that pit attackers and defenders against each other. CyberBattleSim is a training environment that supports the training of red agents, i.e., attackers. We added the capability to train blue agents, i.e., defenders. The paper describes our changes and reports on the results we obtained when training blue agents, either in isolation or jointly with red agents. Our results show that training a blue agent does lead to stronger defenses against attacks. In particular, training a blue agent jointly with a red agent increases the blue agent's capability to thwart sophisticated red agents.

Analyzing Attacks On Client-Side Honeypots from Representative Malicious Web Sites

Brian A. Foley, Neil C. Rowe, Thuy D. Nguyen
Department of Computer Science, Naval Postgraduate School, Monterey, California, USA

Abstract - Client-side cyberattacks are becoming more common relative to server-side cyberattacks. This work tested the ability of the honeyclient software Thug to detect malicious or compromised servers that secretly download malicious files to clients, and tested its ability to classify these exploits. We tested Thug's analysis of delivered exploits in different configurations. Results on randomly generated Internet addresses found a low rate of maliciousness of 5.6%, and results on a blacklist of 83,667 suspicious Web sites found 163 unique malware files. Thug demonstrates the usefulness of client-side honeypots in analyzing harmful data presented by malicious Web sites.

Performance Evaluation of Partially Homomorphic Encryption Algorithms

George Dimitoglou, Carol Jim
Department of Computer Science and IT, Hood College, Frederick, Maryland, USA

Abstract - Homomorphic encryption allows the computation of encrypted data without decryption or key exchanges. In this work, we focus on the performance characteristics of three additively homomorphic algorithms, Paillier, Damgard-Jurik and ElGamal. Our main contribution is the run-time analysis and comparison of these algorithms under variable key lengths during encryption (including key generation) and homomorphic addition operations. There is extensive theoretical coverage of the cryptographic strength of these algorithms but limited performance benchmarking in the literature. By implementing and exercising these algorithms under various loads, we observe how the theoretical underpinnings of each algorithm's strength affect their speed. Our findings align with the expected performance penalties of complex key generation and encryption processes while demonstrating promising performance speeds on the homomorphic operations. The latter is of particular importance given that the computationally expensive key generation and encryption would be performed once or very few times in a production environment. At the same time, the homomorphic operations would be performed numerous times and with high frequency.

Ada-Thres: An Adaptive Thresholding Method to Mitigate the False Alarms

Jayesh Soni, Nagarajan Prabakar, Himanshu Upadhyay
Florida International University, Florida, USA

Abstract - In a wide variety of domains, the advanced intrusion detection system consists of a learning-based detection method and a signature-based analysis approach. Such a system scans the incoming data, performs the analytics on it by using an anomaly detection algorithm, and finally transfers the report of suspicious activity for

further analysis if found. The major problem of such a current system is the high false-positive rate (FPR), specifically in the case of a highly complex system with a large dataset. Such high FPRs, which are non-crucial, can easily overwhelm the user of the system and can further increase the likelihood of ignoring such indications. Therefore, mitigation approaches aim to develop a technique to reduce high FPR without losing any potential harmful threats. Thus, in this study, we develop an adaptive thresholding algorithm that can mitigate the issue of high FPR. The proposed algorithm applies three scoring mechanisms. They are Anomaly Pruning, Sequence Scoring, and Adaptive Thresholding. The model is trained on sequential data. Anomaly Pruning gives a score to an individual data point. It either rejects or accepts the data points to be considered for Sequence Scoring. This Sequence Scoring will give a score to an individual sequence. Finally, an Adaptive Thresholding is applied to the cumulative score of all the sequences to detect the anomalous nature of the analyzed data. Multiple experiments have been conducted using various optimizers to access our proposed approach. Using the proposed approach, we train a deep learning-based LSTM algorithm widely adopted for sequential data. Furthermore, we validate it with three different datasets of various sizes. From the experimental results, we infer that the proposed approach allows the model to train faster and reach the minimum loss at a faster rate.

Data Behind the Walls - An Advanced Architecture for Data Privacy Management

Amen Faridoon, M. Tahar Kechadi
School of Computer Science, University College Dublin, Dublin, Ireland;
Insight SFI Research Centre for Data Analytics, Ireland

Abstract - In today's highly connected society, we are constantly asked to provide personal information to retailers, voter surveys, medical professionals, and other data collection efforts. The collected data is stored in large data warehouses. Organisations and statistical agencies share and use this data to facilitate research in public health, economics, sociology, etc. However, this data contains sensitive information about individuals, which can result in identity theft, financial loss, stress and depression, embarrassment, abuse, etc. Therefore, one must ensure rigorous management of individuals. privacy. We propose, an advanced data privacy management architecture composed of three layers. The data management layer consists of de-identification and anonymisation, the access management layer for re-enforcing data access based on the concepts of Role-Based Access Control and the Chinese Wall Security Policy, and the roles layer for regulating different users. The proposed system architecture is validated on healthcare datasets.

When AI Facilitates Trust Violation: An Ethical Report on Deep Model Inversion Privacy Attack

Mahdi Khosravy, Kazuki Nakamura, Antoine Pasquali, Olaf Witkowski, Naoko Nitta, Noboru Babaguchi
Cross Labs, Cross Compass Ltd., Tokyo, Japan;
Tokyo University of Science, Tokyo, Japan;
Mukogawa Women's University, Hyogo, Japan;
Institute for Datability Science, Osaka University, Osaka, Japan;
Department of Management and Information Sciences, Fukui University of Technology, Fukui, Japan

Abstract - This article raises concerns about the considerable capability of artificial intelligence in boosting privacy violations and motivates the necessity of AI ethics. Despite all AI advantages like the efficiency and accuracy of recent techniques, and its positive effects on our life quality, when it comes to security and privacy the facilities with AI empowerment have been met with anxiousness and distrust in the public. This article is an ethical view of the AI role in a recent work wherein AI considerably facilitates privacy violation in a gray-box attack on a deep face recognition system. While the user identities. data is fully secured and just the recognition deep model is accessible, AI-boosted model inversion reveals the faces of the identities via high-accuracy

generated clones. An analytical and subjective evaluation of the generated face clones with and without AI integration in model inversion illustrates a big gap from non-clear noise face clones to crystal clear face clones which efficiently reveal the identity of a targeted user by their high-level naturalness, similarity, and recognizability amongst many users.

SMS Malware Detection: A Machine Learning Approach

Heather Broome, Yaju Shrestha, Nyzaireyus Harrison, Nick Rahimi
School of Computing Sciences & Computer Engineering, University of Southern Mississippi, Hattiesburg, Mississippi, USA

Abstract - In recent years, reliance on technology for communication has risen exponentially. Perhaps the most responsible component for this growth has been the Short Message Service (SMS), which supports text message exchange across devices. While SMS technology has provided unparalleled benefits for ease of communication, users with intentions to spread malicious software have been unrelenting in their use of it. The accessibility of chatbots and the advancement of conversational artificial intelligence (AI) have made SMS malware a critical threat for unsuspecting victims. Furthermore, conversational AI is rapidly improving. This has made the deep analysis of messages'properties - rather than reliance on recognizing suspicious language - imperative for building a reliable threat detection system. In this paper, we will first present the initial part of our research which used supervised machine learning (ML) to build a highly effective SMS malware detection program. This research was done by programming 5 machine learning algorithms to analyze a database containing 1,048,575 messages, each with 85 attributes. By integrating optimal attributes and algorithms, our final program can accurately detect malware in SMS messages with 94.6% accuracy. Next, we will analyze the effectiveness of different classifiers on datasets of 11 SMS malware families respectively.

Security Analytics Framework Validation based on Threat Intelligence

Viktor Sowinski-Mydlarz, Vassil Vassilev, Karim Ouazzane, Anthony Phipps
London Metropolitan University Cyber Security Research Centre, London, United Kingdom (UK)

Abstract - Logical analysis of the ontology of digital security in banking helps us to identify the possible entry points for illegal access. The threats described in the ontology are detected by Machine Learning engines. The theoretical analysis is validated by verifying the framework and Machine Learning algorithms. Intelligence Graphs (original term) which are adding the actions to knowledge graphs to form workflows, are a base for validation of the framework through simulated execution of the scenarios specified in them. The output is a method for analysing live network traffic data (machine learning algorithm) combined with semantic model to give a hybrid framework for threat intelligence in digital banking, leading to a complete threat detection platform. The model is validated using operation workflows, namely 12 scenarios of banking "journeys" under the duress of various threats. In this work we are presenting the validation of the framework by simulation of the banking operations and transactions stemming from the Ontology of Digital Banking used as a model of the banking infrastructure (assets, vulnerabilities and threats included). This validation is based on the use of the Intelligence Graphs to demonstrate the capability of the framework to deal with typical scenarios by detecting the threats through ML, identifying them by checking the situations in which they appear in the ontology of threats, selecting appropriate counteraction using the security rules and planning their execution as containerized services. For better understanding of the subject matter, the authors would like to refer the reader to a previous article on general framework we developed.

An Experimental Evaluation of a Privacy-Preserving Authentication and Authorization Scheme

Kai Wang, Tyler Phillips, Xukai Zou
Department of Computer Science, Georgia Southern University, Statesboro, Georgia, USA;
Department of Computer and Information Science,
Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, USA

Abstract - We conduct an experimental evaluation of a PrivacyPreserving Authentication and Authorization Scheme based on an earlier work. The scheme is flexible in its design and allows itself to be incorporated into various biometrics-based authentication systems. In this study we use face images as biometrics data in the authentication system but protect system users' privacy by utilizing the scheme's Biometric-Capsule based security mechanism. We also employ additional state-of-the-art deep-learning based face recognition methods to help achieve high accuracy of the system.

Implementation of an Enhanced Security Algorithm for Detecting Distributed Denial of Services Attacks in Cloud Computing

Coster Baloyi, Topside E. Mathonsi, Deon P. Du Plessis, Tshimangazdo M. Tshilongamulenzhe
Tshwane University of Technology, South Africa

Abstract - Cloud Computing has the benefit of offering on-demand scalable services to its customers without having to invest much on hardware infrastructure, resources and software. Most private and public sectors are moving to the Cloud, as a result Cloud Computing has become an ideal option due to its flexibility, scalability and cost efficiency. The existence of vulnerabilities in the network systems hosting Cloud services have raised an opportunity for attackers to launch attacks in Cloud Computing. Cybersecurity invaders target public and private sector's key business applications such as webservers, financial servers, legacy servers, and other servers that are hosted in Cloud Computing. The Cybersecurity invaders attack these business applications exploiting Distributed Denial of Service (DDoS) attacks. This research study proposed an enhanced security algorithm to detect DDoS attacks using real-time dataset to mitigate the challenges of large amount of network traffic data transmitted in the network. The MATLAB simulation tool was employed to simulate the proposed algorithm. The proposed algorithm has achieved good results of 99.2% detection rate, 99.5% classification of DDoS attacks, 0.9% connectivity time out and less than 18% false positive.

Unmanned Aerial Vehicle Forensics Investigation Performance under Different Attacks

Taiwo Ojo, Hongmei Chi, Samuel Kofi Erskine
Department of Computer & Information Sciences, Florida A&M University, Tallahassee, Florida, USA

Abstract - Unmanned Aerial Vehicles (UAVs), also called drones, have grown tremendously in recent years and have been adopted in various sectors. With the continuous development of machine learning algorithms to detect various attacks on UAVs. The attackers also focused on utilizing this machine-learning algorithm to disrupt the UAV's activities by using predictions to generate the attack route. In digital forensics, the forensics experts focus on the communication data between the controller and the drones, the multimedia files, and the flight data to make a complete cybersecurity report. This paper reviews various ways the drone may be compromised and its performance evaluation using the existing machine learning detection algorithm to make the forensics report about the drones. The machine learning algorithms investigated in this study include Multilayer Perceptron (MLP) for the detection technique.

Comparative Analysis of LSTM and CNN for Efficient Malware Detection

Frances Osamor, Briana Wellman
Computer Science and Information Technology,
University of the District of Columbia, Washington, DC, USA

Abstract - Intrusion-based Detection Systems are recognized as a crucial element in the safety of an organization's network infrastructure. It is the base responsible for detecting any potential threat. To detect different attacks, many IDS tools incorporate the analysis of the network traffic using the flow-based network. Research in network safety is a developing field, and IDS-specific is the current emphasis, with numerous studies and methods developed and proposed. In this study, we propose and compare the use case of a deep learning model, Long Short-Term Memory, and Convolutional Neural Network for efficient malware detection. We use the open-source benchmark dataset, ADFA-LD, for training and assessment. The experimental analysis shows that the proposed models can achieve promising results in detecting the malware concerning the recall, accuracy, and false positive rate.

Roles of Information Security and Benefits in Forming User Satisfaction of Metaverse

Hyeon Jo, Sangsun Park
Department of Planning, Realsecu, Busan, South Korea;
Division of Business Administration, Sungkonghoe University, Seoul, South Korea

Abstract - The metaverse is a type of virtual world that has been steadily drawing the attention of Internet users. As its utility continues to increase, it is now used academically and commercially. This study suggests the theoretical framework to identify the leading factors of satisfaction of metaverse users. Data were gathered from the actual users who had experienced the metaverse. Partial least square (PLS) technique was applied to conduct structural equation modeling (SEM). The results of the study found that trust is significantly related to satisfaction. The empirical results triggered that utilitarian value has a significant influence on satisfaction. The findings revealed that hedonic value is the precursor of satisfaction. The results of this paper will offer meaningful implications for the security and marketing fields.

The Need for Proactive Digital Forensics in Addressing Critical Infrastructure Cyber Attacks

April Tanner, Frances C. Dancer, Janne Hall, Naukee Parker, Russell Bishop, Tony McBride
Department of Electrical & Computer Engineering and Computer Science,
Jackson State University, Jackson, Mississippi, USA;
Division of Information Technology, Jackson State University, Jackson, Mississippi, USA

Abstract - Currently, cyberattacks are increasing and have negatively impacted the operation of critical infrastructures globally. Consequently, the organizations overseeing these infrastructures are seemingly allocating more resources toward investigating cyberattacks after they have occurred instead of planning for and preventing potential attacks. Global supply chains, financial and health services, industrial, nuclear power plants, water treatment facilities, pipelines, and transportation are increasingly being targeted by threat actors utilizing ransomware and other destructive attacks. This research explores the need for proactive digital forensic modeling approaches that focus on Supervisory Control and Data Acquisition (SCADA), Industrial Control Systems (ICSs), and automation in the preparation and defense of critical infrastructure networks.

A Design Approach for Minimizing Data Theft

George O. M. Yee
Computer Research Lab, Aptusinnova Inc., Ottawa, Canada;
Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada

Abstract - Not a day goes by without some organization in the world being victimized by the theft of private data, resulting in heavy losses to both the organization and the owners of the data. For an organization, these losses include large expenditures to resume normal operation and damages to its reputation. For data owners, the losses may be financial, and may include identity theft. Security controls (e.g., encryption) help to secure vulnerabilities, but they are not foolproof. This paper proposes a design approach that minimizes the theft of private data from an organization by limiting as much as possible the quantity of such data that is stored on-premise or in the cloud, preferring instead to let the data remain on the user's computing device. This paper additionally presents an example of applying the approach to e-commerce.

Cyber Warfare and National Security: Imperative for the Nigerian Navy

Abdullahi Muhammad Ahmed, Adamu Hussaini, Alhassan Abdulhamid
Nigerian Navy, Nigeria;
Towson University, USA;
University of Bradford, United Kingdom (UK)

Abstract - The increase in the reliance on an application of digital technologies in the 21st Century has triggered several cyber threats that threaten national security in various dimensions. To keep at pace with the dynamics of the global information domain, nations are increasingly developing cyber capabilities and integrates these capabilities into hybrid warfare in accordance with the changing character of warfare. The interplay of defensive and offensive cyber capabilities brought to fore the new paradigm of safe guarding own digital technologies against adversaries. attack and at the same time exploiting the vulnerabilities. The employment of cyber-warfare within the paradigm of national security involves the activities in cyberspace which centred around the use of Information and Communication Technology (ICT) weapons for offensive and defensive operations by state and non-state actors against other nation-states or nonstate actors. Accordingly, Advanced navies have developed their capacities to defend maritime assets and infrastructure against cyber-attacks by state and non-state actors. The Federal Government of Nigeria (FGN) has, over the years, made concerted efforts to develop the capacity of the Nigerian Navy (NN) to defend the maritime assets against any threats, including cyberrelated attacks, for enhanced national security. However, the cyber warfare capabilities of the NN remain limited, thereby making Nigeria.s Maritime Environment (NME) vulnerable to cyber-attacks with negative implications on national security. The desire to explore ways of developing the NN capabilities in cyber warfare for naval operations is the motivation of this research. The major problem is that the cyber warfare capability of the NN to defend the nation from the maritime sector is low due to inadequacies or outright lack of requisite cyber tools. The study was descriptive research and integrated both qualitative and quantitative data. The data were collected from primary and secondary sources. The research proffered strategies including formulation of NN cyber warfare policy, creation of a cyber warfare directorate and special recruitment and training of naval cyber warriors.

Gender Differences in Perceptions of Information Security, Usefulness, and Enjoyment Towards Metaverse

Hyeon Jo, Sangsun Park
Department of Planning, Realsecu, Busan, South Korea;
Division of Business Administration, Sungkonghoe University, Seoul, South Korea

Abstract - Metaverse is leading new markets and industries as a virtual space. Users trade goods and make payments on the metaverse. They also introduce it to be useful in their work. Some consumers use the metaverse because it is fun. In this context, information quality, usefulness, and enjoyment may be very important determinants for the sustainable development of the metaverse. On the other hand, demographic information has a significant influence on the behavior of information technology users. In particular, gender plays a key role in the adoption and use of information technology. It would provide useful guidance for academia and industry to identify how the perception of information quality, usefulness, and enjoyment of metaverse differs according to gender. Therefore, this study examines the differences in cognition towards metaverse targeting metaverse users. Data were collected from 180 users. To make comparisons between groups, we performed an analysis of variance (ANOVA). The study result shows that there is a significant difference in the perception of information security between men and women. The analysis partially supports differences in perceptions of usefulness by gender. The findings partially support the significant differences in the enjoyment by gender. The results of this paper will suggest meaningful implications for researchers and practitioners.

Qualitatively Evaluating a Four-Phase Process for Managing Solution of the United States' Biometric Identity Repository

James Merrill, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract - The problem addressed in this qualitative exploratory study is, there lacks a supporting process which permits a committee of cybersecurity experts, legal professionals, and citizen representatives to select, review, assess and approval any solution of a United States sanctioned human identity repository from a cybersecurity perspective. To mitigate this problem, we have proposed a four-phase process, which consists of “select”, “review”, “assess” and “approve” that permits a committee of cybersecurity experts, legal professionals, and citizen representatives responsible for the decisions in managing the technology for a United States sanctioned human identity repository. The main contribution of this study is that we have found a group of cybersecurity experts, legal professionals, and citizen representatives to evaluate the proposed process via answering a set of pre-defined and open-ended survey questions.

Development of Threat Hunting Model Using Machine Learning Algorithms for Cyber Attacks Mitigation

J. E. T. Akinsola, E. A. Olajubu, G. A. Aderounmu
Department of Computer Sciences, First Technical University, Ibadan, Nigeria;
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

Abstract - Threat hunting has become very popular due to the present dynamic cyber security environment. As there remains increase in attacks’ landscape, the traditional way of monitoring threats is not scalable anymore. Consequently, threat hunting modeling technique is implemented as an emergent activity using machine learning (ML) paradigms. ML predictive analytics was carried out on OSTO-CID dataset using four algorithms to develop the model. Cross validation ratio of 80:20 was used to train and test the model. Decision tree classifier (DTC) gives the best metrics results among the four ML algorithms with 99.30% accuracy. Therefore, DTC can be used for developing threat hunting model to mitigate cyber-attacks using data mining approach.

An Intelligent Approach for Intrusion Detection using Snake Optimizer and Random Forest

Salahaldeen Duraibi
Department of Computer and Network Engineering, Jazan University, Jazan, Saudi Arabia

Abstract - In recent years, many researchers have used various Machine Learning (ML) models that have demonstrated the power of such methods on Intrusion Detection (ID) and thus helped classifying network packets as normal system behavior or an attack. This paper presents a novel SO-RF model that combines Snake Optimizer (SO) and Random Forest (RF) for ID. The SO Meta-Heuristics (MH) algorithm is employed to select Optimal Feature Subset (OFS) from large datasets and the resulted OFS is used by the RF model to improve learning process and classification accuracy. The SO-RF is validated on two datasets for ID: KDD CUP99 and NSL-KDD. Results show that the introduced SO-RF achieves better performance outcomes compared to the RF, SVM, and several other reported models in the literature for ID.

Preventing Data Breaches: Utilizing Log Analysis and Machine Learning for Insider Attack Detection

Haydar Teymourlouei, Vareva E. Harris
Department of Technology & Security, Bowie State University, Bowie, Maryland, USA;
Criminal Justice Adminstartion & Center for Cyberseucrity, Benedict College, Columbia, South Carolina, USA

Abstract - Insider attacks are rapidly growing and expanding in complexity. As a result, security measures must be able to strengthen intrusion detection and prevention. These challenges are intensified as data becomes unbalanced, behavior is irregular, and ground truth is limited. We propose a machine learning approach that uses random forest to predict insider attacks. Network activity is collected from servers to analyze user behavior and make predictions about future insider attacks. The goal of this method is to prevent future data breaches by making accurate predictions. The results show that a machine-learning based system with the random forest classifier is accurate with highquality training data. The methodology shows 99% accuracy. Feature selection identified several feature subsets that were redundant.

Implementation of an Enhanced Cybersecurity Algorithm for Smart Grid

Kholofelo Hope Halefose, Topside E. Mathonsi, Tonderai Muchenje, Deon P. Du Plessis
Department of Information Technology, Tshwane University of Technology, Pretoria, South Africa

Abstract - Smart Grids are imminent and without which it would be impossible to manage and control modern ways of power Generation, Transmission and Distribution. In order to have effective and efficient supply of electricity without any forms of destructions and interruptions, it is imperative for the Smart Grid to be protected against cyber-attacks. Therefore, this paper proposes an enhanced cybersecurity algorithm to improve and strengthen cybersecurity for Smart Grid. The proposed algorithm is expected to provide end-to-end SCADA security, improve efficiency of the encryption and decryption technique and ensure secure authentication and communication during data transmission.

Multi Class Classification of Online Radicalization using Transformer Models

Chesta Sofat, Shabeg Singh Gill, Divya Bansal
Cyber Security Research Centre, Punjab Engineering College, Chandigarh, India;
Indraprastha Institute of Information Technology, Delhi, India

Abstract - Online Radicalization is a major security threat to a nation and has the power to influence young minds through online blogs and articles present on social media. It can occur in various forms such as political, social, criminal radicalization etc depending upon the intention and target of the propagator. Each type of radicalization aims to harm a different section of society and thus, requires a different type of treatment and mitigation plan. In our paper we have proposed the use of transformer based models such as BERT etc. to identify the type of radicalization in online text. We have also presented a comparative analysis of several transformers based classifiers for multi class classification of radicalization on social media. The results show that DistilBERT outperforms the other transformer models and has achieved an accuracy of 96.3 percent in this text classification task. As per our knowledge, this is a first of its kind study where the type of radical behaviour in text is being detected.

Assessing the Attacks Against the Online Authentication Methods Using a Comparison Matrix: A Case of Online Banking

Munir Majdalawieh, Anoud Bani-Hani, Mohammed Hussain, Aisha Alshamsi
Information Systems and Technology Management, Zayed University, Dubai, UAE;
Applied Information Technology, Zayed University Dubai, UAE;
Quality Asurance Office, Zayed University, Dubai, UAE

Abstract - This paper aims to examine commonly existing online authentication methods, review the current authentication threats and attacks against these methods, and discuss and analyze the methods used in banks operating in several countries. The paper established a new comparison matrix with 11 characteristics of authentication methods. The discussion and analysis section are based on the data collected by sending a questionnaire to more than 25 people in several countries. The data is captured through observation while navigating the targeted banks' internet banking services from the user's perspective who has the privilege (credentials) to access banks. online services. A total of 16 international banks were analyzed in this research. This research shows that authentication methods such as security questions, smart cards, PINs, and virtual keyboards will provide strong authentication to protect customers' data confidentiality and integrity. The results contribute to the practice by confirming the importance of having strong authentication methods to protect the confidentiality and integrity of the customer's data. The particularity of our study helps to identify the common authentication methods used in banks operating in different countries.

Detection of Malicious Activities Using Machine Learning in Physical Environments

Tzu-Yen Hsu, Cheng-Yuan Tang
Department of Mechatronic Engineering, Huafan University, New Taipei, Taiwan;
Department of Information Management, Huafan University, New Taipei, Taiwan

Abstract - With the increasing number of information security incidents in recent years, most organizations focus on defending against cyber-attacks but ignore the importance of physical security. Nowadays, the problem in the current defense systems is that the defense system operates independently. Therefore, if a single defense system

fails, attackers will quickly enter the organization to carry out malicious activities. In order to effectively reduce this risk, we propose a method to integrate three general types of defense systems, including an access control system, a surveillance system, and a host defense system, to achieve physical access control detection. In the access control system, the deep learning method is used to detect the abnormal login behavior. In the surveillance system, the object detection method is used for real-time tracking. The distance measurement methods, including the Intersection Over Union (IOU) and the Intersection Over Area (IOA), are used to detect whether people enter the unauthorized areas. In the host defense system, Word2vec and deep learning is used to detect whether there is an abnormal command execution. Our proposed integration methods and through theoretical analysis and experiments can effectively reduce the malicious activity in the physical environment.

Cyber-Criminals Metrics Engine for Attack Determination on Web Technologies Platforms

Olumide Malomo, Shanzhen Gao, Adeyemi A. Adekoya, Ephrem Eyob, Weizheng Gao
Department of Computer Information Systems, Reginald F Lewis College of Business,
Virginia State University, Petersburg, Virginia, USA;
Department of Mathematics and Computer Science, College of Natural Sciences and Mathematics,
West Virginia State University, Institute, West Virginia, USA

Abstract - Web technologies have created a worldwide web of problems and cyber risks for individuals and organizations. In this paper, we evaluate web technologies and present the different technologies and their positive impacts on individuals and business sectors. Also, we present a cyber-criminals metrics engine for attack determination on web technologies platforms. weaknesses. Finally, this paper offers a cautionary note to protect Small and Medium Businesses (SMBs) and make recommendations to help minimize cyber risks and save individuals and organizations from cyberattack distress.

Linguistic Casualties: Cybersecurity Degrees must Require Foreign Languages

Gary L. Wallace
Southern Utah University, Cedar City, Utah, USA;
Commander and Director of Information Technology and Security,
United States Navy Space and Naval Warfare Systems Command (retired), USA

Abstract - Unlike Russia and China, the United States has no mandatory foreign language requirement at any educational level. The lack of multilingual cybersecurity training poses a threat to national security. Current higher education curricula limit the number of foreign language courses taken by students pursuing cybersecurity degrees. Assessing these limitations and prioritizing foreign language courses for cybersecurity students will improve national security and prevent linguistic casualties.

Detection and Isolation Malware by Dynamic Routing Moving Target Defense with Proxies

Kouki Inoue, Hiroshi Koide
Graduate School of Information Science & Electrical Engineering, Kyushu University, Fukuoka, Japan;
Research Institute of Information Technology, Kyushu University, Fukuoka, Japan

Abstract - In recent years, many companies and organizations have introduced internal networks. While such internal networks propose availability and convenience, there have been many cases in which malicious outsiders have intruded on these local networks, and leaked customer information through cyber attacks. In addition, there have recently been reports of a type of attack called "Advanced Persistent Threats (APT)". Unlike conventional cyber attacks, these attacks target specific objectives. And they use sophisticated techniques to penetrate the target's system. Once malware successes to intrude into the system, malware does not immediately attack the target but hides for a long time to investigate the system and gather information. Moving Target Defense, MTD is a technology that dynamically changes the configurations of systems targeted by cyber attacks. In this study, we implemented a model using a proxy-based network-level MTD to detect and quarantine malware in internal networks. And we can confirm that the proposed method is effective in the detection and quarantine of malware.

Privacy-Preserving Collaborative Intrusion Detection Systems: A Federated Learning Framework

Chukwuebuka Ezelu, Ulrich Buehler
Department of Applied Computer Science, Fulda University of Applied Sciences, Germany

Abstract - Security concerns are one of the significant obstacles to collaborative systems. With the ever-increasing sophistication of attacks in networked environments, it is impossible to overlook the need for collaboration. Collaborative Intrusion Detection Systems (CIDS) presents a solution to combating such attacks but also introduces some privacy concerns. The need for information transmission within CIDS architecture raises some privacy issues, which can be deterring to organizations with sensitive data. This work presents a forest-based Federated Learning framework that allows organizations to collaboratively build a forest without compromising the security and privacy of their data. Two strategies were proposed for building the forest and were evaluated against a centralized dataset. The result shows that the proposed framework outperforms the centralized dataset framework based on model reliability and accuracy of prediction.

Considering the Implications of Artificial Intelligence, Quantum Computing, and Cybersecurity

Dominic Rosch-Grace, Jeremy Straub
Department of Computer Science, North Dakota State University, Fargo, North Dakota, USA

Abstract - This paper considers the impact of using artificial intelligence and quantum computing as part of a cyber defense strategy and the deterrence, defense, and offensive capabilities that it provides. A model for this is presented and discussed.

How Can We Make Law Firms Less Vulnerable to Cyber-Attacks?

