CSCI 2021 BOOK of ABSTRACTS

The 2021 World Congress in Computer Science, Computer Engineering, and Applied Computing
CSCE 2021

https://www.american-cse.org/csce2021/
July 26-29, 2021
Luxor Hotel (MGM Property), 3900 Las Vegas Blvd. South, Las Vegas, 89109, USA

Table of Contents

Note that the titles of papers, the authors' names and the abstracts that appear in this book ("Book of Abstracts") were extracted from the papers that were submitted to the EVALUATION web portal (i.e., extracted from the first draft submissions). The official published proceedings/book will have any and all changes/revisions that authors may have done.

KEYNOTE ADDRESSES:
(The Keynote lectures are open to all participants)

CONGRESS WELCOME REMARKS:

Dr. Hamid R. Arabnia
Chair, Steering Committee & Coordinator
University of Georgia, USA; Editor-in-Chief, The Journal of Supercomputing

CSCE Congress Keynote:

THE COMING PROBLEM OF SELF-AWARE ROBOTS
Dr. James Crowder
Engineering Fellow at Colorado Engineering, Inc., USA
(Formerly at Raytheon Technologies, USA)

SAM & CSCE Keynote:

CYBER SECURITY- PRIVACY - HOW CAN I PROTECT MYSELF IN CYBER SPACE? CAN I GET SOME PRIVACY, PLEASE?
Prof. Levent Ertaul
Department of Computer Science, California State University, East Bay, California, USA

SAM & CSCE Keynote:

ZERO-TRUST: SECURING THE NEW PERIMETER
Dr. Nader Nassar
Program Director & Senior Technical Staff member, IBM CIO, Assured Identity & Cyber Ops Organization;
Master Inventor at IBM; IBM, USA

ICDATA & CSCE Keynote:

BADGES & NEURAL NETWORKS: IMPROVING LAW ENFORCEMENT WITH AI
Andrew Johnston
Consultant with Mitiga, CEO and Co-Founder of Recluse Laboratories;
Adjunct Professor at Fordham University, USA

CSCE (ON-LINE) Keynote:

ENHANCING SECURITY OF COMMUNICATION NETWORKS WITH RANDOM NEURAL NETWORKS
Prof. Erol Gelenbe
FIEEE FACM FIET FRSS FIFIP; Professor,
Institute of Theoretical and Applied Informatics (IITIS-PAN),
Polish Academy of Sciences;
Coordinator (PI) of the 5M Euro H2020 SerIoT Research and Innovation
project of the European Union;
Inventor of the random neural network and the eponymous G-networks.
Fellow of the National Academy of Technologies of France and of
the Science Academies of Belgium, Hungary, Poland and Turkey

ICDATA (ON-LINE) Keynote:

DATA SCIENCE: THE FRONTIER OF DIGITAL ECONOMY
Dr. Peter Geczy
National Institute of Advanced Industrial Science and Technology (AIST), Japan

The 5th International Conference on Applied Cognitive Computing (ACC 2021)
https://www.american-cse.org/csce2021/conferences-ACC

Transformers: Malware in Disguise

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

Abstract: We introduce a new application of a Transformer architecture, typically used in Natural Language Processing (NLP), for the purpose of classifying malicious stack traces. Our adoption makes use of conventional tokenization schemes used in NLP, as well as propose an experimental tokenizer which is specifically tailored to parse function names from the Windows API library. Our approach aims to provide semantic analysis of the sequence of API calls - largely based on the innate ability of humans to cognitively extract the logical aspects of meaning from text - from the Windows OS in user mode and kernel mode. We train the Transformer architecture from scratch, allowing for the model to be fine-tuned as needed for future use-cases. Our application is one of the first of its kind, and represents a unique application of transfer learning between very different domains. Our results demonstrate quick training and inference time due to the minimal number of trainable parameters required (order of 103) and the small size of the corpus of function calls (approx. 1,100-1,500) relative to the vocabulary in NLP applications. Our model scored 95+ F1 on the classification of 9 Malware enterprise threats from benign activity, obtained from Registry and File System stack trace event activity - with select models achieving F1 scores as high as 96.5.

Design and Implementation of Water Quality Management Middleware based on Big Data and Neural Networks

Sung-Hoon Park, Young-Chul Seo, Su-Chang Yoo
Department of Computer Engineering, Chungbuk National University, Chungju, Chungbuk, Korea

Abstract: As there is a growing concern that the scale of damage caused by water pollution accidents due to urbanization and industrialization is increasing, the demand for water safety in society is increasing. In the last five years, many water pollution accidents have occurred in the four major river basins, and this has caused direct and indirect damage to public health, such as stopping water supply and drinking contaminated tap water. Therefore, there has been a constant demand for a water quality environment management system that can minimize the uncertainty of the water quality environment. In the Ubiquitous Sensor Network environment, the water quality management system transmits data measured in real time from the terminal node to the server, and the middleware of the received system has secured and stored data integrity and redundancy. However, the biggest problem in this processing is that the generated source data contains many errors and cannot be used as it is. Therefore, the assistance of experts is required to ensure the integrity of the water quality data, and there is a difficulty in paying money economically. As a solution to this problem, building a multilayer neural network using a machine learning model is the best solution. In this study, we design a real-time water quality data verification middleware based on multi-layer neural networks and propose to develop it as a monitoring system.

Solving Logistics Distribution Center Location Problem by New Spider Monkey Algorithm

Guangjun Liu, Xiaoping Xu, Feng Wang, Long Liu
School of Sciences, Xi'an University of Technology, Xi'an, P. R. China;
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, P. R. China;
School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, P. R. China

Abstract: In order to solve the location problem of logistics distribution center, this paper proposes a new spider monkey algorithm. On the basis of the basic spider monkey algorithm, three parts of the algorithm are improved: phase factor is used in the local leader stage to realize the iterative updating from coarse to fine; adaptive learning factor is adopted in the global leader stage to balance the global exploration and local mining capacity; the mutation operation of differential evolution algorithm is applied in the local leader decision-making stage to find the potential solution in the local region of the current solution and improve the search efficiency and solution accuracy of the algorithm. Finally, the location problems of logistics distribution center are calculated using the improved spider monkey algorithm, and compared with other algorithms, the results show that the proposed algorithm in solving the location problems of logistics distribution center is effective.

Problem Recognition, Explanation, and Goal Formulation

Sravya Kondrakunta, Venkatsampath Raja Gogineni, Matthew Molineaux, Michael T. Cox
Wright State University, Dayton, Ohio, USA;
Parallax Advanced Research, USA

Abstract: Goal reasoning agents can solve novel problems by detecting an anomaly between expectations and observations, generating explanations about plausible causes for the anomaly, and formulating goals to remove the cause. Yet, not all anomalies represent problems. This paper addresses discerning the difference between benign anomalies and those that represent an actual problem for an agent. Furthermore, we present a new definition of the term "problem" in a goal reasoning context. This paper discusses the role of explanations and goal formulation in response to the developing problems and implements it; the paper also illustrates the above in a mine clearance domain and a labor relations domain. We also show the empirical difference between a standard planning agent, an agent that detects anomalies, and an agent that recognizes problems.

Autonomous Goal Selection Operations for Agent-Based Architectures

Sravya Kondrakunta, Michael T. Cox
Wright State University, Dayton, Ohio, USA

Abstract: An intelligent agent has many tasks and goals to achieve over specific time intervals. The goals may be assigned to it or the agent may generate its own goals. In either case, the number of goals at any given time may exceed its capacity to act upon concurrently. Therefore, an agent must prioritize the goals in chronological order as per their relative importance or significance. We show how an intelligent agent can estimate the trade-off between performance gains and resource costs to make smart choices concerning the goals it intends to achieve as opposed to selecting them in an arbitrary basis. We illustrate this method within the context of an intelligent cognitive architecture that supports various agent models.

Artificial Intelligence-Assisted Situational Awareness: An Air Force Perspective

Arslan Munir, Alexander Aved, Erik Blasch
Kansas State University, Manhattan, Kansas, USA;
Air Force Research Laboratory (AFRL), Rome, New York, USA;
Air Force Office of Scientific Research (AFOSR), Arlington, Virginia, USA

Abstract: Situational awareness (SA) is defined as the perception of entities in the environment, comprehension of their meaning, and projection of their status in near future. From an Air Force perspective, SA refers to the capability to comprehend and project the current and future disposition of red and blue aircraft, and surface threats within an airspace. In this article, we propose an artificial intelligence-assisted model for SA and dynamic decision-making that incorporates artificial intelligence (AI) and dynamic data-driven application systems to adapt measurements and resources in accordance with changing situations. We then discuss a few AI-related techniques and technologies that help improve SA ranging from cognitive Internet of things and unmanned aerial vehicles to automated vision systems. Finally, we conclude this article with insights into the future of AI-assisted SA.

A Dynamic Active Noise Control System for Live Music Attenuation

Elliot Krueger, Lichuan Liu
Department of Electrical Engineering, Northern Illinois University, DeKalb, Illinois, USA

Abstract: This paper proposes a system design that will be suitable for applying active noise control (ANC) effectively to live musical instruments. The design consists of three parts: a signal separation section, an instrument classification section, and the active noise control section. The two instruments of focus will be the trombone and tuba for their low frequency and ability to be quite loud in a typical music ensemble. The proposed system can dramatically reduce the high-level volume music compared with traditional passive noise control solution. It performed slightly better than a general active noise control scheme that does not include a classifier.

The "Room Theory": A Computational Model to Account Subjectivity into Natural Language Processing

Carlo Lipizzi, Dario Borrelli, Fernanda de Oliveira Capela
School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, New Jersey, USA

Abstract: This work introduces a novel method to consider subjectivity and context dependency in text analysis. The method uses a computational version of the Framework Theory by Minsky leveraging on text vectorization to create a numerical representation of the knowledge of the context. its components are: a framework/“room” representing the point of view of the individual or the collective; a benchmark/set of keywords representing the criteria for the analysis; and the document to be analyzed. By measuring the similarity between the vectors representing words/semantic elements, we extract the relevance of the elements in the benchmark for the document to be analyzed. This method could be applied where subjectivity is relevant to understand the relative value or meaning of a text.

The 22nd International Conference on Bioinformatics & Computational Biology (BIOCOMP 2021) https://www.american-cse.org/csce2021/conferences-BIOCOMP

Accurate and Robust Trypan Blue-Based Cell Viability Measurement Using Neural Networks

Adele Peskin, Joe Chalfoun, Steven P. Lund, Chenyi Ling, Laura Pierce, Sumona Sarkar, Firdavs Kurbanov, Michael Halter, John Elliott
National Institute of Standards and Technology, NIST, Maryland, USA;
National Institute of Standards and Technology, NIST, Colorado, USA

Abstract: Trypan blue dye exclusion-based cell viability measurements depend on image quality and consistency. To enable reproducible measurements, reliable in-focus image capture with respect to cell features is required. Current software uses human-selected features to classify cells. Instead, we use neural networks to determine whether cells in a bright field image are dead or alive, placing no limit on the range of features used and removing the need to locate the sharpest image in each new sample. Our viability measurements can be made over a wide range of focal planes (up to 150 µm), and viability levels (0 to 100 % viability in test sets), while keeping the viability estimates within the range of manual identification of cells by several experts.

Multimodal Analysis of Plant Root Composition as a Function of Drought

Daniela Ushizima, Esther Singer, Elizabeth Carpenter, Dilworth Parkinson
Lawrence Berkeley National Laboratory, Berkeley, California, USA;
University of California Berkeley, Berkeley, California, USA

Abstract: Due to climate change droughts are increasingly becoming a threat to biofuel crop production. Water limitation can modify plants both phenotypically and chemically. However, we have limited knowledge on the impact of drought on root architecture and chemical composition, which impact crop and soil health. We cultivated Panicum hallii, a perennial grass, in a tightly controlled climate chamber, called EcoPOD, that allows monitoring and manipulation of interactions among plants, soil, microbes, and atmosphere simultaneously. This work focuses on using multimodal experimental data from confocal microscopy and microtomography to perform multiscale multimodal root analysis, including measuring fluorescent intensities for plants in different humidity conditions. These measurements rely upon new efforts on computational methods for assessing roots properties and potential functionalities. Preliminary investigations show high correlation between cellulose and lignin concentration in response to drought conditions.

A Machine Learning Approach for Predicting Furin Cleavage Sites of Viral Envelope Proteins

Christine Gu, Wendy Lee
San Jose State University, San Jose, California, USA

Abstract: Furin is a ubiquitous proprotein convertase that cleaves and activates a wide range of proprotein substrates, including growth factors, receptors, and pathogenic agents. Glycoproteins from various enveloped viruses, such as HIV, Ebola virus, influenza virus, and novel SARS-CoV-2, exploit furin for viral entry and infectivity. Accurate identification and prediction of furin cleavage sites is crucial to understanding furin cleavage-mediated viral infections and may support future synthesis of furin inhibitors. We developed a novel ensemble machine learning approach using

the physicochemical properties of amino acids to predict furin cleavage sites in enveloped viruses, including hydrophobicity, volume and chain flexibility. Our method currently achieves an accuracy of 86% in a 10-fold cross-validation analysis and 83% in an independent test set.

Machine Learning Prediction for Mortality of Non-Diabetic Patients with Hyperglycemia Hospitalized in the Intensive Care Unit (ICU) with Covid-19

Lida Y. Ramirez Reatiga, Wilson Rojas Reales, Ernesto Barcelo
Biomedical Informatics, Universidad El Bosque, Bogota D.C., Colombia

Abstract: This work will examine the relationship between hyperglycemia as an independent mortality risk factor in non-diabetic patients admitted to the intensive care unit (ICU) due to Covid19. We will conduct a multicenter retrospective study using data obtained from the health records of Covid-19 patients admitted to the adult ICU of two medical centers in Barranquilla (Colombia) between 01 July and 31 December 2020. First, we will perform statistical analysis of the demographic and metabolic variables and conduct the mortality risk assessment. Second, a prototype will be developed to predict mortality in the group studied, using a machine learning algorithm and thus seeks to contribute to other prediction studies, in this case considering the blood glucose level.

Predicting the Risk of Colon Cancer with Machine Learning

Chao Zhou, Hong Wang
Department of Mathematics and Statistics, the University of Toledo, Toledo, Ohio, USA;
Department of Engineering Technology, The University of Toledo, Toledo, Ohio, USA

Abstract: Colon cancer is often fatal; however, early diagnosis greatly reduces this risk. Many diagnosis and treatment methods are available. In this study, machine-learning methods were used to predict the patients' outcome. Correlation coefficients of two variables are present in terms of multicollinearity. After dividing the data set into train and test sets, seven base models were used to determine a baseline. Lightgbm appeared to be the best basic model. This report also used grid search for super parameter tuning. After tuning, with L1 normalization setting as 0.15 and constant of max depth at 5, the AUC score eventually reached 0.94. This result is statistically significant, but limitations exist. The sample size and universality of samples may require more cross validation and data from different hospitals.

Survival Prediction Based on Deep Learning Extracted Histopathology Imaging Features

Mohamedelfatih Eltigani Osman Abaker, Saloni Agarwal, Ovidiu Daescu
The University of Texas at Dallas, Richardson, Texas, USA

Abstract: Survival prediction has many implications in the treatment and monitoring of patients and the use of histopathological images in it is a growing area of research. Current machine learning approaches rely on expertannotated regions of the images or methods like phenotype clustering based on heuristics or randomly sampling patches from the images. We propose a method that uses the entire tissue region of the histopathology images, we experiment with three deep learning models to extract useful prognostic features from them, we then combine these features with clinical features to get a multimodal survival prediction model. We evaluate the proposed models on 10 public datasets and the best performing model achieved 2.6% c-index improvement over the baseline model.

Bioinformatics Metadata Extraction for Machine Learning Analysis

Zachary Tom, Wendy Lee
San Jose State University, San Jose, California, USA

Abstract: Sequencing errors in conventional NGS workflows are key confounding factors for detecting mutations. Various steps of the workflow are tracked in the sequencing metadata, such as sample handling and library preparation, and these steps can introduce artifacts that affect the accuracy of calling rare mutations. This metadata is often scarce and challenging to filter and extract. This paper presents SRAMetadataX, a new tool that enables researchers to easily extract crucial metadata from NCBI SRA submissions. The tool was used to find sequencing runs that utilized various enrichment processes. Potential sequencing artifacts were identified, and machine learning models were trained on the data to determine a relationship between metadata and artifacts.

The Information and Complexity Analysis of COVID-19 RNA Sequence

Shijun Tang
Department of Math and Science, Alvernia University, Reading, Pennsylvania, USA

Abstract: In this paper, we analyse one COVID-19 RNA sequence via utilizing the LZ complexity method and Shannon information entropy. LZ complexity of a DNA/RNA sequence can be used in extraction of the useful information from complicated DNA/RNA sequence and in modern structural analysis of complete genomes. LZ complexity of a DNA/RNA sequence can also be used to find new motifs/patterns and mutations which have important biological functions. The special structure and patterns in the studied RNA would have a great significance for the clinic diagnosis and treatment of COVID-19 and some similar diseases (e.g. SARS).

The Role of Background Colour in Pollen Recognition Task using CNN

Fernando C. Monteiro, Cristina M. Pinto, Jose Rufino
Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politecnico de Braganca, Campus de Santa Apolonia, Braganca, Portugal

Abstract: Pollen recognition is a crucial but challenging task in addressing a variety of questions like pollination or palaeobotany, but also for other fields of research, e.g., allergology, melissopalynology or forensics. State-of-the-art methods mainly use deep learning CNNs for pollen recognition, however, we observe that existing published approaches use original images without study the possible biased recognition due to pollen's background colour. In this paper, we evaluate the DenseNet model trained with original images and with segmented images (remove background) and analyse network's predictive performance under these conditions using a cross evaluation approach. An accuracy of 97.4% was achieved that represents one of the best successes rate when weighted with the number of taxa of any attempt at automated pollen analysis currently documented in the literature. From these results, we confirm the existence of background specific influence in the recognition task.

Towards Genomics Data-Based Disease Prediction with Graph Neural Networks

Laura Hernandez-Lorenzo, Vanesa Pytel, Jordi A. Matias-Guiu, Jorge Matias-Guiu, Jose L. Ayala
Department of Computer Architecture and Automation, Complutense University of Madrid, Madrid, Spain; Department of Neurology, Hospital Clinico San Carlos, San Carlos Institute for Health Research (IdiSSC), Complutense University of Madrid. Madrid, Spain

Abstract: The study of genomics data, and specially genomic variants, is relevant in neurodegenerative diseases to contribute to an early diagnosis, being an affordable source of biomolecular data and a relatively non-invasive technique. This work presents a comprehensive methodology for the integration of genetic variants into different types

of biological networks for disease prediction by means of Graph Neural Networks. The methodology was applied to the ADNI cohort for classification of its three diagnosis groups. The obtained results, although preliminary, highlight the complexity of the sparse nature of genetic data, as well as the potential of integrating genomic data in a graph-structured way.

Mechanism and Learning of Motion Direction Detection Using Dendritic Neuron Model based Direction Selective Ganglion Cells

Yuki Kobayashi, Yuki Todo, Zheng Tang
Graduate School of Natural Science & Technology, Kanazawa University, Japan;
Faculty of Electrical Information and Computer Engineering, Kanazawa University, Japan;
Department of Intelligence Information Systems, Toyama University, Toyama, Japan

Abstract: In this paper, we present a novel motion detection mechanism and train dendritic neuron model based ganglion cells to learn the motion direction detection. First, we propose a novel motion detective mechanism which uses local motion direction detective neurons to infer global motion direction. Secondly, we assume that these motion direction detective neurons have arbitrary morphology of the dendrites and there be an interaction among all synapses on the dendrites. Additionally, these neurons acquire motion direction detection function by learning to eliminate unnecessary branches or synapses, strengthen useful ones and finally form correct branches and synapses. We show by simulations that learned neurons are capable of detecting the motion direction of objects with different shapes, sizes and positions.

Tridimensional Structure Prediction and Purification of Human Protein GPN2 to High Concentrations by Nickel Affinity Chromatography in Presence of Amino Acids for Improving Impurities Elimination

Jorge Juurez Lucero, Maria del Rayo Guevara-Villa, Anabel S Sanchez Sanchez,
Raquel Diaz Hernandez, Leopoldo Altamirano Robles
Instituto Nacional de Astrofisica, Optica y Electronica, Mexico;
Universidad Politecnica de Puebla, Mexico

Abstract: The possible tridimensional structure of hGPN2 was modeled using both Protein Structure Prediction Server and Python Molecular Software. Also, recombinant protein of hGPN2 was expressed from E. coli and purified with different concentrations of Imidazole or lysine-glutamic acid or arginine-glutamic acid buffers. The use of lysine-glutamic acid in equimolar concentration (100 mM) was enough to get the hGPN2 pure, homogeneous and monodisperse, which was verified using dynamic light scattering. The final concentration of hGPN2 was 30 mg/mL of hGPN2, conditions required to get a tridimensional model using X-ray crystallography.

Signal Processing and Machine Learning Pipeline for EEG Manipulation in the Diagnosis of Dementia

Carlos Moral Rubio, Vanesa Pytel, Jordi A. Matias-Guiu, Jorge Matias-Guiu, Jose L. Ayala
Department of Computer Architecture and Automation, Complutense University of Madrid, Madrid, Spain;
Department of Neurology. Hospital Clinico, San Carlos, San Carlos Institute for Health Research (IdiSSC),
Complutense University of Madrid. Madrid, Spain

Abstract: There is strong clinical, imaging and pathological evidence that neurodegeneration is associated with altered brain connectivity. Electrophysiological (EEG) recordings provide direct measures of neural activity with excellent temporal resolution, opening an exciting diagnosis approach based on these recordings. In this work, we propose a full preprocessing pipeline and a set of transformations and methods, based in both signal transformations and machine learning processes, that allows physicians to gain a deeper understanding of the signal and a more simple interpretation of the disease, achieving a procedure to perform early diagnosis from EEG data.

Representation of Boolean Genetic Regulatory Networks by Hypothesis Logic

Pierre Siegel, Andrei Doncescu, Vincent Risch, Sylvain Sene

Aix-Marseille Universite, France;
Universite des Antilles, France; Universite Publique, France

Abstract: Boolean networks (BNs) are particularly used in the context of Boolean Genetic Regulation Networks modeling. This article studies the representation of BNs using a modal Hypothesis Logic (H). A dynamic of a BN organizes the entities updates over time. The dynamic is fundamental for the representation of BNs. An important part of the studies done on BNs have focused on the analysis of their attractors, particularly the stable configurations and stable/unstable cycles/oscillations. These attractors allow in particular, to model the cellular specialization. We introduce representations for BNs in H which aim at making possible to discriminate between stable configurations, limit cycles and unstable cycles. Ghost extensions, defined in H, play here a key role.

Detection and Classification of Tandem Repeats in Protein Structures: A Machine Learning Approach

Luis Muroya, Layla Hirsh
Department of Sciences, Pontificia Universidad Catolica del Peru, Peru;
Department of Engineering, Pontificia Universidad Catolica del Peru, Peru

Abstract: Tandem Repeat (TR) regions can be found in many proteins that are involved in a number of diseases. Therefore their classification is keystone in finding ways to treat them. There are two approaches to this end: to use either the sequence or 3D structure information. Here, we present a tool that exploits the latter to classify closed-loop repeats (sub-classes within class IV). It has two modules: filtering and classification. The former was built upon a CNN already trained to detect rotation symmetries and had an error of 22% over class IV chains reported in RepeatsDB. The latter transformed the structure information into an image and fed it into a DenseNet CNN for classification with 95% accuracy.

First Theory of Cancer Gene Data Analysis by 169 Microarrays and Four Universal Data Structures for Big Data

Shuichi Shinmura
Professor Emeritus Seikei University, Sakasai Kashiwa City, Chiba, Japan

Abstract: By our new discriminant theory, we discriminate six microarrays having two-classes. Revised IP-Optimal LDF finds six minimum number of misclassifications are zero, indicating that the six data are linearly separable. Program3 and 4 coded by LINGO can decompose data into many Small Matryoshkas and Basic Gene Sets. SMs and BGSs are multivariate cancer genes set to discriminate two normal and cancer patients (or two cancer classes) by few genes. We confirmed these facts by 163 microarrays registered on GSE after 2007. The 100-fold CV validates these cancer gene sets and finds many averages of error rates for validation samples (M2) are zero and useful for cancer gene diagnosis. Thus, we can establish the first theory of gene data analysis.

Seventeen Serious Mistakes of Cancer Gene Data Analysis Since 1995

Shuichi Shinmura
Seikei University, Sakasai Kashiwa City, Chiba, Japan

Abstract: From 1995, microarray could measure protein expression values of over ten thousand genes. Medical researchers started new studies to find multivariate oncogenes different from over a hundred legacy oncogenes, such as driver genes and tumor suppressor genes, discovered microscopically based on biological knowledge. Many

researchers from many fields, such as statistics, bioengineering, machine learning (ML), AI, pattern recognition, and others, started to study a new research theme (high-dimensional data analysis; Theory2) using microarrays released from medical projects. However, they could not solve this new frontier because they lacked the basic knowledge and data analysis skills. We can quickly solve Theory2 as same as the pass/fail determination of exam data by the basic principle of the simultaneous equation. We find four universal data structures of discriminant data that open a new frontier of data analysis, including small data, high-dimensional gene data, Big data, and other data. We explain seventeen serious mistakes made by many researchers from our 169 microarray studies based on our successful results of 169 microarrays.

Infer Key Cellular Regulatory Networks in Chronic Myeloid Leukemia

Jialu Ma, Nathan Pettit, Mary Qu Yang
MidSouth Bioinformatics Center & Joint Bioinformatics Graduate Program of University of Arkansas at Little Rock, USA; University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA;
Department of Philosophy & Interdisciplinary Studies, University of Arkansas at Little Rock, Little Rock, USA

Abstract: Chronic myeloid leukemia (CML) is a myeloproliferative disease characterized by a unique BCR-ABL fusion gene. Tyrosine kinase inhibitors (TKI) were developed to target the BCR-ABL oncoprotein, inhibiting its abnormal kinase activity. TKL treatments have significantly improved CML patient outcomes. However, the patients could develop drug resistance and relapse after therapy discontinues due to intratumor heterogeneity. Single-cell RNA sequencing (RNA-seq) measures gene expression in individual cells, offering an unprecedented opportunity to decipher tumor heterogeneity. In this study, we analyzed single-cell RNA-seq data of chronic-phase CML stem cells for investigating the genetic variations underlying drug response. The CML stem cells were separated into four groups according to BCR-ABL status (positive or negative) and TKI molecular response (good or poor). We performed pairwise differential expression analysis between each cell group and normal hematopoietic stem cells. Consequently, 21 common differentially expressed genes (DEGs) across four groups were revealed including several known drug resistance genes. The hierarchy clustering analysis utilizing the expression changes of the shared DEGs showed that cells with the same molecular response tend to be clustered together, while the clusters were less dependent on BCR-ABL status. We further applied SCENIC (Single-cell regulatory network inference and clustering) to infer gene regulatory networks. We found that most common DEGs were directly regulated by one or more transcript factors (TFs). Interestingly, the DEGs appeared to have more TF regulators in the cells with good responses than poor responses. We studied the gene expression alteration patterns and corresponding transcriptional regulatory networks in the CML stem cell groups having different drug responses and fusion gene status. Our work provided new insight into the mechanism driving therapy resistance in the CML.

The 7th International Conference on Biomedical Engineering & Sciences (BIOENG 2021)
https://www.american-cse.org/csce2021/conferences-BIOENG

Partial Contrast Enhancement Method for the Classification of Ultrasound Breast Cancer Images

Mingue Song, Yanggon Kim
Department of Computer & Information Sciences, Towson University, Towson, Maryland, USA

Abstract: Complete analysis and early detection of tumor types are an essential procedure of breast cancer diagnosis and treatment. However, the ultrasoundbased breast cancer images in computer aid diagnosis (CAD) systems still insufficient to perform discrimination of benign and malignant perfectly. In addition, along with additional problems such as speckle noise, classification of breast cancer tumors still remains a challenging task. In this work, we propose a partial contrasting and partial denoising of region of interest (ROI) to enhance the discrimination of tumor types. The proposed method is designed to extract highlevel features from the correlation information between tumor and background rather than only focus on the shape of tumor. To measure the experimental results, pre-trained deep-learning models were fine-tuned to the dataset. The results demonstrates that the proposed approach is outperformed than approaches performing image contrasting and denoising against the entire image in terms of accuracy. Furthermore, it also shows that sufficient performance improvement is possible without data augmentation.