Liam Cosgrave, Lila Rajabion
School of Graduate Studies, SUNY Empire State College, New York, USA

Abstract - In today's digital world, cyber-attacks are increasing daily, and many institutions and organizations understand the critical need to secure their system and hire cybersecurity professionals. Cybersecurity threats for law firms are also growing. These firms usually Work with sensitive and confidential information. In today's digital economy, information is power. The volume and type of sensitive data and information the law firms have makes them an attractive target for hackers. The American Bar Association's 2019 Legal Technology Survey cites that 26% of responders reported that their law firms had experienced some security breach. Many also reported needing to find out if they have had security breaches at their organization. With the rise of remote working in law firms and law practices, there has been an increase in the demand for security to maintain the safety of information and the firm's reputation. Throughout this research, we will discuss some key points that make a law firm secure while working in the office and from home/remotely. Such topics can include VPN software, multi-factor authentication, cybersecurity awareness and training, Document Management Systems, Finance, and payroll systems, and much more. This study conducts a systematic literature review to examine how can law firm be less vulnerable to cyber-attacks. The challenges and suggestions and strategies are summarized and discussed.

Contemporary Cybersecurity Challenges in Metaverse using Artificial Intelligence

Fadi Al-Ayed
Department of Electrical Engineering and Computer Science, The Catholic University of America, Washington D.C., USA

Abstract - There are several difficulties and challenges in Metaverse from the viewpoint of cybersecurity employing Artificial Intelligence (AI). One of the most critical concerns is the use of artificial intelligence to produce synthetic media, commonly known as "Deep fakes". Deep fakes are computer-generated photographs or movies that appear to be actual individuals. They can be used maliciously such as spreading false information or causing harm to someone's reputation. There are numerous methods for reducing the hazards posed by deep fakes. This paper examines vulnerabilities and the possible threads occur in the Metaverse environment and protection approaches which can be a great assistance of developing designated algorithms that can detect them. Another target is to establish rules and regulations governing the usage of AI generated media. Finally, it is critical to raise awareness about the hazards of deep fakes and how to recognize them. As technology advances, it is critical to be aware of the hazards and take actions to reduce them.

Unsupervised Methods for Detecting Cyber-Physical Attacks in Industrial Control Systems

Corwin Stanford
Department of Mathematics & Statistical Sciences, and Department of Information Technology, Jackson Sate University, Jackson, MS, USA

Abstract - Detection of cyber-physical attacks typically relies on maintaining a list of known exploits, and examining real time data for evidence of those exploits. However, this is often cost prohibitive, and still remains vulnerable to zero-day attacks. Thus, development of models that can accurately detect attacks without a priori knowledge of what those attacks look like is an important research question. Here we examine a common unsupervised learning method, and discuss techniques for refining its application to intrusion detection in industrial control systems (ICS) data. Principal component analysis (PCA) has been shown to be effective in this endeavor,

although problems can arise in implementation. Here we discuss a collection of techniques for improving this method. The techniques are applied to the Hardware-in-the-Loop-Based Augmented Industrial Control System (HAI) dataset, and the results compared.

Machine Learning for Detecting Phishing on Websites

Kagiso Matlaila, Tonderai Muchenje, Topside Ehleketani Mathonsi, Vusimuzi Malele
Department of Information Technology, Tshwane University of Technology, South Africa

Abstract - Phishing is a sort of cyber-attack that employs deceit to trick unwary online users into giving personal information such as credentials, social security numbers, and credit card details. Attackers fool Web users by posing as a reputable or legitimate webpage in order to gain confidential information. The use of fraudulent tactics to get personal information is growing more common these days. Whenever users access a phishing site, it detects whether or not the site is fake. This work presented a phishing detector technique for detecting banned URLs, also known as phishing sites, such that consumers may be notified while browsing or visiting a given website. As a result, it can be used for identification and verification, as well as to keep individuals from being duped. The goal of this research is to recognize phishing Websites using classification algorithm by extracting and assessing various features of authentic and phishing URLs. The study's intended outcome is the development of a machine learning model that will employ a Decision Tree, Random Forest, and xgboost.

Arithmetic Optimization Algorithm based Feature Selection Approach for malicious URL Detection

Nawaf Abdualaziz Almolhis
Computer Science Depertment, College of Computer Science and Information Technology,
Jazan University, Jazan, Saudi Arabia

Abstract - Today, web applications are the most costeffective manner of software delivery and they are used by a wide range of businesses and organizations to provide their services over the internet. However, due to the huge number of users who can access these websites; make them vulnerable and easily targeted by different forms of attacks. This paper uses Arithmetic Optimization Algorithm (AOA) as a Feature Selection (FS) based approach to select optimal and informative Feature subset. The AOA is validated on an open source dataset for the detection of malicious URLs. Overall results show that the AOA gained better performance outcomes and faster than other compared to Particle Swarm Optimization, Multi-Verse Optimizer and Salp Swarm Algorithm FS methods.

Data Approach to Biometrics in Cybersecurity with Related Risks

Angelo Arteche, Cody Asher, Caroline Bull, Henry Dare, Isha Datey, Erich Elshoff, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - This paper will research the implications that biometrics have on cybersecurity. The paper will introduce topics such as: what biometrics are, recent advancements in biometrics and how biometric data is stored. The paper will then proceed to discuss privacy implications of biometric authentication, how biometric authentication is implemented at an organizational level, the cost of biometric authentication and the performance impacts of using biometric authentication. Subsequently, the paper will delve into what models and systems are currently used for biometric authentication and the risks posed by using biometric authentication. The risks posed in paragraphs prior will be addressed in their own section, and the advantages will be presented to be weighed against the risks. The paper will also include discussions on how biometric data can be stored securely and processed using technology like cancellable biometrics. The paper will present the findings, and recap arguments brought forward throughout the paper.

CSecPrivAF - Cybersecurity and Privacy Assessment Framework for IoT Systems

Sergio Luis Ribeiro, Joao Pedro de Lima Cassiano, Alana Almeida
Security and Privacy, CPQD - Centro de Pesquisa e Desenvolvimento em Telecomunicacoes, Campinas/SP, Brazil

Abstract - This article presents a proposal for an “CSecPrivAF - Cybersecurity, and Privacy Assessment Framework for IoT Systems" for real-time security management, which is one of the results of the TecSEG Project (Security Technologies Project), a long-term Brazilian program focused on Research, Development and Innovation in Cybersecurity Assessment. The existence of numerous architectures and frameworks related to security assessment is a reality, however, a framework with the objective of dealing with heterogeneous concepts and architectures, which is one of the main motivators of the IoT, is not yet a reality. To respond to this challenge, the TecSEG project was proposed, with the main objective of developing and applying a set of security assessment and investigation methodologies, and the framework presented in this article is one of the first practical results.

Unknown Malicious Domain Detection Based on DNS Query Analysis using Word2Vec

Hiroto Yamada, Daiki Nobayashi, Akihiro Satoh, Yutaka Nakamura, Takeshi Ikenaga
Kyushu Institute of Technology, Fukuoka, Japan

Abstract - Malware infection of network user devices is a critical threat to users on the Internet. A major countermeasure against malware infection is a signature-type detection method that uses a blacklist to prevent communication with a Command and Control (C&C) server when the malware attempts a malicious attack. However, if the malicious domain name is unknown to the C&C server, it is difficult for this method to detect the malware. This paper focuses on the behavior of devices infected with malware. We propose a method for unknown malicious domain detection based on domain name system (DNS) query analysis using Word2Vec, a popular machine learning method. We then evaluate the effectiveness of our proposed method using an actual DNS query log.

CSCI-RTMC:
RESEARCH TRACK ON MOBILE COMPUTING, WIRELESS NETWORKS, & SECURITY

Modeling the Resource Allocation in 5G Radio Access Networks with Network Slicing

Daniel Haro-Mendoza, Luis Tello-Oquendo, Vicent Pla, Jorge Martinez-Bauset,
Luis A. Marrone, Shih-Chun Lin
College of Engineering, Universidad Nacional de Chimborazo, Ecuador;
LINTI, Universidad Nacional de La Plata, Argentina;
iWN Lab, Department of Electrical and Computer Engineering, North Carolina State University, USA;
Communications Department, Universitat Politecnica de Valencia, Spain

Abstract - Network Slicing (NS) is one of the technologies considered a pillar of 5G networks. It allows the division of the physical infrastructure of a network into several isolated logical networks (slices). The slices can have different sizes and be offered to other use cases. We analyze the radio resource allocation problem through a random access channel model considering the radio access network (RAN) with NS in a steady state. We perform an in-depth study of the random access procedure (RAP) to optimize resource allocation in a 5G RAN with NS. We focus on assigning subsets of preambles for each slice depending on the service's priority. The main contributions of our work are the following: i) A model for a scenario of n slices; that is, it has no limitation for the number of use cases. ii) An efficient RAP resource allocation policy to maximize the probability of successful access by UEs in each slice.

On the Resource Allocation for Radio Access Network Slicing in Cellular IoT with Massive Traffic

Daniel Haro-Mendoza, Luis Tello Oquendo, Vicent Pla, Jorge Martinez-Bauset,
Luis A. Marrone, Shih-Chun Lin
College of Engineering, Universidad Nacional de Chimborazo, Ecuador;
LINTI, Universidad Nacional de La Plata, Argentina;
iWN Lab, Department of Electrical and Computer Engineering, North Carolina State University, USA;
Communications Department, Universitat Politecnica de Valencia, Spain

Abstract - The limited capacity of the random access channel (RACH) represents a challenge for adequate resource allocation in 5G radio access networks with network slicing. Furthermore, a fair division of scarce radio resources is required to simultaneously support many users with heterogeneous service requirements. In this work, we look at the problem of uplink radio resource allocation to slices on the radio interface of one cell in a non-stationary regime with mMTC, eMBB, and H2H traffic. We analyze four resource allocation policies for efficient random access to improve each slice's capacity in terms of successful access probability, the number of preamble transmissions, and access delay. Besides the number of available preambles in the RACH, we also consider the limitation of uplink grants in the radio access network.

A Comparison of Power Allocation Mechanisms for 5G D2D Mobile Communication Networks

Dimitris Georgiadis, Elena Politi, George Dimitrakopoulos
Department of Informatics and Telematics, Harokopio University, Athens, Greece

Abstract - The ever-increasing demands for higher data transmission speeds, channel reliability and less energy consumption push the scientific community to discover new methods to manage resources of modern networks, without degrading the quality of communications. Modern cellular systems of the fifth generation (5G) are internationally recognized for their excellent performance and for the innovative technologies that they apply. The Device-to-Device (D2D) protocol has been established among the technologies of 5G, for the fast and secure direct connections between terminals that it offers. The protocol is able to offload the network and reorganize connections in order to save system resources, while also supporting decentralized communications. The present work aims to highlight the importance of D2D communications in 5G cellular networks by assessing the performance of four well known energy resource allocation mechanisms in scenarios with environmental constraints.

Optimizing the Transmission of Multimedia Content over Vehicular Networks

Pablo Garrido, Pablo Pinol, Miguel Martinez-Rach, Otoniel Lopez,
Hector Migallon, Manuel P. Malumbres
Universidad Miguel Hernandez, Elche, Spain

Abstract - The multi channel operation mechanism of the IEEE 1609.4 protocol, used in vehicular networks, may impact network performance if applications do not care about its details. Packets delivered from the application layer to the MAC layer during a Control Channel time slot have to wait to be transmitted until the following Service Channel time slot arrives. The accumulation of packets at the beginning of this time slot may introduce additional delays and higher collision rates when packets are transmitted. In this work we propose a method, which we call SkipCCH, that deals with this issue in order to make a better use of the wireless channel and, as a consequence, increase the overall network performance. With our proposal, streaming video in vehicular networks will provide better reconstructed quality at the receiver side under the same network conditions. Furthermore, this method has particularly proven its benefits when working with QoS techniques, not only by increasing the received video quality, but also because it avoids starvation of the lower priority traffic.

Android Malware Detection Using Machine Learning Techniques

Priya Raghuvanshi, Jyoti Prakash Singh
Department of Computer Science and Engineering, National Institute of Technology Patna, Patna, India

Abstract - The exponentially growing use and popularity of the Android Operating System made it more prone to malware attacks. Easy to use android smartphones attracted intruders which resulted in the need for a novel malware detection method. Several works are proposed which makes use of machine learning techniques to detect malign and benign applications. To provide the solution of addressed problem of detection of malware in android, in this paper, we have provided a decent approach based on static analysis of the android apk for feature extraction like API Calls, Intents, Permissions and Command signature. We have used the Drebin [2][26] and Malgenome [35] datasets, both the datasets have a good amount of goodwares and malwares to train the machine learning classifier. After feature extraction from android APK (Android Package Kit), we performed various experiments taking into consideration all the combined extracted features as well as generated the twosome combination of permission. By taking all four extracted features our model achieved a high classification accuracy of 98.19% with Drebin Dataset and with the twosome permission combination it achieved an accuracy of 96.27% with the RF

model. Additionally, in the case of Malgenome Dataset, we have an accuracy of 98.84% considering all four features and 97.63% in the case of permission twosome combination SVM with PCA (Principal Component Analysis). Both experiments have given a decent result compared to some of the existing work considering the same scenario in the android malware domain.

GRASSMARLIN-based Metadata Extraction of Cyber-Physical Systems Intrusion Detection in CyberSCADA Networks

John E. Efiong, Bodunde O. Akinyemi, Emmanuel A. Olajubu, Ganiu A. Aderounmu
Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

Abstract - Cyber-Physical Systems such as those in the SCADA architecture and the entire industrial control system networks have now become essential components of the cyberspace due to their integration with modern IT networks. Cyber attackers have taken advantage of these vulnerable critical assets that have suddenly found themselves in an unprepared situation enabled by the Internet of Things Paradigm. Intrusions into the new CyberSCADA networks seem to go unabated as the proprietary communication protocols in these systems lack security mechanisms to help intrusion detection and analysis. In addition, existing network analyzers are weak in identifying granular details from traffic in these legacy networks that could help track, detect and investigate intrusions. This study performed a passive fingerprinting exercise on captured network data using Grassmarlin and identified useful metadata of critical network devices. The results demonstrate that fingerprints, metadata and critical assets' inventory on Grassmarlin would help industrybased cybersecurity personnel improve intrusion detections on CyberSCADA networks.

Hybrid Modulation Scheme for System Performance Evaluation and Energy Efficiency Measurement in 5G Technology

Bertille Auyane Ngouessy ep Ovono, Temitayo Otunniyi, Marcel Ohanga Odhiambo
Vaal University of Technology, South Africa;
Mangosuthu University of Technology, South Africa

Abstract - The expected exponential increase in data traffic has an impact on power consumption. Thus, many multicarrier modulation schemes are being investigated as part of the 5G physical layer solution to meet the requirements of 5G. On the other hand, Multicarrier modulation has a high peak-to-average power ratio, which reduces energy efficiency. This paper proposes a hybrid modulation scheme based on a Constant Envelope called Spatial Modulation Orthogonal Frequency Division Multiplexing Access (SM-CE-OFDMA). This method compares the hybrid's components by examining and evaluating parameters such as peak-to-average power ratio (PAPR), signal-to-noise ratio (SNR), and bit error rate (BER). In terms of performance, the MATLAB simulation revealed that the SM-CE-OFDMA outperformed the conventional OFDMA and CE-OFDMA in terms of BER performance. The findings also showed that CE-OFDMA has a lower PAPR than conventional OFDMA and a similar amount with SM-CE-OFDMA. The low PAPR led to improved energy efficiency.

Steganalysis on Covert Channel in IPv6 Flow Label

Da-Chih Lin, Quincy Wu
Department of Computer Science and Informance Engineering,
National Chi Nan University, Nantou, Taiwan

Abstract - Modern communication networks provide rich channels for steganographic tools to hide secret data. In addition to payload, an emerging trend is to utilize protocol headers to hide data as well. This paper studies the IPv6 packets with data hidden in their Flow Label field, and proposes an algorithm which utilizes chi-square test to detect the existence of hidden data by the conformance of uniform distribution.

Performance Optimization of Concurrent VOIP Calls Across Wireless Mesh Networks

Olusola Olorunnisola, Topside E. Mathonsi, Deon P. Du Plessis
Department of Information Technology, Tshwane University of Technology, Pretoria, South Africa

Abstract - Over the past years we have seen an increasing number of mobile users with the need and desire for optimized concurrent VOIP calls across Wireless Mesh Networks (WMN). Thus, this study sought to develop a packet aggregation and compression algorithm in order to improve the performance of VoIP calls across WMN. The algorithm was designed by integrating ROHC and end-to-end aggregation. This integration minimised the amount of packet loss and gave a higher throughput. The algorithm was tested against some existing algorithms and performed better.

Mobile AR Application for Navigation and Emergency Response

Nishith Reddy Mannuru, Mounica Kanumuru, Sharad Sharma
Department of Information Science, University of North Texas, Denton, Texas, USA

Abstract - Emergency response, navigation, and evacuation are key essentials for effective rescue and safety management. Situational awareness is a key ingredient when fire responders or emergency response personnel responds to an emergency. They have to quickly assess the layout of a building or a campus upon entry. Moreover, the occupants of a building or campus also need situational awareness for navigation and emergency response. We have developed an integrated situational awareness mobile augmented reality (AR) application for smart campus planning, management, and emergency response. Through the visualization of integrated geographic information systems and real-time data analysis, our mobile application provides insights into operational implications and offers information to support effective decision-making. Using existing building features, the authors demonstrate how the mobile AR application provides contextualized 3D visualizations that promote and support spatial knowledge acquisition and cognitive mapping thereby enhancing situational awareness. A limited user study was conducted to test the effectiveness of the proposed mobile AR application using the mobile phone usability questionnaire (MPUQ) framework. The results show that the mobile AR application was relatively easy to use and that it can be considered a useful application for navigation and evacuation.

Improved Encryption Algorithm for Public Wireless Network

Christopher Khosa, Topside E. Mathonsi, Deon P. du Plessis, Tshimangazdo M. Tshilongamulenzhe
Department of Information Technology, Tshwane University of Technology, Pretoria, South Africa

Abstract - Security mechanisms such as cryptographic algorithms play an important role in Wireless Fidelity (WiFi). However, these algorithms consume a lot of memory and power. In this paper, we proposed an improved encryption algorithm that reduces the heavy consumption of power and storage to effectively protect public Wi-Fi networks. The proposed encryption algorithm was based on a hash-based message authentication algorithm. The proposed cryptographic algorithm was evaluated against the current cryptographic algorithm. We used the Network Simulation 2 (NS-2) tool to evaluate different settings for each algorithm, such as data block size, different platforms, different and decryption speeds. Keywords. Wireless Fidelity, encryption algorithm, Triple Data Encryption Standard, Data Encryption Standard, Advanced Encryption Standard.

Locating and Extracting Digital Evidence from Mobile: Malware Context

Abdullah Mujawib Alashjaee
Department of computer science, Faculty of Computing and Information Technology,
Northern Border University, Arar, Saudi Arabia

Abstract - In this paper, the author tries to locate and extract evidence that can lead to the existence of malicious software in a mobile device. The area as a source of evidence and the time evidence that can be found are discussed. These steps are the most critical steps investigators need to make sure of when conducting any kind of digital forensics. As such, the author has conducted an experiment to further demonstrate how and where to look for evidence once a mobile device is suspected to have had a malicious attack. The dataset employed in the experiment part of the paper is DERBIN dataset which contains a number of Android malware software, and the author was able to locate some of the malware where verified the hashes of the located malware in a public malware repository. Around 50 search engines of the well-known anti-virus vendors have identified the malware as Android malware.

A Survey of Intelligent Traffic Control Algorithm in Wireless Sensor Networks

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

Abstract - Traffic control algorithms have been comprehensively studied and applied in WSNs, however, this is the first attempt to put in viewpoint the numerous efforts in form of a survey. In order to provide an inclusive survey of the existing literature. This paper examines current state of the art traffic control algorithms according to their common methodology applied in order to reduce packet loss and packet delay in WSNs. Additionally, this paper presents a clear understanding of the active research area, by identifying a clear classification and guidelines used when traffic control algorithms are designed. Furthermore, this paper outlines the strength and weakness of existing solutions and highlight the key open research areas for further development. This comprehensive survey act as a basis starting point and a guide for everyone enthusiastic to explore into research on traffic control algorithms in WSNs. Finally, the study presents the conclusion.

Cost-Effective Telecommunication Technologies for Rural Areas in Developing Countries like South Africa

Marcel Ohanga Odhiambo, Weston Mwashita
Mangosuthu University of Technology, Durban, South Africa;
International University of Management, Windhoek, Namibia

Abstract - Rural South Africa has always been underserved in terms of telecommunications services. For numerous reasons, this condition has persisted into the era of broadband connectivity. Market failure, the slow pace of implementing SA Connect (the national broadband policy adopted by Cabinet in 2013), the splitting of the old Department of Communications, and a period of political leadership changes may all be blamed for the problem, despite the National Development Plan's vision of a "dynamic and connected vibrant information society and a knowledge economy that is more inclusive, equitable, and prosperous." Poverty, inequality, and unemployment continue to be a harsh reality for the rural poor. To address all aspects of the digital ecosystem, not only access, decisive and holistic initiatives are required. The scenario necessitates stakeholders rethinking how the digital divide might be bridged. From a policy and regulatory standpoint, it necessitates novel approaches and creative thinking. Despite persistent policy efforts to address these obstacles, South African rural areas remain excluded from the digital communication podia, resulting in digital inequality. This report attempts to address telecommunication system penetration in South African rural areas by utilisng 5G and infrastructure sharing.

Implementation of a Model to Amplify Transmission Quality of Satellite Television

Lebohang Maja, Topside E. Mathonsi, Deon P. du Plessis, Tshimangazdo M. Tshilongamulenzhe
Department of Information Technology, Tshwane University of Technology, Pretoria, South Africa

Abstract - Satellite Television broadcasting is one of the many uses of communication satellite technologies. Communications satellites are used for marine applications, large-scale telephone connections, business systems, internet, and broadcasting television programming. In television broadcasting, satellite communication is the easiest way to transmit many services and offers a variety of choices across a varied region, thereby overcoming the need for the complex infrastructure of terrestrial transmitters that a terrestrial network needs to broadcast its signals throughout a wide range area like countries or continents and providing quality digital television viewing. However, Satellite television broadcasting has a shortcoming of outage effect caused by rain fade. These interruptions instigate due to bad raining weather, which at once will cut signal transmission from the transmitter satellite to the receiver dish preventing customer picture viewing. To overcome this shortcoming, this study will be undertaken to explore the challenges that satellite television broadcasting faces, which is caused by the rain fade effect. Thereafter, a model to amplify the transmission quality of satellite television will be designed. The proposed model will reduce outages that are caused by bad weather condition such as rain, storm , and many more. The Matlab simulation tool was used to evaluate the proposed model.

Crowd-Funded Earthquake Early-Warning System

Hudson Kaleb Dy, Hsi-Jen James Yeh
Department of Engineering and Computer Science, Azusa Pacific University, Azusa, California, USA;
Walnut High School, Walnut, California, USA

Abstract - Earthquake early warning systems has been proven to save countless lives in Japan, Mexico, and Chile, where earthquake warnings are often broadcast live on TV up to a minute before residents experience shaking. Unfortunately, traditional early warning systems require extensive capital investment. The high cost of traditional earthquake early-warning systems and limited budgets prevent earthquakeprone developing countries like the

Philippines, Indonesia, Afghanistan, India, Burma, Ghana, Nigeria, Columbia, Venezuela, and Bolivia from building traditional earthquake warning systems. This project describes repurposing old Android smartphones into affordable dedicated seismometers to detect tremors. These smartphones have become disposable items and are continuously "upgraded" and replaced. Yet every one of these devices includes everything needed to act as a dedicated seismometer: Wi-Fi capability, GPS, and an accelerometer. The software developed for this project converts these smartphones into dedicated seismometers and uses existing web technologies for telemetry services. This system would also trigger alerts to all devices that has the software installed whenever a tremor is detected, effectively making each seismic detection station double as an earthquake early-warning alarm. A large network of these seismic detection stations will effectively create an affordable earthquake early-warning system that can be rapidly implemented at an extremely low cost. It would provide developing nations an affordable life-saving alternative to expensive traditional earthquake early-warning systems. This solution is cheap, keeps old smartphones from landfills, and will save lives.

Vehicle Ad Hoc Network System for Intelligent Transport Systems: A Review

Ndzalama Ngobeni, Topside E. Mathonsi, Tonderai Muchenje, Vusumuzi Malele
Tshwane University of Technology, Pretoria, South Africa

Abstract - In the past few years, the road of South Africa is accumulated by the vehicles manufacturing increase day by day, with more and more vehicles being manufactured there is less opportunity of having better traffic. However, the increase in vehicle manufacture increases accident rates and congestion which increase transmission delay between two vehicles as Global Positioning System (GPS) technology does not have feature for collision avoidance due to its poor accuracy of position tracking, which increases the number of stolen vehicles. Therefore, in this paper an enhanced Radio Frequency Identification (RFID) based Vehicle Ad Hoc Network system for Intelligent Transport Systems is proposed. The proposed algorithm will be designed in such a manner that it will be effective in maintaining traffic through communication among vehicles to vehicles, to the Road Site Unit (RSU). The RSU would generate all possible routes by Ad hoc On-demand Distance Vector (AODV) routing protocol and control the signals within the accessible limit. RFID tracking would begin once the information regarding the stolen automobiles or the accident is given to the appropriate team. We will use NS2 simulations to evaluate the efficiency of the proposed algorithm.

A Review of Authentication Algorithms for Mobile Ad Hoc Networks

Thandiwe W. Mahlalela, Topside E. Mathonsi, Tonderai Muchenje, Vusumuzi Malele
Tshwane University of Technology, Pretoria, South Africa

Abstract - MANETs are infrastructure-less network and they are more vulnerable to security threats than conventional wired or wireless networks. Because of their nature, MANETs are vulnerable to attacks such as lack of security infrastructure, constantly changing topology and open wireless medium. Nodes can freely join and leave the network in dynamic period. Authentication is required to verify the identity of nodes to ensure data transfer and communication amongst the nodes is secure, establishes trust relationship amongst the nodes in the network and also to be able to prevent impersonation. This paper proposes an Authentication algorithm to authenticate every node joining and leaving the network. Mathematical modelling will be used when designing the proposed Authentication algorithm. A computer network simulator will be used to evaluate the effectiveness of the proposed algorithm against existing algorithm.

Collision Avoidance Method for Broadcast Communication over Multi-hop HD-PLC Network

Yuki Furuhama, Daiki Nobayashi, Kazuya Tsukamoto, Takeshi Ikenaga
Kyushu Institute of Technology, Fukuoka, Japan

Abstract - High Definition Power Line Communication (HDPLC) is a communication technology that enables high-speed communications with a maximum PHY rate of 240 Mb/s using power lines. HD-PLC with multi-hop technology has been developed as a means of enabling long-distance communications. Multi-hop HD-PLC is expected to be used in BACnet for factory equipment and building management. However, the communication performance of BACnet is degraded by frequent data collisions caused by simultaneous broadcast message transmissions from multiple nodes. In this paper, we propose a transmission method to avoid collisions and reduce packet loss by using RTS/CTS control only in sections where collisions occur frequently. Simulation results show that the proposed method can reduce packet loss.

CSCI-RTOT:
RESEARCH TRACK ON INTERNET OF THINGS & INTERNET OF EVERYTHING

RSSI and Machine Learning-Based Positioning System with Sub-Meter Accuracy

Hanyuan Huan, Varick L. Erickson
Department of Computer Science, California State University, East Bay, California, USA

Abstract - Accurately monitoring the location of objects remains an important problem in many applications, such as inventory management, asset tracking, and Internet of Things (IoT). Solutions such as Bluetooth and RFID generally require significant infrastructure and high setup cost. Developing accurate positioning systems with lower cost, simpler infrastructure, and lower power consumption is becoming an increasingly important challenge. In the present work, we develop a lowpower, low-cost indoor positioning system. The system leverages an inexpensive ESP32 MCU board, WiFi, RSSI (Received Signal Strength Index) measurements, and Machine Learning to monitor location. The system runs with low power requirements and leveraging existing WiFi infrastructure, reducing overall cost and complexity. We demonstrate the ability to track objects with sub-meter accuracy by utilizing machine learning to process RSSI data.

User Acceptance of a Modular Interactive Digital Twin of the Production Department for Virtual Reality

Matej Dvorak, Jan Kubr, Miroslav Bednar, David Krakora, Miroslav Malaga,
Petr Horejsi, Michal Simon
Faculty of Mechanical Engineering, University of West Bohemia, Pilsen, Czech Repubic

Abstract - This thesis deals with the creation and verification of a modular interactive digital twin of the assembled production section of a Fischertechnik kit in virtual reality. The experimental reference physical modular section was primarily designed for the training of industrial engineers and has proven to be very useful during practice. In this paper, a method for creating a digital twin (DT) of the reference physical model and its implementation for virtual reality (VR) will be demonstrated. Furthermore, an initial pilot study will be described to preliminarily verify the potential of the solution for virtual training and further follow-up research. A randomized group of students was selected to be exposed to the actual physical model and then to the same model in virtual reality. The surprising result was that when validated immediately after the exposures, there was a higher rate of recall in the virtual reality work than in the physical model work. These results are discussed. In a follow-up study, we want to focus on the synchronization issue, which proved to be the most difficult to implement.

Delay-Based Dynamic Clustering Method for the IoT Cluster

Seo Jin Chang, Boyeon Kim, Yunseok Chang
Department of Computer Science, University of California, Irvine, California, USA;
Department of R&D, Sekyung BMT Co., Seoul, Republic of Korea;
Department of Computer Engineering, Daejin University, Kyunggi, Republic of Korea

Abstract - A dynamic clustering method can reconstruct IoT nodes into logical group units when using several groups of IoT nodes. However, massive control at a logical group level does not guarantee optimal control performance. To efficiently deliver and execute a massive control command within a certain communication delay, it is necessary to cluster each logical group into subgroups and execute massive control commands by each subgroup. This paper proposes the DBDC (Delay-Based Dynamic Clustering) method, a clustering method based on DBSCAN that can optimize the communication delay of a massive control within a logical group. Through several simulations, the parameters of the DBSCAN algorithm create different clustering cases, and the DBDC method can find the optimal parameter value that satisfies the optimization condition. Therefore, this study shows that the DBDC method can effectively design a cluster system so collective IoT clusters have the best massive control performance.