Formulating the Correlation between Speech Recognition Tests and Word Recognition Test

Mai Trinh, Sarah Beaver, Eduardo Colmenares
University of Houston, Texas, USA; University of North Texas, Texas, USA;
Midwestern State University, Texas, USA

Abstract One of the major obstacles that currently exists within the healthcare system and the cochlear implant industry is the adoption of different metrics obtained through different types of tests. In some cases, a clinic may conduct a test that does not necessarily meet the criteria requested by the insurance company. For example, the audiology clinic prefers Consonant Nucleus Consonant (CNC), while multiple insurance companies prefer Arizona Bioindustry Association (AzBio). This leads to data inconsistencies which in many cases may result in additional expenses for the user and insurer. Although money is an issue, the most negative and prominent impact is the lack of a fast and accurate conversion among the adopted tests. This paper presents the result of phase 1 of the research which acts as a predictive model between different types of hearing test scores and provides an easy deployable and reliable solution to the problem. This research focuses on the speech recognition test (AzBio) and the word recognition test (Bamford-Kowal-Bench (BKB), CNC).

Non-invasive Diabetes Detection using Gabor Filter: A Comparative Analysis of Different Cameras

Christina A. Garcia, Patricia Angela R. Abu, Rosula S. J. Reyes
Department of Electronics, Computer, & Commun. Engin., Ateneo de Manila University, Quezon City, Philippines;
Department of Information Systems & Computer Science, Ateneo de Manila University, Quezon City, Philippines

Abstract: This paper compares and explores the performance of both mobile device camera and laptop camera as convenient tool for capturing images for non-invasive detection of Diabetes Mellitus using facial block texture features. Participants within age bracket 20 to 79 years old were chosen for the dataset. 12mp and 7mp mobile cameras,

and a laptop camera were used to take the photo under normal lighting condition. Extracted facial blocks were classified using k-NN and SVM. 100 images were captured, preprocessed, filtered using Gabor and iterated. Performance of the system was measured in terms of Accuracy, Specificity and Sensitivity. Best performance of 96.7% accuracy, 100% sensitivity and 93% specificity were achieved from 12mp back camera using SVM with 100 images.

Indirect Bilateral Filter with Sharpening Capability to Dynamic PET Images Enhancing

P. Gonzalez, A. Ubilla, C. Tauber
Departamento de Computacion e Industrias, Universidad Catolica del Maule, Talca, VII Region, Chile;
UMR U1253 iBrain, Universite de Tours, Tours, Region Centre, France

Abstract: Positron tomography (PET) images are inherently affected by noise and low spatial resolution, leading to incorrect estimates of the real uptake of the radiotracer in the tissue. In this work, a new method can simultaneously reduce noise and increase the contrast between functional regions to enhance dynamic positron emission tomography images. The neighborhood information defined by the bilateral filtering framework, the local geometry of the signal, and the oriented laplacian, enables an adaptation of the filtering process to denoising inside regions, and at the same time, enhancing the contrast between functional regions. The results obtained in realistic simulated images assess the proposed method's potential compared to other methods proposed in the literature.

CT Image Analysis in Sinograms

Hong Zhang
Georgia Southern University, Savannah, Georgia, USA

Abstract: Machine learning techniques such as convolutional neural networks have been applied to object detection and analysis of computed tomography images. The reconstruction of CT images from sinograms involves the process of inverse Radon transform which is known to be an ill-posed problem and certain information loss is unavoidable. In this paper a novel approach to preform convolutions is proposed. Instead of applying convolutions in the reconstructed CT images, the proposed method will perform the equivalent operation in the domain of sinograms. This method avoids the artifacts associated with the inverse Radon transform and could potentially improve the performance for low-dose CT scans.

Effect of Software-Created Ground Truth Images in Photoacoustic Tomography

Ivan Gomez-Arista, Raquel Diaz-Hernandez, Leopoldo Altamirano-Robles, Elsi V. Mejia-Uriarte,
Oleg Kolokoltsev, Cesar L. Ordonez-Romero, Marco A. Canchola-Chavez, Maria del Transito Borraz-Arguello
CONACYT-INAOE, Tonantzintla, Puebla, Mexico; INAOE, Tonantzintla, Puebla, Mexico;
ICAT-UNAM, Ciudad Universitaria, Mexico; IF-UNAM, Ciudad Universitaria, Mexico;
Universidad Politecnica de Puebla, Cuanala, Puebla, Mexico

Abstract: One of the most relevant problems of PAT systems that are widely studied is the artifacts produced during the reconstruction process of photoacoustic images. Artifacts can be caused by low sampling, limited view, and/or narrow bandwidth. Under these conditions, the use of some deep learning algorithm focusing on the domain of the images can eliminate artifacts in the reconstructed images, however, using deep learning commonly requires an image that has better quality than the input images. These images are known as Ground Truth images. In the case of photoacoustic tomography, there is no always a better image than the reconstructed images. This work explores the alternative of using Ground Truth images generated by software from experimental data. The image reconstruction algorithm is based on the fast Fourier transform and uses interpolation between the spatial and temporal domain coordinates. First, an interpolation using the nearest neighbor is used to obtain the reconstructed images. Then a Linear

interpolation algorithm is used for the generation of Ground Truth images. The feasibility of the proposal is demonstrated with the use of a convolutional neural network applied to the reduction of artifacts in a PAT system and the use of image quality measures for the evaluation of the improved images.

A Mind-Controlled Drone for Virtual Tour in Mixed Reality

Ziming Liu, Fengpei Yuan, Xiaopeng Zhao, Miao Wang
Department of Mechanical, Aerospace & Biomedical Engineering, University of Tennessee, Knoxville, USA;
Department of Electrical and Computer Engineering, Miami University, Oxford, Ohio, USA

Abstract: In this paper, we develop a virtual tour platform, which enables a user to control a quadrotor drone in a hybrid Mix Reality (MR)-Brain Computer Interface (BCI) environment. The platform integrates MR and BCI technologies to assist users to control a quadrotor drone, giving them a sense of free moving in an immersion simulation. Motor imagery is utilized as the signature of classifying the user's mental states. The characteristics of the users' motor imagery are calculated using the electroencephalogram (EEG) and driving the movement of the drone. MR is utilized to generate the virtual tour environment through recording from a webcam on the drone. Simulation results demonstrate potentials of this platform as an alternative approach in enhancing severely paralyzed patients' quality of life and further promoting their interactions with family members or caregivers.

The 19th International Conference on Scientific Computing (CSC 2021)
https://www.american-cse.org/csce2021/conferences-CSC

Two Dimensional Scientometric Analysis of Smart Healthcare

Abdulaziz Aldaej
College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Saudi Arabia

Abstract: The advances in information technology have directed different nations' efforts towards a health-oriented society. Also, a new concept called smart healthcare (SHC) has been coined in this evolution. In recent years, this significant interest and prevalence of SHC among countries have resulted in large numbers of research publications in this area. Henceforth, research's high significance and development involve the overall analysis of the structure and development in this area, which can be clearly understood through quantitative research methods. From the perspective of healthcare and research fields, this paper provides a two-dimensional scientometric analysis of SHC research. This analysis is focused on comprehensive bibliographic records from the Web of Science from 2011-2020. To analyze the origin, status quo, and performance of the SHC study, it explores different methodological approaches. This paper offers detailed insights into the SHC research's publishing patterns, citation trends, and co-occurrence keywords for technological advancements. From different perspectives, this scientometric research defines key categories, research domains, countries, and institutions for SHC analysis. On the whole, this paper provides an overall and deeper understanding of diverse patterns, developments, and other aspects as the basis for future research directives and collaboration in the SHC study field.

Pharmacy Medications Inventory Component based System using RFID Technology in Saudi Arabia

Ashwag Maghraby, Munirah Alotaibi, Aisha M. Al-Asmari
Department of Computer Science and Information Systems College, Umm-AlQura University, Saudi Arabia

Abstract: Nowadays, pharmacies in the public hospitals in Saudi Arabia manually count inventory through a periodic system. Manually inventory methods faces some problems such as inconsistency in data entry; medicines shortages; and lost, outdated and damaged medicines. This research proposed and developed a component based software using Radio Frequency Identification technology to solve problems of manually inventory methods. It aims to increase the accuracy and efficiency of the inventory and helps pharmacies to efficiently manage inventory and improve patient safety. The proposed system was developed using Eclipse environment and Derby embedded database. It has been tested and validated which has been proven that the proposed software have better efficiency.

New Generation of Interoperable Artefacts in Medical Informatics

Filipe Nunes, Regina Sousa, Antonio Abelha, Jose Machado
ALGORITMI Research Center, School of Engineering, University of Minho, Gualtar Campus, Braga, Portugal

Abstract: A healthcare institution produces huge amounts of data that came from n sources and have m different formats. The exchange of clinical information between information systems is crucial for the effective provision of care, significantly improving the performance of the institutions. The work described here aims to fill the gap existing in the dentistry speciality, of the Hospital da Santa Casa da Miseric´ordia de Vila Verde, where it would not be possible to verify the patient's status regarding their rights to Sickness Assistance to State Servants (ADSE). This gap was filled with the development of a platform for the validation of dental legislation through ADSE, using the Agency for the Interoperation and Dissemination and Archive of Medical Information (AIDA) and ADSE web services.

Solving 0-1 Knapsack Problem based on Improved Monkey Algorithm

Xiaoping Xu, Feng Wang, Jingjing Rong, Long Liu
School of Sciences, Xi'an University of Technology, Xi'an, P. R. China;
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, P. R. China;
School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, P. R. China

Abstract: To overcome the shortcomings of the existing method in solving the 0-1 knapsack problems, an improved monkey algorithm is proposed. After a mathematical model of the 0-1 knapsack problem is given in, an improved monkey algorithm is proposed to solve it. The main idea of the improved algorithm is as follows. The initialization of the basic monkey algorithm is solved by using Kent chaotic map to make the initial position more uniform. The induction factor in climb process is also introduced to make the solution accuracy and the search speed reach a certain balance. The resolving results of the 0-1 knapsack problems show that the improved monkey algorithm has higher precision and more stable performance

Increased Security for Users with Password Complexity

Navninderjit Singh, Kanwal Gagneja
California State University, Bakersfield, California, USA;
Punjabi University, Patiala, India

Abstract: Password security standards are respected by the industry, but still need to be tested intermittently. By running brute force, dictionary, and cryptanalysis attacks against passwords implementing known security standards, this research works to demonstrate the effectiveness of these standards in defending against these attacks. The end

results show that common words do not seem to have an impact on the results of brute force attacks, and that the inclusion of special characters does indeed raise the complexity of passwords.

Fileless Keystroke Logging and Security Vulnerabilities

Andrew A. Su, Bryson Payne
Department of Computer Science and Information Systems, University of North Georgia, Dahlonega, Georgia, USA

Abstract: The past year during the COVID-19 pandemic has seen an exponential increase in cyberattacks, with public companies, private companies, institutions, nonprofits, individuals, and various parts of the industry around the world experiencing attacks ranging from intellectual property theft, stolen information, identity theft, forgery, ransomware, and many other forms of attacks as everything shifts to virtual format in response to the pandemic. Methods to exploit information such as keystroke logging, have been used to take advantage of negligent users and retrieve their information, giving organizations and individuals more problems in conjunction to adapting to the pandemic. This research analyzes the history of keystroke logging, the usage and ramifications of keystroke logging, the logistics of this technology, and how attackers use this against unsuspecting users to reveal potential security vulnerabilities with implementation of a keystroke logger and how it works. Vulnerabilities include details revealed by victims and how attackers can use these details to fulfil their ulterior motivations. Sensitive information, such as URL's, usernames, passwords, or other relevant information will also be examined to determine how keystroke logging can be a root cause to exploits and data breaches.

Comparing Machine Learning Techniques for Alfalfa Biomass Yield Prediction

Jonathan M. Vance, Christopher D. Whitmire, Hend K. Rasheed, Christian Adkins, Ali Missaoui, Khaled M. Rasheed, Frederick W. Maier
Department of Computer Science, University of Georgia, USA;
Institute for Artificial Intelligence, University of Georgia, USA;
Department of Crop & Soil Sciences, Institute of Plant Breeding Genetics & Genomics, University of Georgia, USA

Abstract: The alfalfa crop is globally important as livestock feed, so highly efficient planting and harvesting could benefit many industries, especially as the global climate changes and traditional methods become less accurate. Recent work using machine learning (ML) to predict yields for alfalfa and other crops has shown promise. Previous efforts used remote sensing, weather, planting, and soil data to train machine learning models for yield prediction. However, while remote sensing works well, the models require large amounts of data and cannot make predictions until the harvesting season begins. Using weather and planting data from alfalfa variety trials in Kentucky and Georgia, our previous work compared feature selection techniques to find the best technique and best feature set. In this work, we trained a variety of machine learning models, using cross validation for hyperparameter optimization, to predict biomass yields, and we showed better accuracy than similar work that employed more complex techniques. Our best individual model was a random forest with a mean absolute error of 0.081 tons/acre and R2 of 0.941. Next, we expanded this dataset to include Wisconsin and Mississippi, and we repeated our experiments, obtaining a higher best R2 of 0.982 with a regression tree. We then extended the work by investigating its eligibility for domain adaptation (DA). DA, which has shown promise in alleviating the problem of scarce training data, involves training models using input from a source domain to predict or classify samples in a target domain with a different data distribution. We obtain surprisingly positive results by simply excluding the target data while training on the source, with R2 as high as 0.986 using a regression tree. These results suggest that DA could be a promising technique in solving the problem of data sparsity in the domain of crop biomass yield prediction, so we plan to research more complex DA techniques in forthcoming work.

Provoking the Mobile Payment Intention to Actual Use Behaviour

A. R. M. Musallam, P. S. Josephng, H. C. Eaw
Graduate Business School, UCSI University, Malaysia;
Institute of Computer Science & Digital Innovation, UCSI University, Malaysia;
Faculty of Business and Management, UCSI University, Malaysia

Abstract: Financial Payment has evolved into mobile Payment for easy access and convenience. This resulted in transforming how the banking sector functions and enabled the customer to transact and interact remotely. Despite the benefits of mobile Payment, the adoption rate is unsatisfactory worldwide. Therefore, understanding the customer's adoption intention factors is essential for both researchers and practitioners. This study examines mobile payment intention and actual use by adapting the integrated framework, which combines Mobile Technology Acceptance Model and the Theory of Planned Behaviour with additional new constructs found to be related. The integrated new framework named Mobile Payment Technology Acceptance Model and Behaviour (MPTAMB). The new model consists of 12 variables is examined via a structured equation mixed-mode methodology. Based on past studies, the determined constructs for the revised model found reliable after running the pilot test. MPTAMB factors that influence mobile payment intention can enhance the possibility of provoking the actual use of mobile Payment towards the newly revised theoretical contribution. Knowing mobile payment influence will shape the industry strategic decision and socially contribute to transforming society into a cashless society.

Innovative Pivot-Based Elimination of Unknown (Nuisance) Parameters from Underlying Models for Constructing Efficient Statistical Decisions under Parametric Uncertainty

N. A. Nechval, G. Berzins, K. N. Nechval
BVEF Research Institute, University of Latvia, Riga, Latvia;
Transport and Telecommunication Institute, Riga, Latvia

Abstract: The big question in statistics is: How can we eliminate the unknown (nuisance) parameter from an underlying model? Eliminating unknown (nuisance) parameters from an underlying model is universally recognized as a major problem of statistics and has been formally studied in virtually all approaches to inference. A surprisingly large number of elimination methods have been proposed in the literature on this topic. The classical method of elimination of unknown (nuisance) parameters from the model, which is used repeatedly in the large sample theory of statistics, is to replace the unknown (nuisance) parameter by an estimated value. However, this method is not efficient when dealing with small data samples. The Bayesian approach is dependent of the choice of priors. In this paper, a new method is proposed to eliminate the unknown (nuisance) parameter from the underlying model. This method isolates and eliminates unknown (nuisance) parameters from the underlying model as efficiently as possible. Unlike the Bayesian approach, the proposed method is independent of the choice of priors and represents a novelty in the theory of statistical decisions. It allows one to eliminate unknown parameters from the problem and to find the efficient statistical decision rules, which often have smaller risk than any of the well-known decision rules. To illustrate the proposed method, some practical applications are given.

Zero Trust Model: Autonomic Data and Network Security

Haydar Teymourlouei, Lethia Jackson
Department of Technology & Security, Bowie State University, Bowie, Maryland, USA

Abstract: Zero Trust Model is about one's security approach as well as about specific tools and implementation strategies. The Zero Trust network model is a modern alternative to traditional perimeter network security and is gaining popularity the last several years. The main principle of Zero Trust Model's architecture is that everything inside or outside the network is not reliable until verified. The Zero Trust network model is a modern network security model that aims to solve challenges. This paper discusses multi-factor authentication (MFA), network segmentation, real time monitoring, inspect event log, access control, policies and Virtual private network (VPN) as the effective methods to develop Zero Trust. In addition, these methods can help protect data and the entire network.

Securing an IoT Network: Intrusion Detection and Malware Identification

Haydar Teymourlouei, Daryl Stone
Department of Technology & Security, Bowie State University, Bowie, Maryland, USA

Abstract: Computing and communication has significantly progressed with devices such as smartphones and household appliances. Because of the evolution of the Internet of Things (IoT), these devices must be developed on a large-scale and deliver better-quality services to society. Malware detection from IoT devices has proven to be a challenge due to the lack of intrusion detection. However, with an effective intrusion detection system in place, malware threats in IoT devices can be identified. Security methods such as whitelisting, greylisting, and blacklisting are effective tools. They can secure systems and control malware detection. Another tool is real-time monitoring, which constantly conducts intrusion detection in search of threats and weaknesses. Real-time monitoring can be used to prevent malware attacks against IoT devices.

Performances Evaluation of LoRaWAN Tracking Devices

Thierry Antoine-Santoni, David Araujo, Alexandre Mameli, Bastien Poggi, Jean-sebastien Gualtieri, Antoine Aiello
UMR CNRS SPE 6134, University of Corsica, Corte, France; SITEC, Biguglia, France;
UMS CNRS Stella Mare 3514, University of Corsica, Biguglia, France

Abstract: Animal tracking is an important issue for science but also for farms. Numerous devices exist, but they are often poorly configurable, with a very limited lifespan and a fairly high purchase cost. In this paper, we propose a set of tests on low-cost long-range communication devices with the aim of improving GPS accuracy and energy consumption. The results show a real feasibility in terms of performance with an algorithm optimisation energy but limited in some cases.

Design Space Exploration of Concurrency Mapping to FPGAs with OpenCL: A Case Study with a Shallow Water Model Kernel

Moteb Alghamdi, Graham D. Riley, Mike Ashworth
Department of Computer Science, University Of Manchester, Manchester, United Kingdom, UK

Abstract: FPGAs are emerging as a viable accelerator for compute-intensive workloads on HPC systems. The adoption of FPGAs for scientific applications has been stimulated by the emergence of better programming environments such as High-Level Synthesis (HLS) and OpenCL available through the Xilinx SDSoC design tool. The mapping of the multi-level concurrency available within applications onto HPC systems with FPGAs is a challenge. OpenCL and HLS provide different mechanisms for exploiting concurrency within a node leading to a concurrency mapping design problem. In addition to considering the performance of different mappings, there are also questions of resource usage, programmability (development effort), ease-ofuse and robustness. This paper examines the

concurrency levels available in a case study kernel from a shallow water model and explores the programming options available in OpenCL and HLS. We conclude that the use of SDSoC Dataflow over functions mechanism, targeting functional parallelism in the kernel, provides the best performance in terms of both Latency and execution time, with a speedup of 314x over the naive reference implementation.

An Intrusion Detection System in Mobile Ad Hoc Network using RSA and Block Based Algorithm

S. Venkatesan, M. Ramakrishnan
Department of Computer Applications, Madurai Kamaraj University, Madurai, Tamilnadu, India

Abstract: Link error and malicious packet dropping are two sources for packet losses in one-hop wireless ad hoc network. While observing a sequence of packet losses in the network, we are interested in determining whether the losses are caused by link errors only, or by the combined effect of link errors and malicious drop. A one-hop wireless network, nodes cooperate in relaying/routing traffic. An adversary can exploit this cooperative nature to launch attacks. The adversary may first pretend to be a cooperative node in the route discovery process. It is including in a route the adversary starts dropping packets. The malicious node is stop forwarding every packet received from upstream nodes, completely disrupting the path between the source and the destination. Detecting selective packet-dropping attacks is extremely challenging in a highly dynamic wireless environment. The difficulty comes from the requirement that we need to not only detect the place (or hop) where the packet is dropped, but also identify whether the drop is intentional or unintentional. The main challenge in our mechanism lies in how to guarantee that the packet-loss bitmaps reported by individual nodes along the route are truthful, i.e., reflects the actual status of each packet transmission. Truthfulness is essential for correct calculation of the correlation between lost packets. The challenge is not trivial because it is natural for an attacker to report false information to the detection algorithm to avoid being detected. For example, some packets may have been dropped by the node but the node reports that these packets have been forwarded. The typical wireless device is resource-constrained and also requires that a user should be able to delegate the burden of auditing and detection to some public server to save its own resources.

Economic and Environmental Optimization of Demand Side Management in the Presence of Multi-DG and DR

H. Shayeghi, H. A. Shayanfar, M. Alilou
Energy Management Research Center, University of Mohaghegh Ardabili, Ardabil, Iran;
College of Technical & Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran;
Electrical Engineering Department, Urmia University, Urmia, Iran

Abstract: In this study, the combination of multi-objective particle swarm optimization and fuzzy decision-making method is utilized to find the best schedule for demand side management. The distribution system is considered in the presence of multi distributed generation (renewable and nonrenewable) and demand response. Economic and environmental indices are evaluated during the optimization. For better evaluation the results, the daily performance of grid and DG is optimized with and without consideration of demand response. Finally, the proposed algorithm is implemented on test distribution system. The obtained numerical results indicate the impact of demand side management on improving the economic and environmental indices of distribution system.

An Essence based Framework using a Domain Driven Design Approach to Address Microservices Lifecycle from Identification to Implementation

Parthasarathi Ray, Pinakpani Pal
Applied Statistics Unit, Indian Statistical Institute, Kolkata, India;
Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata, India

Abstract: To counter the digital disruption in today's world, the organizations are opting for digital transformation. While there is a marked propensity to turn towards microservices architecture to re-imagine the application landscape, there are also challenges involved. This is where adoption of the Domain Driven Design (DDD) approach can prove beneficial. In our endeavour to develop a framework to address the microservices lifecycle using DDD, we leverage Essence, the Software Engineering language and kernel advocated by the SEMAT (Software Engineering Method and Theory) community. This paper builds on a previous work proposing three new alphas by describing the related work products and activities and their inter-relationships as we model the Microservices practice using DDD based on Essence.

Fuzzy Evaluation Modeling and its Combination with Factorial Analysis

Rajan Alex
School of Engineering, Computer Science, and Mathematics, West Texas A&M University, Canyon, Texas, USA

Abstract: The resolution of fuzzy regression analysis has been presented through many well-known publications and has been widely used in fuzzy set theory and its application. The author has also introduced a new kind of fuzzy regression, which indeed is not a linear fuzzy regression, but a convenient, simple, quick and adaptable method for fuzzy regression. In this work, we will re-name the method as enveloping regression. Then a consistency between the enveloping regression modeling and factorial analysis will be presented. Factors space theory is an offshoot of mathematical cognition, which provides a universal coordinate framework for description and cognition of objects or things. It applies two kinds of information resources, quantitative and qualitative, simultaneously in practical prediction problems. This work introduces factorial analysis to accompany enveloping regression in learning and estimation. By means of comparing and combining, we illustrate how the enveloping regression using factorial analysis can be applied in a modeling problem for a set data.

Formal Specification of Aptitude Architecture for Recommendation and Adaptation of Learning Contents and Activities Based on Learning Analytics

Adelina Aleksieva-Petrova, Milen Petrov
Faculty of Computer System and Technologies, Technical University of Sofia, Sofia, Bulgaria;
Faculty of Mathematics and Informatics, Sofia Univesity "St. Kl.Ohridski", Sofia, Bulgaria

Abstract: The adaptation and recommendation of learning content and learning activities are two processes which have important roles in each learning system. When the learning analytics has used in these processes, the software system has complex architecture. The paper goal is to propose the formal specification of such type of system. The study method is to define a meta-model, detail system constructs using UML notation and to provide proof-of-concept implementation of the model which allows to validate the system architecture.

Steady Viscous Flow in a Cavity for Large Rayleigh Numbers and Small Prandtl Numbers

Saeed Dubas, Nihal Siriwardana, Saroj Pradhan
Engineering and Computer Science Division, University of Pittsburgh at Johnstown, PA, USA;
Department of Mathematics, Prairie View A & M University, Prairie View, Texas, USA

Abstract: A finite difference scheme is developed for the flow of a viscous incompressible fluid in a square cavity. The top wall of the cavity is held at a higher temperature than the bottom wall. This problem has received considerable interest in literature due to its suitability in testing techniques that may be applied to a wide range of practical problems. The method has converged for all parameters of physical interest. The results are presented and compare favorably with some of the most accurate results available from other studies. The stability of the method is shown in the convergence for a wide range of Rayleigh numbers, Prandtl numbers and mesh sizes.

Quantum Machine Learning Foundations and Applications: A Succinct Literature Review

Javier Orduz-Ducuara, Pablo Rivas, Erich Baker
School of Engineering and Computer Science, Department of Computer Science, Baylor University, Texas, USA

Abstract: The advantages of leveraging machine learning with quantum computing theory are still under study and so far are very promising. The new field of quantum machine learning has great, transformative, potential in our field. However, one of the major difficulties presented for those in either the machine learning community and the quantum community is the knowledge gap that naturally poses a barrier preventing both fields from joining forces without a significant amount of preparation, time, and space for professional development. With this in mind we prepared this succinct literature review that provides an introduction to Quantum Machine Learning, involving selected topics in quantum mechanics, mathematics, and computing. We focus on fundamental concepts and a few interesting applications. Upon reading this review, the reader will be exposed to interesting topics about the technological revolution that Quantum Computing and Machine Learning have brought.

Population Protocol for the Multi-objective Knapsack Problem

Chihiro Kaneko, Akihiro Fujiwraa
Graduate School of Computer Science & Systems Engineering,
Kyushu Institute of Technology, Iizuka, Fukuoka, Japan

Abstract: A multi-objective optimization problem involves a number of objective functions to be maximized or minimized, and no single solution exists for the problem because there is no solution that simultaneously satisfies all objective functions. In the present paper, we consider the multi-objective knapsack problem, and propose an optimization algorithm based on a population protocol (PP). PP is a computation model based on mobile sensor networks with limited computational power, in which computation is performed by agents communicating with each other. The experimental results show that the proposed algorithm obtains a better set of Pareto solutions than existing algorithms.

Trajectory Degradation of the Tent Map Chaotic Signal and the Implications on Statistical Independence

Alexandru Vaduva, Adriana Vlad, Bogdan Badea
Faculty of Electronics, Telecommunications & IT, POLITEHNICA University of Bucharest, Bucharest, Romania;
The Research Institute for Artificial Intelligence, Romanian Academy, Bucharest, Romania;
Facebook, California, USA

Abstract: The paper studies the effects of the numerical precision used when iterating the tent map chaotic signal, phenomenon further referred to as "trajectory degradation". The experimental results indicate a relationship between the tent map parameter and the minimum required number of bits for an iterative calculus without encountering the trajectory degradation phenomenon. The paper evaluates the quasi-independence distance between two tent map samples and compares the results of using double precision against arbitrary extended precision.

Role of Cardiac Output during Sleep and Exercise of Human Respiratory Control System Model with Multiple Transport Delays

Saroj P. Pradhan, Nihal J. Siriwardana, Saeed M. Dubas
Department of Mathematics, Prairie View A & M University, Prairie View, Texas, USA;
Engineering and Computer Science Division, University of Pittsburg at Johnstown, Johnstown, PA, USA

Abstract: We study the effects of cardiac output ($\dot{Q}$) on stability analysis of human respiratory control system for congestive heart failure (CHF) and during exercise with two control loops (peripheral control (GP) and central control (GC)) and two delays (peripheral control delay (τp lung to carotid body) and central control delay(τb) lung to brain). CHF have small $\dot{Q}$, high τp and τb and small stability regions. During exercise, heart rate and cardiac output increases. It decreases τp, increases GP , increases peripheral ventilation VP , and increases instability. It also decreases the τb, increases GC , increases central ventilation VC , and decreases instability. Our conclusion is that stability region decreases in CHF case in sleep and gradual increases of exercise is more stable than the sudden increase of exercise in normal case.

Hash Function Visualization Language

Andrew Chabot, Hans-Peter Bischof
Department of Computer Science, Rochester Institute of Technology, Rochester, New York, USA

Abstract: Hash functions are essential in today's computing world, showing up everywhere from data security to digital fingerprinting. However, describing how these functions work has a daunting task in the past because of their complex inner workings. This paper describes a language that can be used to solve this problem through creation of visualizations. This language is called the Hash Function Visualization Language and is intentionally simple to use by design. The design and syntax of the language are presented along with its features. Examples from visualizations created in the language for SHA-1, SHA-2 256 and SHA-3 256 are also shown to give some context as to what visualizations in the language look like.