Implementation of a Multisensors Fire-Fighting Monitoring System for Forest Protection

Giovanni Pettorru, Marco Bertolusso, Michele Spanu, Matteo Anedda, Daniele Giusto
DIEE Department of Electrical and Electronic Engineering, University of Cagliari

Abstract - Monitoring and control of natural environments is becoming increasingly sensitive in the face of climate change. This work proposes a solution for the protection to safeguard forests and woodlands against natural or arson fires. Currently, existing solutions have critical issues in terms of early detection of outbreaks, or rely on expensive solutions if implemented on a large scale. Electronic noses for fire smoke detection have been developed. The data is transmitted in real time to a Social Internet of Things (SIoT) cloud platform that analyzes the data and detects potential critical situations. Preliminary results obtained show the effectiveness of the system in detecting the change in detected parameters during the occurrence of a fire front.

Review Paper to Enhance Authentication of IoT Devices in Industrial Control Systems

Jahid Hasan Miah, Tonderai Muchenje, Topside Ehleketani Mathonsi, Vusimuzi Malele
Department of Information Technology, Tshwane University of Technology, South Africa

Abstract - Industrial Control Systems are hardware and particular manipulating structures used to automate and modify diverse industrial operations and procedures. ICS incorporates devices, networks, and configurations applied to operate and optimise more than one strategy withinside the manufacturing chain. However, this layout no longer meets trendy commercial enterprise necessities for working with cutting-edge technologies like big data analytics. Therefore, to meet the industry requirements, various ICSs had been linked to agency networks that permit commercial organisation customers to get the right of entry to real-time statistics generated using power plants. Which led to the extinction of the air gap in ICS; thus, many ICS security challenges arose. This paper reviewed cyber security issues related to ICS and decided to solve one specific problem. Weak authentication of IoT devices in ICS is the only problem proposed to be solved in this research by implementing an algorithm using the MATLAB simulator to enhance the authentication of IoT devices in ICS.

Low-Cost Brain-Computer Interface Design Using Deep Learning for Internet of Brain Controlled Things Applications

Mokhles M. Abdulghani, Wilbur L. Walters, Khalid H. Abed
Department of Electrical & Computer Engineering and Computer Science,
Jackson State University (JSU), Jackson, MS, USA

Abstract - Huge efforts have been made so far aiming to classify human thoughts. Controlling machines using the concept of Brain-Computer Interface (BCI) is a practical method that opens the way toward a fully synchronized method between human thoughts and controlled objects. Furthermore, a reliable design of a BCI-based Internet of Things (IoT) system is still in the early stages because it still has several challenges, such as the issue of accurately implementing the individual's intention. This paper presents a method for brain waves recognition using deep learning based on shapes and colors for use in merging concepts of the IoT and BCI, which we defined as the "Internet of Brain Controlled Things (IoBCT)". We used a low-cost 8-channel Electroencephalography (EEG) headset, and the results showed an acceptable accuracy of 93% in the brain waves pattern recognition, which opens the way for designing a reliable IoBCT.

Smart Prediction System for Classifying Mirai and Gafgyt Attacks on IoT Devices

Sapiah Sakri, Rawan Aldawod, Noura Alsaleh, Nojod Aldalbahi, Reema Alqahtani
College of Computer & Information Science, Princess Nourah bint Abdulrahman University,
Riyadh, Saudi Arabia

Abstract - The proliferation of Botnet attacks on IoT devices indicates that IoT network traffics is more vulnerable than other IT-based device network traffic. Mitigating this threat has led to new techniques for identifying attacks initiated by infected IoTbased devices. The study proposed an intelligent system (Maltect) that could predict botnet attacks on IoT-based devices by utilizing the machine learning model as the prediction engine. During the implementation process, there are two main parts; firstly, to perform the model classification using machine learning (ML) algorithms to determine the best-fitted model, and second, to develop the web-based system. For model classification, the study deployed the N-BaIoT dataset to train two renowned superior performance classifiers based on previous studies. Support-Vector Machine (supervised learning algorithm) and Extreme Gradient Boosting (ensemble learning algorithm) were the deployed classifiers. The study evaluated the models based on the Accuracy, Precision, Recall, and F-measure performance metrics for each attack type according to three types of IoT devices: the doorbell, security camera, and thermostat. The results denote that Extreme Gradient Boosting was the most performing model, achieving 99.9% accuracy in predicting attacks on all the IoT devices.

Humanoid Robot Acceptance: A Systematic Review of Literature

Reilly Moberg, Arshia Khan
Department of Computer Science, University of Minnesota Duluth, Minnesota, USA

Abstract - As a relatively new area of study, the field of humanoid robotics is a fascinating topic of discussion. This concise literature review analyzes and discusses the different papers and studies from within the field of humanoid robot acceptance. We look at the different types of robots used in each study, the roles of participants, the cultural acceptance across different countries, and the trends in methodology and technology over the course of the 21st century. Overall, we found that the approach towards the acceptance of humanoid robots must be specialized by taking into account the user's age, gender, culture, and experience with modern technology, among other factors, due to the varying attitudes towards humanoid robots that accompany different human characteristics.

Towards a Smarter IoT Environment with Ethereum Virtual Machine Enabling Ledgers

Sandi Gec, Dejan Lavbi., Vlado Stankovski, Petar Kochovski
Faculty of Computer and Information Science, University of Ljubljana, Slovenia

Abstract - Distributed Ledger Technology (DLT), from the initial goal of moving digital assets, allows more advanced approaches as smart contracts that are executed on distributed computational enabling nodes such as Ethereum Virtual Machines (EVM) initially available only on Ethereum ledger. From the launch of new cryptocurrencies it was reasonable to enable EVM-based smart contracts and thus incentive the integration of smart contract on other domains such as IoT environments. In this paper, we analyze the most IoT environment expedient quantitative metrics of various popular EVM-enabling ledgers to provide an overview of potential EVM-enabling characteristics.

Security Model over Mobile Internet of Things in Wireless Networks: A Review

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

Abstract - Over the past years, the use of Internet of Things (IoT) technologies such as physical objects, electronics, sensors, and links has increased. These technologies connect objects in order to exchange information using wireless networks (WNs) which includes different objects such as serves, connected devices and centralized systems which are based on different communication infrastructure. The packets in IoT network are collected using different sensors, nodes, and collectors which transfer the information into the cloud via the internet. The consumers, healthcare, businesses as well as governments utilizes the IoT to gather information from different environments. The IoT devices are being introduced each day all over the world which are used in WNs. This raised more concern about the security issues in IoT such as spoofing attack which most of the current algorithm are failing to prevent in WNs.

Study of Ultrasonic Data Collection System for Anomaly Detection on Mechanical Equipment

Rio Shigyo, Daiki Nobayashi, Kazuya Tsukamoto, Mitsunori Mizumachi, Takeshi Ikenaga
Kyushu Institute of Technology, Fukuoka, Japan

Abstract - IoT-based acoustic data collection systems have been proposed for detecting anomalies in equipment. However, such systems do not work well in outdoor environments due to interference caused by various sound sources (e.g., wildlife sounds and traffic noise). Outdoor noise has unique frequency characteristics, with energy localized in the low-frequency domain. In contrast, mechanical equipment such as motors mainly generates harmonics, including ultrasonic components, during operation. This paper proposes a noise-robust ultrasonic data collection system for outdoor use. To reduce the size of collected data, which increases in proportion to the sampling frequency, the proposed system uses bandpass filtering and data compression.

5G Reference IoT Board for Smart Farm

So-Young Kwon, Young-Hyung Kim, Dong-Soo Jung, Ki-Ryong Cha, Yong-Hwan Lee
School of Electronic Engineering, Kumoh National Institute of Technology, Gumi, Korea;
Department of IT Convergence, Kumoh National Institute of Technology, Gumi, Korea;
Raonsolution, Gumi, Korea;
Sungun Materials, Gumi, Korea

Abstract - Recently, smart farms are being introduced to solve problems in farms such as severe manpower shortage and global warming. Currently, 4G and LTE are being used in smart farms. Using 5G, which is about 20 times faster than 4G and LTE, can build a smart farm with improved quality. In this paper, we propose a 5G reference IoT board for smart farms. Sensors that measure temperature, carbon dioxide and illuminance are used to measure the state of the smart farm. In our system, each sensor seamlessly exchanges data with the reference IoT board via 5G communication.

CSCI-RTSC:
RESEARCH TRACK ON SMART CITIES AND SMART MOBILITY

Fallacies in the Implementation of a Local Micro Smart Grid

Steffen Spathe, Peter Conrad
Faculty of Mathematics and Computer Science, Friedrich Schiller University Jena, Germany

Abstract - The implementation of local micro smart grids has become relevant for many people and poses a milestone on the way to a more sustainable future via efficient usage of locally generated renewable energy. Building a micro smart grid requires integration of decentralized production capabilities, public and private charging infrastructure for electric vehicles, as well as smart control mechanisms to respond to energy demand in realtime. Due to the variability and individuality of such systems, there is as yet no codified way to set them up. Although implementation and operation may seem simple, some pitfalls need to be considered when theory meets reality. In this research paper, we would like to share our experiences from building a micro smart grid in a real-world scenario and describe some of our faulty assumptions and unforeseen technical challenges we encountered during the process. This article is intended to serve in particular as a guide for anyone planning to build a micro smart grid, helping them to avoid annoying and potentially costly mistakes right from the start by planning ahead. In this way, it is especially meant to serve as a thought-provoker and an additional checklist.

Study on Q-Learning and Deep Q-Learning in Urban Roadside Parking Space Search

Hung-Chin Jang, Chun-Yee Lee
Department of Computer Science, National Chengchi University, Taipei, Taiwan, R.O.C.

Abstract - In recent years, most research on urban roadside parking space search has focused on improving the prediction of the vacancy of roadside parking spaces. One simple but expensive practice is setting up sensors in each parking space to provide drivers with real-time parking space information so drivers can find suitable parking spaces. Although providing real-time information on each parking space can help drivers when choosing a driving route, there is a possibility that other drivers take the parking space during the time of getting to the specific parking space. . better approach to the parking space search is to find a suitable parking area rather than specific parking spaces. Predicting the probability of an available parking area can reduce the time the vehicle lingers in search of a parking space. In this study, we proposed to use Deep QLearning with fewer sensors to solve the problem. Besides, we used LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) models to improve the accuracy in estimating the Q value of the Deep Q-learning. Finally, we compared the performance of Q-Learning and Deep Q-Learning using simulated traffic flow data.

Data Driven Characterization and Predictive Classification of Energy Economy for Public Transport

Roman Sennefelder, Ruben Martin-Clemente, Ramon Gonzalez-Carvajal
Research & Development, EVO Engineering GmbH, Munich, Germany;
Electronics Engineering Department, Universidad de Sevilla, Seville, Spain

Abstract - The transformation towards sustainable mobility in public transportation sector requires thorough understanding of the use case, thus real world driving data and transparency in vehicle energy economy. Uncertainty about the energy demand leads to conservative design which lasts in inefficiency and high costs. Predicting the energy economy precisely upfront would significantly reduce costs and enhance fleet operations. Within this paper, we introduce a data driven approach for characterization and predictive classification of electric city buses by powerful machine learning algorithms. The presented framework facilitates the design and operation planning of alternative fleets in urban environment.

Development and Implementation of a Communication Network in a Formula SAE Car

Guilherme P. de Lima, Luiz Carlos G. Sacramento Junior, Cristiano de S. de Carvalho,
Paulo Roberto F. Junior, Luis B. de Assis Junior, Fabricio L. e Silva, Diego N. Brandao
CEFET-RJ, Nova Iguacu, Brazil; NUPEM/CEFET-RJ, Nova Iguacu, Brazil;
CEFET-RJ, Rio de Janeiro, Brazil; PPCIC/CEFET-RJ, Rio de Janeiro, Brazil

Abstract - This paper presents the development and implementation of a communication network to compose a data acquisition and telemetry system for a Formula SAE race car using lowcost components. In addition, it presents a brief review of articles on the theme. In order to reach a feasible and robust system, a combination of two protocols was used in the proposed framework. They collected data through several sensors installed in the vehicle, treated the data in a decentralized way with Electronic Control Units (ECUs), and shared them in a single bus through the CAN protocol. The information is sent via a wireless transmission with the Zigbee protocol to a remote monitoring station. The system performs each of these tasks with low implementation cost and outstanding reliability of shared data.

Methods and Tools for Monitoring Driver's Behavior

Muhammad Tanveer Jan, Sonia Moshfeghi, Joshua Conniff, Jinwoo Jang, Kwangsoo Yang,
Jianna Zhai, David Newman, Ruth Tappenm, Borko Furht
College of Engineering and Computer Science, Florida Atlantic University, Boca Raton, Florida, USA;
Christine E. Lynn College of Nursing, Florida Atlantic University, Boca Raton, Florida, USA

Abstract - In-vehicle sensing technology has gained tremendous attention due to its ability to support major technological developments, such as connected vehicles and self-driving cars. In-vehicle sensing data are invaluable and important data sources for traffic management systems. In this paper we propose an innovative architecture of unobtrusive in-vehicle sensors and present methods and tools that are used to measure the behavior of drivers. The proposed architecture including methods and tools are used in our NIH project to monitor and identify older drivers with early dementia.

An Empirical Investigation on Linking Ethics with the Acceptance of Autonomous Vehicles

Konstantina Karathanasopoulou, Ilias Panagiotopoulos, George Dimitrakopoulos, Konstantina Marra
Department of Telematics and Informatics, Harokopio University of Athens, Greece

Abstract - Autonomous vehicles (AVs) are planning to be available on market, in full scale, in the next few years. The advent of AVs aims to bring substantial benefits like accessibility to transport, traffic flow, safety, fuel use, emissions, and comfort. All these potential societal benefits will not be achieved unless AVs are accepted and used by a critical mass of people. Among others, AVs technology acceptance is also influenced by ethics, which constitutes an extremely significant factor that differentiates things from what we knew up to now and can influence the acceptance of AVs. Addressing these challenges, the present study aims at investigating whether or not ethical issues, in the scope of unavoidable traffic accidents, influence the acceptance of AVs by the general public, and, if so, to what extent and in which manner. For that purpose, this paper: (a) proposes a technology acceptance modeling process by extending the original Technology Acceptance Model (TAM) to explain and predict consumers. intensions towards AVs programmed according to certain moral codes, (b) based on the proposed TAM-extended framework, a 30-question survey was conducted in order to obtain a better understanding of perspectives, concerns and the factors influencing consumers. intensions to use and accept AVs programmed with ethical rules. Results show that the constructs of perceived usefulness, perceived ease to use, perceived trust and social influence, are all useful predictors of behavioral intentions to have or use ethical AVs, with perceived usefulness having the strongest impact.

Independent and Non-Invasive State Detection for Elevator Based on Total Current Consumption

Pirmin Held, Daniel WeissHaar, Djaffar Ould Abdeslam, Dirk Benyoucef
Furtwangen University, Furtwangen, Germany;
Universite de Haute-Alsace, Mulhouse, France

Abstract - Elevators contribute significantly to the electricity consumption of residential buildings, office buildings and commercial enterprises. In this paper, the electricity consumption is investigated using an elevator system and its individual operating states as example. In addition to analyzing and allocating the energy demand, this work examines how the individual operating states can be determined solely on the basis of the power consumption of the elevator. The knowledge gained from this, such as the usage behavior, the travel profile, or load, is determined independently of the elevator control system. A subsequent installation on any system can be easily realized. In this work, the investigated elevator requires a substantial part of the total annual power consumption in standby (>90 %). This shows an enormous potential for energy savings. The individual elevator states, as well as the load, can be detected very well on the basis of the measured total power consumption. The work thus shows exemplary the potential of an intelligent measurement system for the state detection of elevator systems.

A Contrastive Learning based CNN-GRU Model for Time Series and Its Application for Water Quality Prediction

Yahong Lian, Jing Yun, Yuncheng Wang, Zhiwei Xu
College of Data Science and Application, Inner Mongolia University of Technology, Hohhot, P. R. China

Abstract - Water quality monitoring is of great importance in sustainable development of intelligence city. One of the main challenge of this data is that the lack of labeled data. Therefore, the unsupervised representative framework which can produce reliable results could be of great value. It is key to applying contrastive learning, which can be used in environmental data, where there exists lots of unlabeled sequential data. This work proposes a contrastive learning based CNN -GRU model which exploits a data augmentation scheme in which new instances are generated by mixing two data instances with a mixing component and utilize CNN and GRU model to extract more useful representation from time series data. To show the effectiveness, we conduct comparative experiments on groups of time series dataset, and apply our model to a real-world water quality prediction task where the data comes from Erdaohezi, Tongliao, China.

Modeling of DC House Distribution Network

Lauren Rotsios, Taufik Taufik, Siddharth Vyas, Majid Poshtan, Ali Dehghan Banadaki
Electrical Engineering Department, California Polytechnic State University,
San Luis Obispo, California, USA

Abstract - This paper presents modeling and simulation of a distributed network of DC Houses using MATLAB Simulink. The model will allow for sharing of power between houses within the network. The developed model consists of five separate DC House branches with local power generation. Each branch consists of a PV MPPT charge controller subsystem, a resistive load, and a bidirectional buck-boost converter subsystem. The performance of the individual components of the model is verified before constructing the network. The power sharing capability of the network was evaluated by measuring the efficiency transmission at varying wire gauges, distances, and high-end voltages. Results of the study show that for the most part, higher transmission voltage resulted in higher efficiency. However, this was not the case at some configurations due to different methods of power sharing.

Waste Management System

Mesk Saad Aldossari, Taghreed Assaf Almarshadi, Reem Abduallah Alwasel,
Rahaf Saeed Alqahtani, Ala Saleh D. Alluhaidan
Information Systems Department, College of Computer and Information Science,
Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

Abstract - Waste management is one of the most critical issues that must be addressed in the public and private sectors, given the recent increase in population and the importance of the environment protection. Waste management has its role in numerous areas, including the economic field. It plays an essential role in preventing environmental pollution. Unsafe disposal or improper disposal can create community havoc and disrupt the environment ecology. This project's contribution is manifested in creating a waste management system application which users will use to create a claim when waste is found, the claim will be analyzed to identify the risk to help organizations interested in managing waste and keeping the green environment. The system's expected outcomes are 1/ reducing waste management resources such as: time and money, 2/ helping to make the best decisions based on monthly and annual reports that the system will provide. All of this plays an important role in maintaining the ecosystem.

Reaching Sustainable Development Goals Through Tax Reductions and Incentives within a Trustworthy Decentralized Autonomous Organization

Petar Kochovski, Vlado Stankovski, Sandi Gec, Maja Klun, Iga Kotnik
Faculty of Computer and Information Science, University of Ljubljana, Slovenia;
Faculty of Public Administration, University of Ljubljana, Slovenia

Abstract - Fulfilling the Sustainable Development Goals (SDGs) relies on the proper implementation of various governmental policies, among which are also financial and taxation policies. The current public administration in Slovenia is based on a fully digitalized data management process. Our goal is to achieve new data integrity methods based on state-of-the-art technologies of the ONTOCHAIN software ecosystem. ONTOCHAIN includes technologies to achieve decentralised knowledge graphs, smart oracles, smart contracts, and tokenomics principles that rely on high integrity and trustworthiness. Consequently, we study the possibility to realize flexible incentives and taxation policies that contribute to the seamless achievement of SDGs.

GIS Applications in Air Pollution Modeling

Daniel Isakovic, Josh Jordan, Connor Bowyer, Isabella Jamil, Elias Qamar, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - Geographic Information Systems (GIS) is a system tool that started in the early 1960s. GIS is used to map a wide range of information and analyze different types of data. GIS incorporates information with a wide range of expressive data. This data is used to recognize geographical patterns and trends so that we can deepen our understanding of the environment and plan for the future. The advantages incorporate proficiency and superior navigation as well as a stronger administration. In this paper, we will discuss the many different applications of GIS in air pollution modeling, the advancements within air pollution modeling and answer challenges that GIS currently face.

Detecting Anomalous Energy Consumptions in Smart Buildings - an Overview of two Unsupervised Techniques as a Starting Point for our Research

Cindy Mund, Lena Charlotte Altherr
Electrical Engineering and Information Technology,
FH Aachen University of Applied Sciences, Aachen, Germany

Abstract - In context of the increasing importance of monitoring our energy consumption due to our energy reserves running short, this paper gives an overview of existing possibilities to detect anomalous energy consumptions in smart cities based on unsupervised machine learning. After defining and presenting different types of anomalies, practical applications and methods for anomaly detection in context of smart cities are discussed.

CSCI-RTCC:
RESEARCH TRACK ON CLOUD COMPUTING AND DATA CENTERS

Container-based VoIP Failover Design on Kubernetes Overlay Networks

Cai-Sian Jhuang, Quincy Wu
Department of Computer Science and Informance Engineering,
National Chi Nan University, Nantou, Taiwan

Abstract - This paper studied how to utilize a popular container scheduling orchestrator - Kubernetes (K8s) - in today's cloud computing scenario to support the Voice over IP (VoIP) applications. In VoIP, two major protocols - Session Initiation Protocol (SIP) and Real-time Transport Protocol (RTP) - usually require network traffic to obey the following pattern: the packets from both end devices must be directed to the same server during a call. However, in a K8s network environment, the automatic LoadBalancer mechanism only ensures that traffic is directed to a healthy Pod but does not consider the need for both sides of the session to be handled by the same server. Therefore, after VoIP is introduced into the container, some subtle design must be proposed to integrate SIP and RTP into the existing K8s network architecture to handle two-way communication and failover so that the above problem can be properly addressed. In this paper, we use the K8s Label selector, K8s leader-election, and a few opensource software to solve the problem of failover, and measure the pressure of failover on K8s and the underlying operating system.

Response-Time Tail Mitigation in Cloud Systems

Asaad Althoubi, Hassan Peyravi
Department of Computer Science, Kent State University, Kent, Ohio, USA

Abstract - Response time is one of the most important metrics in cloud data centers that measure application performance and availability. The response time depends on how the scheduler distributes tasks belonging to a job to various cores. Even with at a maximum level of parallelism, the slowest task (the straggler) dominates the response time, thereby increasing the average and extending the tail of the response time distribution. In this paper, we propose a distributed scheduling scheme where tasks are systematically assigned to randomly selected servers based on the expected response time of each server and the size of the tasks. Consequently, straggling tasks are less likely to become head-of-line blockers and are more likely to finish in a shorter period of time. In light of the bursty nature of task arrivals, the instantaneous response time is not an accurate measure of a server's load. Therefore, we developed a Truncated Exponentially Weighted Moving Average (TEWMA) for the selection of servers. To adapt to the new schemes, we augmented the Sparrow simulation and compared its performance to Sparrow's. Our approach probes fewer servers and selects most suitable ones based on a simple matching algorithm. The results have shown that the tail of response time distribution has been declined considerably, and consequently the average response time has been reduced accordingly.

Evaluation of Security Issues in Cloud-Based File Sharing Technologies

Sikhumbuzo Ngwenya, Munienge Mbodila, Omobayo A. Esan, Femi Elegbeleye
Department of Information Technology Systems, Walter Sisulu University, Eastern Cape, South Africa

Abstract - Cloud computing is one of the most promising technologies for backup and data storage that provides flexible access to data. Cloud computing plays a vital role in the remote backup. It is so unfortunate that this computing technique has flaws that thrilled and edgy end users in implementing it effectively. These flaws include

factors like lack of integrity, confidentiality, and privacy of information. A secure cloud is impossible unless the computer-generated environment is appropriately secured. In any form of technology, it is always advisable that security challenges must be prior identified and fixed before the implementation of that technology. Primarily, this study will focus on identifying security issues in cloud computing with the objective of identifying concerns like credential theft and session management in the "Cloud". Main arguments like HTTP banner disclosure, Bash "ShellShock" Injection, and password issues were discovered during the stages of study implementation. These challenges may provide information that will permit hackers in manipulating and exploiting the cloud environment. In identifying credential theft and session management in cloud-based file sharing technologies a mixed method approach was implemented throughout the course of the study due to the nature of the study and unity of analysis. Penetration tests were performed as security testing techniques. Prevention and guideline of security threats lead to a friendly and authentic world of technology.

Serverless Cloud Functions - Opportunity in Chaos

Madhavi Jaival, Katya Mkrtchyan, Adam Kaplan
Department of Computer Science, California State University (CSU) Northridge, California, USA

Abstract - Due to its cost-effectiveness and limited scope of administration, Serverless Computing has fast become a favorite cloud computing execution model. Meanwhile, with the rise of distributed cloud architectures and microservices in the last decade, many development teams have adopted the principles of Chaos Engineering. This allows them to assess the effects of random failures or delays on an application. In prior literature, serverless developers measured and reported cold-start penalties and transaction latency, whereas Chaos Engineers have studied security and resiliency in cloud infrastructure. In this work, we combine these approaches to measure the performance of a set of serverless cloud functions which implement common server-side file and database operations. We study each function's performance response under a set of controlled chaos experiments, wherein we emulate various client load conditions, as well as inject random delays into the function execution. We find that under heavy 1000-client load, the longest-latency operations can provide as much as 36.5% improvement to overall response time by failing early.

Performance Assessment for Scheduling Algorithms in Cloud Computing

Mimouna Alkhonaini
Department of Computer Science, Prince Sultan University, Riyadh, Saudi Arabia

Abstract - The concept of cloud computing describes a new category of network-based computing that has taken over the internet technology. The primary advantage of cloud computing technology is application scalability. Cloud computing offers great benefits for the application that share their resources on nodes that are different. Scheduling the tasks can be quite challenging in a cloud computing environment. However, there is need to propose new scheduling strategies to counter the challenges that are presented by network properties between the user and resources. The new strategies can make use of the available conventional scheduling concepts to join them with other network strategies to bring solutions for more effective and efficient job scheduling. Job scheduling is a core technology in cloud computing. This paper aims at providing a study on performance assessment of scheduling algorithms and systemizes the schedule challenges in both existing and proposed cloud computing scheduling algorithms. A brief analysis of numerous scheduling parameters considered in these algorithms is also presented. The results of the experiments show that the best performance level can be achieved by using HPC approach.

Approach of Migrating SAAS Applications to Password-less Authentication

Iehab Al Rassan, Hala Alajlan
Computer Science Department, King Saud University, Riyadh, Saudi Arabia

Abstract - Passwords are the most used authentication method. even after two decades of attempts to replace passwords with other alternatives, it is still dominating over other authentication mechanisms. Passwords are full of weaknesses and vulnerabilities, they are hard to remember, and users tend to reuse passwords for different services. As passwords are full of weaknesses, a more secure alternative is needed. Password-less authentication is a great alternative to passwords, as they replace something the user know with something the user has and it cannot be forgotten. Many password-less authentication method were proposed over the years, including magic link, one time password and cryptographic based. in this work we propose migrating Software as a service application-s to use password- less authentication using Fast Identity Online standards. Migrating to password-less is not easy, and it requires time and skills. in our proposal, we propose a middleware that will implement the password-less authentication and act between the user and the Software as a service application, without any modification to the Software as a service side.

A Review of Security Models to Detect and Prevent Advanced Persistent Threats over Cloud Computing

Lawrence Mashimbye, Topside E. Mathonsi, Vusumuzi Malele
Tshwane University of Technology, Pretoria, South Africa

Abstract - Cloud computing makes it easier and more efficient for businesses and individuals to access data and services. As the number of users and data increases, cloud computing becomes more vulnerable to targeted attacks. In complex IT systems like the cloud, advanced persistent threat (APT) attacks steal data from users daily and are extremely difficult to detect. Using continuous monitoring and defense systems to detect and prevent APT attacks has proven very effective for detection and prevention. These systems, however, have limitations since they rely on known attack signatures, and attackers constantly change characteristics to escape detection. In this paper, we propose a security model to defend cloud data against these types of attacks and overcome the limitations of existing techniques. This model simplifies classification by identifying complex relationships between database features. Attackers can scan the network for open network services, security perimeter systems, and individuals with access to targeted data. Honeypots combined with Moving Target Defense (MTD) will provide the best reconnaissance protection. To enhance the security of cloud information systems against attacks, we propose Multifactor authentication with a one-time password.

CSCI-RTPD:
RESEARCH TRACK ON PARALLEL & DISTRIBUTED COMPUTING

A Concurrent Relaxed Queue for Handling Unordered Parallel Accesses on GPUs

Mengshen Zhao, David Troendle, Byunghyun Jang
Computer and Information Science, University of Mississippi, Univeristy, Mississippi, USA

Abstract - We propose a relaxed concurrent queue for Graphic Processing Units (GPUs) which supports groups of unordered enqueue and/or dequeue operations. When the group size is one, it becomes a conventional strict First-In-First-Out (FIFO) queue. We call these groups Parallel Operations Groups (POGs) and leverage a persistent thread model to support the processing of an arbitrary number of POGs. To minimize thread contention and synchronization overhead, queue operations in each POG are processed as a group then committed to the queue in a single update process. The experiment compares our proposed queue with a non-blocking concurrent queue (baseline) on synthetic inputs that simulate the diverse configurations of POGs present in realworld applications. The experiments show that our relaxed queue achieves a significant speed up over the baseline when processing a large number of POGs while retaining good scalability.