Soft Body Collisions for Ring Simulations with Rust

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

Abstract: We present a detailed analysis of the mathematics and computational aspects of soft-sphere collisions with an implementation in Rust. This work is part of an effort to rewrite and update an existing C++ ring simulation code. We compare our approach to the approach implemented in PKDGRAV and find that our altered derivations of the parameters for the impact forces provide more stable results for the coefficient restitution. However, attempts to use a higher-order integrator and smooth the transition from gravity to intersecting forces did not improve the numerical stability as desired. We also found that this approach often produces incorrect results when the colliding particles are of very different sizes.

Photometric Rendering of Dust and Freed Regolith in Ring Simulations

Elizabeth Ruetschle, Lance Ellis, Mark C. Lewis
Department of Computer Science, Trinity University, San Antonio, Texas, USA

Abstract: This paper presents continuing work by our group to produce scientifically accurate synthetic observations of planetary rings, based on N-body simulations, to compare with observations. The primary focus of this paper is the inclusion of dust and small bodies in synthetic images produced by photometric renderings. Actual ring systems inevitably have an abundance of small bodies from centimeter sized particles down to micron-scale dust. N-body simulations can't include material at this scale for numerical reasons. The work presented here is two fold. First, we look at N-body simulations of collisions between meter sized particles that have a regolith layer to see how the regolith is released by collisions of various velocities and orientations. Second, we model this ejecta as ellipsoids of "dust" that can be included in photometric synthetic observations.

MedSensor: Medication Adherence Monitoring Using Neural Networks on Smartwatch Accelerometer Sensor Data

Chrisogonas Odhiambo, Pamela Wright, Cindy Corbett, Homayoun Valafar
Computer Science & Engineering Department, University of South Carolina, Columbia, South Carolina, USA;
College of Nursing, University of South Carolina, Columbia, South Carolina, USA

Abstract: Poor medication adherence presents serious economic and health problems including compromised treatment effectiveness, medical complications, and loss of billions of dollars in wasted medicine or procedures. Though various interventions have been proposed to address this problem, there is an urgent need to leverage light, smart, and minimally obtrusive technology such as smartwatches to develop user tools to improve medication use and adherence. In this study, we conducted several experiments on medication-taking activities, developed a smartwatch android application to collect the accelerometer hand gesture data from the smartwatch, and conveyed the data collected to a central cloud database. We developed neural networks, then trained the networks on the sensor data to recognize medication and non-medication gestures. With the proposed machine learning algorithm approach, this study was able to achieve average accuracy scores of 97% on the protocolguided gesture data, and 87% on natural gesture data.

UVMBench: A Comprehensive Benchmark Suite for Researching Unified Virtual Memory in GPUs

Yongbin Gu, Wenxuan Wu, Yunfan Li, Lizhong Chen
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon, USA

Abstract: The recent introduction of Unified Virtual Memory (UVM) in GPUs offers a new programming model that allows GPUs and CPUs to share the same virtual memory space, which shifts the complex memory management from programmers to GPU driver/ hardware and enables kernel execution even when memory is oversubscribed. Meanwhile, UVM may also incur considerable performance overhead due to tracking and data migration along with special handling of page faults and page table walk. As UVM is attracting significant attention from the research community to develop innovative solutions to these problems, in this paper, we propose a comprehensive UVM benchmark suite named UVMBench to facilitate future research on this important topic. The proposed UVMBench consists of 32 representative benchmarks from a wide range of application domains. The suite also features unified programming implementation and diverse memory access patterns across benchmarks, thus allowing thorough evaluation and comparison with current state-of-the-art. A set of experiments have been conducted on real GPUs to verify and analyze the benchmark suite behaviors under various scenarios.

Extended General Malfatti's Problem

Ching-Shoei Chiang, Hung-Chieh Li, Min-Hsuan Hsiung, Fan-Ming Chiu
Computer Science and Information Management, Soochow University, Taipei, Taiwan, ROC

Abstract: The malfatti's problem ask for 3 circles (called malfatti circles) tangent to each other and two side of a right triangle. 7 Ajima extend this problem to arbitrary triangles and call the new problem the general Malfatti problem. In this paper, we want generalize the general problem to find 3, 6, 10, …$\sum i n$ 8 1 circles inside the triangle, so that the corner circle tangent to two edge 9 of the triangle, boundary circle tangent to one side of the triangle, and 4 other circles (at least two of them are boundary or 10 corner circles), and the inner circles tangent to 6 other circles. We call this problem extended general Malfatti Problem (EGMP). 11 We proposed an O(n) algorithm, where n is the number of circles inside the given triangle, to solve EGMP in this paper.

Machine Learning for Predicting Protein Functions

Benedetta Pelosi
Molecular Biosciences, Stockholm University, Sweden

Abstract: N/A - POSTER

Application of Linear Mixed-Effects Models in Sleep Study: A Comparison and Assessment with Linear Models for Large Data

Mahmoud Affouf, Rebecca Robbins
School of Mathematical Sciences, Kean University, Union, New Jersey, USA;
Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Massachusetts, USA; Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts, USA

Abstract: In this paper, we study prospective observational data obtained from sleep tracking applications. We focus on statistical modeling of the variations of three outcomes: sleep duration, onsetlatency, and bedtime variability. We compare linear mixed-effect models and linear regression models and their explanatory power.

Communication in Agile Software Development - A Mapping Study

Suddhasvatta Das, Kevin Gary
Arizona State University, Tempe, Arizona, USA

Abstract: Software industry is a fast-moving industry and to keep up with this pace the development process also needs to be fast and efficient and Agile software development (ASD) is the answer to thisproblem. Even though ASD has been in there for over two decades there are still multiple unknown questions tied to ASD that need to be addressed. In this study we are going to address one of the most critical factors of ASD i.e. Communication. We conducted a review of 14 studies and found the areas under ASD communication that the community is interested in as well as research gaps.

Extracting Data Element Names from Natural Language Definitions

Bader Alshemaimri, Ramez Elmasri, Tariq Alsahfi, Abdullah Almoqbil
King Saud University, Saudi Arabia; University of Texas at Arlington, Texas, USA;
Jeddah University, Saudi Arabia; University of North Texas, Texas, USA

Abstract: In this paper, we design and develop rulebased natural language processing (NLP) techniques to automatically extract data element names from data element definitions written in American English. The goal is to study how using NLP techniques can improve the accuracy of extracting standardized data element names in a domain-independent context. To achieve this, we first identify heuristic patterns that mine noun phrases and relationships from data element definitions. Then, we use these noun phrases and relationships as input to determine components of data element names. The output of the patterns is reviewed by a domain expert. We apply our method to extract the 5 standard components of a data element name in the Railway and Transportation domains. We first achieved 80% accuracy, then by improving the rules and adding a similarity function, we improved the accuracy to 95% in our final experiments.

A School Redistricting Integer Programming Model for Achieving Connected School Attendance Zones

Joshua C. Detwiler, Layne T. Watson, Naren Ramakrishnan, David S. Bieri
Department of Computer Science, Virginia Tech, Blacksburg, Virginia, USA;
School of Public and International Affairs, Virginia Tech, Blacksburg, Virginia, USA

Abstract: School redistricting is an optimization problem of geographically partitioning students to public schools. A school board chooses the set of criteria used to compare and evaluate redistricting plans, which often include measures and constraints such as compactness of school attendance zones (SAZs), proximity of students to their schools, connectivity of the SAZs, and so on. Connectivity has been heavily explored in redistricting literature. In this work, we develop an integer nonlinear programming model that expresses SAZ connectivity exactly using the Laplacian matrices of SAZs. We also offer another formulation that indirectly applies the connectivity constraint via lazily added cutting planes. The latter formulation may be seen as more computationally viable with existing software.

Design of a Smart Waste Management System for the City of Johannesburg

Beauty L. Komane, Topside E. Mathonsi
Department of Information Technology, Tshwane University of Technology, Pretoria, South Africa

Abstract: Every human being in this world produces waste. Having the population grow in each segment of the world means more waste is produced. South Africa is a developing country with many townships that have limited waste resources. Over increasing population growth overpowers the volume of most municipal authorities to provide even the most essential services. Waste in townships is produced via littering, dumping of bins, cutting of trees, dumping of waste near rivers and overrunning of waste bins. Waste increases diseases, air pollution, environmental pollution and lastly increases gas emissions that contributes release of greenhouse gases. The ungathered waste is dumped widely in the streets and drains contributing to flooding, breeding of insect, rodent vectors and spreading of diseases. Even waste that is collected is often disposed in uncontrolled dumpsites or burned, polluting water resources and the air. Therefore, the aim of this paper is to design a smart waste management system for the city of Johannesburg. The city of Johannesburg contains waste municipality workers and has provided some areas with waste resources such as waste bins and trucks for collecting waste. But the problem is that the resources only are not enough to solve the problem of waste in city. The waste municipality uses traditional ways of collecting waste such as going in each street picking up waste bins, organising and picking up of waste gathered in the corners of each street. The traditional way has worked for years but as the population is increasing more waste is produced which causes various problems for the waste municipalities and the public in large. The proposed system consists of sensors, user applications, and real time monitoring system. This paper adopts the experimental methodology.

Using a Layered Approach for Graphical-based Passwords

Jessy Ayala
Tandon School of Engineering, New York University, Brooklyn, New York, USA

Abstract: This paper proposes a developed approach to graphical authentication relying on previous work already done. More specifically, how a user would be able to utilize an image for a password in hopes of improving security and avoiding an attacker from being able to log-in their device. The existing methods using graphical-based passwords range from connecting dots on an Android device to dragging the cursor across the screen with a pattern. This approach focuses on layering a Bluetooth protocol for the user to be able to upload a specific image, in which will be checked for authentication, in conjunction with a modified cued-click point interface in hopes of improving security in a graphical-based password mechanism.

Threaded Multi-Core GEMM with MoA and Cache-Blocking

Stephen Thomas, Rishi Khan, Lenore Mullin, Kasia Swirydowicz
National Renewable Energy Laboratory, Golden, Colorado, USA;
Extreme Scale Computing, Wilmington, Delaware, USA;
University of Albany, SUNY, Albany, New York, USA;
Pacific Northwest Laboratory, Richland, Washington, USA

Abstract: A threaded multi-core implementation of the high performance dense linear algebra matrix-matrix multiply GEMM kernel is described. This kernel is widely implemented by vendors in the basic linear algebra subroutine BLAS library. The mathematics of arrays (MoA) paradigm due to Mullin (1988) results in contiguous memory accesses by employing outer-product forms. Our performance studies demonstrate that the MoA implementation of double precision DGEMM combined with optimal cache-blocking strategies results in at least a 25% performance gain on the Intel Xeon Skylake processor over the vendor supplied Intel MKL basic linear algebra libraries. Results are presented for the NREL Eagle supercomputer. The multi-core DGEMM achieves over 100 GigaFlops/sec with eight openMP threads.

SESSION: Military and Defense Modeling and Simulation

Co-Chairs: Dr. Douglas D. Hodson (Chair), Dr. Michael R. Grimaila**, Dr. Ryan D. Engle****
**Computer Science and Engineering Department, US Air Force Institute of Technology (AFIT), USA*
***Head, Systems Engineering & Management Department, US Air Force Institute of Technology (AFIT), USA*
****Systems Engineering, US Air Force Institute of Technology (AFIT), Wright-Patterson AFB, Ohio, USA*

DDS-Cerberus: Data Distribution via Ticketing

Andrew Park, Richard Dill, Douglas Hodson, Wayne Henry
US Air Force Institute of Technology (AFIT), USA

Abstract: N/A

Comparison of Archetypal Entity-Component Systems and Relational Databases

Bailey Compton, Douglas Hodson, Richard Dill, Michael Grimaila
US Air Force Institute of Technology (AFIT), USA

Abstract: The Entity-Component-System (ECS) architectural design pattern separates data from computer logic (i.e., behavior) - components define data, systems define behavior. As such, it embraces the ideas espoused by Data-Oriented Programming. Relational databases can be compared to ECS implementations through emphasis on organization and retrieval of data. Conceptualizing archetypal ECS tables as single-table, standalone databases allows for Data-Oriented programmers to employ database standardization practices to the organization of data in ECS applications.

AFSIM's Entry into Aircraft Hardware in the Loop

Nathaniel Peck, Douglas Hodson, Richard Dill, Michael Grimaila
US Air Force Institute of Technology (AFIT), USA

Abstract: Hardware in the loop (HIL) simulations benefit from a 3D world to visualize their state. This research presents a basic entry point for the Advanced Framework for Simulation, Integration, and Modeling (AFSIM) as a solution to visualize aircraft state. This research highlights AFSIM's XIO capabilities (a network interface) as a potential path to support simulations with integrated hardware.

Understanding Unity's ECS Architecture

Brett Martin, Douglas Hodson, Laurence D. Merkle
US Air Force Institute of Technology (AFIT), USA

Abstract: Previous research has shown that the dataoriented programming paradigm offers a multitude of performance benefits over the traditional object-oriented paradigm. In particular, the entity component system (ECS) design pattern provides multi-threading capabilities, optimal memory utilization, and a means of separating data from logic from a data-oriented approach. This research is focused on the implementation of a flight simulator that uses a simple flight dynamic model (FDM) using Unity's new Data-Oriented Technology Stack (DOTS) package. In this simulation, we implement a flight simulator using the ECS design pattern, but also make use of some of the more traditional, object-oriented approaches native to the Unity development platform.

A Data Oriented ECS-Based Design to Send DIS Packets

Noah Scott, Douglas Hodson, Richard Dill, Michael Grimaila
US Air Force Institute of Technology (AFIT), USA

Abstract: The Entity-Component-System (ECS) is an architectural design pattern that separates data from computer logic (i.e., behavior) - components define data, systems define behavior. It is based the data oriented programming paradigm and is increasingly being used by game engines to improve execution efficiency. The Distributed Interactive Simulation (DIS) protocol is an IEEE military-oriented standard for exchanging simulation data within a distributed system to achieve real time execution performance to support human and/or hardware interaction. This effort explores implementing a interoperability interface using a Rust-based, ECS-architected simulation to send DIS packets.

Quantifying DDS Network Overhead

Nathaniel Peck, Douglas Hodson, Richard Dill, Michael Grimaila
US Air Force Institute of Technology (AFIT), USA

Abstract: The Data Distribution Service (DDS) is an emerging distributed communication middleware which aims to simplify and enhance the communication between distributed applications. While DDS may simplify the creation of such communication mechanisms that include quality of service (QoS) controls, some developers have expressed concern about introduced runtime overhead. This research seeks to measure some of the network overhead paid as a penalty for supporting this type of communication style with QoS characteristics. We examine the communication traffic associated with DDS over a fixed period of time in various configurations.