An Approach to Parallelization of Respiratory Disease Spread Simulations in Emergency Rooms

Martin Paradiso, Lucas Maccallini, Agustina Vericat, Fernando Romero, Diego Encinas
Informatics Research Institute LIDI. Universidad Nacional de La Plata, Argentina;
SISDIC, Hospital San Roque de Gonnet (MSPBA), La Plata, Argentina;
SimHPC-TICAPPS, Universidad Nacional Arturo Jauretche, Florencio Varela, Argentina

Abstract - The agent-based modeling (ABM) is a flexible simulation model that can be easily adapted to the simulation of different problems. A disadvantage of some tools for ABM implementation is the low performance obtained. This paper presents a re-implementation of an in-hospital disease simulator, developed in Repast Simphony, in its high-performance alternative Repast for High Performance Computing. This tool allows multinode execution, allowing execution on multi-core machines as well as on physical and virtual clusters. The objective of this work is to analyze the performance of the new implementation in physical and virtual clusters, and to determine in which scenarios it is justifiable to implement a system in a lower-level framework.

Towards a Generic Framework for GPU-Parallelized Simulations of Light-Driven Particles

Florian Fey, Sergei Gorlatch
University of Muenster, Germany

Abstract – N/A

Distributed Tire Parameter Optimization Using CREATE-GV

Jeremy Mange, Jacob Brendle
US Army, Ground Vehicle Systems Center, DEVCOM GVSC - Analytical Methods, Warren, Michigan, USA

Abstract - Automated optimization of vehicle components according to a set of criteria is desirable in a number of contexts. Modeling and simulation can enable this process; however, it is often computationally intensive, particularly when a large number of potential designs are involved. In this paper, we present a project using high-performance computing (HPC) resources to perform distributed optimization of tire parameters through the CREATE-GV mobility analysis software. We demonstrate that this highly parallel computing process can be used effectively to analyze large sets of model alternatives in order to optimize component design for specific vehicle mobility goals.

Performance Evaluation on the Parallel Processing System with the Raspberry Pi 4

Hyunseung Han, Kyungha Kim, Seungwoo Jung, Yunseok Chang
Department of Computer Engineering, Daejin University, Pocheon, Republic of Korea

Abstract - Cost performance is one of the essential perspectives in designing and building parallel processing systems. Building a high-performance parallel processing system at a low cost is complicated work. Still, the recent development of computer and network technology provides a foundation for a rapid increase in the cost-performance ratio. This study verified whether the Raspberry Pi 4 clusters could show scalability and performance similar to other cluster systems from the cost-performance viewpoint. To evaluate the performance, we build a cluster system with 16 Raspberry Pi 4s and a high-speed network switch and perform the HPL benchmark. The experiment showed that the Raspberry Pi 4-based cluster system has parallel scalability according to cluster size and similar processing performance to other parallel processing systems. Therefore, it can expect that a cluster system with a low-cost processing device such as Raspberry Pi 4 could be applied to the high-performance cluster system.

CSCI-RTPC:
RESEARCH TRACK ON SIGNAL & IMAGE PROCESSING, COMPUTER VISION & PATTERN RECOGNITION

Spatial Resolution Enhancement for Halftone Images using Convolutional Neural Networks

Guilherme Apolinario Silva Novaes, Hae Yong Kim
Dept. Eng. Sistemas Electronicos, Escola Politecnica, Universidade de Sao Paulo, Sao Paulo, Brazil

Abstract - To print a grayscale image on a printer with current technology (laser or inkjet), it is necessary to convert it into a binary image called halftone, so that when a human being looks at it from a distance, the binary image visually resembles the original image. In this work, we present a technique to increase the spatial resolution of halftone images based on convolutional neural networks (CNNs). To our knowledge, this is the first work that seeks to solve this problem using CNNs. We use our algorithm to increase the resolution of binary halftones and show that our algorithm is considerably better the previous decision-tree based method. We also use our algorithm to increase the resolution of scanned old comics pages, and show that our algorithm is better than conventional resampling methods.

Fast Key Points Detection and Matching for Tree-Structured Images

Hao Wang, Xiwen Chen, Abolfazl Razi, Rahul Amin
School of Computing, Clemson University, South Carolina, USA;
MIT Lincoln Laboratory, Massachusetts Institute of Technology, Massachusetts, USA

Abstract - This paper offers a new authentication algorithm based on image matching of nano-resolution visual identifiers with tree-shaped patterns. The algorithm includes image-to-tree conversion by greedy extraction of the fractal pattern skeleton along with a custom-built graph matching algorithm that is robust against imaging artifacts such as scaling, rotation, scratch, and illumination change. The proposed algorithm is applicable to a variety of tree-structured image matching, but our focus is on dendrites, recently-developed visual identifiers. Dendrites are entropy rich and unclonable with existing 2D and 3D printers due to their natural randomness, nano-resolution granularity, and 3D facets, making them an appropriate choice for security applications such as supply chain trace and tracking. The proposed algorithm improves upon graph matching with standard image descriptors. It also improves [1], which faces various problems when deploying in real-world applications. For instance, image inconsistency due to the camera sensor noise may cause unexpected feature extraction leading to inaccurate tree conversion and authentication failure. Also, previous tree extraction algorithms are prohibitively slow hindering their scalability to large systems. In this paper, we fix the current issues of [1] and accelerate the key points extraction up to 10- times faster by implementing a new skeleton extraction method, a new key points searching algorithm, as well as an optimized key point matching algorithm. Using minimum enclosing circle and center points, make the algorithm robust to the choice of pattern shape. We show that our algorithms outperform standard key descriptors such as SIFT, FAST, and ORB in terms of accuracy (with a margin of 6% to 10%), while utilizing much fewer key points (about 20% to 80% reduction). In contrast to [1] our algorithm handles general graphs with loop connections, therefore is applicable to a wider range of applications such as transportation map analysis, fingerprints, and retina vessel imaging.

Foreign Object Debris Detection for Airport Pavement Images based on Self-supervised Localization and Vision Transformer

Travis Munyer, Daniel Brinkman, Xin Zhong, Chenyu Huang, Iason Konstantzos
Department of Computer Science, University of Nebraska Omaha, Omaha, Nebraska, USA;
Aviation Institute, University of Nebraska Omaha, Omaha, Nebraska, USA;
Durham School of Architectural Engineering and Construction, University of Nebraska Lincoln, Lincoln, Nebraska, USA

Abstract - Supervised object detection methods provide subpar performance when applied to Foreign Object Debris (FOD) detection because FOD could be arbitrary objects according to the Federal Aviation Administration (FAA) specification. Current supervised object detection algorithms require datasets that contain annotated examples of every to-be-detected object. While a large and expensive dataset could be developed to include common FOD examples, it is infeasible to collect all possible FOD examples in the dataset representation because of the openended nature of FOD. Limitations of the dataset could cause FOD detection systems driven by those supervised algorithms to miss certain FOD, which can become dangerous to airport operations. To this end, this paper presents a self-supervised FOD localization by learning to predict the runway images, which avoids the enumeration of FOD annotation examples. The localization method utilizes the Vision Transformer (ViT) to improve localization performance. The experiments show that the method successfully detects arbitrary FOD in real-world runway situations. The paper also provides an extension to the localization result to perform classification; a feature that can be useful to downstream tasks. To train the localization, this paper also presents a simple and realistic dataset creation framework that only collects clean runway images. The training and testing data for this method are collected at a local airport using unmanned aircraft systems (UAS). Additionally, the developed dataset is provided for public use and further studies.

Perspective Transformation Layer

Nishan Khatri, Agnibh Dasgupta, Yucong Shen, Xin Zhong, Frank Y. Shih
Department of Computer Science, University of Nebraska Omaha, Omaha, Nebraska, USA;
Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey, USA

Abstract - Incorporating geometric transformations that reflect the relative position changes between an observer and an object into computer vision and deep learning models has attracted much attention in recent years. However, the existing proposals mainly focus on the affine transformation that is insufficient to reflect such geometric position changes. Furthermore, current solutions often apply a neural network module to learn a single transformation matrix, which not only ignores the importance of multi-view analysis but also includes extra training parameters from the module apart from the transformation matrix parameters that increase the model complexity. In this paper, a perspective transformation layer is proposed in the context of deep learning. The proposed layer can learn homography, therefore reflecting the geometric positions between observers and objects. In addition, by directly training its transformation matrices, a single proposed layer can learn an adjustable number of multiple viewpoints without considering module parameters. The experiments and evaluations confirm the superiority of the proposed layer.

Transform Decomposition Switching for Efficient Attribute Compression of 3D Point Clouds Using Neural Networks

Reetu Hooda, W. David Pan, Bernard Benson
Department of Electrical & Computer Engineering, University of Alabama in Huntsville, Alabama, USA;
McLeod Software Corporation, Birmingham, Alabama, USA

Abstract - An adaptive technique to switch between RAHT and Dyadic RAHT using 3D Sobel filter has been found to improve the compression in 3D point clouds by offering substantial cumulative compression gains. However, the drawback of this switching scheme is its need for tuned thresholds. To this end, we propose to use neural networks to resolve the threshold dependency issue so that the switching becomes truly adaptive. Two publicly available point cloud datasets were used to test the effectiveness of the proposed method. We achieved significant gains on MVUB and minor gains on 8iVFB dataset over all Dyadic approach.

Smartphone-Based Turbidity Estimation with Inherent Calibration

Ding Zhang, Alan Marchiori, Joshua Stough
Department of Computer Science, Bucknell University, Lewisburg, Pennsylvania, USA

Abstract - Water is one of the most vital natural resources; it is essential to daily life. In many developing countries, however, access to clean and safe water is a crucial issue due to a lack of economic and infrastructure resources. The most broad and universal measure of water quality is turbidity. Traditional labuse turbidimeters, though highly accurate, may be prohibitively expensive for wide-scale use. Thus, proposing cost-effective water turbidity estimation systems has long been a trend. In this paper, we introduce an innovative water turbidity estimation system. Different from other approaches which have only one water sample for analysis, our system consists of a low-cost illuminated cuvette holder containing a test and a control sample, where the control always contains clear water. Acquired images of the two samples are input into an image processing chain to estimate the sample's turbidity. We evaluated our approach in both lab and in-situ environments and found that in the lab environment our approach achieved mean error of 19.07 NTU over the range 16.1 to 417 NTU compared to 69.86 NTU without using the control sample. In outdoor real-world use, we found mean error of 20.68 NTU in the shade and 57.34 NTU in full sun.

Interpretable Generative Modeling using a Hierarchical Topological VAE

Etienne Desticourt, Veronique Letort, Florence d'Alche-Buc
Centre d'Excellence Systeme, ALMS, Air Liquide, Antony, France;
MICS, CentraleSupelec, Universite Paris-Saclay, Gif-sur-Yvette, France;
LTCI, Telecom Paris, IP Paris, Palaiseau, France

Abstract - Generating realistic datasets with fine-grained control over their properties can help overcome challenges linked to the scarcity of data in many domains, such as medical applications. To that end, we extend Variational Autoencoders by using a hierarchical and topological prior consisting of a sequence of Self-Organizing Maps (SOM), which are stacked in the latent space and learned without supervision, jointly with the parameters of the variational autoencoder. We induce a hierarchy between the codes of the SOM sequence, each SOM corresponding to a different hierarchical level and learning increasingly finer-grained representations of the data. Our model combines the power of deep learning with the interpretability of hierarchical and topological clustering and produces competitive results when evaluated on three well-known computer vision benchmarks and a custom medical dataset.

Fixed-point Iterative Computation of Gaussian Barycenters for Some Dissimilarity Measures

Alessandro D'Ortenzio, Costanzo Manes, Umut Orguner
Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Italy;
Department of Electrical and Electronics Engineering, Middle East Technical University, Turkey

Abstract - In practical contexts like sensor fusion or computer vision, it is not unusual to deal with a large number of Gaussian densities that encode the available information. In such cases, if the computational capabilities are limited, a data compression is required, often done by finding the barycenter of the set of Gaussians. However, such computation strongly depends on the chosen loss function (dissimilarity measure) to be minimized, and most often it must be performed by means of numerical methods, since the barycenter can rarely be computed analytically. Some constraints, like the covariance matrix symmetry and positive definiteness can make non-trivial the numerical computation of the Gaussian barycenter. In this work, a set of Fixed-Point Iteration algorithms are presented in order to allow for the agile computation of Gaussian barycenters according to several dissimilarity measures.

Attentional Convolutional Neural Network for Automating Pathological Lung Auscultations Using Respiratory Sounds

Md. Motiur Rahman, Shiva Shokouhmand, Miad Faezipour, Smriti Bhatt
School of Engineering Technology, Electrical & Computer Engineering Technology,
Purdue University, West Lafayette, Indiana, USA;
Department of Computer and Information Technology, Purdue University, West Lafayette, Indiana, USA

Abstract - Respiratory diseases are one of the most prevailing diseases which can be limited by ensuring continuous monitoring and treatment through automating lung auscultations. In this paper, an attentional deep learning model has been proposed with a new data augmentation technique named as homogeneous padding to predict the respiratory diseases using the audio data of the ICBHI-2017 dataset. Since the dataset is imbalanced over the classes as well as the devices used to collect the data, we have performed ablation studies over the loss functions, and reported the appropriate loss function which performs better for this dataset. We have also performed experimental analysis to report the appropriate position of attention, and observed that the attention mechanism on high-level features has worked better in comparison with the low-level features. We have found that our developed model has outperformed most of the recent convolutional neural network (CNN)-based models, and the inclusion of an attention mechanism has contributed significantly to improve the accuracy of the model.

Sentence-Level Sign Language Recognition Framework

Atra Akandeh
Circuits, Systems, and Neural Networks (CSANN) Laboratory, Computer Science and Engineering,
Michigan State University, East Lansing, Michigan, USA

Abstract - We present two solutions to sentence-level SLR. Sentence-level SLR requires mapping videos of sign language sentences to sequences of gloss labels. Connectionist Temporal Classification has been used as the classifier level of both models to avoid pre-segmenting the sentences into individual words. The first model is an LRCN-based model and the second model is a Multi-Cue Network. In the first approach, no prior knowledge has been leveraged. Raw frames are fed into an 18- layer LRCN with a CTC on top. In the second approach, three main characteristics (hand shape, hand position, and hand movement information) associated with each sign have been extracted using Mediapipe. 2D landmarks of hand shape have been used to create the skeleton of the hands and then are fed to a CONV-LSTM model. Hand locations and hand positions as relative distances to head are fed

to separate LSTMs. All three sources of information have been then integrated into a Multi-Cue network with a CTC classification layer. We evaluate the performance of proposed models on RWTH-PHOENIXWeather. After performing an excessive search on model hyperparameters such as the number of feature maps, input size, batch size, sequence length, LSTM memory cell, regularization, and dropout, we achieve relatively low Word Error Rate.

HPGER: Integrating Human Perception into Group Emotion Recognition

Abolfazl Farahani, Navid Hashemi Tonekaboni, Khaled Rasheed, Hamid R. Arabnia
School of Computing, University of Georgia, Athens, Georgia, USA;
Department of Computer Science, College of Charleston, Charleston, South Carolina, USA

Abstract - Over the past few years, deep neural networks have been widely employed for representation learning and achieved remarkable success in many computer vision tasks such as visual sentiment analysis and emotion recognition. However, identifying the image sentiments similar to what humans do is challenging due to the complexity of raw images and the intangible nature of human visual perception. Besides, training a deep learning model from scratch for a complex task such as group emotion recognition is very time-consuming and requires a large amount of labeled data representing the total population which is practically infeasible. In addition, creating annotated data for each specific task is costly and sometimes impossible. So instead, we can use the knowledge extracted by a model trained on related existing labeled datasets. To address the above challenges, we propose an end-to-end group emotion recognition framework that integrates human perception learned from human eye fixation data to efficiently and effectively extract the sentiment of input images and classify them as positive, negative, or neutral. The proposed architecture aims to leverage the information in eye fixation data to learn how humans perceive and interpret the visual world in a free-viewing task. Our framework can be utilized in any free-viewing task, potentially reducing the need for a large amount of labeled data and speeding up the training phase. In the following sections, we outline the initial steps and describe different parts of our architecture.

Performance Analysis of Geo-fencing Effect for Intruder Monitoring

Young-il Kim, Park Seong Hee, Heo Tae-Wook, Lee Tae-Chong, Yong-Hyun Yeun, Kim Seung Tae
Electronics and Telecommunication Research Institute, Daejeon, Korea;
MGT, Daejeon, Korea; MSOTEK, Daejeon, Korea;
Agency for Defense Development, Daejeon, Korea

Abstract - In order to protect major facilities from external intruders, a method for establishing and operating an image surveillance system that detects, classifies, and tracks intruders using an image deep learning technology has been applied in various fields. Since the image deep learning performance depends on the number of pixels of the image object and the quality of the image, it is very important to acquire a good quality of images in a large monitoring area. Based on YoLo (You look once), the detection accuracy according to the image size change of the intrusion object was analyzed. A large number of cameras must be used to monitor large areas, but reducing the number of cameras used to reduce budgets has become an important issue. When using a small number of cameras, a blind spot is created as the surveillance direction of the camera must be shifted according to the movement of the intruder. In this paper, the modeling parameters of the camera were analyzed to establish the intruder monitoring area, additionally, in an environment where Geo-fencing is established using a small number of cameras, the number of cameras required according to the parameter settings of the cameras, and the effect on the handover area between intruder surveillance cameras required to track the movement of intruders were analyzed.

Smart Observer for Distant Water Fishing in Taiwan

P. Karthikeyan, Tan-Khang Luong, Chih Chun Chang, Maram Soni Reddy,
Hsiu-Chun Hsu, Pao-Ann Hsiung
Department of Computer Science and Information Engineering,
National Chung Cheng University, Chiayi, Taiwan

Abstract - Investigations in recent years have revealed alarming incidents of human rights violations in the distant water fishing industries in Taiwan. Although the Taiwanese government recently changed fundamental rules to protect migrant fishermen, progress has been insufficient. The government and fishing vessel owners continue to fail to protect the human rights of migrant fishermen working in Taiwan. Taiwan.s government needs smart observers for the distant water fishing industry to detect whether the fishermen are over-working and forced to work high-intensity work with a low salary. Developing a smart observer for distant water fishing is challenging because the videos are captured at night and in a volatile environment. The deep-learningbased tracking of humans also becomes quite challenging due to the multiple occlusions in the working areas, such as fishing equipment and fish processing tools. We proposed a statistical assessment method to check whether the fisherman is overworking. We have implemented the labor intensity assessment method and applied distance correlation to calculate the work intensity of a fisherman. Experimental results show that the statistical assessment method gives a more reasonable prediction than the counting assessment method. The labor intensity assessment method is suitable for differentiating fishermen's low-intensity and high-intensity work.

Multi-Instance Contingent Fusion for the Verification of Newborn and Infant Fingerprints

Tiwalade Odu, Joke Badejo, Tokunbo Ogunfunmi, Aderemi Atayero
Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria;
Department of Electrical and Computer Engineering, School of Engineering,
Santa Clara University, Santa Clara, California, USA

Abstract - There is a need to recognize newborns (1 - 28 days old) and infants (29 days - 12 months old) automatically, with the ubiquitous 500ppi fingerprint reader, to combat newborn swapping, aid identification of missing children, vaccination tracking, medical history, etc. This research seeks to show the possibility of future identification of babies with fingerprints acquired with a 500ppi fingerprint reader, through multi-instance (left thumb and right index fingers) fusion. The fingerprints were acquired from babies who were 1 day - 6 months old at enrollment (Session 1). The sum score fusion algorithm was employed. This method produced verification accuracies of 73.8%, 69.05% and 57.14% for time lapses of 1 month, 3 months and 6 months respectively, between enrollment and query fingerprints. To the best of our knowledge, this study is the first to attempt to verify newborn infant fingerprints with multi-instance contingent fusion. The modelling of the growth of infant fingerprints is recommended to be incorporated to improve verification accuracy.

Weed Recognition in Sesame Field Using Computer Vision Algorithms

Uchechukwu Ngozi Nwachukwu, Cristian Rodriguez Rivero
Cardiff School of Technology Cardiff Metropolitan University, Wales, United Kingdom (UK)

Abstract - Deep learning is one aspect of machine learning that is growing rapidly in today's research, and one of its major applications is the use of computer vision in object detection especially in agricultural sector. In agricultural production, weeds have become a crucial factor affecting it because of its menace to crops and the environmental pollution herbicides used in weeding causes, but with the advancement in agricultural practices, identifying and distinguishing weeds from crop is paramount. To achieve the precise targeting of weeds in a sesame

farm which will go a long way in helping farmers in their decision making, comes the application of effective object detector and image classifier using deep learning. In this paper, machine learning model built with Tensor flow/ keras, and a bounding box model was used as a convolutional neural network in detecting the object in images from the sesame farmland. The images served as an input into the layers of the neural network where the image size was rescaled, max pooled, and important features extracted and then exiting in the two outputs: classification for object class and localization for the bounding box. The train history with the losses that occurred in the train, validation class and bounding box was analyzed, and the classification result of the proposed model showed a satisfactory performance.

Segmentation of Skin Cancer and Intensity Classification Using DCNN

Hanan Aljuaid, Rafaf Alnoghaimshi, Arwa Alhammad, Fay Aldhafer, Noura Alsakran
Department of Computer Sciences, College of Computer and Information Sciences,
Princess Nourah bint Abdulrahman University (PNU), Riyadh, Saudi Arabia

Abstract - Skin cancer is an excessively common type of cancer. It occurs when mutations appear in the DNA of skin cells. The four main forms of skin cancer are Basal cell carcinoma (BCC), Squamous cell carcinoma (SCC), Merkel cell carcinoma (MCC), and the deadliest form Melanoma. Skin cancer is usually diagnosed late because of the unnoticeable symptoms. Therefore, Reliable automatic detection of skin tumors is needed to help increase the accuracy and efficiency of pathologists. In this paper the DCNN method is used which is designed to perform complex analysis of 2594 images and 2594 of corresponding ground truth (response masks) for training and 1000 images for testing of data using image segmentation and classification for creating model that detect skin cancer in early stages. The models testing produced positive outcome with accuracy 0.95 for classification and 0.895 for segmentation. The results are promising for future enhancement.

Lung Cancer Prediction System using CNN, ANN, and Naive Bayes

Hanan Aljuaid, Raida Alotaibi, Hialh Almatrafi, May Aldhobaie,
Najla Alsowayed, Tina Alshalan, Leen Alfuraih
Computer Sciences Department, Princess Norah bint Abdullrahman University, Riyadh, Saudi Arabia

Abstract - For both men and women, lung cancer continues to be the primary cause of cancer-related mortality, and its prevalence is rising worldwide. we are creating a system that indicates the tumor location and predicts the type of the lung cancer by using three advanced and powerful models which are CNN, ANN and Naïve bayes. The core objective is to come into an accurate ad an efficient result that helps doctors to diagnose and predict correctly. Therefore, CT scan is going to be the leading point of the process to use the three algorithms and to get images. The dataset from kaggle consists of CT scan images of three types of lung cancer. We use CNN as feature extraction and as classifier we use ANN and NB. An accuracy in ANN is 0.95% and in Naïve Bayes is 0.78% have been achieved.

Mean Maximum Raw Thresholding Algorithm for Detecting Suspected Regions of Interest of Tumors in Breast Magnetic Resonance Imaging

Ali Al-Faris
Computer Science Department, Worcester State University, Worcester, Massachusetts, USA

Abstract - In this study, the regions of interest of the suspected tumor in the breast magnetic resonance imaging (MRI) are detected using a new automatic global thresholding approach. As a key pre-processing stage alongside the tumor segmentation process, image thresholding is one of the most significant image processing techniques. In order to prepare the image for more precise tumor detection in the following stage of segmentation, it is crucial to identify the areas of any suspected tumor in the breast MRI. The proposed Mean Maximum Raw Thresholding approach (MMRT), which uses the sub-window frames concept to automatically seek the best threshold value that separates the image into two classes - suspected tumor locations and the background - considers the tumor intensity features. On 40 test images from the RIDER breast MRI dataset, the approach is implemented and tested. The results are then analyzed and recorded in relation to the dataset's Ground Truths using two assessment strategies: pixel-based assessment (Jaccard = 0.643 and Dice = 0.683 measures), and quality assessment (PSNR = 69.97 and MSE = 0.011 measures). The evaluation findings demonstrate that the suggested approach outperforms five common thresholding methods that were tried on the same dataset.

Maritime UAV Patrol Tasks Based on YOLOv4 Object Detection

Yi-Yu Chang, John A, Pao-Ann Hsiung
Department of Computer Science and Information Engineering,
National Chung Cheng University, Minxiong Township, Chiayi, Taiwan

Abstract - The search and rescue operations conducted by designated maritime patrol troops usually employ ships and helicopters in their search and rescue work at sea, making it a resource-intensive undertaking both in terms of patrolling equipment and trained personnel. Oftentimes, when the victims send an SOS or call for help, the rescue unit in question can easily obtain the ship's information but cannot access prevailing maritime images of the sea, thus hindering them from swiftly controlling the accident site as effectively. This work endeavors to address this deficiency and proposes an automated patrolling operation based on small UAVs and a diverse set of maritime search patterns corresponding to different sea areas. Maritime images are collected via a deep neural network-based UAV and promptly returned to the search and rescue system. The object detection model in this case is trained using extensive images of maritime accidents collected from the network. This object detection model based on the YOLOv4 framework is thus used in the timely detection of maritime accidents. Post detection, an alarm system is set up to classify the danger level of the targets and to trigger rescue operations corresponding to the level of danger, followed by an alarm message sent to the rescuers by Short Message Service (SMS). The proposed system has been found to detect humans in the sea and at ship collision sites with an accuracy of 94.20% (mAP), which is 11% more accurate than the Single Shot MultiBox Detector.

Deep-RSI: Deep Learning for Radiographs Source Identification

Farid Ghareh Mohammadi, Ronnie Sebro
Department of Radiology, Mayo Clinic, Jacksonville, Florida, USA;
Center for Augmented Intelligence, Mayo Clinic, Jacksonville, Florida, USA

Abstract - In forensics, the authenticity of digital images is of the utmost importance, considering that modern technology makes it incredibly simple and quick to alter and generate fake but convincing images. As a result, digital image credibility has decreased, making it difficult to demonstrate the source of images. Prior studies have

shown that magnetic resonance imaging (MRI) scans can be traced back to their sources, but radiographs have not. In this paper, we propose the Deep-RSI algorithm, an algorithm that identifies the source (manufacturer and model) of the device used to create radiographs. This is the first time that a medical forensics investigation of this kind will be accomplished to declare and confirm radiograph sources. Researchers in information forensics, security, and medical imaging can use this data to determine scientific fraud, like fake radiographs made from unreliable sources or cut-and-paste fakes. This proposed solution describes how non-content pixels in images enable us to discover the manufacturer and model of a radiographic machine. Since radiographs are obtained from different sites of the body, source recognition has to be sensitive and free of any content-specific information. This will prevent the convolutional neural networks (CNN) from detecting content-specific details and instead identify fingerprints that are unique to the source. CNNs start with low-level features and, in the convolutional blocks, generate high-level features to identify the radiographic machine sources. This proposed solution reports the source (manufacturer and model) of each image. We obtain the highest AUC of 0.97 and a prediction accuracy of 98.54% for radiographic machine manufacturer detection. Our results show that forensic assessments of radiographs can be done with a high level of certainty.

Nostrils and Mouth Detection for Drivers Using Convolutional Neural Networks With Automatically Generated Ground Truth Data

Sorin Valcan, Mihail Gaianu
West University of Timisoara, Romania;
Continental Automotive Romania, Timisoara, Romania

Abstract - Driver monitoring systems are becoming more important in the automotive industry and will be mandatory in the future years to improve driving safety. Nostrils and mouth are the second most important features to be detected on driver faces after the eyes. Training neural networks for detecting facial features requires a manual labelling process to obtain the ground truth data for training and testing. This paper aims to present a method for automatically generation of such ground truth datasets and the results obtained with convolutional neural networks. This process removes the manual labelling effort and provides a simplified way of training neural networks using datasets with real life scenarios.

Fast One-Stage Human Pose Estimation by Polar Coordinate Representation

Jiahua Wu, Hyo Jong Lee
Division of Computer Science and Engineering, Jeonbuk National University, Jeonju-si, Republic of Korea

Abstract - The typical bottom-up human pose estimation methods leverage two-stage approaches to detect individual instances and predict keypoints, which leads to high computation costs and low efficiency. In this paper, we present a novel singlestage model via the Polar Coordinate Representation (PCR). In PCR, one human pose is integrated by a root joint, the angle of joint displacement, and the length of joint displacement. To locate the position of a person instance, previous one-stage methods required the creation of an additional root joint confidence map. To further reduce the computation cost of the one-stage method, we directly exploit the output map of the angle in the polar coordinate to replace the root joint confidence map. In our model, the redundancy prediction head of this root joint confidence map is dropped, which decreases the parameters of the network. For inference, our proposed network only needs one simple decoupling process to generate final poses. Our method achieves comparable performance to previous single-stage bottom-up methods, with lower computation cost and higher speed.

Group-level Analysis of Relations Between Resting-state Functional Connectivity and Arithmetic Ability using CONN

Grace Metri, Xin Yang
Department of Computer Science, Middle Tennessee State University, Murfreesboro, Tennessee, USA

Abstract - This study investigates the relationship between restingstate functional connectivity and arithmetic performance. The goal is to explore the functional connectivity presented for arithmetic ability. We used preprocessing and denoising methods provided through CONN to clean the raw fMRI data. We also implemented a first level analysis and second level analysis on both the resting state and the arithmetic task data. Results revealed that arithmetical ability is correlated with functional connectivity between the left Juxtapositional Lobule Cortex, the right Inferior Frontal Gyrus, the right Frontal Orbital Cortex, the right frontal pole, and the right Lingual Gyrus. These results suggest that the neurocognitive mechanisms of arithmetic processing are complex and involve various regions of the brain, which require coordination and cooperation within specialized functional brain regions.