The 20th International Conference on e-Learning, e-Business, Enterprise Information Systems, & e-Government (EEE 2021)
https://www.american-cse.org/csce2021/conferences-EEE

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

Anastasia Biggs
School of CS and IT, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract: The purpose of this qualitative case study was to explore why and how corporate training managers can adopt mobile technology for employee training. The case study method explored the learning processes to determine if a learning model is appropriate for the use of mobile technology as a training tool (De Zan, De Toni, Fornasier, and Battistella, 2015, p. 341). This qualitative case study utilized interviews to explore how the use of mobile technology can be adopted to train employees. The use of interviews examined the degree of employee growth from mobile training (Alberghini, Cricelli, and Grimaldi, 2014, p. 260). Case study methodology answered how mobile technology through cause-effect relationships explored the lack of mobile technology adoption interventions between corporate managers and the organization (De Zan et al, 2015, p. 335) (Tsang, 2013, p. 197).

An Online Database System for Money Management

Matthew Wiberg, Zizhong John Wang
Department of CS and IT, Virginia Wesleyan University, Virginia Beach, Virginia, USA

Abstract: This research project is to design a website that included a functioning budgeting tool and secure login system. It is designed with the idea in mind that many people have poor financial habits and therefore, this website could benefit those who wish to help organize their finances and learn to budget. Users can create a personal profile and enter income and expense amounts that will be stored and allow them to login back in at a later time for use once again. The project is implemented with HTML, CSS, PHP and MySQL.

Determinants Influencing Intention to Use Social Commerce for Shopping in Developing Countries: A Case Study of Oman

Shamma Thabrit Alharizi, Mariam Al Areimi, Abdul Khalique Shaikh
Department of Information Systems, Sultan Qaboos University, Muscat, Oman

Abstract: Social media has had a significant impact on our individual lives, including our behavior regarding the purchasing of daily products. This study investigates the factors influencing Omani nationals' intentions to obtain products via social commerce. The researcher surveyed 202 participants and utilized the Technology Acceptance Model to develop the theoretical framework. The data collection was analyzed statistically using an appropriate testing mechanism. Statistical methods, including Cronbach's alpha and multiple linear regression, were utilized for reliability and hypotheses testing. After analyzing the collected data and testing the hypotheses, the findings indicated that perceived usefulness, enjoyment, and ease of use of social commerce affect positively on Omani nationals' intentions to utilize social commerce for shopping. The independent variables had a statistically significant impact on the intention to use social commerce shopping for products; these explain 69.9% of the variation on customers' intention to utilize social commerce for shopping.

Software Licenses: Security and Privacy in Terms of Dynamic Link Library

Kanwal Gagneja, Navninderjit Singh
California State University, Bakersfield, California, USA;
Punjabi University, Patiala, India

Abstract: Cracking some Adobe products is as simple as replacing a .dll file in the respective app's folder. These software range from hundreds of dollars in price to thousands of dollars in price. Thus, making sure users have a legitimate copy and are not using pirated software is of utmost importance to many distributed software companies. There is a lot of information concerning this topic on the Internet, since it is related to piracy, most of the technical information is kept behind closed doors. This is where much of our research will have to take place. We plan to solve this problem by inspecting the .dll (dynamic link library) files that regulate the product's license. After analyzing the files, we will try and find a way to see if a product's legitimacy can be verified by testing scenarios such as if the .dll has been modified in any way.

A Determination of the Impact of Training on the use of Tablets in Primary Schools from Students' Perspectives

Ioana Chan Mow, Edna Temese, Fiafaitupe Lafaele, Tara Patu, Misioka Tanielu
National University of Samoa, Apia, Samoa

Abstract: This paper describes a study which evaluated the effectiveness of training sessions which accompanied a rollout of 1389 tablets and Aptus devices to selected primary schools in Samoa. The study evaluated adequacy of the training, level of usage of tablets, Aptus and programs and applications, as well as ease of use and usefulness to

students. Findings of the study showed that the level of usage of both tablets and the Aptus needed to be improved, there was a need for more training, technical and pedagogical support as well as a community of practice of both teachers and learners.

Cut Or Cute: Towards the Gamification of Eye-Tracking and Teaching English Concepts to Children

Reham Ayman, Youssef Othman, Nada Sharaf
Department of CS, Faculty of Media Engineering and Technology, German University in Cairo, Egypt

Abstract: Phonetics is always a crucial step in teaching children. The same letter can, however, affect the phonetic differently making the task harder. For example, the "E" in "cute" lengthens the pronunciation of the letter "U" more than in "cut". The aim of the work presented in this paper is introduce a 3D maze game to teach children the effect of the vowel "E" on pronouncing the preceding vowel. The game also allowed for incorporating a reading task within making it more engaging for children and this allows for the collection of eye-tracking data. The game successfully achieved this goal with an average learning gain of 90%. The game had an average engagement level of 3.7 and considered to have good flow, and a good level of usability.

Red Flags Detection in Public Procurements using a Business Rules Management System

Elias Alberto Mora Gonzalez, Matias Ariel Rivas Tellez, Julio Paciello, Juan Pane
Universidad Nacional de Asuncion, Paraguay

Abstract: Public procurements represent a major source of corruption in governments. Organizations such as Transparency International have shown that public procurement represents a significant percentage of the countries GDP, and therefore they require e-government tools to prevent corruption. Business Rules Management Systems are world widely solutions applied to separate business logic from the code itself, allowing a domain expert to write rules for searching red flags as collusion indicators. This work proposes the use of a business rules engine for detecting red flags in public procurements in Paraguay. More than 45000 red flags were found using paraguayan open data from 2016 to 2020, thus serving as an important e-government tool for preventing corruption for Governments.

The Digital Transformation of Higher Education Institutions (HEIs) in Saudi Arabia (SA): A Study for Evaluation & Development - a Research Plan

Fahd Almutari, Sung-Chul Hong
Department of Computer and Information Sciences, Towson University, Towson, Maryland, USA

Abstract: This research supplements the current literature on measuring and evaluating the Digital Transformation of Higher Education in Saudi Arabia (SA) and explores the challenges, and weakness that could develop, and thus could help improve the Digital Transformation as well as help reach other goals of Saudi Higher Education Institutions (HEIs). The present study also calls for further work evaluating and developing strategies for digitization of education in SA. Higher Education in SA could benefit greatly by this study because it evaluates and compares the digital transformation in their universities with the digital transformation model that is highly valuable and helps to achieve the objectives of improving and transforming higher education for the future.

Optimising an Ecological Farming Plan using Linear and Goal Programming Methods

Nadja Damij, Talib Damij, Liljana Ferbar Tratar, Janez Grad
University of Northumbria, Newcastle Business School, Newcastle upon Tyne, United Kingdom, UK;
University of Ljubljana, School of Economics and Business, Ljubljana, Slovenia

Abstract: This paper explores how linear and goal programming may be used to determine the optimal yearly farming plan. Therefore, a linear programming model was developed with an objective function to maximise production and a goal programming model to minimise the deviations of five goals such as: maximising production and profit, while minimising the capital invested and the number of human and machine working hours. In addition, the constrained optimisation and implementation of the Lagrangian function was used to analyse the objective function's sensitivity to changes in the constraints. While taking into the account the farm owner's strict limitations concerning the land parcel sizes, optimal linear and goal programming solutions were found that enable the farm owner to achieve his vision.

Evaluating the Adoption and Use of Moodle at the National University Of Samoa: The Perspectives of Foundation Computer Studies Students

Ioana Sinclair, Ioana Chan Mow
Computing Department, National University of Samoa, Apia, Samoa

Abstract: This paper describes a study which evaluated the adoption and use of Moodle based on the Foundation Computer studies class at the National University of Samoa. The study, based on the Unified Theory of Acceptance and use of technology (UTAUT) evaluated adoption and use of Moodle based on the variables of performance expectancy, effort expectancy, attitude & behavior, social influence, facilitating conditions, self-efficacy, and anxiety. Findings from the study were highly positive and indicated positive student perceptions and valuations on all these factors and a willingness of students to continue the use of Moodle in the future. Generally there were no age differences but there were gender differences and programme differences across some probes with males and Foundation Science programme having higher valuations across several aspects of Moodle.

Securing RESTful Endpoints in Containerized e-Government Microservices using OAuth 2.0

Dimitrios Posnakides, Harald Gjermundrod, Ioanna Dionysiou
Department of Computer Science, School of Sciences and Engineering, University of Nicosia, Nicosia, Cyprus

Abstract: Secured data is a vital part of any system that operates in a privacy-aware manner. This is especially true for e-Government platforms where a security breach or incorrect implementation of security mechanisms might have a tremendous impact on privacy. A baseline security framework must be adopted to ensure the correct practical deployment of confidentiality, integrity, availability, authentication, non-repudiation and authorization policies. One of the current industry trends in building complex systems is the migration to containerized microservices, abandoning the conventional monolith type of systems. The RESTful microservices approach is becoming the de facto technology to design and implement microservices-based systems, mainly due to itssimplicity; microservices communication is established without the support of additional infrastructure. In this paper, a prototype e-Government microservices-based system is presented. As several microservices expose REST endpoints, solutions are discussed to secure these endpoints.

Frontiers in Cognition for Education: Coherent Intelligence in e-Learning for Beginners Aged 1 to 3 Years

Igor Val Danilov, Sandra Mihailova, Iegor Reznikoff
Academic Center for Coherent Intelligence, ACCI, Rome, Italy; Riga Stradins University, RSU, Riga, Latvia; Department of Philosophy, University of Paris X - Nanterre, Paris, France

Abstract: The essential goal of e-learning research is an online curriculum for children with developmental disorders. The study verifies the hypothesis about the insensitivity of sensory perception to shared intentionality, measuring the ability of children to solve incomprehensible problems without any communication. The online experiments on developing numerical competence of children aged 12 months, 18 months, 28 months, 31 months, and 33 months show their unexpected ability to interact with mothers (excluding the youngest student) creating the bond between sounds of spoken numbers and the appropriate set of items. This insight in numerosity succeeded in children during the uplift of emotion-motion ongoing social dynamics in these dyads–organisms in social entrainment. This interpersonal psychophysiological coherence contributes to cognitive development.

Teaching Network Security, Digital Forensics, and Mathematical Topics in the Pandemic

Eamon P. Doherty, Elly Goei
School of Public and Global Affairs, Fairleigh Dickinson University (FDU), Hackensack, New Jersey, USA

Abstract: This paper discusses the use of videoconferencing, online platforms, and student behavior during the COVID-19 pandemic. The paper hypothesizes that students with more time and in need of diversions have more efforts into schoolwork thus resulting in better outcomes. The paper also discusses the need for online practical exercises during a time when computer lab visits are not possible. There are also some examples of how mathematical topics can be inserted in the digital forensics and network security curricula thus increasing quality in STEM programs.

Integrating Digital Stories in Early Childhood Education using Alternative Authoring Tools

Marianthi Manousi, Jim Prentzas
Department of Education Sciences in Early Childhood, Democritus University of Thrace, Alexandroupolis, Greece

Abstract: Digital stories have been used in a variety of educational settings involving students of different ages. This is facilitated by the existence of alternative available tools for the creation of digital stories. Digital stories may be used effectively in early childhood education. Tools convenient for young children may be exploited for this purpose. This paper discusses the integration of digital stories in two Greek kindergarten classrooms. Five alternative authoring tools were used for the creation of digital stories. Young children worked in groups. They were able to create their own stories using the first four tools. The assistance of the teacher was required to create a digital story with the last tool. The overall results were positive.

The Role of Data Mining in a Post-COVID World from Business Perspective

Amitava Karmaker
University of Wisconsin-Stout, Menomonie, Wisconsin, USA

Abstract: The pandemic caused by COVID-19 (SARS- CoV-2) has been a major game-changer in an unprecedented way, and the relentless pace of the world-wide spread of the disease has forced the world economy to come to a standstill., With the whole world being stuck in their homes, retail businesses had to come up with an innovative way in order to remain in operation (Lind, 2020). Customer demand for a lot of things had diminished, and therefore there

were surpluses in inventory. Consequently, businesses needed to create and target their moves to survive, and that is where data mining plays a critical role. Based on historical data and trend analysis, it can help identify the strategic steps to go ahead, and then come to a conclusion based on what it discovers and then work on top of it. In this paper, I will highlight how data mining has impacted business operations. Furthermore, the combination of artificial intelligence and machine learning with data mining algorithms has played a significant role.

Assessing the Performance and Usability of e-Learning Platforms for Effective e-Learning in Ghana in the Wake of COVID-19

Patricia Ghann, Emmanuel Dortey Tetteh, Kobby Asare Obeng
Computer Science Department, Koforidua Technical University, Ghana

Abstract: Information Communication Technology (ICT) is advancing with rapid development aimed at offering quality education among learners to ensure satisfaction as well as convenience. With the devastating effects of Coronavirus, many institutions are relying on e-learning technologies to carry out both administrative and academic activities to promote social distancing as well as curb the spread of the novel coronavirus. E-learning allows students from different geographical locations to learn as if in the classroom through the internet; providing tools that enhance effective teaching and learning. The question we ask is: are e-learning platforms performing as expected in terms of the ability of users to effectively use them via their interfaces? To find out the answer to this question, five e-learning platforms of recognized institutions in Ghana were tested using heuristic evaluation, cognitive walk-throughs and user testing,This paper establishes a performance evaluation model for e-learning platforms, by elaborating some critical factors that are comprehensive when compared to previous approaches and establishes a criterion for performance evaluation based on parameters such as usability in terms of, the design of the e-learning platforms, accessibility and response/request time, It is anticipated the findings of this research will ensure that e-learning platforms perform better, are accessible and usable for effective teaching and learning.

Web-based Piano Learning System using Visualized Numbered Notation

Kai-Chin Yang, Keh-Ning Chang, Herng-Yow Chen
Department of Computer Science and Information Engineering, National Chi Nan University, Taiwan

Abstract: Recent advances in multimedia and Web technologies have made controlling MIDI in Web possible easily. We can now learn music online with the help of the MIDI technologies and the progress of the support of browser to MIDI. This paper aims to develop a Web-based piano learning system with visualized numbered notation to help beginners. For piano beginners, try to play music on the different keys which may involve different black notes on the keyboard might be very challenging. They tend to feel confused by the new key where their fingers should press in a different scale. However, music learners do not always have a teacher beside. It would be helpful to have a learning system as a supporting tool along their musical journey. We use numbered musical notation as a tool because the numbers can be associated with the notes and music theory. The number system is more readable and can be memorized easily when playing a piano. A successful piano keyboard player may have a magic power, they can see some notes more important than others. It's a mystery. MIDI becomes more popular, and we now can use MIDI events in the Web easily. We can use MIDI keyboards to send MIDI events to our computers. In our system, we can compute the MIDI data to slow down or speed up the song or change the key of the song. This paper presents a prototype through Web for everybody who wants to practice transposition. Everyone can connect to our website. They can choose a song to practice. They can change the key and instruments they want. They can change the label detail to test them whether they can see the key numbers in their brains or not.

Cybersecurity for Modern American Healthcare Institutions

Kyla Konen, Lemuel Cheon, Michael Demetriou, Michael DePalma, Tara Jubran,
Lucas Schleben, Fady Nissan, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA

Abstract: Ever since technology has been integrated into healthcare organizations there have been issues with cybersecurity. Every year the amount of security breaches in the healthcare industry increases, which ends up costing the companies money and in some cases disrupting patient care. In 2017 a company named Merck & Co. was attacked with malicious ransomware. This incident ended up costing the company approximately $870 million in damages. In addition to instances like this, there are also personal data breaches. More than 3000 healthcare data breaches and over 500 patient records have been compromised between 2009 and 2019. In 2019 alone there were 572 healthcare data breaches which involved 41.4 million Americans. With the steady increase in data breaches in the healthcare industry it begs the question, what are they doing to stop it? In this paper we will discuss what cybersecurity and healthcare are, the current state of cybersecurity in the American healthcare system and what healthcare organizations can do to prevent and further mitigate the security risks in the industry.

SESSION: Agile IT Service Practices for the cloud - a Customer Perspective

Chairs: Dr. Manuel Mora, Dr. Jorge Marx Gomez**, Dr. Raul Valverde****
**Autonomous University of Aguascalientes, Mexico*
***University of Oldenburg, Germany*
****Concordia University, Canada*

Role of the IT Department Culture on Agile Methods Adoption - an Exploratory Ethnographic Study

Olayele Adelakun, Tiko Iyamu, Manuel Mora, Fen Wang
Informatics, DePaul University, Chicago, Illinois, USA;
Information Technology, Cape Peninsula University of Technology, Cape Town, South Africa;
Information Systems, Autonomous University of Aguascalientes, Aguascalientes, Ags, Mexico;
IT & Administration Management, Central Washington University, Ellensburg, Washington, USA

Abstract: The adoption of agile methods is influenced by many factors such as the size of the organization, age of the organization, the culture of the organization and the organization industry. The information technology (IT) department culture is recognized to have one of the most influencing factors on agile methods adoption by the IT team members in organizations. There is significant body of knowledge on IT culture but despite the numerous studies and literature, challenges remain, because the manifestation of culture and agile adoption has not been extensively and empirically explained. The focus of this pilot study is to explore the relationships between IT cultures and agile methods adoption among eleven companies in the USA using on-premises. Participant observation research method was used in this study. One important finding was that IT leadership culture has a major influence on IT adoption consistently across all the eleven US organizations that participated in the study.

Lean UX Cloud-Based Model for the Generation of Virtual Reality Environments Applied in Health

Hector Cardona-Reyes, Jaime Munoz-Arteaga, Jose Eder Guzman-Mendoza
CONACYT Research fellow, CIMAT Zacatecas, Mexico;
Autonomous University of Aguascalientes, Aguascalientes, Mexico;
Center for Research in Mathematics, Quantum: Knowledge City, Zacatecas, Mexico

Abstract: This work presents a model based on Lean UX with cloud services, in order to offer an approach focused on optimizing and streamlining the generation of user-centered virtual reality (VR) environments. Cloud services enable the model to facilitate the creation and sharing of these low-cost, lowcomplexity VR environments by enabling the delivery of applications that are easy to access and manage through a simple interface. The model consists of components for the design of VR environments, the cloud services involved under a specific health context in which user interaction strategies are included to achieve their specific objectives. The model is tested through the generation of VR environments to assist in healthcare in adults.

Agile Software Architecture Practices in the SOA and MSA Styles - a Selective Comparative Review

M. Mora, H. Duran-Limon, O. Karam, Y. Reyes
Information Systems, Autonomous University of Aguascalientes, Aguascalientes, AGS, Mexico;
Informatics, University of Guadalajara, Guadalajara, JAL, Mexico; AWS, Atlanta, Georgia, USA

Abstract: Since about one decade, the Service-Oriented Architecture (SOA) design style emerged with promissory benefit, from the Service-Oriented Computing (SOC) paradigm. However, its original realization has been limited, and thus SOA has received critiques in the literature and rejections from the practical communities. In contrast, Microservices Architecture (MSA), a still debatable concept between a new software architecture style or a particular implementation of SOA, has emerged in the practical communities for an agile software design and its lightweight implementation deployed on-premises, cloud, or hybrid data centers. However, software architecture design using MSA, and agile practices are emergent and systematic methods are still missing. In this research, thus, we review the main literature upon SOA and MSA to elaborate an initial framework about agile software architecture design practices. We consider, this contribution useful towards for the full realization of systematic agile software design methods using Microservices.

Review the State of the Art on Agile Development of IT Services for Analytics - Data Science Projects

Beatriz Torres, Raul Valverde, Stephane Gagnon
University of Quebec at Outaouais, Quebec, Canada;
Supply Chain and Business Technology Department, Concordia University, Montreal, Canada;
Departement des sciences administratives, University of Quebec at Outaouais, Quebec, Canada

Abstract: Currently the new trend in IT projects is cloud technology for its implementation. However, for this resources, methodologies and expertise are needed. Technology advances by leaps and bounds so it is necessary to adapt to market demands and offer new solutions according to these needs. This article is based on the literature review analysis for project management in data science with agile methodology focused on the cloud technology and the most recent research on these issues and the main concepts. The purpose of this paper is to review the literature about agile development of IT Services for Analytics – Data Science projects. An approach integrating computerized natural language processing (NLP) with a systematic literature review methodology was adopted. A large corpus of 18 700 titles and abstracts of agile, data sciences and IT Services articles were analyzed using NLP in order to identify the relevant literature using a ranking. The identified literature was subjected to a systematic review guided by a well-established method in Agile methodology in cloud with IT Services.

Agile vs Rigor-Oriented IT Service Design Practices - A Comparative Analysis

Manuel Mora, Jorge Marx-Gomez, Olayele Adelakun, Fen Wang
Information Systems, Autonomous University of Aguascalientes, Aguascalientes, AGS, Mexico;
Informatics, University of Oldenburg, Oldenburg, Lower Saxony, Germany;
Computer Science, De Paul University, Chicago, Illinois, USA;
IT & Administration Management, Central Washington University, Ellensburg, Washington, USA

Abstract: IT Service Design has been a critical phase for all rigor-oriented IT Service Management (ITSM) frameworks and standards, such as ITIL v3.v2011, CMMI-SVC, and the ISO/IEC 20000:2010. IT Service Design phase has guided ITSM practitioners to design useful, cost-beneficial, and warranted IT services. However, the current business environment has changed from stable and mid-term user demands to highly dynamic and short-term ones. Consequently, ITSM frameworks and standards with an agile perspective are now emerging. This paper reports a comparative analysis of rigor-oriented vs agile IT Service Design processes/practices, for providing theoretical and practical insights on the different tenets found in the main two rigor-oriented ITSM frameworks/standards (i.e. ITIL v3, ISO/IEC 20000:2010) and two emergent proffered agile ones (i.e. ITIL v4, and ISO/IEC 20000:2018). We use a rigororiented and an agile IT Service Design Framework to evaluate the four IT Service Design processes/practices. Our conceptual findings indicate that the rigor-oriented IT Service Design process differs substantially from agile ones, and thus ITSM practitioners are alerted to identify correctly the ITSM paradigm to be implemented.

Scrum Process Model to support Inclusive e-Learning in Regular Education during COVID-19 Contingency

Jaime Munoz-Arteaga
Universidad Autonoma de Aguascalientes, Aguascalientes, Mexico

Abstract: Inclusive teaching is practiced in Mexico that help schools to make students with specific learning needs be included in regular education they provide the necessary help to enhance the learning process. Additionally, The COVID-19 contingency has made attitudinal changes of students, together with the long academic brake, and the forced technologyintensive on-line learning. This work preconizes a Scrum model process as a feasible solution, making easy and stronger the collaboration, role definition, and goals prioritization. Together, consultants, teachers, parents, and technologist have analyzed gaps in the intervention process of the supporting units, related with collaboration, teamwork, adaptations in activities and knowledge acquisition, and proposed a solution to it. Current work presents how the proposed model was used to support inclusive elearning in regular elementary education during COVID-19 contingency.

The 19th International Conference on Embedded Systems, Cyber-physical Systems, & Applications (ESCS 2021) https://www.american-cse.org/csce2021/conferences-ESCS

An Efficient B-tree Implementation for Memory-Constrained Embedded Systems

Nadir Ould-Khessal, Scott Fazackerley, Ramon Lawrence
Department of Electronic Engineering Technology, Okanagan College, Kelowna, BC, Canada;
Department of Computer Science, University of British Columbia, Kelowna, BC, Canada

Abstract: Embedded devices collect and process significant amounts of data in a variety of applications including environmental monitoring, industrial automation and control, and other Internet of Things (IoT) applications. Storing data efficiently is critically important, especially when the device must perform local processing on the data. The most widely used data structure for high performance query and insert is the B-tree. However, existing implementations consume too much memory for small embedded devices and often rely on operating system support. This work presents an extremely memory efficient implementation of B-trees for embedded devices that functions on the smallest devices and does not require an operating system. Experimental results demonstrate that the B-tree implementation can run on devices with as little as 4 KB of RAM while efficiently processing thousands of records.

Throughput Enhancement via Exploting Parallelism for Hyperledger Fabric Blockchain

Ju Won Kim, Jae Geun Song, Hyeon Woo Shin, Ju Wook Jang
Department of Electronics Engineering, Sogang University, Seoul, Korea

Abstract: We proposes a method that improves parallelism in Hyperledger blockchain by reordering transactions. When transactions arrive, our system identifies the dependency between transactions and reorders them according to three factors: Tree level, Height, and Timestamp. By creating a dependency graph, the relationship between transactions in which conflict occurs is represented in a hierarchical tree structure. Also, our system receives information such as block generation information, block size, block generation time so that our system adjusts the submission time of each transaction. Our system improves performance compared to the existing method in terms of block storage capacity, TPS, and total number of executed transactions. By solving real-time obstacles and errors, our system prevents transaction error which may occur in the existing Hyperledger Fabric.

Achieving Reliable Automated Off-grid Indoor Vertical Farming by FMEA and Digital Twin

Muzaffer Hosen Akanda, Stephen Smyth, Fatema Tuz Zohra, Otto Randolph, Bahram Asiabanpour
Ingram School of Engineering, Texas State University, San Marcos, Texas, USA

Abstract: The classic farming process proves to be inefficient as food requirements increase due to growing population size. To grow enough food to feed a country requires massive tracks of fertile land, large amounts of freshwater, and high labor costs. Many countries throughout the world do not have access to these resources. This, along with classic farming's susceptibility to weather volatility, blights, infestations, and other forms of possible crop damage, require the world to find new solutions to the growing food supply problem. Project Evergreen looks to solve this problem with an Indoor Vertical Farming (IVF) system by cutting down on water, land, and labor costs through

automation and self-sufficient off-grid (water and energy) design. This approach is not without issues though. An automated off-grid system requires several interconnected systems, any of which could fail at any time if poorly implemented and overlooked. With the system designed to be entirely off-grid, any potential failure could lead to total crop loss, as it does not have labor employed to check on and maintain each component. To mitigate this, the project looks to employ Failure Mode and Effects Analysis (FMEA) tools through the design process to increase the overall quality of the off-grid IVF system. Additionally, as part of failure prevention and optimization efforts, digital twin technology has been employed in this project. Results including the framework for the improvement in the reliable supply of water, energy, and communication as well as consistent ambient conditions are presented.

Design of a Health Monitoring System for an e-Bike

Eli Gabriel Avina Bravo, Christophe Escriba, Jean Yves Fourniols,
Fabien Pillard, Georges Soto-Romero
LAAS-CNRS, Toulouse, France; INSA-Toulouse, Toulouse, France;
Centre Hospitalier Universitaire Toulouse, University Sports Clinic, Toulouse, France;
ISIS-Castres, Castres, France

Abstract: N/A

Accelerated Evaluation of Autonomous Drivers using Neural Network Quantile Generators

Edward Schwalb
MSC Software, USA

Abstract: Whereas autonomous driver engineering has limited control on outcomes for individual scenarios, engineering processes must exert control over the performance statistics. A key challenge impeding statistical confidence is the need to test for a long list of infrequent hazardous events. Using ”smart miles” promises to reduce the evaluation cost by increasing the frequency of those infrequent hazardous events. We propose a simulation based Bayesian approach which increases the frequency of those infrequent events by many orders of magnitudes as compared to naturalistic miles. We represent the Operational Design Domain (ODD) using a population of scenarios, and provide methods for sampling the ODD. We propose a quantile function based sampling approach which is able to generate a single sample with thousands of instances using a single forward pass of a deep neural network (DNN). We develop a practical sampler training method for an “inverted DNN” architecture. The resulting sampler is capable of generating a skewed distribution comprising of “smart miles“ in which hazardous events of interest occur at a frequency >95%. To gauge the quality of the generated sample we propose the quality metrics of histogram diversity, histogram homoscedasticity and average sample distance.

Smart Control of Commercially Available Atmospheric Water Generators for Reduced Power Consumption

Natalie Ownby, Mark Summers, Bahram Asiabanpour
Ingram School of Engineering, Texas State University, San Marcos, Texas, USA

Abstract: The availability of drinking water is a major issue in some countries, especially during the summer season due to the lack of rain. Atmospheric Water Generation (AWG) is an alternative solution for collecting fresh water directly from the condensation of air particles, however, it requires more energy to produce water than most other clean water sources. This study applies basic power management principles and psychrometric analysis to implement a smart control system for a commercially available AWG. This system only allows the AWG to run when the atmospheric conditions are favorable for water production. This control system increased efficiency and decreased the cost of the water produced by the AWG systems by an average of 5.9% in the summer season.