Deep Learning Based Multimodal Image Retrieval Combining Image and Text

Md Imran Sarker, Mariofanna Milanova
Department of Computer Science, University of Arkansas at Little Rock, Arkansas, USA

Abstract - Multimodal learning is omnipresent in our lives. Human absorbs features in different ways, whether through pictures or text. Combining these features in computational science, especially in Image retrieval problems, poses two significant challenges: how and when to fuse them. Most image retrieval systems use images or text data associated with the image. In this paper, we study the image retrieval task, where the input query is an image plus text sentence that describes the image. The system starts a query triggered by input image and text while taking the help of the Transformer model, which puts attention on both modalities and combines embedded features through the feature fusion technique. We proposed a feature fusion layer using modified Text Image Residual Gating in our work. We have used two methods based on the features extracted from the fusion layer. First, we trained K Nearest Neighbor (KNN) algorithm on the training data, and later we used test data to find a similar image. Second, we used the clustering technique and a support vector machine to compute the nearest neighbor points and cluster the center to see a similar image. We found that SVM (Support vector Machine) is more effective from the results, giving an overall accuracy of 92%.

Transfer Learning Approach for Multilabel Fine Grain Image Attribute Extraction

Gaurav Sharma, Kuruvilla Abraham, Bagyalakshmi V
Tata Consultancy Services, India

Abstract - In the era of artificial intelligence, industries are taking big leap towards transforming overall process of product manufacturing, business measures, retail & customer experience. AI is contributing to small as well as big businesses to enable the automation throughout the pipeline. For steps like product onboarding, attribute identification, PDP enrichment, catalogue generation. Small & mid-level eCommerce industries, in early stage, tends to adopt a low-cost solution for transforming the services. The Literature provides unique and affordable deep learning based multilabel classification approach fore extracting fine grained image level multiple attributes. Converting the objective into multilabel classification problem and using existing stable deep learning models to further train some fine level features, transfer learning is one of the affordable steps to consider. It provides efficient

architecture with optimum results, validated on best resources. Extracting product image level fine grained properties is a complex task which requires dense networks with high computing resources. Existing state of the art architectures for classification problems, after tweaking few layers, used as a baseline models. These architectures have proven one of the best solutions for multilabel classification problems with minimal resources required.

UNet++ with Attention Mechanism for Hippocampus Segmentation

Xueting Cui, Thony Yan Liang, Marymossadat Aghili, Micheal Adeyosoye,
Rosie E. Curiel Cid, David Lowenstein, Ranjan Duara, Malek Adjouadi
Electrical and Computer Engineering, Florida International University, Miami, Florida, USA;
Microsoft, San Francisco Bay Area, California, USA;
Department of Neuropsychology, University of Miami, Miami, Florida, USA;
Center for Cognitive Neuroscience and Aging, University of Miami, Miami, Florida, USA;
Wien Center for Alzheimer's Disease and Memory Disorders, Mt Sinai Medical Center, Miami, Florida, USA

Abstract - Analyzing the hippocampus in the brain through magnetic resonance imaging (MRI) plays a crucial role in diagnosing and making treatment decisions for several neurological diseases. Hippocampus atrophy is among the most informative early diagnostic biomarkers of Alzheimer.s disease (AD), yet its automatic segmentation is extremely difficult given the anatomical structure of the brain and the lack of any contrast in between its different regions. The gold standard remains manual segmentation and the use of brain atlases. In this study, we use a well-known image segmentation model, UNet++, and introduce an attention mechanism called the Convolutional Block Attention Module (CBAM) to the UNet++ model. This integrated model improves the feature weights of our region of interest, and hence increases the accuracy in segmenting the hippocampus. Results show averages of 0.8715, 0.8107, 0.8872, and 0.9039 for the metrics of Dice, Jaccard, Precision, and Recall, respectively.

On the Design and Development of a Smart Safety Warning System for Slow Moving Vehicles

Hongjun Su
Department of Electrical & Computer Engineering, Georgia Southern University,
Savannah, Georgia, USA

Abstract - This paper explored different technologies for designing an intelligent safety warning system for slow-moving vehicles. With the recent advancements in edge computing devices and sensors, it is now possible to develop such a system with a reasonable cost, especially for human-powered small vehicles - such as bicycles. In our design, we utilized the recent advancements in computer vision, AI, machine learning, and Convolutional Neural Networks (CNNs) technologies to design and implement a prototype of such an intelligent safety warning system for slow-moving vehicles. We discussed the pros and cons of different technologies we tested and gave an in deep discussion of the design we chose. We also discussed possible future research directions in this area and the lessons learned from this project.

Citizen Science Mobile Apps with Machine Learning for Recyclable Objects

Chelsea Yeh, Fahim Hasan Khan
Walnut High School, Walnut, California, USA;
Department of Computer Science and Engineering,
University of California Santa Cruz, California, USA

Abstract - We describe the application of citizen science mobile applications (also known as "apps") with integrated machine learning for the real-time identification of recyclable objects. Citizen science mobile apps enable the collection and dissemination of scientific data to a broad non-scientific community by utilizing existing sensors and devices on ubiquitous mobile phones. In this research, we created mobile apps with integrated pre-trained machine-learning models to identify common recyclable objects captured with mobile phone cameras. The models were trained with images of recyclables on a server and loaded into the mobile phones for client-side identification. The mobile app has applications for quickly identifying recyclable items that have been discarded, providing recycling regulations and information based on location, and gathering data regarding the density and frequency of improperly discarded recyclables.

The Applications of GIS in Water Resources

Kirollos Girgis, Nikki Clouse, Olyana Mushe, Parker Morgan, Terrell Brooks, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - In basic terms, a Geographic Information System (GIS) is a system made up of computer hardware, software, and data that can be used to collect, manage, analyze, and display all types of geographical information. GIS enables users to view, understand, question, analyze, visualize, and interpret their data in many ways. GIS software is also able to create reports and charts that reveal relationships, patterns, and trends. Using a GIS allows users to view data in a way that is easily understood and shared, and to answer questions that users may have. Because of this, GIS technology can easily be integrated into a variety of enterprise information systems. As public awareness grows, stricter measures are being enacted, and new legislation has been enacted concerning the use of water resources, which has significantly increased the importance of the use of advanced technologies. In the area of water resources management, GIS are useful tools that can be used to manage, store, and display spatial data. It is no secret that GIS is being applied increasingly in the field of water resources. Several applications related to this area are addressed and evaluated in the present study in order to highlight the importance of GIS for the management of water resources in the future. It is possible to summarize the fundamentals of GIS and examine the evolution of GIS in relation to water resources. In this paper, we will discuss a variety of ways in which GIS can drastically benefit agricultural needs, based on its applications in fields such as surface hydrologic and groundwater modeling, water supply and sewage system modeling, stormwater and pollutant characterization of urban and agricultural areas, along with other related topics.

Automatic Arrhythmia Detection Using Deep Learning

Saeka Rahman, Miad Faezipour
School of Engineering Technology, Electrical and Computer Engineering Technology,
Purdue University, West Lafayette, Indiana, USA

Abstract - Arrhythmia is a common cardiac disease that can be a potential life threat. Arrhythmias are detected using an Electrocardiogram (ECG) signal record. ECGs are usually performed with 12 lead channels. However, this arrangement requires extensive equipment and expertise to handle. The doctors also need much time to find problematic instances when there are many records. A single-channel system is necessary to make a portable,

wearable embedded system to detect arrhythmias. Such a system could assist the patients for frequent monitoring and the caregivers to follow up treatment plans and/or immediate actions. Therefore, an automated arrhythmia-detecting mechanism using single-channel information is required to improve cardiac treatment. In so doing, a Convolutional Neural Network (CNN) based lightweight deep learning model is developed to detect arrhythmia. The dataset is prepared from the MIT-BIH arrhythmia database of ECG records to validate the model. 2-D images are taken from the snapshots of R-R peaks, including regular, ventricular premature beats, and atrial premature beats of the MLII lead. The proposed classifier's overall performance on test data to detect arrhythmia and the normal rhythm is 0.9167, with 8,658 parameters.

An Unsupervised Object Localization Based on Topological Data Analysis

Hong Cheng
Department of Math and Computer Science, Southern Arkansas University, Magnolia, Arkansas, USA

Abstract - In this paper, an unsupervised object localization based on topological image processing is proposed to deal with image localization problem. We use the method to process images with the intention to identify the objects of interest and draw bounding boxes to surrounding the objects in the images. We experi

The Practical Use of GIS in Agriculture

Aaron Butora, Brent Soloniewicz, Cameron Schwartz, Christian Aziz, Sichang Su, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - Agriculture is a business sector ideally suited for the application of GIS. It is natural resource based, requires the movement, distribution, and utilization of large quantities of products, goods, and services, and is increasingly required to record details of its business operations from the farm field to the supermarket. Agricultural agencies typically use GIS for support of pesticides and food safety regulations in the country. It is suited to show economic impacts of policy alongside revealing environmental health issues that show up. It can help in mapping an area to record that data and arbitrate land use conflicts to help farmers achieve an increase in production and reduce costs by enabling better management of land resources. GIS has been a proven effective technology for any government. The goal of this paper is to research and discuss the use of GIS in agricultural production. Important components of GIS in agriculture that we will talk about is the use of GPS and remote sensors to manage the large fields. The value of GIS to agriculture continually increases as advances in technology also increase. The increase of technology assists farmers in increasing production, reducing costs, and providing an effective means of managing land resources.

Synthesizing a Reference Image from the Macro Images of a Painting for Vignetting Correction

Soonchul Jung, Jae Woo Kim, Yoon-Seok Choi, Hyeong-Ju Jeon, Jin-Seo Kim
Content Research Division, Electronics and Telecommunications Research Institute, Republic of Korea

Abstract - Vignetting is an unwanted reduction in the brightness of an image toward the edges compared to the image center caused by camera settings or lens limitations. Conventional vignetting correction methods based on reference images have shown satisfying results. We cannot however use the conventional methods when the reference images are not available. There have been several studies to estimate the vignetting function from

microscopic images without any reference images. Most of those studies are based on the fact that it is easy to distinguish foreground and background pixels in a microscopic image. In this paper, we propose a simple method for synthesizing a virtual reference image from macro painting images to carry out vignetting correction. It creates a virtual reference image from many macro images, as obtained by imaging a gray target. Then we apply the existing reference image-based vignetting correction algorithm as it is. According to our experimental results, the proposed method produced satisfying results that are comparable to the results from the conventional methods using a real reference image. Furthermore, the proposed method could more consistently produce the corrected images for unseen input images.

Deep Learning-based Synthesized Image Attribution Using Frequency Distribution Information

Junbin Zhang, Yixiao Wang, Hamid Reza Tohidypour, Panos Nasiopoulos
Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada

Abstract - As the deepfake technology emerges at a breathtaking pace, it threatens to become a destructive political and social force with unpredictable impact on society. Therefore, detecting deepfakes and even figuring out which deep learning generative models (GMs) created such images is of extreme importance. There are already several methods that find and categorize artifacts left by GMs, with the latest efforts focusing on utilizing the frequency domain to achieve these goals. In this paper, we propose a deep learning-based solution with a learnable coefficient layer that highlights GMs. artifacts to achieve high accuracy on the synthesized image attribution task. Evaluation results have shown that our proposed method not only has comparable performance to state-of-the-art methods, but it also outperforms them on unseen image types, showing great generalizability.

Design a Hardware of Image Stitching using LBP Algorithm

Chang-Yong Lee, Yong-Hwan Lee (Minjun Cho)
Department of Electronic Engineering, Kumoh National Institute of Technology, Korea;
School of Electronic Engineering, Kumoh National Institute of Technology, Korea

Abstract - Image stitching algorithms are often used in real life and are mostly handled by software. The proposed method is a hardware design level of image processing method using a local binary pattern. To solve the background matching problem during stitching, the proposed method is used instead of the existing stitching algorithm. Designed with Verilog HDL, it minimizes multipliers to reduce resource consumption. Parallel processing and pipeline architecture are exploited to increase processing speed. We implement the proposed algorithm into a hardware with similar accuracy to existing algorithms.

CSCI-RTHI:
RESEARCH TRACK ON HEALTH INFORMATICS AND MEDICAL SYSTEMS

Leveraging Data Pathways for Next Generation Safety Monitoring of Medicines and Vaccines

Jeffery L. Painter, Laurie Girard, Michael Glaser, Andrew Bate
GlaxoSmithKline (GSK), Durham, North Carolina, USA;
GlaxoSmithKline (GSK), Upper Providence, PA, USA;
GlaxoSmithKline (GSK), London, United Kingdom (UK)

Abstract - Safety evaluation of medicines and vaccines is critical to ensure patient safety and maintain confidence in treatment and disease prevention strategies. Leveraging data pathways for next generation pharmacovigilance (PV) requires the creation of new platforms that seamlessly integrate both structured and unstructured data. Here, we describe the design of a novel data environment that provides enhanced data mining, information retrieval, and data governance to improve PV processes and activities. The goal of which is to further inform the knowledge of potential safety issues during the life cycle of medicines, from routine healthcare delivery to informing future drug and vaccine development.

Diagnosis of Pediatrics Epilepsy Based on Graph Analysis of Scalp EEG Applying Mutual Information

Elaheh Zarafshan, Parisa Forouzannezhad, Hoda Rajaei, Ulyana Williams,
Mercedes Cabrerizo, Ilker Yaylali, Alberto Pinzon, Malek Adjouadi
Center for Advanced Technology and Education, Department of Electrical and Computer Engineering,
Florida International University (FIU), Florida, USA;
Department of Radiation Oncology, School of Medicine, University of Washington, Seattle, Washington, USA;
Massachusetts Institute of Technology, Massachusetts, USA;
Oregon Health & Science University, Portland, Oregon, USA;
Department of Neurology, Neuroscience Center, Baptist Hospital of Miami, Florida, USA

Abstract - Epilepsy is a brain disorder that causes seizures, affecting nearly half a million children in the US alone. In this study, we aimed to use a nonlinear driven method to characterize scalp EEG recordings of pediatric epilepsy patients (PE: n=7) compared to pediatric control subjects (PC: n=7) in a clinical environment. A time-varying approach was used to construct functional connectivity networks (FCNs) of all subjects. Next, the FCNs are mapped into the form of undirected graphs that are subjected to the extraction of graph theory-based features. An unsupervised clustering technique based on K-mean is used to delineate the PE from the PC group. Our findings show a statistically significant difference in the mean FCNs between PC and PE groups ($t(340)=-15.9899$, $p << 0.0001$). Performance results showed an accuracy of 92.5% with a sensitivity of 90% and a specificity of 95.3%. This approach can help improve and validate the early diagnosis of PE by applying non-invasive scalp EEG signals.

A Bioengineering System for Assessing Children's Cognitive Development by Computerized Evaluation of Shared Intentionality

Igor Val Danilov
Academic Center for Coherent Intelligence, Riga, Latvia

Abstract - The article creates a conceptual framework for translational research on the bond between shared intentionality magnitude in caregiver-child dyads and scores of children's cognitive development trajectory. The current study assessed the shared intentionality magnitude in 30 subjects (neurodivergent (ND) and neurotypical (NT) children) aged 2 to 10 years. Disclosing the relation of shared intentionality magnitude across different stages of children's development with their diagnoses allows for developing an assessing method of cognitive development trajectory in preverbal children. The article proposes directions for future research regarding (i) evidence of shared intentionality; (ii) proof of a shared intentionality assessment method validity, (iii) conditions for quantitative measurement to satisfy dependability in the cognitive development assessment.

A Topological Characterization of DNA Sequences based on Chaos Geometry and Persistent Homology

Dong Quan Ngoc Nguyen, Phuong Dong Tan Le, Lin Xing, Lizhen Lin
Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Yanqi Island, Huairou District, Beijing, P. R. China;
Department of Applied Mathematics, University of Waterloo, Waterloo, Canada;
Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, Indiana, USA

Abstract - Methods for analyzing similarities among DNA sequences play a fundamental role in computational biology, and have a variety of applications in public health, and in the field of genome analysis. In this paper, a novel geometric and topological method for analyzing similarities among DNA sequences is developed, based on persistent homology from algebraic topology, in combination with chaos geometry in 4-dimensional space as a graphical representation of DNA sequences. Our topological framework for DNA similarity analysis is general, alignmentfree, and can deal with DNA sequences of various lengths, while proving first-of-the-kind visualization features for visual inspection of DNA sequences directly, based on topological features of point clouds that represent DNA sequences. As an application, we test our methods on three datasets including genome sequences of different types of Hantavirus, Influenza A viruses, and Human Papillomavirus.

Comparison of Dimensionality Reduction Methods for Multimodal Classification of Early Stages of Alzheimer's Disease

Luana Okino Sawada, Ulyana Morar, Robin Perry Mayrand, Christian Yaphet Freytes, Micheal Adeyosoye, Mercedes Cabrerizo, Rosie E. Curiel Cid, David Loewenstein, Ranjan Duara, Malek Adjouadi
Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, Florida, USA;
Electrical and Computer Engineering, Florida International University, Miami, Florida, USA;
Center for Cognitive Neuroscience and Aging, University of Miami Miller School of Medicine, Miami, Florida, USA;
Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami, Florida, USA

Abstract - Early diagnosis of Alzheimer.s Disease (AD) is challenging due to its progressive nature. This study proposes a comprehensive comparison of four classifiers combined with different dimensionality reduction methods to discriminate normal controls (CN) from pre-mild cognitive impairment (pMCI) and early MCI (EMCI) using multimodal datasets including MRIs, PETs, SUVr, clinician amyloid visual reads, and subjects demographics. The most robust classifier for CN vs. MCI is the Mutual Information Best Percentile - Bagging Classifier combination, with 73.91% accuracy and a 4.82% standard deviation (SD). The best performance of 65.23% (11.84% SD) accuracy for CN vs. EMCI was DTC with ANOVA. In comparing CN with pMCI the best classification accuracy was ANOVA-DTC 51.06% (14.19% SD). An accuracy of 56.34% (10.67% SD) was achieved by bagging with ANOVA for multiclass classification of CN vs. pMCI vs. EMCI.

Evaluation of an Automated Mapping from ICD-10 to SNOMED CT

Patrick Philipp, Julia Veloso de Oliveira, Arno Appenzeller, Tobias Hartz, Juergen Beyerer
Vision and Fusion Laboratory IES, Karlsruhe Institute of Technology KIT, Karlsruhe, Germany;
Klinisches Krebsregister Niedersachen, Hannover, Germany;
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Fraunhofer Center for Machine Learning, Karlsruhe, Germany

Abstract - The amount of data in the medical field is constantly increasing. But it is not only the sheer amount of information that is important, but also its quality and type of representation. While nomenclatures such as SNOMED CT (Systematized Nomenclature of Medicine and Clinical Term) are suited for finegrained documentation and modern analysis, much information is also bound in classifications. This fact is often historical, as billing systems are typically based on classifications such as ICD10 and then also had been used for documentation. Leveraging this information automatically is subject of this paper - i.e. enabling an automatic mapping from ICD-10 to SNOMED CT. Because this mapping provides a large set of SNOMED codes for each ICD-10 concept, the approach is non-trivial. In order to pick the best possible code, we propose to take advantage of the hierarchical structure of the SNOMED system to find the concept which lies closer to all candidates in the target system. In other words, our algorithm searches the lowest common ancestor (LCA) of all candidates. For evaluation, we studied 1692 codes from a real-world dataset. The results are promising and show that the proposed approach achieves good results in the majority of cases.

On Creating a Comprehensive Food Database

Lexington Whalen, Brie Turner-McGrievy, Matthew McGrievy, Andrew Hester, Homayoun Valafar
College of Computing and Engineering, University of South Carolina, Columbia, South Carolina, USA;
Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, USA

Abstract - Studies with the primary aim of addressing eating disorders focus on assessing the nutrient content of food items with an exclusive focus on caloric intake. There are two primary impediments that can be noted in these studies. The first of these relates to the fact that caloric intake of each food item is calculated from an existing database. The second concerns the scientific significance of caloric intake used as the single measure of nutrient content. By requiring an existing database, researchers are forced to find some source of a comprehensive set of food items as well as their respective nutrients. This search alone is a difficult task, and if completed often leads to the requirement of a paid API service. These services are expensive and noncustomizable, taking away funding that could be aimed at other parts of the study only to give an unwieldy database that can not be modified or contributed to. In this work, we introduce a new rendition of the USDA.s food database that includes both foods found in grocery stores and those found in restaurants or fast food places. At the moment, we have accumulated roughly 1.5 million food entries consisting of approximately 18,000 brands and 100 restaurants in the United States. These foods also have an abundance of nutrient data associated with them, from the caloric amount to saturated fat levels. The data is stored in MySQL format and is spread among five major tables. We have also procured images for these food entries when available, and have included all of our data and program scripts in an open source repository that anyone can access, for free.

Comprehensive and User-Analytics-Friendly Cancer Patient Database for Physicians and Researchers

Ali Firooz, Avery T. Funkhouser, Julie C. Martin, W. Jeffery Edenfield, Homayoun Valafar, Anna V. Blenda
College of Engineering and Computing, University of South Carolina, Columbia, South Carolina, USA;
School of Medicine Greenville, University of South Carolina, Greenville, South Carolina, USA;
Prisma Health Cancer Institute, Greenville, South Carolina, USA

Abstract - Nuanced cancer patient care is needed, as the development and clinical course of cancer is multifactorial with influences from the general health status of the patient, germline and neoplastic mutations, co-morbidities, and environment. To effectively tailor an individualized treatment to each patient, such multifactorial data must be presented to providers in an easy-toaccess and easy-to-analyze fashion. To address the need, a relational database has been developed integrating status of cancer-critical gene mutations, serum galectin profiles, serum and tumor glycomic profiles, with clinical, demographic, and lifestyle data points of individual cancer patients. The database, as a backend, provides physicians and researchers with a single, easily accessible repository of cancer profiling data to aid-in and enhance individualized treatment. Our interactive database allows care providers to amalgamate cohorts from these groups to find correlations between different data types with the possibility of finding "molecular signatures" based upon a combination of genetic mutations, galectin serum levels, glycan compositions, and patient clinical data and lifestyle choices. Our project provides a framework for an integrated, interactive, and growing database to analyze molecular and clinical patterns across cancer stages and subtypes and provides opportunities for increased diagnostic and prognostic power.

Construction of Disease-pathological Feature Knowledge Graph Based on Entity Recognition and Relation Extraction

Jun Li, Honghai Feng, Pengcheng Zhou, Ruihui Hou, Yaju Wei, Yanyan Chen
Henan University, Kaifeng, P. R. China

Abstract - The pathological features of the diseases are helpful to the diagnosis of disease. There are no one-to-one correspondences between most of pathological features and diseases. If the detailed relations between all diseases and all pathological features are constructed, they can be used to assist doctors and the machine to reason from the huge knowledge relations. In this paper, an unsupervised entity recognition and relation extraction algorithm based on semantics is proposed, it can extract a large amount of relation data from massive pathological report texts, and with it the corresponding disease-pathological feature knowledge graph has been constructed. By processing 3 million web pages, 11755 effective disease entities, 83511 pathological features and 533 relation words are obtained. The correct rate of extracting relation words is 83.17%, and the recall rate is 73.26%. The experimental results show that the knowledge graph constructed in this paper is relatively complete and reliable.

Applying Systems Engineering and Decision Analysis to Healthcare: a Perinatal Operations Center

John Celona, Louis Halamek
Decision Analysis Associates, LLC, San Carlos, California, USA;
Division of Neonatal & Developmental Medicine, Department of Pediatrics,
Stanford University, Stanford, California, USA

Abstract - Despite space age technology, healthcare missed the Industrial Revolution of designed, engineered, and standardized processes; thus tremendous potential remains to improve the quality of care and reduce costs. Applying systems engineering and decision analysis to create an operations center can realize this potential. For a number of reasons, perinatal (maternal and newborn) care the right scope for introducing this methodology and realizing the potential benefits.

Summarizing Behavioral Health Electronic Health Records using a Natural Language Processing Pipeline

Tristram Dacayan, Daniel Ojeda, Daehan Kwak
Department of Computer Science and Technology, Kean University, Union, New Jersey, USA

Abstract - Doctors and nurses have limited time between patients to analyze and review a patient's documents to provide a quality assessment. This problem is supplemented by the existence of Electronic Health Records (EHR), which are essentially digital files regarding the patient. However, the length and content of each document vary greatly, reducing the effectiveness. Therefore, this research aims to reduce the need for medical professionals to manually search for crucial information about the patient's health history. We intend to accomplish our objective using various natural language processing (NLP) techniques to break down digital documents into smaller subtasks, such as event extraction and abstractive summarization, to provide a concise summary. Our proposed system intends to streamline the process and nullify the issue of having to read lengthy documents to locate essential information, which can affect the overall efficiency and quality of the patient's care. In the future, we intend to migrate to a closeddomain event extraction model and implement a timeline for easier visualization.

Subject Skin Tone Classification with Implications in Wound Imaging using Deep Learning

Masrur Sobhan, Daniela Leizaola, Anuradha Godavarty, Ananda Mohan Mondal
Computing and Information Sciences, Florida International University, Miami, Florida, USA;
Biomedical Engineering, Florida International University, Miami, Florida, USA

Abstract - Chronic wound healing is inconsistent on an individual basis, leading to large treatment costs. The effectiveness of any treatment approach is typically assessed by visual inspection of the wound. Optical imaging technologies have recently been developed to objectively assess wound physiology to complement the subjective visual assessment. One such device is a low-cost SmartPhone Oxygenation Tool (SPOT), which can measure the tissue oxygenation of the wounds via non-contact imaging and assessing healing status. The varying skin tones impact tissue oxygenation measurements due to the different melanin concentrations in the epidermis of the skin. Hence, it is essential to consider melanin-related attenuation in the epidermis and account for it during tissue oxygenation measurements. This study aims to implement a machine learning algorithm to classify the skin tones using in-vivo measurements from control subjects towards a future correction for these skin tones during imaging studies using SPOT. In this study, we developed a benchmark dataset of 75,348 samples of 28 X 28 RGB images of human subjects' hands. The images were then converted to 28 X 28 grayscale images and were flattened to attain 784 pixels or features for each sample. We also developed a deep learning-based pipeline to classify the FST skin types, producing high accuracies (> 98%). The deep learning model can be incorporated into the SPOT device as an additional feature to verify or correct the melanin concentration during near-infrared (NIR) imaging of wound regions.

Automatic Tuberculosis Detection using Binary Pattern of Phase Congruency

Afonso Ueslei da Fonseca, Juliana Paula Felix, Gabriel da Silva Vieira,
Deborah Fernandes, Fabrizzio Soares
Universidade Federal de Goias, Goiania - GO, Brazil;
Instituto Federal Goiano, Campus Ceres, Ceres - GO, Brazil

Abstract - Computational solutions to assist the automatic analysis of radiographic images are already a reality. Commercial solutions for the diagnosis of tuberculosis have shown exceptional results similar to those of radiologists. However, these solutions are still far from regions of greater vulnerability, given their high costs. Researchers have shown a growing interest in developing increasingly efficient and effective solutions that require the least resources. Thus, in this work, we explore the use of chest X-ray and Phase Congruence Binary Pattern features to build a minimalist, low computational cost, and high-efficiency Feed-Forward Neural Network model to aid the diagnosis of tuberculosis. Our results showed high performance compared to related research, placing our solution as a viable alternative.

Exploring Emotional Stimulation to Better Address BPSD: Analysis of Heart Rate Variability, Electrodermal Activity, Mood Alterations and its Ethical Implications of in Dementia Care

Yumna Anwar, Arshia Khan
Department of Computer Science, University of Iowa, Iowa, USA;
Department of Computer Science, University of Minnesota Duluth, Minnesota, USA

Abstract - Dementia is a growing problem and will continue to grow with the aging population, causing an increase in economic burden on the affected person, the family members and the society. In addition, the behavioral and psychological symptoms (BPSD) of dementia are causing additional burden on all the stakeholders. Mood can play a major role in stimulating the BPSD issues. Emotional stimulation techniques can be applied to influence the mood. The emotional stimulation phenomenon can be tracked and observed for further analysis in exploring the implications of emotions in ailments such as Alzheimer's, vascular dementia and more. This study investigates the variation of Heart Rate Variability (HRV) and Electrodermal activity (EDA) in response to emotional stimuli under an ethical lens. In this experiment, 26 people volunteered to participate and had their data collected for two phases; the baseline phase during which the participants were simply resting, and the emotional stimuli phase during which participants were shown emotionally evocative media. Time-domain and non-linear analysis of HRV was then performed on the collected HRV data. Moreover, for analysis of EDA, it was decomposed into tonic and phasic components. This paper comprehensively covers the experimental protocol used, data analysis performed, the results obtained and the observations from the conducted study along with some of the ethical aspects. This study can help reduce the burden of the caregivers of people affected with dementia.

eVision: Forecasting the Spread of Tuberculosis in India with Deep Learning

Juan Zuluaga, Michael Castillo, Divya Syal, Andres Calle, Navid Shaghaghi
Ethical, Pragmatic, & Intelligent Computing (EPIC) Laboratory, in collaboration with the Healthcare Innovation & Design (HID) Program, Departments of Computer Science & Engineering (CSEN), Information Systems & Analytics (ISA), Mathematics & Computer Science (MCS), and Bioengineering, Santa Clara University, Santa Clara, California, USA

Abstract - Humanity has battled tuberculosis for all of recorded history, as a matter of fact some studies estimate that Mycobacterium tuberculosis may have been around as long as 3 million years; but it was only in 1834 when Johann Schonlein officially presented the characteristics of it. Even though the TB epidemic has touched all corners of the world, Africa and Asia are the regions that currently suffer the worst consequences. The purpose of this study is to construct a model within the eVision forcasting environment, capable of forecasting the number of Tuberculosis cases in India, as India is the country that accounts for the largest percentage of TB cases and deaths. Being able to do the prediction for India could also possibly lead to successful results for other regions in Asia and Africa. In order to do so, this study presents different test cases that show the effectiveness of the model, varying the number of steps for each one of the data sets created. It's important to note, that these datasets are combinations of data gathered from the states with the most TB cases in India in the last years, as well as the total data for India, and supplemental data from Google Trends, as a way to facilitate the learning. Even though the final results were respectable compared to past research done on India and other countries, the model nevertheless has a limitation on the number of weeks in which the predictions are still considered to be good, 7 being the optimal number.

Blockchain-Based Internet of Medical Things (IoMT) for Healthcare Management

Hamed Taherdoost
Department of Arts, Communications and Social Sciences, University Canada West, Vancouver, Canada

Abstract - When it comes to the use of the Internet of Things (IoT), the healthcare sector is set to become the next frontier of the digital revolution thanks to Internet of Medical Things (IoMT). Due to their weight, importance, and sensitivity, these files must be protected in the strictest manner. Now that blockchain is becoming more widespread, scientists are concentrating on how to use blockchain tactics within healthcare management to improve data security. Nevertheless, owing to the differing needs of these two technologies, such integration is exceptionally complex and demanding. In order to help users, take full control of their data, this study provides an overview of the current state of blockchain platforms for the IoMT by focusing on the difficulties presented by combination systems. This article will examine blockchain's use in healthcare IoT, including supply chain transparency, health data arrangement, smart contracts, and IoT security for remote monitoring. The final portions focus on challenges and potential developments in the future.

M-Health System Framework for Diagnosing Inflammatory Breast Cancer with Fuzzy Logic

Buket D. Barkana, Ahmed El-Sayed, Mark Pitcher, Ruba Deeb, Ruth Pfeiffer, Marilyn Roubidoux, Rana H. Khaled, Maha Helal, Hussein Khaled, Catherine Schairer, Amr Soliman
Electrical Engineering Department, University of Bridgeport, Bridgeport, Connecticut, USA;
Biostatistics Branch, National Cancer Institute,
The National Institutes of Health (NIH), Bethesda, Maryland, USA;
Department of Radiology, University of Michigan, Ann Arbor, Michigan, USA;
National Institute of Cancer, Cairo University, Cairo, Egypt;
City University of New York, Medical School, New York, USA

Abstract - Inflammatory breast cancer (IBC) is an aggressive and fatal breast cancer. The American Joint Committee on Cancer defines the clinical symptoms of IBC as erythema, edema, peau d'orange that is over at least a third of the breast with a duration of the first symptom to the diagnosis of fewer than six months, and histopathologic diagnosis. Unfortunately, these signs are not present in many cases and make it challenging to diagnose IBC. In this work, we proposed a pioneering framework of an M-Health system for early diagnosis of IBC to improve survival rates and presented the supporting preliminary experimental results. Particular emphasis is given to the system framework and bilateral mammography images. Since IBC is a rare type of cancer, there is currently no public-domain mammography image dataset. Our work evaluated the system performance using the bilateral mammography images of six IBC and eight non-IBC breast cancer cases provided by the National Cancer Institute (NCI) - Cairo, Egypt. The proposed model extracts features and combines them in the feature bank to send to the Fuzzy logic Type 1 classifier for diagnosis. The system achieved promising performance with an accuracy of 92.3%, sensitivity of 83.3%, and specificity of 100%.

Neuro-Respiratory and Mechanical Ventilation Methodologies for Exploring Human Cognitive Overload

Mokhles M. Abdulghani, Wilbur L. Walters, Khalid H. Abed
Department of Electrical & Computer Engineering and Computer Science,
Jackson State University, Jackson, MS, USA

Abstract - This paper explores the deep understanding of the respiration system in the human body for the purpose of improving the design of mechanical ventilators used in intensive care units and improving the research about human operators cognitive overload challenges. We discuss the neuro-physiological and cardio-physiological mechanism of respiration. We describe the two respiration processes. The first process is the external respiration or .ventilation,. which is the uptake of oxygen and excretion of carbon dioxide in the lungs. The second process is the internal respiration or "gas exchange" between the pulmonary alveolus and the blood. Then, we discuss the mechanical ventilation methods and the human-ventilator synchrony problem. The results of this research provide insights about deep understanding for the mechanism of the ventilation and gas exchange in the normal respiration process and help with designing a reliable mechanical ventilation able to achieve the synchronization between the mechanical ventilation unit and the patient's efforts for breathing in an intensive care unit. This research also provides additional insights to human operators cognitive overload challenges in human-machine teaming applications such as pilots of autonomous systems.

Implementation of Predictive Model for Diarrhea among Afghanistan Children based on Medical and Non-Medical Attributes

Kittichai Lavangnananda, A. Zahirzada
School of Information Technology, King Mongut's University of Technology Thonburi, Bankok, Thailand

Abstract - Childhood diarrheal illness is still prevalent in many low-income countries and Afghanistan is no exception to this. In spite of the fact that the cause of diarrhea is well understood in the field of medicine, other non-medical factors may attribute to this illness directly or indirectly. Nevertheless, studies of diarrhea in children, especially in Afghanistan, have ignored non-medical factors. The objective of this study is to implement predictive model for diarrhea in Afghanistan children where both medical and non-medical factors are considered. The dataset used is Afghanistan's Demographic and Health Survey (AfDHS) 2015. Information Gain and Correlation-based Feature Selection algorithms are utilized during the preprocessing. The five well-known machine learning algorithms, Support Vector Machine (SVM), Random Forest, Neural Network, NaÃ¯ Bayes and Decision Tree are investigated during implementation. Predictive models ar evaluated by four common metrics, Accuracy, Area Under Curve (AUC), Precision, and Recall. It is found that the predictive model implemented by Random Forest yields the best overall performance. The predictive model in this work can be used as a preventive measure to detect the possibility of diarrhea in children at early stage.

The Research Activities and Development Trends of Antineoplastics Targeting PD-1/PD-L1 based on Scientometrics and Patentometrics

Ting Zhang, Juan Chen, Yan Lu, Dongzi Xu, Xiaoyi Yang, Zhaolian Ouyang
Institute of Medical Information, Chinese Academy of Medical Sciences and
Peking Union Medical College, Beijing, P. R. China

Abstract - Objectives: This paper aims to discuss the research activities and development trends of antineoplastics targeting PD1/PD-L1 based on scientometrics and patentometrics. Methods and Results: Publications and patents related to antineoplastics targeting PD-1/PD-L1 were searched and collected from the Web of Science (WoS) and

the Derwent Innovation Index (DII) respectively. Totally, 11244 publications and 5501 patents were obtained. The publications were analyzed from the annual number, the top countries/regions and organizations to describe the scientific research trends in this field. The patents were analyzed from the annual number, the top priority countries and patent assignees to reveal the characteristics and status of technological development. As well as the identification of scientific research focus and technological development focus was based on the title and abstract of the publications and patents, using the freely available computer program VOSviewer for clustering and visualization analysis. The number of scientific publications and patent applications showed obvious increase of 29.84% and 33.46% in recent ten years (2009-2018), respectively. Results suggested that the most productive countries/regions publishing on antineoplastics targeting PD-1/PD-L1 were USA and China, and the top three productive organizations were all from USA, including Harvard University, VA Boston Healthcare System (VA BHS) and University Of California System. There were four scientific research focus and five technological development focus in the field of antineoplastics targeting PD-1/PD-L1, and both the scientific research focus and technological development focus between China and USA were a little different. Conclusions: The results of this study presents an overview of the characteristics of research status and trends of antineoplastics targeting PD-1/PD-L1, which could help readers broaden innovative ideas and discover new technological opportunities, and also serve as important indicators for government policymaking.

College Spread of COVID-19 in Ohio

Akinkunle Akinola, Akoh Atadoga, Kimberlyn Brooks, Vibhuti Chandna, Robert C. Green II
Data Science Program, Bowling Green State University, Bowling Green, Ohio, USA;
Department of Computer Science, Bowling Green State University, Bowling Green, Ohio, USA

Abstract - To determine whether Ohio college re-opening plans were effective in controlling the spread of COVID-19, cumulative case counts by county were gathered to compare various metrics related to the spread of COVID-19 cases between counties with NCAA colleges and counties without NCAA colleges. Various non-parametric statistical tests were used to determine if the samples were similar, and the analysis found the differences were statistically significant. Metropolitan and non-metropolitan groupings were also added to further subdivide the data set, but the analysis found no statistically significant differences in this case.

An Intelligent Zigbee Algorithm for Healthcare Monitoring System using Wireless Sensor Networks

Nthabiseng Pertunia Mosana, Tonderai Muchenje, Topside Ehleketani Mathonsi, Vusimuzi Malele
Department of information Technology, Tshwane University of Technology, South Africa

Abstract - Wireless Sensor Network could be implemented as a communication technology for sensory devices. In WSN applications such as homes, hospitals, clinics and other industrial areas, the number of wireless sensor devices continues to increase. End to end delay of packets, loss of packets, packets collision and bandwidth are some of the major issues experienced in WSNs. In this study, ZigBee routing protocols are studied and analysed. High network latency is identified as one of the shortcomings of ZigBee Tree Routing protocol. This shortcoming is caused by the end-toend delay of packets as they are routed on a Tree topology. In response to this shortcoming, this study will propose an algorithm that attempts to improve the original ZigBee Tree Routing protocol. New Tree Routing Protocol (NTRP) is an output of merging ZigBee Tree Routing algorithm and Kruskal.s Minimum Spanning Tree Protocol. NS-2 simulator will be used to validate the proposed algorithm.

Medication Nonadherence in TB Patients: Innovative Technology Use and Related Legal Framework

Oluwaseun Adetunji, Olayinka Adeniyi, Adele Botha, Marlien Herselman
School of Computing, University of South Africa, Johannesburg, South Africa;
Centre for Intellectual Property and Information Technology, Strathmore University, Nairobi, Kenya;
CSIR, School of Computing, University of South Africa, Pretoria, South Africa

Abstract - Tuberculosis is a deadly disease that has been at the forefront of many deaths around the world, but patient nonadherence to medication has been a foremost contributor to this result. Nonetheless, there could be innovative, technological intervention to aid patients. adherence to medication and prevent the many deaths. However, some of the medication nonadherence causes as well as technological interventions could have legal consequences, but the impact of this has not been investigated, especially with regards to the impact on the human rights of the patients and on the society. The article is written to add to the body of expertise on the topic and will be useful for policymakers and all stakeholders in the related area of using technology to improve patient health in the bid to attain the Sustainable Development Goals of bringing an end to the tuberculosis epidemic by 2030 (SDGs).

Towards a SaaS Application Providing an Ethical Support for Human Caregiving of People Living at Home

Pierre Bouchet, Sophie Jacquot, Christian Toinard
Laboratoire d'informatique fondamentale d'orleans (LIFO), France;
Institut Denis Poisson (IDP), France

Abstract - With an aging population, we have to prepare for a growing number of people with a loss of autonomy. Many papers focus on digital solutions for assisting and monitoring those people. In contrast, solutions are missing for monitoring human caregivers. This paper suggests using digital solutions towards a support for ethical reasoning through data collected among various participants including the recipient, the caregivers, the family and may be the physicians. Our purpose is thus to ease the cooperation among the participants. We start from an existing SaaS application marketed in France called SIPAD Connect. This paper proposes several additions in order to better support cooperation. First, new data and indicators are suggested for the recipient. Second, the participants can give their opinions for those data and indicators. Those opinions can be computed in order to raise alerts. The proposed alerts include not only the situation of the recipient but also feedbacks aiming at improving the SaaS application. Thus, both the cooperation and the computation open discussions regarding the recipient, the efficiency of the care and the quality of the SaaS application.

Current Cybersecurity Challenges of Applying Blockchain in Healthcare

Nahla F AL Hamad, Jing-Chiou Liou
Department of Computer Science and Technology, Kean University, Union, New Jersey, USA

Abstract - The most significant problems faced in electronic healthcare are data protection, sharing, and interoperability. These problems may be reliable by using Blockchain. With blockchain, transparency and trust can be established in transactional systems by using a peer-to-peer (P2P) distributed ledger technology. This technology enhances security, data exchange, interoperability, integrity, and real-time updating and access when correctly implemented. Blockchain regulates accessibility to the database, and transfers of rights among individuals based on certain situations and it facilitates access to the information of user profiles, the user will have full access to his information and control how his data will be shared. A blockchain would also securely store access control policies, and only the users could change them. This creates an environment of transparency and

allows the user to make all decisions as to what data is collected and how the data is shared. However, the most difficult challenges for blockchain healthcare are the data backup and compliance with regulation. There are national and international privacy laws such as HIPAA, EU's General Data Protection Regulation, and the GDPR-like California Consumer Privacy Act. Decentralization in blockchain makes it impossible to have a full data backup.

Explainable A.I. in Healthcare

Hritik Rao, Arshia Khan
Computer Science Department, University of Minnesota Duluth, Minnesota, USA

Abstract - Explainable AI in healthcare is an emerging field with enormous potential to change the structure of the modern healthcare industry. The automation of assistive healthcare and its evolution with the help of ML algorithms and Deep neural nets will play an essential role in advancing the current healthcare infrastructure in the future. Explainable AI can help solve the lack of trust in these near blackbox neural networks in an industry like healthcare where the decision making process to achieve the result holds as much importance as the results themselves. Explainable AI models like LIME, SHAM, decision sets, etc plays a vital role in developing in depth understanding of a prediction model and the factors that lead to a predictions made by it, with this constant learning and feedback not only is it possible to create a transparent model but also improve on the results with human feedback to fine tune the model for a higher degree of accuracy than before.

Assessments Tools and Types of Therapies for Motor Skills Enhancement in the Elderly with Dementia

Isra Hassan, Arshia Khan
Computer Science Department, University of Minnesota Duluth, Minnesota, USA

Abstract - The literature review aims to consolidate the different assessment tools and types of therapies that are used to enhance motor skills in elderly people living with dementia. This paper follows the PRISMA formatting and checklist. Dementia is a clinical state occurring at later stages of human life whereby the elderly experience not only a decline in cognitive function, such as loss of memory and judgment, but also in complex motor skills and functionality, leading to a decline in independent performance in activities of daily living and therefore, a deterioration in the quality of life. Our focus for this literature review is on the decline of motor skills, with the objective of identifying effective tools for assessment and therapies that can be implemented for improving the motor and process skills of elderly people living with moderate to severe dementia.

Effects of Individual Cognitive Stimulation Therapy on People with Dementia: A PRISMA- Guided Systematic Review

Alfia Pervez, Arshia Khan
Computer Science Department, University of Minnesota Duluth, Minnesota, USA

Abstract - Background: Individual Cognitive stimulation therapy (iCST) has been used to improve cognitive function and quality of life. Aim: To analyze the efficacy of the CST among people with dementia. Design: Systematic review and meta-analysis. Results: A total of 17 studies were included in the review. Individual Cognitive stimulation therapy can be used as an intervention to improve cognition and quality of life. Conclusions: Individual Cognitive Stimulation Therapy improves the quality of life of the participants including people with dementia and their carer. It can potentially be used to improve cognition, but further research may consider the implementation of a standard protocol with a rigorous study design to provide sufficient evidence to address this area.

Improved Navigation for Social Robots Through Process Offloading

Maryam Kameli, Thomas Zurales, Arshia Khan
Department of Computer Science, University of Minnesota Duluth, Minnesota, USA

Abstract - The aging population warrants new mechanisms to help care for the elderly. Robotics is one such mechanism that requires the robots to navigate to the residents. Pepper robot is used in nursing homes to navigate to the residents. The sensors that help Pepper navigate are less than par in helping Pepper navigate to the residents. To help Pepper map the nursing home for navigation we equipped it with a LiDAR detector and an NVIDIA Nano. This paper discusses the design and the enhancements to Pepper to help navigate. Limitations and struggles will also be shared along with some use cases.

Graph OLAP-based Influence Analytics for Complex Gene and Disease Influence Network

Mira Kim
IBM Automation, International Business Machines Corporation Costa Mesa (IBM), California, USA

Abstract - Genes provide vital information in determining disease occurrences as pattern analytics of gene-disease association became critical in diagnosing, treating and preventing diseases. It is estimated that humans have between 20,000 and 25,000 genes, and hence the influence network between gene and disease can be highly complex. Then, there is a key technical challenge of devising time and resource efficient algorithms for identifying relevant genes or plausible diseases. We present an efficient Graph OLAP-based framework for analyzing gene and disease influences. We propose 3 classes of algorithms to efficiently explore their relevance.

Effects of Emotion Therapies on People with Dementia: A PRISMA-guided Systematic Review

Sakina Rao, Arshia Khan
Department of Computer Science, University of Minnesota, Duluth, Minnesota, USA

Abstract - Background: Emotion therapies are known to have positive effects on emotions and behavior. However, not many studies have been conducted to verify their effectiveness for neuropsychiatric symptoms among people with dementia. Aim: To analyze the efficacy of emotion therapies among people with dementia. Design: Systematic review. Methods: Articles related to emotion therapies for dementia patients were searched on PubMed. Studies were selected for the analysis based on the inclusion and exclusion criteria. Results: A total of 18 studies were included. Overall, emotion therapies showed a positive effect on behavior, assessments, and quality of life among people with dementia. Conclusion: Emotion therapies are an effective way to improve the quality of life of people with dementia. Integrating emotion therapies in dementia care is recommended as it improves neuropsychiatric symptoms of dementia.

Covid-19 Prevention Online Information Systems

Tyson Howard, Zizhong John Wang
Department of Computer Science, Virginia Wesleyan University, Virginia Beach, Virginia, USA

Abstract - The purpose of this paper is to present a class of online systems that provide users with the necessary information that they need to know about Covid-19 and give users the ability to check their own symptoms. The

platform also allows users to click a variety of links to helpful websites such as the CDC website, John Hopkins website, and local health department website. The key feature to this project is the self-assessment survey, which allows users to see if they may have contracted with COVID-19. The system is implemented by using HTML, PHP, CSS, and MySQL.

Free Technical Solutions for Ecological Momentary Assessments - Searching GitHub Plus Google

Michael Stach, Manfred Reichert, Fabian Prasser, Harald Baumeister,
Winfried Schlee, Peter Heuschmann, Matthias Rose, Rudiger Pryss
Institute of Databases and Information Systems, Ulm University, Ulm, Germany;
Berlin Institute of Health, Charite - Universitatsmedizin Berlin, Berlin, Germany;
Department of Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany;
Eastern Switzerland University of Applied Sciences, St. Gallen, and
University of Regensburg, Regensburg, Germany;
Institute of Clinical Epidemiology and Biometry, University of Wurzburg, Wurzburg, Germany;
Department of Psychosomatic Medicine, Charite Universitatsmedizin Berlin, Berlin, Germany

Abstract - Smartphones are increasingly being used to deliver healthcare services and to conduct medical research using apps. The emerging field is also called mobile health (mHealth). Many specific data collection strategies have been developed for medical data to mitigate biases of more traditional methods (e.g., the recall bias). In the larger context, the aim is to collect data in realtime and real-life, also referred to as data with higher ecological validity. One strategy that is particularly prominent is called Ecological Momentary Assessments (EMA). Even if other data collection strategies are used, they are usually combined with EMA. For this purpose, scientists are faced with the problem of having to locate a technical mHealth platform supporting EMA. To simplify this task, we conducted a search on GitHub and applied the PRISMA guidelines, with the goal of identifying current platforms and frameworks that can be used freely and support on-premise installations. In addition, we did a Google search to see if there are frameworks that meet the above criteria, but are not published on GitHub. The results show that despite great popularity in the scientific community, only few freely available EMA platforms and frameworks exist.

Data Engineering to Support Intelligence for Precision Medicine in Intensive Care

Nasim Sadat Mosavi, Manuel Filipe Santos
Information Systems and Technology (Algoritmi Research Center), University of Minho, Guimaraes, Portugal

Abstract - This paper aimed to present the unique data engineering work for dealing with fragmented and infrequent data collection and to integrate data from Intensive Care (ICU) with other resources. The outcome of this data processing supports the development of an Intelligent Decision Support System for Precision Medicine (IDDS4PM) by providing the possibility to analyze all the clinical events in one platform from the date/time of admission. Thus, to obtain the precise treatment, whole clinical data will be considered regardless of the diversity of data sources and frequency of data creation.

Computational Analysis of a Light-Weight SUVr Processing Technique for Neuroimaging Alzheimer's Disease

Robin Perry Mayrand, Christian Yaphet Freytes, Luana Okino Sawada, Micheal Adeyosoye,
Rosie E. Curiel Cid, David Lowenstein, Ranjan Duara, Malek Adjouadi
Electrical and Computer Engineering, Florida International University, Miami, Florida, USA;
Department of Neuropsychology, University of Miami, Miami, Florida, USA;
Wien Center for Alzheimer's Disease and Memory Disorders, Mt Sinai Medical Center, Miami, Florida, USA;
School of Computing & Information Sciences, Florida International University, Miami, Florida, USA;
Cognitive Neurosciences and Aging, University of Miami, Miami, Florida, USA

Abstract - The Standard Uptake Value (SUV) is conventionally calculated using the ratio of the injected PET radiotracer dose and subject body weight (Binj). SUVs are used to obtain SUV ratios (SUVr), an important metric in many Alzheimer.s Disease (AD) neuroimaging studies. However, SUVr can be obtained using only neuroimaging data, bypassing the need for Binj. This paper proposes the SUVr-LightWeight (SUVr-LW) algorithm which is not reliant on clinical data and instead focuses on PET intensity values. The SUVr-LW was evaluated using the Centiloid Project Florebetaben (FBB) subject cohort and reached a linear regression slope of 0.98, while the healthy control subjects produced a slope of 0.87.

Inflammation Assessment of Burn Wound with Deep Learning

Saeka Rahman, Miad Faezipour, Guilherme Aramizo Ribeiro, Elika Ridelman,
Justin D. Klein, Beth A. Angst, Christina M. Shanti, Mo Rastgaar
School of Engineering Technology, Purdue University, West Lafayette, Indiana, USA;
Department of Surgery, Division of Pediatric Surgery
Children's Hospital of Michigan & Wayne State University, Detroit, Michigan, USA

Abstract - Accurate, reliable, and on-time diagnosis, treatment, and follow-up are crucial for effective burn wound management. Currently, the burn injury classification is performed by visual inspection by experienced clinicians and is based on a wound's size and depth. However, clinical inspection presents many shortcomings, including inter-rater variability and poor prognostic accuracy. An automated burn assessment framework could address these challenges and therefore improve burn wound outcomes. This research proposes a convolution neural network (CNN) based deep learning model to assess burn wound. In so doing, the proposed model detects the degree of inflammation of the burn wounds undergoing skin grafting treatment. The dataset used to validate the model was prepared from the 2-D images collected from the Children.s Hospital of Michigan/Wayne State University, USA. Experienced burn providers performed the labeling. Based on the ground truth of the labels, the model's accuracy on the test dataset is found to be 0.8750.

Question Formulation and Transformer Model Resilience

Ashraf Elnashar, William Schreiber, Jules White, Douglas C. Schmidt
Department of Computer Science, Vanderbilt University, Nashville, Tennessee, USA

Abstract - Question-answer is a paradigm that seeks to provide automated responses to queries posed in natural language utilizing a body of textual content as the source of the answers. A key research challenge is how the changes in question formulation affect the stability of current question-answer transformer models. This paper conducts a preliminary analysis of the stability of question-answer transformer models in the medical domain when the same question is asked in different orders or with other semantically identical variations. The results from our experiments demonstrate that the arrangement of words influences the outcome and consistency of answers from transformer models.

Generalized Pandemic Equation for Monitoring COVID-19

Michael Shur
Electric Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, New York, USA

Abstract - The generalized Pandemic Equation describes multiple waves of the COVID-19 pandemic, models the space dependence of the infection rate, and compares different pandemic evolution scenarios by extrapolating the pandemic evolution curves for the periods of time on the order of the Pandemic Equation instantaneous characteristic time constant. The parameter extraction for multiple locations and time periods could be used for uncertainty quantification of the Pandemic Equation predictions.

The Applications of Machine Learning in Medicine

Sailesh Bk, Nicholas Schmidberger, Louis Bentil, Shukri Ali, Steven Westmoreland, Mohammed Mahmoud
Department of Computer Science & Information Systems,
Minnesota State University, Moorhead, Minnesota, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - The area where Machine Learning can be helpful in medicine is the prediction of outbreaks. We have seen the importance of outbreak predictions in the past two years. This paper presents an in-depth study of different applications of Machine Learning in medicine. We will discuss AI based technologies that can prevent future outbreaks. Digital diagnosis that is based on Machine Learning and the difference between supervised and unsupervised learning will be covered.

Internet-of-Things Monitoring of Physical Restraint Patients

Chelsea Yeh, Alexander W Lee, Melissa S Lee, Karin C Li
Walnut Valley Research Institute, Walnut, California, USA;
University of California Riverside School of Medicine, Riverside, California, USA

Abstract - In this paper, we describe the monitoring of physically restrained patients with Internet-of-Things (IoT) technology. Physical restraints are used in nursing homes and hospitals for confused, agitated patients to prevent falls, injuries, and other bodily harm. However, physical restraints have been shown to have many potential negative outcomes, and their use has been strongly discouraged. Physical restraints can increase the length of hospital stays, develop injuries, and even cause deaths. Physical restraints require constant monitoring to avoid negative outcomes. The proposed IoT physical restraint management will continuously monitor patients who are put under physical restraint. IoT sensors detect the motion of the physical restraint or the equipment or device on which the patient is restrained. The sensor data is processed for each physically restrained patient, and the staff is alerted to take immediate remedial action if a high-risk event is detected. The research created the IoT sensors and devices used to capture the physically restrained patient's motion and developed the staff alert application that runs on mobile phones and consoles located in the nurses' stations.

Real-Time Monitoring of Urine Output with Internet-of-Things Connected Foley Catheters

Alexander W. Lee, Hsi-Jen James Yeh
Department of Engineering and Computer Science, Azusa Pacific University, Azusa, California, USA;
Walnut High School, Walnut, California, USA

Abstract - We describe the use of Internet-of-Things (IoT) connected Foley catheters that enables real-time monitoring of urine output. Urine output is an important indicator of the level of renal or urinary impairment in patients. A Foley catheter is a device that allows the collection of urine from patients and has volume gradation that allows the reading of urine volume periodically by the nursing staff. The readings are typically performed once or twice per day with low levels of accuracy and time resolution. We proposed and developed an IoT-enabled system that captures urine output in real time by continuously monitoring the weight of the urine collected by the Foley catheter. The data is sent via wireless networking to a data collection and analysis server which provides accurate analyses of the patient urine output. The data can then be used by the physician to measure real-time urine response to medication and other medical interventions.

Development of a Healthcare Monitoring Diabetes Mobile Application for Community

Kanyarat Bussaban, Kanyarat Bussaban, Nareenart Ruksuntorn, Jaruwan Chutrtong, Chanyapat Sangsuwan
Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand;
Computer Science Program, Suan Sunandha Rajabhat University, Bangkok, Thailand;
Robotics Engineering Program, Faculty of Industrial Technology,
Suan Sunandha Rajabhat University, Bangkok, Thailand

Abstract - The purpose of this project is to develop the mobile application, by applied Machine learning, for analyzing, collecting, monitoring, and retrieving information between patients with diabetes especially diabetes type 2 and village public health volunteers and to study the impact of using mobile application based on self-learning and self-management in diabetes information. This is a research and development mobile application and the sample consisted of 30 diabetes patients and 5 village health volunteers participated in this research. The project has demonstrated the effectiveness of using mobile application to support patients and village health volunteers. The results showed that user satisfaction has a high level.

HCI for Mobile Healthcare During the COVID-19 Pandemic

Mitchell Macuga, Nicholas Wysocki, Rick Dymond, Terrell Brooks, Travis Bauman,
Zachary Layman, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - Human-Computer Interaction (HCI) is a field of study within technology that can encompass multiple disciplines in practice. The general focus being on the design of computer technology and the interaction between humans and that technology. HCI has been a continuously evolving field since the 1980.s. The concept of HCI itself has expanded since its inception and now covers many aspects of information technology and design. The ideal designer of an interactive system now requires expertise in a broad range of topics such as psychology and cognitive science. These skills can be used to obtain knowledge of a user's perceptual, cognitive, and problem-solving skills during system development. Through this paper we will be researching the concept of HCI and how it has helped in the development and advancement of mobile healthcare. This is especially relevant in recent years given the current state of health care regarding the COVID-19 pandemic and how HCI has developed since the beginning of its spread worldwide. Topics such as remote healthcare consultation, accommodations for individuals

with disabilities, compliance and security concerns, as well as mobile healthcare application development will also be discussed in this paper.

Modelling Companion Animals' Well-being based on Activity Recognition and Biological Rhythms Monitoring

Hazar Zilelioglu, Ioana Ocnarescu, Ghazaleh Khodabandelou, Clement Bataille, Solene Le Bars
LISSI Laboratory, University of Paris Est-Creteil (UPEC), Capgemini Engineering, Paris, France;
Robotics by Design Lab, Strate school of Design, Paris, France;
Frog Lab, Frog Design, Paris, France;
Future of Healthcare Program, Capgemini Engineering, Paris, France

Abstract - Human-animal companionship is a particular relationship where the well-being of a companion animal entirely depends on their human partner's care. However, proper care of an animal companion can become challenging with aging due to autonomy loss (i.e., motor or cognitive disabilities). In this regard, this very early-stage study proposes a data collection protocol centered on pet dogs, over several days in the context of an elderly-pet relationship. The goals of this study include dog activity recognition (e.g., resting, running, eating, etc.) and behaviour modelling. In order to do so, machine learning techniques will be used and daily behaviour patterns will be modeled over several days using Markov models. This research consolidates a data base on animal behaviour while bringing insights on how to measure the well-being of animals at home.

Steady-state Visual Evoked Potential Classifiers for their use in Industrial Brain-Computer Interfaces

Dmitri Bryzgalov, Gaspard Andre
Capgemini Engineering, Paris, France;
Departement Mathematiques, Universite Paris Saclay, Paris, France

Abstract - Compared to brain-computer interfaces (BCIs) developed in the laboratory settings, real-world BCIs are generally constrained by several factors: low resolution of the signal, its low signal-to-noise ratio and a lack of well-labeled data. In these conditions, good performance of BCI classifiers become even more challenging. In the current project, we have set up an EEG study to enrich the repertoire of BCI classifiers of steadystate visual evoked potentials (SSVEPs) and adapt them to industrial and clinical settings. In the early state of the project, we have tested existing solutions, such as deep convolutional neural network, canonical correlation analysis (CCA) and task-related component analysis (TRCA). Based on the performance of each algorithm, analysis of their perspectives was made, and the program of further research was outlined to create adaptive and flexible repertoire of SSVEP classifiers.

Event Abstraction in Medical Context Using Self Organizing Maps

Patrick Philipp, Marie Bommersheim
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB,
Fraunhofer Center for Machine Learning, Karlsruhe, Germany

Abstract - With the digital transformation in medicine, enormous amounts of data are being generated and are available for analysis. Process mining techniques can be utilized to extract process models from this data. On the one side these process models typically provide detailed, fine-granular activitiy descriptions. But on the other hand these models become increasingly less recognizable. Therefore, in this contribution we explore the use of Self Organizing Maps for an event abstraction in the medical context. Our approach achieved promising results on a publicly available sepsis data set.

Nutrient Profiles of Food Items in Adolescent Diet: A Cluster Analysis of Data Collected from a High School Citizen Science Project

Jonah T. Treitler, Senait Tekle, Jennifer Ushe, Qing Zeng
Jefferson High School for Science and Technology, Alexandria, Virginia, USA;
Department of Clinical Research and Leadership, George Washington University, Washington, DC, USA;
Alexandria City Public Schools, Alexandria, Virginia, USA

Abstract - A healthy diet during adolescence is essential to youths. physical and mental development. This study performed a cluster analysis on a dietary record dataset collected through a high school citizen science project. The analysis was performed based on nutrients rather than food groups to characterize the adolescent dietary patterns better. The resulting nine food clusters have distinct nutrient profiles and vary significantly in size.

Monitoring Circadian Rhythms of Night-shift Operators with Hidden Markov Models

Rodrigo Balp, Solene Le Bars
Altran Technologies, Capgemini Engineering, Meudon, France;
Altran Lab, Capgemini Engineering, Meudon, France

Abstract - Chronotherapy refers to the delivery of therapeutic treatments following a person's physiological cycles, also known as circadian rhythms. These cycles may be thought as an individual's "biological clocks", having an approximate periodicity of twenty-four hours. Recent research has found that coordinating with and adapting to circadian rhythms can maximize treatment effectiveness, minimize treatment side effects, or both. However, disturbance of circadian rhythms is not limited to sick patients. For example, sleep perturbation may induce a disruption of circadian rhythms, similar to acute or chronic stress. In the present article, we discuss the interest of gathering information concerning circadian rhythms of night-shift workers and applying Hidden Markov Models to such data in order to extract important information about these rhythms.

Intelligent Tutoring Systems in Healthcare for Low Literacy Population: A survey and the next steps

Ryan Ackerman, Francisco Icobelli, Renu Balyan
Math and CIS Deptartment, SUNY Old Westbury, Old Westbury, New York, USA;
Department of Computer Science, Northeastern Illinois University, Chicago, Illinois, USA

Abstract - Intelligent tutoring systems (ITSs) can offer a scalable alternative to instruction for underserved populations which can help in healthcare related fields, such as breast cancer, where education can greatly bridge inequity gaps. Breast cancer is one of the leading causes of cancer-related deaths in the Hispanic population. Breast cancer survivors often experience ongoing diagnosis and treatment-related symptoms that negatively impact their health-related quality of life. In this paper we review existing ITSs in healthcare that cater to low-literacy population education with the goal of building a prototype that improves the quality of ITSs in healthcare for Hispanic breast cancer survivors.

CSCI-RTSE:
RESEARCH TRACK ON SOFTWARE ENGINEERING

Towards Specification Completion for Systems with Emergent Behavior based on DevOps

Mohamed Toufik Ailane, Adina Aniculaesei, Christoph Knieke, Andreas Rausch, Fauzi Sholichin
Institute for Software and Systems Engineering, Clausthal University of Technology,
Clausthal-Zellerfeld, Germany

Abstract - Software systems may experience multiple emergent behaviors during their operation time. These emergent system behaviors occur when system engineers develop their system under the closedworld assumption, but this assumption is not met during its operation. This means that system engineers work on the basis that they have complete knowledge of the system and its environment during its design, when the system specification that has been created is actually incomplete. In this paper, an observation of an emergent behavior is considered to be a solid proof that the system model specification is still incomplete. A conceptual framework is proposed to harness the emergent behavior and complete the system specification that is provided during the its design. The framework consists of two parts, system development and system operations. It is built on a modeldriven approach in order to provide a clear definition of the emergent behavior and a concrete development scheme. The framework exploits the DevOps paradigm as a successful paradigm to achieve the ultimate goal of developing complete system models through the continuous specification completion based on the observed emergent behavior. The goal of this framework is to help develop high-quality and reliable emergent systems based on the specification derived from the emergent behavior that occurred at run time.

A Research Agenda for Embedding 4IR Technologies in the Leadership Management of Formal Methods

John Andrew van der Poll
Digital Transformation and Innovation, Graduate School of Business Leadership (SBL),
University of South Africa (Unisa), Midrand, South Africa

Abstract - The use of Formal Methods (FMs) in software development holds much promise for constructing provably correct, or at least highly dependable software. That said, the use of such techniques remains controversial - advocates of the use of FMs cite high-quality software as a major selling point, while critics point to the steep learning curve in acquiring the necessary mathematical skills. In this paper we investigate aspects around ambiguity of semi-formal specifications with respect to the Quality 4.0 framework. Following these we argue that the opportunities in the Fourth Industrial Revolution (4IR) coupled with leadership-management support may facilitate the use of FMs as a software methodology, not only for missioncritical software, but equally for Business ICTs. Following our analyses, we propose a research agenda for investigating the leadership-management, FMs and 4IR triad, aimed at facilitating the adoption of FMs in the new industry.

Test Case Prioritization for Mobile Apps

Zeinab Abdalla, Kerstin Haring, Anneliese Andrews
Department of Computer Science, University of Denver, Denver, Colorado, USA

Abstract - Like for any software, Mobile Applications (Apps) are modified during their specification, implementation, and maintenance phases with the goal to satisfy new requirements, fix defects, and change or add functionality. There is a need to regression test for and detect faults in every phase. However, resource and time constraints may lead to Mobile Apps not being tested. In this paper we present a model-based test approach to prioritize test cases based on the input complexity for each test path of the Mobile App. We argue that this novel approach will significantly improve the efficiency and effectiveness of current techniques.

Guidelines for Combining Regression Testing Approaches

Zeinab Abdalla, Kerstin Haring, Anneliese Andrews
Department of Computer Science, University of Denver, Denver, Colorado, USA

Abstract - Regression testing is an important activity in software maintenance and software enhancement as it can help developers to determine if changes made to the system are handled appropriately without compromising efficiency. While not employed often, it is possible to combine several regression testing techniques. Combining them can lead to a more efficient and effective regression test suite. The three types of regression testing, selective, minimization, and prioritization, can be combined in four different ways: There are three ways to combine two of the approaches and one way to combine all three. However, to efficiently and effectively employ regression testing, it is crucial to select the appropriate combination for the software that is to be tested. For example, the expected quality of the software, the kind of changes made, or the criticality of the software heavily influence which strategies to combine and in what order. This paper presents guidelines for combining regression testing approaches based on a systematic approach. We outline all possible situations that can occur and show how each of them influences which combination to use.

Uncertainty and Dependency-wise Requirements Prioritization

Norah Ibrahim Alfassam, Mohammad Abdullah-Al-Wadud, Mubarak Rashed Alrashoud
College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

Abstract - With the increase of complexity in software systems, optimizing the requirements management has become an important task in Software Engineering. The selection and prioritization of software requirements among plenty of requirements for each release of a software is a common problem. Several methods have been introduced to solve this common problem. Literature shows that the majority of existing requirement prioritization (RP) approaches ignore fundamental factors such as uncertainty in cost and interdependency between requirements. The objective of this paper is to present an approach to solve the RP problem that considers, with other commonly utilized factors, uncertainty and interdependences between requirements.

On the Educational and Professional Implications of Integrating Mind Mapping in Software Testing

Pankaj Kamthan, Nazlie Shahmir
Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada;
IS Quality Engineering, Canadian Pacific Railway Limited, Calgary, Canada

Abstract - The increasing societal dependence on software, and the negative consequences of undesirable behavior of software, calls for ever more attention to software quality. The purpose of software quality control is activities that support the achievement of desirable levels of software quality, and software testing is one such activity. In order to preserve the intent of software testing, it must be deliberated carefully, and given necessary time and due diligence, before being acted upon. In that regard, this paper adopts a human-centered approach to software testing, embraces mind mapping as a lightweight technique for early software testing-related commitments, provides elements of a theoretical basis for mind mapping, and offers a guided tour of software testing-related mind maps using practical examples relevant in academia as well as industry.

GATE II: Visualizing Semantic Web Search

Nisha Kottekat, Kevin Gary
School of Computing and Augmented Intelligence, Arizona State University, Tempe, Arizona, USA

Abstract - The semantic web is a mesh of information linked in a way that can be easily shared and reused to make inferences for the end user. The semantic web attempts to find and access web sites and web resources not by keywords but by descriptions of their contents and capabilities. This has been made possible by adding structure to the content of web pages and developing an environment where software agents can perform sophisticated functions for users. Semantic web customization of JMaPSS uses the Java Marker Passing Search System (JMaPSS) which applies a spreading activation search algorithm known as marker passing to significantly improve search results. This research project focuses on visualizing the semantic network and displaying marker propagation to distinctively illustrate various elements of the semantic web ontology in the form of a graph structure. The tool displays the results of marker propagation by highlighting the active nodes and the propagation path.

A Process Model for Ensuring Diffusion and Adhesion of Cybersecurity Innovations in Software Development Organizations

Gregory Vsevolozhsky, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract - Optimizing existing business processes through centralization and uniformity of solutions is the preferred strategy for cyber risk reduction. However, the analysis of cybersecurity needs that reflect the objectives, design, and architecture of the respective missions or products is necessary to ensure the best alignment of cybersecurity innovations with various facets of the overarching organizational objectives. By employing the design science methodology, this study developed an artifact that incorporated the closed-loop Innovation Adhesion Evaluation Process (IAEP) to be utilized by subject matter experts (SME) in assessing the feasibility, timeliness, suitability, and sustainability of cybersecurity-related innovations during the operational phase of a project life cycle. The study found that a group of SMEs can use the IAEP model to evaluate the proposed innovation's potential impact on missions or products.

Towards the Development of Text-based Format Converters for Object Representation

Carlos R. Jaimez-Gonzalez, Betzabet Garcia-Mendoza
Departamento de Tecnologia de la Informacion, Universidad Autonoma Metropolitana, Unidad Cuajimalpa, Mexico City, Mexico

Abstract - Interoperability is the ability to communicate between applications that are written in different programming languages. The communication is established through the exchange of data, which has normally a representation in a standard format. The interoperability problem can be solved by implementing serializers from objects to standard formats, which can then be deserialized in any programming language. This paper presents a project that proposes a solution to address the interoperability problem, which is closely related to the implementation of object serializers. The project proposes to address two important aspects: 1) the representation of objects in standard text formats, independent of the programming language to allow interoperability between applications; and 2) the teaching of the processes involved to represent objects, through the incorporation of these topics in some of the courses of an undergraduate degree in Information Technologies and Systems.

Improvement Quality Software Requirements by using a Triplet Structure

Bruel Gerancon, Sylvie Trudel
Department of Computer Science, University of Quebec at Montreal, Montreal, Canada

Abstract - The development of a quality software system is a priority in the domain of software engineering. Quality requires to define unambiguous and consistent requirements that are not conducive to various interpretations. Indeed, the success of the realization of a software system depends largely on the phase of software requirements specification. The requirements specification phase consists, among other things, in describing in a precise and unambiguous manner the characteristics of the system to be developed. Moreover, the techniques for writing software specification documents used in the industry don.t facilitate to define unambiguous and consistent requirements. In industry, software requirements are often written in natural language, and no technical details are specified. Thus, software requirements are incomplete, inconsistent and prone to ambiguities, and therefore interpretation errors can easily be made by analysts. This article introduces a new technique for writing and developing software requirements that could help to reduce the ambiguities and inconsistencies in the document of specification software requirements. Our technique is validated by the development of a new tool that detects the ambiguities and inconsistencies in the software requirements, and generates the potential methods and classes from requirements written in natural language. Our tool integrates a set of techniques in natural language processing (NLP), and in artificial intelligence which helps to improve the software requirements quality.

Measuring Influencing Factors of API Usability

Jeff Tian, Sultan Alanazy, Abdullah Bokhary, Sultan Alharthi, Sakher Ghanem
Department of Computer Science, Southern Methodist University, Dallas, Texas, USA;
College of Computer Science and Information Technolog,
Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia;
Department of Software Engineering, College of Computer Science and Engineering,
University of Jeddah, Jeddah, Saudi Arabia;
Department of Computer Science and Artificial Intelligence,
College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia

Abstract - In this paper, we identify the influencing factors of API usability by extending a comprehensive framework for measuring API usability we proposed earlier, through a systematic examination of entities involved and artifacts produced in API development and usage. Existing research in software metrics are adapted to define metrics capturing relevant metrics dimensions and at the appropriate levels. Our current set of identified influencing factors includes metrics defined on API documentation and other entities/artifacts, in addition to the code metrics. Some preliminary results from actual measurement of such influencing factors are included to demonstrate the viability of our approach.

Hospitality Industry 4.0 - Customize and Optimize Real-Time Allocation of Rooms

Vijayarangan Natarajan, Rashmi Kumari, Sakthi Priya M
Travel & Hospitality-Strategic Initiative group, TCS Ltd, India

Abstract - In the modern era of Industry 4.0, where customization and offerings play a major role in the commercial world, different industries have implemented various realistic approaches to enable their customers to choose the products with available preferences and priorities. Industries like retail have enabled their customers to choose their products based on their preferred colors, sizes, and materials; automobile industry has enabled their customers to pick their own colors, materials for interiors, add-ons for safety features etc., for their chosen vehicles. Equivalently, different entities in the hospitality industry have also made several attempts to meet the customers' expectation of customizing their products (i.e., rooms) with the available preferences and priorities (i.e., room amenities and hotel properties), but their systems are being unable to cope with the complex operational challenges in accommodating rooms at business or transit hotels that have a large room inventory and an average occupancy of more than 50% per day. To overcome the functional challenges like the transfer process of guests (customers), their timeline and turnaround process of rooms, we have invented an efficient system having 3 main components: 1. The Room Type Assignment System helps guests to get into their desired room types along with their preferred amenities. 2. The Check-In Check-out (CICO) Machine Learning Prediction Model helps hotels to get all the rooms that are available to be mapped to the guests at the time of check-in. It also provides those rooms whose booking gets canceled and are now available to get mapped again to any other guest. 3. Looping Method for Room Number Assignment helps in real-time allocation of a preferred room type to the guest along with the room number based on his/her preferences and priorities. Hence, our proposed system will not only offer magnificent experiences to our guests and improve hotel room occupancy rates but also commencement of doors to more enhanced operational models for the hospitality industry.

How We Develop an Online Kanban Board Game in Two Months?

Yu-Hsiang Hu, Chien-Tsun Chen, Yu Chin Cheng
Computer Science and Information Engineering, National Taipei University of Technology, Taipei, Taiwan;
Teddysoft, Taipei, Taiwan

Abstract - In the Agile and Lean Software Development course at Taipei Tech, students play an in-person board game called getKanban to learn the essential concepts of Kanban. The COVID19 outbreak turned instruction online and made playing the inperson game impossible. Since an online substitute was not yet available, we decided to develop our own online version of the board game based on ezKanban, a kanban management system we previously developed. We were able to produce the game in two months and put it in use for online instruction. We attributed the fast development to reuse and extension enabled by adopting Domain-Driven Design, following the guidelines of Clean Architecture in the development of ezKanban, and keeping the game related logic in frontend.

Web System for Storing and Visualizing Web Objects in XML

Jose M. Hernandez-Salinas, Carlos R. Jaimez-Gonzalez, Betzabet Garcia-Mendoza
Departamento de Tecnologia de la Informacion, Universidad Autonoma Metropolitana, Unidad Cuajimalpa, Mexico City, Mexico

Abstract - Web Objects in XML (WOX) is a framework for creating object-based distributed applications, it supports the interoperability among different object-oriented programming languages, it uses XML as the format representation for objects, and it uses HTTP as its transport protocol. This paper presents a web system that was developed to complement the functionality of WOX, which allows the storage of distributed objects and the visualization of their state and methods. The web system has a repository, in which objects can be visualized and their methods can be executed through a web interface where the user can provide values for each of their parameters.

GreenWebAdvisor: Discovering and Recommending Hidden Eco-Design Best-practices

Moussa Aboubakar, Moubarak Zoure, Ines Ben Kraiem, Mouna Ben Mabrouk
Sogeti, France

Abstract - Every day, web applications proliferate on the internet, waste critical quantities of energy, and emit a substantial mass of CO2. Developers of these web applications often fail to implement software eco-design tenets, either because they are unaware of such tenets, or because they lack practical tools to assist them. Thus, in this abstract paper, we introduce GreenWebAdvisor, an intuitive eco-design framework that assists developers in building green web applications. GreenWebAdvisor 1) monitors the carbon emissions of web applications, 2) assesses their eco-designess, and 3) provides developers with coding best practices recommendations to optimize the energy use and the carbon footprint of Web applications. Compared to existing tools, GreenWebAdvisor goes beyond static practice recommendations by leveraging artificial intelligence algorithms to discover hidden good/bad practice patterns and ultimately provide richer ecooptimization suggestions to developers.

Identifying Impact Criteria for Software Requirements Verification and Validation Tools Selection

Khalid Aldousari, Mohammad Abdullah-Al-Wadud
Department of Software Engineering, King Saud University, Riyadh, Saudi Arabia

Abstract - In the past few decades, the software requirement gathering phase as part of the software development life cycle has become of great interest to software project managers, developers, and clients. It has become a critical factor in project success and defining its quality. The requirement gathering phase is how developers gather the software projects. functional and nonfunctional requirements. Therefore, developers need to go through intensive testing activities to ensure that these requirements are correct and sufficient. In addition, project managers employ software tools for requirements verification and validation. However, there is a lack of research studies for identifying criteria affecting the selection of software tools used for requirements verification and validation. In this work, we focus on identifying some criteria to choose software tools for requirement verification and validation to serve the requirements gathering phases. The criteria can be utilized as a selection guide for IT employees and managers to select appropriate verification and validation tools.

Automating File Operations via Python

Andrew Miller, Nicholas Kobylski, Elias Qamar, Jimmy Xiao, Nicholas Veal,
Richard Kenney, Nicholas Wysocki, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - Python is a computer programming language which is used for many purposes, including automating tasks. To reduce cost and increase the level of operating efficiency, network engineers made this possible by developing network automation techniques to automate recurring tasks, such as file backup. File backups are a necessity for maintaining data in the local storage. Using Python and a collection of modules and functions, such as shutil to copy data from one file to another, and os and sys to obtain the file path allows even a beginner to develop this automation. Up until now, legacy methods used by network engineers were not only time-consuming, but they required relevant knowledge about the related protocols and technologies. We are now lucky to have support from most major network companies which resulted in a large open-source community where you can find key information to use the power of automation within many different applications. In this paper, we will focus on task automation, such as automating file backups, the different ways it can be implemented through techniques and technologies, and why it is essential for personal and business use.

CSCI-RTED:
RESEARCH TRACK ON EDUCATION

Improving Autonomous Systems Education: A Literature Survey

James I. Maher, Shannon Beck, Adrian A. de Freitas, Troy Weingart
Department of Computer and Cyber Science, United States Air Force Academy,
USAF Academy, Colorado, USA

Abstract - As an undergraduate-only institution, we found value in teaching Autonomous Systems (AS) to our Computer Science majors. In order to improve our teaching of AS, we conducted a literature survey on the current state of ProjectBased Learning (PBL) used in AS education. Additionally, we survey current AS hardware platforms and provide recommendations that reduce extraneous course setup costs and limit course preparation frustrations; resulting in improved student success and learning rates. We identified five major trends in successful AS course PBL literature. We also observed professors inconsistently collecting survey data from students, which created difficulty in assessing the effectiveness of various AS teaching techniques. Our proposed future work encourages designing a learning assessment framework for AS post-course surveys.

A MOOC on Computational Thinking for All: Pedagogical Principles, Challenges, and their Application

Noa Ragonis, Orit Hazzan
Faculty of Education, Computer Science Department, Beit Berl College, Beit Berl, Israel;
Faculty of Education in Science and Technology, Technion, Haifa, Israel

Abstract - For the last two decades, the world of pre-academic and academic education has been occupied with ways of developing computational thinking. Responding to the common perception to develop those thinking skills in all learners, and indeed in the entire population, we developed a MOOC on computational thinking. The course advocates the development of computational thinking skills in every human being, at any age, and in any subject matter, emphasizing the common, essential computational thinking skills: problem decomposition, abstraction, and generalization. Based on our belief in the importance of the application of computational thinking skills when using computerized environments, and the importance of its significant application in developing an understanding of any discipline, the course is based on the development of simulations of computational processes in any area of knowledge. We widely discuss the pedagogical challenges of developing a MOOC on thinking skills without teacher-learner interaction, and share how we overcame these challenges by implementing the Four Pedagogies for Developing Computational Thinking (4P4CT) framework, which integrates the pedagogies of active learning, project-based learning, product-based learning, and contextbased learning. We present preliminary findings from research we conducted during the first two MOOC cycles with about 1,600 learners. Specifically, quantitative and qualitative data analysis of students' learning processes are discussed, reflecting the students. multi-faceted and deep engagement in the MOOC.

Leveraging Continuous Feedback in Education to Increase C Code Quality

Max Schrotter, Bettina Schnor
Department of Computer Science, University of Potsdam, Potsdam, Germany

Abstract - This paper presents a framework for students learning C programming. CoFee is a modular framework focusing on code security and code robustness by using state-of-the-art software analyzers. Further, error messages are supplemented by meaningful hints suited for novice students. It also follows the theory of situated learning by exposing students to typical software engineering workflows using Gitlab for version control, continuous integration and code quality reports. To check code quality CoFee supports well-established open-source tools which were tested on a purpose build test suite. Its modular architecture allows easy integration of future analyzers. The evaluation within an operating systems course shows that CoFee enhances the code quality of the students. submissions and that it is well-suited for novice students.

'WHO DONE IT?' A Multidisciplinary Mobile Device Forensics Framework: From Theory to Practice with Intermediate Students

Frances C. Dancer, April L. Tanner, Mya N. Thomas, Mary M. Thomas
Department of Electrical and Computer Engineering & Computer Science,
Jackson State University, Jackson, Mississippi, USA;
Shelby County Schools, Memphis, Tennessee, USA

Abstract - It is essential to initiate a curriculum concerning mobile device forensics and law enforcement in the intermediate grades in Mississippi using a facet of multidisciplinary subjects: biology, chemistry, computer science, criminology, and physics. The students participated in engaged teams who gathered the evidence (critical thinking), analyzed the evidence (deductive reasoning), and drew conclusions (inference) in a story-based scenario entitled "Cyberbullying Mobile Device Criminal Investigation". This research presents cyberbullying while bringing awareness to the law enforcement community by understanding digital forensics; furthermore, it shows how STEM and Criminology are presented to intermediate students by solving a middle school mystery based on a multidisciplinary framework. By including a multidisciplinary curriculum in this process, students developed an appreciation of the interrelatedness between the STEM Sciences and Liberal Arts Education.

Co-designing Immersive and Inclusive Virtual Museum with Children and People with Disabilities: A Pilot Study

Aldo Caldarelli, Stefano di Tore, Silvia Ceccacci, Michele Domenico Todino, Lucia Campitiello, Catia Giaconi
Educational Neurosciences and Sports Lab - HERACLE, Niccolo Cusano University of Rome, Rome, Italy;
Department Humanities, Philosophy and Education, University of Salerno, Fisciano (SA), Italy;
Department of Education, Cultural Heritage and Tourism, University of Macerata, Macerata, Italy;
Department Humanities, Philosophy and Education, University of Salerno, Fisciano (SA), Italy

Abstract - Immersive virtual environments represent a great opportunity for museums to enhance visitor experience through edutainment. However, to provide an enjoyable entertainment and learning experience for all visitors, including people with disabilities, the virtual museum must not only be accessible, but also inclusive: they must provide greater equality and cultural and learning opportunities for all social groups. To achieve this goal, the concept of Universal Design needs to evolve into a user-centered approach where people are involved in codesigning the virtual museum experience. In this context, the article describes a pilot study conducted at the University of Macerata, which explores the possibility of using high-fidelity prototyping in a virtual laboratory to support the co-creation of an immersive virtual museum environment with relevant target users, including children and people with disability, from the earliest design stages. The paper presents the results of the codesign process and discusses its implications in defining design requirements to ensure the accessibility of immersive solutions for cultural heritage.

Developing Effective Cybersecurity Labs: Initiating the Conversation

Glen Sagers
Department of Computer Science and Information Security,
Southern Utah University, Cedar City, Utah, USA

Abstract - Hands-on labs are critical to cybersecurity education, but are often based on one instructor's preferences, knowledge, and skills. The chosen topics may not align with established curriculum guidelines, and even when they do, labs may not be maximally effective in teaching the topics, partly because the professor is not an instructional designer. This paper represents a first attempt to draw up guidelines for cybersecurity labs, based on pedagogical principles and best practices in related fields.

Concept and Realization to Automatically Generate Test and Training Data for a Natural Language Processing Algorithm for the Assessment of Free Text Answers for Digital Courses on Work Design in the Context of Psychology Studies in Higher Educational Institutions

Ramona Srbecky, Franz Buhler, Simon Alexander Wetzel, Michael Winterhagen,
Wieland Fraas, Matthias Hemmje
Multimedia and Internet-applications, University of Hagen, Hagen, Germany;
Industrial and Organizational Psychology, University of Hagen, Hagen, Germany

Abstract - In this full paper for the CSCI-RTED research track, we present an approach on how to generate test and training data for a Natural Language Processing (NLP) algorithm to asses free text answers for a digital university course on the topic of work organization and work design in an automated approach without actual data from the learners' as test and training data. For this purpose, this paper first presents the research methodology, questions, and relevant theoretical background. Based on this, the concept and the implementation will be presented. In the end, the evaluation and future improvements are presented.

Exploring the Impact of Student Learning and Programming Pedagogy while Adjusting to COVID-19 (An HBCU Case Study)

Edward Dillon, Krystal L. Williams, Steven Fuller, Elisee Djapa, Abigail Dina,
Makhya Wilridge, Chauncey Rogers
Department of Computer Science, Morgan State University, Baltimore, Maryland, USA;
Louise McBee Institute of Higher Education, University of Georgia, Athens, Georgia, USA

Abstract - Learning to program can be a challenge, but it is an essential skill that all computer science (CS) majors are expected to develop. Much attention has been rendered towards ways to enhance student learning and program pedagogy in the CS classroom. One reason is to improve the retention of CS majors, while also aiding them with the ability to showcase their learned skills in practical and professional settings. Due to COVID-19 pandemic, the way students learn have shifted significantly. One notable shift being transitioning from in-class learning to virtual/online learning. Recent literature indicate that student learning has declined because of this transition. Moreover, emphasis on how COVID-19 has impacted student learning in the CS classroom is needed. This article discusses a case-study conducted on CS majors enrolled at an Historically Black University (HBCU) in the midAtlantic Region of the United States. This study is comprised of pedagogical coding review (PCR) assessments and analyses conducted over a span of six semesters (Fall 2018 to Spring 2021) involving two CS courses, CS2 and an Object-Oriented Programming course. The objective was to determine if a decline in the students. learning

and performance existed from these PCRs during the semesters where online learning was prevalent due to COVID-19. Results revealed that a slight decline occurred during the semesters where COVID-19 was prevalent. These findings contribute to current literature as it pertains to underlying challenges of COVID-19 in higher education. Moreover, this case study directly emphasizes the impacts of COVID-19 on student learning in the CS classroom.

Developing an Interdisciplinary Curriculum for Educational Robotics: Programming, STEM, and Art

Lauren Harter, Alaina Caulkett
Education, VEX Robotics, Pittsburgh, USA

Abstract - Educational robotics provides an interdisciplinary approach to STEM, computer science, and 21st century skill acquisition for students of all ages. Robotics can increase student perceptions of STEM subjects as well as foundational skills through an engaging and authentic learning environment. However, the many benefits of robotics for students and teachers are dependent on the curriculum. In this paper we describe how using the Understanding by Design pedagogical framework supports the development of truly integrated and interdisciplinary educational robotics curriculum. Two examples are provided for a virtual robot that showcase this pedagogical approach for combining STEM, art, and programming in an authentic manner.

Showcasing Native Wildflowers in Virtual Reality

Andy Wagner, Leonidas Deligiannidis
School of Computing and Data Science, Wentworth Institute of Technology, Boston, Massachusetts, USA

Abstract - The goal of this project is to develop a virtual reality program on the Oculus Quest 2 to showcase and educate users on native wildflowers that can be found in North America, specifically in the United States. We have created a virtual environment with framed images of some of America.s native flowers and, using interactive menus, we provide educational information on each plant. In addition to the 2D images, a 360- image viewer was implemented to show some of the native habitats of these plants to create a more immersive learning experience. Using the Oculus Quest.s experimental hand tracking feature, users can switch between using the Oculus controller and using their actual hands as the controller, allowing the users to choose how they interact with the virtual environment. The United States is home to many beautiful native flowers and by showcasing them in virtual reality anyone can experience these flowers and learn about them.

Initial Impact of Evidence-based and Experiment-focused Teaching Approach in a Computer Architecture Course in Computer Science

Md Mahmudur Rahman, Opeyemi Adeniran, Neda Bazyar Shourabi
Computer Science Department, Morgan State University, Baltimore, Maryland, USA

Abstract - The Digital Logic and Computer Architecture related courses are one of the key knowledge areas in Computer Science (CS) used to enhance students' understanding of Boolean algebra, logic gates, registers, arithmetic logic units, etc. and provides insights how software and hardware are related in a computing system. Experimental Centric based Instructional Pedagogy (ECP) with portable laboratory instrumentation might provide real hands-on experience to obtain a practical understanding of the concepts at a lower cost compared to virtual hands-on laboratories. This paper presents the initial adaptation of an evidence-based, experiment-focused teaching approach to introduce the fundamentals of digital logic concepts using the commercial ADALM 1K

Active Learning Module in a Computer Architecture course for the first time in the CS department at a university serving predominantly minority students. To evaluate the impact of the ECP on student performance in the Computer Architecture course, we conducted three different evaluations, which are class observation, signature assignment, and the Motivated Strategies for Learning Questionnaires (MLSQ) survey. The results of the Classroom Observation Protocol for Undergraduate STEM (COPUS) show more student engagement when ECP is implemented; the output of the signature assignment shows an increase in students. learning outcomes; and the MLSQ survey, which tests the students' motivation, critical thinking, curiosity, collaboration, and metacognition, ascertains the effect of the ECP on the CS students who participated in the experiment.

Analysis of Student Learning Outcomes in Data Structures and Algorithms

Dewan Tanvir Ahmed
Department of Computer Science, University of North Carolina at Charlotte, Charlotte, North Carolina, USA

Abstract - Student learning outcomes (SLOs) are the specified knowledge, skills, and abilities that students are expected to achieve by the end of a course or program. Student learning outcomes describe what educators expect students to know and be able to do. SLO's specify an action of a student that must be observable, measurable and demonstrable. On the other hand, grades are objectives measure. This study attempts to quantify SLOs of a graduate level foundation course named Algorithms and Data Structures. Three key aspects of the course are defined to address learning outcomes: a) Algorithm Design Skill, b) Data Structure Selection and Design Skill and c) Algorithm Analysis Ability. For each aspect five achievement classes are specified in order to observe and quantify their knowledge, skills and ability. According to self-assessment, 86% students can a) determine the correct time and space complexities in an elegant and effective way, and b) correctly figure out time and space complexities of the designed algorithm but more elegant analysis exists. The study also shows 86% students can a) design elegant and robust algorithms and b) the designed algorithm is correct and efficient in most cases but it does not scale well in extreme cases. And finally, 79.7% students mentioned that a) data structure is well chosen (i.e., efficient) and well justified, and b) data structure is appropriate, but operations are not described sufficiently clearly. This clearly shows better success because of appropriate course modality, teaching pedagogy and engagement among many others.

Experience in Teaching and Engaging Computer Science and Computer Information Systems Students in Active Learning

Weizheng Gao, Olumide Malomo, Somasheker Akkaladevi, Ephrem Eyob,
Adeyemi A. Adekoya, Shanzhen Gao, Julian D. Allagan
Department of Mathematics and Computer Science, College of Natural Sciences and Mathematics, West Virginia State University, Institute, West Virginia, USA;
Department of Computer Information Systems, Reginald F Lewis College of Business, Virginia State University, Petersburg, Virginia, USA;
Department of Mathematics, Computer Science and Engineering Technology, Elizabeth City State University, Elizabeth City, North Carolina, USA

Abstract - Active learning refers to a broad range of teaching and pedagogical strategies that engage students as "partners-in-progress" participants in their learning during class time with their instructors. We discuss some real-life experiences related to students' engagement in an early introductory course to computer science. In this paper, commonly known best teaching practices such as Active Learning, Problem-Based Learning, and Integrated Learning are implemented and proven to be effective given students' survey responses.

Simulating Turing Machine in Augmented Reality

Wanwan Li
University of South Florida (USF), Florida, USA

Abstract - We present an easy-use interactive user interface for simulating Automaton and Turing machine in Augmented Reality (AR). We design a C-like programming language to describe the automaton and Turing machine, visualize their data structures, and simulate their running process in AR given to the user-specified input string or tape. Our Automaton and Turing machine simulator support three different modes, including Finite Automata (FA), Pushdown Automata (PDA), and Turing Machine (TM). We have deployed our AR Turing machine simulator on HoloLens2, users can interact with the simulator through AR buttons at the same time they can edit the scripts on the desktop where there is a USB cable connecting the desktop and HoloLens2 such that the Turing machine simulator can be updated simultaneously as users update their scripts. This mechanism can enable the user to efficiently design the Turing machine through our interactive interface in augmented reality.

LACGC: Learning Research Model for Knowledge and Contribution

Satpreet Singh
Ardass Corporation, Manteca, California, USA

Abstract - This paper introduces the new Literature, Analysis, Consumption, Gap, Contribution (LACGC) learning model to sustain the academic and non-academic business to future educational and organizational generations. The consumption of knowledge and the creation of new knowledge is a strength and focal interest of all academics and Non-academic organizations. Implementing newly created knowledge sustains the businesses for the next generation with growth without detriment. Existing models like the Scholarpractitioner model and Organization knowledge creation models focus on academic or non-academic, not both. The LACGC model can be used for both Academic and Nonacademic at the domestic and/or international levels. Researchers and scholars play a substantial role in finding literature and practice gaps in academic and non-academic disciplines. LACGC model has unrestricted the number of recurrences because the Consumption, Creation, implementation of new ideas, disciplines, systems, and knowledge is a never-ending process and must continue from one generation to the next.

Student Persistence in the Use of Test Driven Development

John W. Coffey, Steven Bitner
Department of Computer Science, The University of West Florida, Pensacola, Florida, USA

Abstract - The authors performed a prior study of a modified test-driven development approach to software development in a first course in data structures. The current work is an attempt to follow up on the original work to determine if students persisted in utilizing a test-driven development approach in a follow-on course in advanced data structures and algorithm analysis. This paper addresses literature relevant to the use of test-driven development as a pedagogical tool, the design of the current study, results, and recommendations emerging from this work.

Forensic Technologies to Automate the Acquisition of Digital Evidences

David Garcia, Llanos Tobarra, Antonio Robles-Gomez, Rafael Pastor-Vargas
Departamento de Sistemas de Comunicacion y Control, Escuela Tecnica Superior de Ingenieria Informatica, Universidad Nacional de Educacion a Distancia (UNED), Madrid, Spain

Abstract - The main goal of this work is to propose the automatic acquisition of evidences in a remote way. This automated capacity becomes interesting for companies with extensive networks and/or several locations, as it allows them to delegate and centralize the acquisition task at a single point in their structure, while saving time and travel costs. This research has been carried out through the initial implementation of a virtual laboratory made up of a network and different scenarios, by including an experimentation process. The virtual network includes both the machine from which automatic acquisitions are performed and the devices from retrieving the evidence. The group of devices will be made up of various experiments. The aim is to analyze the viability of the acquisition in different scenarios, since distributed networks are not homogeneous in the real world.

Framework for Developing Cybersecurity Activities for Children in Grades K-5

Georgia Tyner, Lila Rajabion
School of Graduate Studies, SUNY Empire State College, New York, USA

Abstract - In today's digital world, cyber-attacks are increasing daily, and many institutions and organizations understand the critical need for cybersecurity professionals. Many universities focus on developing cybersecurity programs at the undergraduate and graduate levels. Universities nowadays try to reach into K-12 education to interest and educate youth in cybersecurity. Through these efforts, universities expect that there will be an increased enrollment in cybersecurity programs at the technical college and university levels, helping to meet the demand for cybersecurity professionals. Research shows, there is 1) a shortage of cybersecurity talent in the workforce and 2) cybersecurity training is needed and should start as early as grades K-12. Most of the existing programs are for middle and high school students, with most courses being technology-related or computer science. K-5 has activity worksheets, books, and a handful of digital games. This study conducts a systematic literature review to examine what currently exists for K-5 cybersecurity education and applied research to create new activities thereby developing a framework that can be used by anyone to develop cybersecurity activities for children in K-5.

Keeping Capstone Project Current with Technology

Hong Wang, Mason Huelskamp, Jakob Stober, Joshua Taylor, Amanda Joyner, Tyler Westhoven
Engineering Technology, The University of Toledo, Toledo, Ohio, USA

Abstract - Integrating current technology to a senior capstone project helps students to accomplish educational objectives. This paper introduces a software system design project on an embedded system that uses face-recognition to implement an automatic door opener (ADO). The system is designed to help avoid transmissible disease via hands-free approach using machine learning technology. Students went far and beyond what they have learned in the curriculum and created an artificial intelligent system with real life application.

The Impact of Online Support on the Performance of Students Transitioning to University Mathematics in a 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 present study compared the achievement in first year mathematics of two demographic groups in an Australian university. Previous work by the present authors found that two changes towards more online support for tertiary mathematics students had each facilitated the transition to university mathematics study. The present study compared the effects of each change on two demographic groups of students whose entrance trajectories differed. That is, international students (Group 1) tended to have had more experience of independent study during their preparation for entrance to the university, as compared with local school leavers (Group 2). Achievement was measured by marks in the final examination for the first two core mathematics subjects taken by science and engineering students (Mathematics 1A and 1B). The first change replaced one of two face-to-face tutorials per week with an online tutorial containing tasks of increasing difficulty, each level designed to provide scaffolding for the next. In Mathematics 1A, the last year before the change, showed significantly higher achievement for Group 1, whereas in the next year there was no significant difference between groups. The second change provided students with unlimited online practice in solving test bank items with minimal right-wrong online feedback. In Mathematics 1A, there was no significant difference between the two groups in the last year before the change. In the first year after the change there was some evidence ($p < 0.1$) for higher achievement in Group 1. For mathematics 1B, there were no significant or near-significant differences. The conclusion was that the two forms of online support each helped students in their early university mathematics study, with the online tutorials more beneficial to Group 2, and the availability of formative online practice giving marginally more benefit to Group 1. The lack of any group differences in Mathematics 1B implies that it was in the transition to university study that the two changes were most important.

Teaching Induction and Deductive Reasoning

Peter M. Maurer
Department of Computer Science, Baylor University, Waco, Texas, USA

Abstract - Discrete mathematics, which includes proof by induction tends to be a difficult subject for many computer science students, especially proof by induction. I have had some success in teaching this subject and I wish to share my experience with others, in the hope that my observations on the subject will prove useful. I begin by giving a basic "what do I do next?" approach to inductive proofs. I then show that induction can be used deductively to obtain new information.

The Use of Technology to Overcome Transitional Challenges of First Year Students from Face-To-Face to Online

Munienge Mbodila, Tembo Francine Kikunga
Department of Information Technology Systems, Walter Sisulu University, South Africa

Abstract - Universities around the world and South Africa in particular have been involuntary forced to transit from face-to-face to online learning and teaching as a result of the nationwide lockdown due to the coronavirus pandemic (COVID-19). Online learning and teaching is classified as the use of digital technology to deliver tuition in synchronous mode, meaning allowing live interaction between the teacher and the learners or asynchronous

mode which is basically delays in time of delivery between teacher and learners. However, numerous challenges hinder First Year Students in transiting to the online platform to realising the full potential of e-learning, especially those from disadvantaged schools background considered as under prepared, educationally underprivileged and had little or no access and skills to technology usage prior to their enrolment at the university. This paper introduced a module called First Experience Computer Literacy (FeCOL) to facilitate students transition from face-to-face to online platform at the university. The main objective of this study was to provide students with basics training skills needed in terms of technology-related used to enable them to engage and participate effectively in the online platform. Data was collected among a group of first year students in the department of information technology systems in one rural institution in South Africa. The results show that the majority of learners have not used computers or had experience on technologies for teaching and learning in their previous schools. However, learners showed interest on basic IT training in terms of hardware functionality and software application to assist them carry out their academic tasks effectively. The study proposed that FeCOL should be used as a kick-start module to facilitate first year learners. transition from face-to-face to online.

Open-Source Educational Tool Development: An Experiential Capstone Report

Joseph Faulkner, Matthew McNulty
Southeastern Louisiana University, Louisiana, USA

Abstract - Southeastern Louisiana University's curriculum is built around preparing students with the project experience they need to enter the workforce well prepared. Concluding the Industry Connect Initiative, part of the curriculum that is focused around preparing students with real world experience, is a capstone project where students will have a real client who will give them a task to accomplish and will be expected to present working code for the task at the end of the semester. Last semester, my client gave my group the task of creating an LMS embeddable application to check basic Java syntax for introductory computer science courses. My group chose to use a technology called H5P and due to several obstacles, primarily based on poor documentation, we were unable to deliver to our client a complete product. This experience was nonetheless a very productive learning experience.

Teacher Practices for Computational Thinking Implementation in Elementary Science

Jennifer Pietros, Sara Sweetman, Minsuk Shim
Education Department, Univeristy of Rhode Island, Kingston, Rhode Island, USA

Abstract - Computational thinking (CT) is a key component of computer science and a foundational thinking process in K.12 classrooms. CT can be easily integrated into science lessons with the right policies, programs, and practices in place. This paper explores the current practices of CT in a northeast US state based on survey data collected from teachers. The research uncovers how much teaching time is spent teaching science and the percentage of lessons that have CT concepts and approaches present in the science instruction while describing what these lessons look like. This paper then discusses the next steps for implementation efforts based on the described CT practices in elementary science.

An Educational Software for Digital Terrestrial Television Broadcasting (DTTB) Systems to Engineering Students using Python

Konstantinos Zarkadas, Apollon Smyrnaios, George Dimitrakopoulos
Department of Informatics & Telematics, Harokopio University, Athens, Greece;
Real Property Department at ERT, S.A. Hellenic Broadcasting Comporation, Greece

Abstract - Digital broadcasting techniques in Digital Terrestrial Television which allow the use of single frequency networks (SFN), are developing rapidly and are useful in telecommunication networks. Consequently, the market demands qualified engineers with full knowledge on this field, particularly on the design and deployment of wireless networks and services. Also, many graduates from Electronic and Electrical engineering departments focus on wireless communications. Yet, the theoretical approach often becomes extremely complicated, and several technical terms might confuse students on this subject. This survey presents the "Array Antenna Designer" (AAD), a GUI software tool for designing and optimizing array antennas in the UHF band, expected to give the students more information and technical skills on this field and help them understand theoretical concepts easier. It could also be a useful tool for specific panel antenna array topologies of various directions and channels and help the students obtain practical knowledge of the way the broadcast communications function in real-world applications.

Knowledge: Making a Quality Choice

Antonio Sanchez
Department of Computer Science, Texas Christian University, Texas, USA

Abstract - We discuss in this paper how we teach an intro to data science course. Knowledge is a matter of choice and choosing the best course of action is through computer programming this is the way which we consider how to teach this course. We begin with an introduction to big data, data mining, machine learning and data visualization. We use python as the language to clean big data files and to select the right attributes for classification when mining the data. We discuss the Bayes method for classification, then we move to java, specifically Weka, to show how to program the algorithms, some applications are also discussed specifically using Bayes for text processing, we then move to other supervised classification algorithm such as J4.8 and Random Forest trees. And then move to clustering, where we discuss both KMeans and EM algorithms. We discuss truth and knowledge and it is where we use cross validation as the most reasonable approach to find truth based on historical data. Later we discuss supervised and unsupervised Learning, briefly talking about Neural Networks collective and reinforced learning. Briefly talk about current deep learning using CNNs. We finish the course talking about Genetic Algorithms and soft computing.

Differences Between Neurodivergent and Neurotypical Learning during Covid-19: Towards the E-tivities Satisfaction Scale

Jonathan Bishop, Kamal Bechkoum
Crocels Research CIC, United Kingdom (UK);
University of Gloucestershire, United Kingdom (UK)

Abstract - This paper presents the results of a study conducted during the Covid-19 pandemic into the differences between neurotypical and neurodivergent learners when it comes to participating in e-tivites. The study, conducted as part of the DWP Kickstart programme, found that neurodivergent participants are much more suited to online learning environments involving e-tivities than neurotypical ones.

A Framework for Promoting Students Experience the Examination of Articles that Alert Dangers of Disaster on Social Media

Tomoki Ishii, Ryo Onuma, Hiroki Nakayama, Hiroaki Kaminaga, Youzou Miyadera, Syoichi Nakamura
Fukushima University, Japan; Tsuda University, Japan; Yamagata University, Japan;
Tokyo Gakugei University, Japan

Abstract - When natural disasters such as heavy rains and landslides are coming, there are an increasing number of cases where articles that appeal to imminent dangers and the need to evacuate (called disaster caution) posted by individuals on social media help people reduce damage. Although such articles should be put to the best use, it is a difficult for inexperienced people to discover valuable articles from a huge number of posts on social media. In this paper, we mainly describe a framework for promoting the students experience examine the disaster caution and methods for extracting the candidates of noteworthy articles. We also describe an experiment using actual posts on Twitter and discuss the characteristics of our methods based on its results.

A Set of Machine-Learning Exercises

Mehdi Zargham, Jie Cheng
Department of Computer Science, University of Dayton, Ohio, USA;
Department of Computer and Information Technology, Miami University, Ohio, USA

Abstract - This paper represents several assignments for teaching a machine/deep-learning course. Students starts with simple assignments to implement neural networks in Python and progress toward more advanced assignments to perform image classification using Google TensorFlow. Students will use "Jupyter Notebook" platform to code the lab assignments and projects. Each assignment represents the main benefits of using neural networks. Visualization of the performance and final results are used to verify the correctness of the results.

Review of Video Games and Simulation in Computer Science Education

Joel Weymouth, Rose Atuah
Morgan State University, Baltimore, Maryland, USA;
Computer and Information Systems, Towson University, Towson, Maryland, USA

Abstract - There have been different attitudes about video games in education. There are many proponents of using video games as a learning supplement. Many studies show that test scores improved using video games. Since games could make learning enjoyable, the arguments for using video games appeared to be self-evident. On the other hand, some studies suggest that games are addictive and cause anti-social behavior. There is research that suggests games create apathy and introversion that is anti-social. However, the later analysis shows that video games are more beneficial than detrimental. This proposal will present evidence that using video games is helpful in learning. The perceived problems with games do not fit the pathology criteria, and using video games in the greater context of life balance will prevent adverse pathology. Some dedicate their lives to studying at the expense of all, which does not make scholarships detrimental. This manuscript first presents studies that show how video games improve learning. The neurological implications make video games a positive force in education and address the concerns through studies that have answered these concerns. Thirdly, examples of educators using video games as a pedagogy. Finally, further research needs to use video games as a comprehensive curriculum and its implications. The implications are to evolve video games from supplemental to integrate them into the curriculum utilizing AI and SOA.

Students Behaviors and Factors Affecting Online Learning during the COVID-19 Situation Based on Machine Learning

Jaruwan Chutrtong, Narumon Boonman, Supatchalee Sirichokworrakit, Kunyanuth Kularbphettong
Suan Sunandha Rajabhat University, Bangkok, Thailand

Abstract - The objectives of this research paper are study online learning behaviors, the success factors of online learning and perspectives on online learning during the Covid-19 situation. It is quasi experimental research. The population is 127 students of the Faculty of Science and Technology, Suan Sunandha Rajabhat university. The sample group enrolled in microbiology courses, semesters 1 and 2 of 2020 academic year. Research tool is a questionnaire. This paper has two main objectives: to analyze the behavior of students and to predict learning behaviors through the online system by using SVM, and KNN. Overall academic success on study is also at moderate levels. The experiment shows that the support vector machine (SVM) model has the best and most stable effect. The average recall rate, precision rate, and accuracy rate reached 80.81%, 84.75%, and 81.83%, respectively.

Information Technology Education: A Multi-Fold Learning Approach

Moh'd A. Radaideh
Department of Software Engineering, Faculty of Computer and Information Technology,
Jordan University of Science and Technology, Irbid, Jordan

Abstract - This paper elaborates on the author's multi-fold learning approach for delivering an introductory course on Information Technology at the Faculty of Computer and Information Technology (FCIT) of Jordan University of Science and Technology (JUST). By the first semester of 2021-22, the course was created in eight modules covering the various FCIT.s specializations, and it was offered at that time as a fully-online course. By the second semester of 2021-22, the course was revised and offered as a blended-online course. By the summer semester 2021-22, the course was revised once again and offered as a blended in-class course. As for the current first semester of 2022-23, the course is being offered as a hybrid course that consists of five folds: (i) compulsory synchronous in-class lectures; (ii) compulsory asynchronous off-class selfreading activities; (iii) compulsory synchronous and/or asynchronous online/off-line quizzes, and exams; (iv) optional asynchronous off-class self-watching activities; and (v) optional synchronous online discussion meetings. The said course was delivered differently on every semester of the past three ones since September 2021. The causes for altering the course.s delivery mechanism during the current and the past semesters include, the following: (i) the ongoing changes in the requirements by the ministry of higher education and scientific research in respect to the online education, particularly the allowed percentage of online courses in each undergraduate degree program; (ii) the FCIT.s requirement of adding new modules to serve its new specializations; (iii) the necessity of maintaining the course.s compliance with the ABET and IET accreditation criteria; and (iv) the ongoing feedback and complaints received from the students and faculty members during the past semesters. Although the participating students. satisfaction figures continued to rise during the past three semesters, and the numbers of participating students have been tremendously increasing, there will always be rooms for improvement.

Implementation of Microsoft's PowerBI into Education

David Zenisek, Michal Simon
Department of Industrial Engineering, University of West Bohemia, Pilsen, Czech Republic

Abstract - The aim of this article is to present a case study of the implementation of Microsoft.s PowerBI into teaching. The article introduces the study course, the aims of BI tool implementation, the process of BI tool selection, its implementation and evaluation after 3 years. Finally, thoughts on further improvement are presented.

An Educational Perspective on Online Learning Platforms

Cameron Schwartz, Jason Wahby, Geko Ezekiel Jimenez, Kemal Koymen, Hunter Doll, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA;
Department of Computer Science, Bemidji State University, Bemidji, Minnesota, USA

Abstract - The World Health Organization has identified the COVID-19 pandemic as a current hazard to humanity. As a result of the pandemic's successful global shutdown of several activities, including educational activities, colleges have migrated dramatically as a crisis response. Due to the cancellation of in-person classes, online learning has become more popular and has enabled students to continue their education. However, the abrupt shift from in-person to online learning has presented many difficulties for students, instructors, administrators and educational leaders. Zoom, Google Meet, Google Classroom and LMS platforms are the most frequently used media. Perceptions of the learning media have an impact on how students use them. This study intended to describe how students perceive the convenience and usefulness of the various learning tools accessible. Overall, the research shows that the advantages of taking classes online exceed the difficulties that students may encounter. After showing the numerous advantages and drawbacks of using virtual learning in higher education, it is clear that there is still much opportunity for the future.

Industry Connect Initiative: A Preliminary Report on Internships and the Addition of Technical Career Manager to University Career Services

Bonnie Achee, Ghassan Alkadi, Sandy Summers, Matthew McNulty, John Burris
Southeastern Louisiana University, Hammond, Louisiana, USA

Abstract - The Industry Connect Initiative (ICI) of the Department of Computer Science at Southeastern Louisiana University aims to prepare undergraduate students for successful careers by utilizing a four-pronged approach which includes an Industry Advisory Board, Real-World-Ready Curriculum, the Distinguished Lecturer Series and Internships. Additional assistance is provided through the Workforce Talent Initiative (WTI) position of Technology Career Manager (TCM), who plays a crucial role in aiding students in securing internships and career placement opportunities. This preliminary report gives an overview of the WTI and discusses the relationship of Career Services and the new Technical Career Manager (TCM) to the fourth-prong, Internships, and the ICI as a whole. Preliminary results from informal observations and secondary data from LinkedIn are presented. Future work with detailed results will commence upon IRB approval.

Industry Connect Initiative: A Real World Ready Case Study in Information Systems

Caleb Martin, Jared Martin, Bonnie Achee, Ghassan Alkadi
Southeastern Louisiana University, Hammond, Louisiana, USA

Abstract - The Industry Connect Initiative (ICI) at Southeastern is a four-pronged approach that involves industry in a meaningful way in the curriculum. One of these courses is Information Systems. Information Systems give students an opportunity to learn and experience how to develop software in a realistic environment. Students work closely with three senior developers from Envoc, a local software development company, to develop a web-based software application. [1,2] The project requirements were to create a software development distribution service, similar to Valve.s Steam. This service allows customers to buy and download software and allows publishers to upload and sell software. This paper details the project development.

Networking and Computing: Creating Opportunities for Low Income Students to Succeed in Computing Careers

Sherri Turner, Anne Hinderliter, Arshia Khan
Department of Educational Psychology, University of Minnesota, Minneapolis, Minnesota, USA;
Chemistry and Biochemistry Department, University of Minnesota Duluth, Duluth, Minnesota, USA;
Department of Computer Science, University of Minnesota Duluth, Duluth, Minnesota, USA

Abstract - With the national need for computer science, computer engineering, and robotics professionals soaring, computer science depts. are struggling to attract, and graduate students from low-income groups. In particular, women, Native Americans, and other ethnic minorities who are highly represented among low-income students, are not entering computer science. Thus, they are not available to provide the variety of perspectives needed for effective problem solving in computer science domains. To respond to this need, we designed a Track 1 project, Networking and Computing (NAC), with the overall objective of enabling low-income academically talented Univ. of Minnesota Duluth computer science students to attain a 4-year baccalaureate degree and enter into a computer science job or graduate program within 1 year of graduation. Mechanisms to attain this goal are to increase students' sense of belonging, professional identity, and networking capabilities. Our Longer-Term Outcomes are to establish a model of student support that is effective and transferrable to other STEM programs, and substantially increase the number of low-income students who enter the computer science field.

Gateway Scholars Program - Reducing Barriers to STEM for Undergraduate Students through Scholarship and Supportive Programs

Julia Thom Oxford / Diane Smith
Department of Biological Sciences, Biomolecular Research Center, Center of Excellence in Biomedical Research, Boise State University, Boise, Idaho, USA

Abstract - This report presents the Gateway Scholars Program, an NSF-S-STEM supported program that recruited academically talented undergraduate students with demonstrated financial need. The objectives of our program included establishing a mentored cohort program, implementing enhanced risk-based advising, integrating evidence-based instructional practices in the curriculum, engaging students in co-curricular experiences, and generating new knowledge about the effect of activities on retention, student success, and degree attainment. Knowledge about broadening participation and effectiveness of evidence-based practices in STEM curricular and co-curricular activities and systems developed through this program have the potential to impact all STEM departments.

Redesigning an Upper-Division Java Elective into a Job Interview Preparation Class

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

Abstract - This paper describes the revision of a Java elective, helping students prepare for job interview and review topics that they have learned in other courses and however still are useful in their future jobs. Initially, this elective focuses on various features of Java. Based on the feedback from the students, it was redesigned, which, other than Java features, also includes a review of the core topics initially introduced in other courses, the application of Java, and the exercises for jab interview. The new design will be implemented next summer. Then, its effectiveness will be evaluated, finding out how to modify the design and better assisting students in their future career development.

Building Framework and Infrastructure to Assist a Minority Serving Institution (MSI) in Cybersecurity Program Development

Swapnoneel Roy, O. Patrick Kreidl, Karthikeyan Umapathy, Zornitza G. Prodanoff, Iman Vakilinia
School of Computing, University of North Florida, Jacksonville, Florida, USA;
School of Engineering, University of North Florida, Jacksonville, Florida, USA

Abstract - The University of North Florida (UNF) offered to leverage resources generated as a result of its CAE-EDU designation to assist Edward Waters University (EWU), a local Minority-Service Institution (MSI), with establishing cybersecurity educational opportunities for their STEM faculty and undergraduate students. This project, funded by NSA as part of their Cybersecurity Education Diversity Initiative (CEDI) program, designs the framework and cultivates the infrastructure by which EWU gains access to UNF instructional resources (e.g., courses, facilities), faculty expertise (i.e., curriculum, consulting) and student activities (e.g., clubs, advising) in cyber defense education. Project plans included (i) outreach activities with support from faculty of the Florida State College of Jacksonville (FSCJ), a local community college, also CAE-designated, as well as (ii) proof-of-concept assessments with support from faculty and students of Edward Waters University (EWU), an MSI with no existing cybersecurity programs. Some of the activities performed under this project and their outcomes are reported in this paper.

Data Science to Enhance Research Capacity of a Research-Intensive Medical College

Qingguo Wang, Aize Cao, Vibhuti Gupta, Ashutosh Singhal, Todd Gary
School of Applied Computational Sciences, Meharry Medical College, Nashville, Tennessee, USA

Abstract - This paper describes our efforts to enhance the research capacity of the Research Centers in Minority Institutions (RCMI) program at Meharry Medical College and foster collaborations between RCMI and a newly established School of Applied Computational Sciences (SACS). These efforts include offering subawards to collaborative research groups, financial support to graduate students for their participation in collaborative research, performing need assessment to understand the needs of the community so as to guide our project implementation, and providing data science training to enhance the data analytics skills of the RCMI investigators, staff, medical residents, and graduate students. This paper presents the progress of this project, which clearly indicates its positive impact on the local community.

Proposal of a System Enabling Adaptive Support Based on Transition of Each Learner's Source Codes

Soichiro Sato, Yoshiki Sato, Shoichi Nakamura, Youzou Miyadera
Tokyo Gakugei University, Japan;
Fukushima University, Japan

Abstract - The difficulties that students face in programming education can be broadly classified into two types: difficulty in conceiving a solution and difficulty during the actual coding process. In this study, we developed a machine learning model that enables adaptive instruction by estimating a solution that will be easy for the learner to understand according to the learner's past work on previous tasks. Our evaluation showed that the developed approach could estimate easily understandable code instances for learners with high accuracy (97.5% AUC).