The 17th International Conference on Foundations of Computer Science (FCS 2021)
https://www.american-cse.org/csce2021/conferences-FCS

An Asynchronous P System using Bitonic Sort

Kosei Nagao, Akihiro Fujiwara
Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Fukuoka, Japan

Abstract: Membrane computing, which is a computational model inspired by living cell activity, has considerable attention as one of new paradigms of computations. Using the membrane computing, several sorting algorithms have been proposed. However, existing sorting algorithms for the membrane computing have different types of inputs. In addition, there is still room for improvement on the complexities of the existing algorithms. In the present paper, we propose an asynchronous P system based on bitonic sort for inputs representing binary numbers. The proposed P system sorts n binary numbers of m bits in O(mn log 2n) sequential steps and O(m log 2n + log 3n) parallel steps.

Mirror Matrices and Symmetric Boolean Functions

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

Abstract: Matrix-based symmetry of Boolean functions has been shown to be a useful tool for analyzing and simulating Boolean circuits. So far, the useful groups have been the set of all n n × permutation matrices for a particular n, and the conjugates of this group. For some time it has been known that there are other matrix groups, not conjugate to the group of permutation matrices, that can be used to characterize Boolean functions. Without understanding the origin of these groups, it has been difficult to use them in practice. This paper provides a powerful tool for tracing the origins of these groups, namely mirror matrices. Using various types of mirror matrices, it is possible to understand these groups and possibly put them to use.

Storing and Processing Information in Technological and Biological Computing Systems

Janos Vegh, Adam Jozsef Berki
Kalimanos BT, Debrecen, Hungary; University of Medicine, Targu Mures, Romania

Abstract: Information is commonly considered as a mathematical quantity that forms the basis of computing. In mathematics, information can propagate instantly, so its transfer speed is not the subject of information science. In all kinds of implementations of computing, whether technological or biological, some material carrier for the information exists, so the information's propagation speed cannot exceed the speed of the carrier. Because of this limitation, for any implementation, one must consider the transfer time between computing units. We need a different mathematical method to take this limitation into account: classic mathematics can only describe infinitely fast and infinitely small computing system implementations. The difference between the mathematical handling methods leads to different descriptions of the behavior of the systems. The correct handling also explains why biological implementations can have lifelong learning and technological ones cannot. The conclusion about learning evidences matches others' experimental evidence, both in technological and biological computing.

A Model for Storing and Processing Information in Technological and Biological Computing Systems

Janos Vegh
Kalimanos BT, Debrecen, Hungary

Abstract: Information is commonly considered as a mathematical quantity that forms the basis of computing. In mathematics, information can propagate instantly, so its transfer speed is not the subject of information science. However, in all kinds of computing implementations, whether technological or biological ones, some material carrier for the information exists, so the information's propagation speed cannot exceed the speed of its carrier. Because of this limitation, we need a model which can consider the transfer time between computing units. For that goal, we need a different mathematical method: classic mathematics can only describe infinitely fast and infinitely small computing system implementations. The more general (completed) model explains the experienced behavior, discovers new features and explains some famous failures. We can describe all known implementations in terms of the model; that enables us to analyze, among others, biology-mimicking technological solutions and biology-mimicking learning methods. The paper, using the correct temporal handling of the computing process, explains why biological implementations can have lifelong learning and technological ones cannot; why technology needs special units for storing information while biology does not; why learning, the mathematical learning methods for artificial intelligence, and technologically implemented machine learning have only their name in common. The conclusion matches others' experimental evidence, both in technical and biological computing.

A Systematic Literature Review on Object-Oriented Databases

Zaeem Anwaar, Muhammad Jawad, Wasi Haider Butt
Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan

Abstract: The database is a compulsory part of any management system as almost every mobile, desktop and webbased applications utilize it for keeping records. An objectoriented database is a pure amalgamation of object-oriented programming and database concepts. The way with which the programming (Encapsulation, Inheritance etc.) and database ideas come together to provide a synchronized output is known as OODB (Object-oriented databases). A Systematic Literature Review (SLR) has been used to categorize 27 research works, bring out through 2010-2020. The purpose of this SLR is to identify different approaches related to different research questions as query optimization techniques in objectoriented databases. Additionally, other research questions are answered with a claim to support Object-oriented databases. We elaborated on the advantages of OODBMS (Objectoriented database management system) over RDBMS (Relational database management system). Moreover, a detailed comparison of existing OODB is explained to get a wider view. It is concluded from this SLR that Object-oriented provides adequate approaches, methods and techniques for developing systems and databases with many advantages (improved flexibility, extensibility, redundancy, response time, and real-world modeling etc.).

A Systematic Review on NoSQL Databases of Big Data

Mehwish Naz, Muhammad Usman Farooq, Aiman Kalsoom, Saqlain Abbas, Wasi Haider Butt
Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan

Abstract: NOSQL databases have arisen as an essential surrogate to RDBMS (Relational Database Management Systems). It possesses significant features such as scalability, availability and fault tolerance. These are primarily called non-relational or distributed databases having capabilities to store large amount of data efficiently. Relational databases were not able to provide the horizontal scalability that is a big requirement of nowadays scenario. NOSQL with its distributed architecture supports high scalability and can hold big volume of data by using its feature of horizontal scalability with database servers sharing the workload therefore reducing the chances of catastrophe loss. The goal of this study is the comprehensive analysis of NOSQL databases by identifying the security challenges in

NOSQL databases, structural issues in big data processing. Moreover, the best NOSQL database on the basis of performance and migration of RDBMS to NOSQL has also been identified. We have selected 31 studies out of 302 existing studies and delivered categorization and quantitative overview as well. We focused mainly on the most recent research studies relevant to NOSQL databases. Moreover, this paper offers the assessment of NOSQL databases which give greater scalability and performance in storage and access of the records. Therefore, it is concluded that NOSQL databases are highly efficient and scalable in managing massive data as compared to traditional RDBMS.

Impact of Data Representation Techniques on Document Compliance Checking

Iqra, Seemab Latif, Rabia Latif
National University of Science and Technology (NUST), Islamabad, Pakistan;
Prince Sultan University, Riyadh, Saudi Arabia

Abstract: One of the most challenging works in text mining is to extract information or objects from an unstructured text document. Learning a meaningful and effective text representation for words and documents is a prerequisite for many machine learning tasks, such as classifying, clustering, and retrieving text. This study proposes to compare the results of the document representation techniques i.e., tf-idf, word embedding, topic modeling using compliance check. In this study, the standard and targeted documents are tokenized via NLP techniques and then compliance checking is done using different document representation techniques. The principle objective of this examination is to find best document representation technique for automated compliance checking.

How the Time Spent Online Correlates with the Homework Scores?

Ana Oprisan, Sorinel A. Oprisan
Department of Physics and Astronomy, College of Charleston, Charleston, South Carolina, USA

Abstract: World Wide Web significantly changed the way we communicate and learn, especially during the COVID-19 pandemic. Physics instructors have multiple offers for online homework systems (PHGA, WebAssign, etc.) Almost all web-based assessment systems offer instant feedback and a convenient grade book. We identified different strategies adopted by students in solving physics problems by analyzing data regarding time management as reflected by PHGA records. We identified different behavioral patterns by correlating the performance reflected by the student's score with the duration of activity measured by the time interval between "start" and "stop" and its timing regarding the deadline.

The 17th International Conference on Frontiers in Education: Computer Science & Computer Engineering (FECS 2021) https://www.american-cse.org/csce2021/conferences-FECS

Using the Comprehensive Exam to Directly Assess Outcomes in a Graduate Program

Daniel S. Spiegel, Tauqeer Hussain, Lisa A. Frye
Department of Computer Science and Engineering, Kutztown University of PA, USA

Abstract: The requirement to pass a comprehensive exam to complete a master's degree is common to most programs. The authors, all members of the Computer Science & Information Technology Department at Kutztown University, concluded that this format was not most conducive to ensuring success for its students or their department's program itself. That, along with the need to assess the program as part of the university's accreditation created a perfect storm where the potential roadblock was removed, and the cumulative work of graduating students became a central basis of assessing program outcomes.

The Study of Peer-Assessment Impact on Group Learning Activities

Zhiyuan Chen, Soon Boon Lee, Shazia Paras Shaikh, Mirza Rayana Sanzana
School of Computer Science, University of Nottingham Malaysia, Malaysia

Abstract: In the University of Nottingham Malaysia, school of computer science, peer assessment has been used for a few modules. However, comparing to lecturer marked assessments, peer assessment is a more comprehensive learning process and many of the associated problems have occurred. The main objectives of this project are to provide a complete and systematic review of this field, to increase the practice and quality of the peer assessment process, and to improve the effectiveness of students learning in groups with a proper peer assessment system. In this research, pilot studies were conducted and took the form of a survey, focus group interview, and questionnaire that examined the impact of using peer assessment to improve students learning within groups.

Unnecessary Hurdles: A Systematic Literature Review Examining the Hiring Process in Computing

Stephanie Lunn, Monique Ross
Knight Foundation School of Computing & Information Sciences, Florida International University, Miami, Florida, USA

Abstract: Whether a fledgling computer science graduate ready to start their career, or an experienced software engineer looking to make a change, the hiring process is necessary to obtain a position. Although the concept of being interviewed before getting hired is not unique to the discipline, the preparation, recruitment process, and technical challenges required are distinctive. This paper presents the results of a systematic literature review conducted to assess the stages of hiring in computing from the applicant's perspective, and that of the employer/industry. In addition, we offer suggestions to improve the process, and to make hiring more equitable to all job seekers. The findings from this analysis provides valuable insights for students and other job applicants, academia, and industry.

Educating Students on Techno-Capitalism's Impact to STEM

Jon Duncan Hagar
Grand Software Testing, LLC, USA

Abstract: This paper crosses the disciplines of economics, governing, technology, software, science, and systems. These topics are vast and complex each unto themselves, but we are faced with a collision between these in our society and the tech industry today. The paper postulates that much of society and student stress that is being experienced today comes in part from the intersection of these in what is called techno-capitalism. The importance for technology students is to recognize the situations we create and become part of the solution. This paper provides a brief overview of 'techno-capitalism', bringing it into the information age as it impacts technologists, educators, and society. It then considers the career and personal choices that future information technology workers may have to make during their education.

Evention: Online Event Management System

Mudasser F. Wyne, Shahrbanoo M. Gargari, Yousuf Diarra, Armen Sherbeitjian
College of Professional Studies, National University, San Diego, California, USA

Abstract: The year 2020 started with the coronavirus crisis that affected human life and had devastating consequences for people all around the world affecting the way business and office work as well as daily life chore are conducted. Almost all of the activities are happening remotely, in some cases, it is believed that this virtual life may be a new normal. Evention is an online event management web application that offers an efficient user interface for users to create, manage, and participate in online events. The system allow only registered user login and new user are allowed to register on the application. Additionally, the system then allows the user to select date, time, videos and place of event. With this application, people will have a better opportunity to manage, search, find, and attend online meetings and events.

Utilizing Test Driven Development in a Data Structures Course

Steven Bitner, John Coffey
Department of Computer Science, University of West Florida, Pensacola, Florida, USA

Abstract: High failure rates in Data Structures and Algorithms courses suggest the need for different approaches to teaching programming. This work contains a description of an effort to introduce Test-Driven Development (TDD) in a first course in data structures. Methods of teaching and utilization of the method are described, as are empirical results regarding student utilization of TDD and outcomes in the course. Instruction and project distribution that focuses on unit testing indicates an adoption of an incremental development methodology by the students in the study. Advantages, pitfalls and issues that need to be addressed in order to utilize the approach successfully in intermediate programming courses are discussed.

Learning Environment, Scaffolding and Learning Outcomes among Tertiary Mathematics Students

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

Abstract: This study compared learning environment perceptions and learning outcomes for two enrolment groups of mathematics students taking a linear algebra unit, whose learning environment differed in observable ways. Syllabus and organisation of teaching were the same, but Group 1 was taught over a full year, and Group 2 over one semester. Additional support for Group 1 was given by special teaching material designed as scaffolding for learning.

Perceptions of environmental factors, labelled Workload, Support, Task orientation and Integration were obtained from questionnaire responses. Assessment of learning outcomes used a method, based on the SOLO taxonomy, which gave a common scale for scoring students' final examination answers. Scores for Workload, Task orientation, and Integration had significant positive correlations with class test marks. Group 1 had higher scores for Support (significant), and Workload (close to significance). Learning outcome scores were compared by group and gender, with firstyear mathematics marks as covariate. Highly significant differences favoured Group 1. The conclusion, therefore, was that the longer time frame and the additional support facilitated learning.

Virtual Reality for Physics Teaching: an Immersive Experience in Kinematics Learning

Jhasmani Tito Cruz, Tania Basso, Regina Moraes
Univeristy of Campinas - UNICAMP - Campinas, Brazil;
University of Coimbra - UC - Coimbra, Portugal

Abstract: Virtual reality (VR) applied to education has started to grow significantly over the last decade thanks to the inclusion of commercial HDM devices. Particularly, the learning in the areas of Science, Technology, Engineering and Mathematics (STEM) gained prominence with the possibility of having their assimilation facilitated in the development of abstract concepts with the support of VR. As physics is one of the areas of STEM, this work describes the methodology defined to test the hypothesis that the use of an immersive environment based on VR facilitates the understanding of the contents of the discipline of physics and increases the engagement of students. Thus, by applying the methodology the authors aims to assess the relationship between some of the elements of VR and their influence on the results of learning specific concepts of kinematics.

Gamifying Data Structures Algorithms

Hadeel Mohammed Jawad, Samir Tout
Oakland University, Michigan, USA; Eastern Michigan University, Michigan, USA

Abstract: Computer science algorithms are not always easy to learn. Educators recommend using games in teaching a complex subject to make it easier and more interesting to students. In this paper a practical game is proposed to learn the different data structure sorting algorithms. The game idea was inspired by class activities that were used to teach a data structures course. These activities and the students' feedback are demonstrated in this paper. Students enjoyed the class activities and found them useful in learning data structures. Our conclusion is that student learning of computer science algorithms and data structures can be enhanced through a practical gamification that helps stimulate active learning.

Addressing Sources of Stress and Distress among Undergraduate Computer Science Students

Kennedy Manley, Laila Arain, Christian Murphy
University of Pennsylvania, Philadelphia, Pennsylvania, USA;
Bryn Mawr College, Bryn Mawr, Pennsylvania, USA

Abstract: The mental health of undergraduate students, particularly those in challenging fields such as Computer Science, has become a growing concern. This paper seeks to raise awareness of the mental health and wellness concerns among CS students, and to explore ways in which CS faculty members can support them. We present the results of a survey of 213 undergraduate CS students, as well as interviews with 15 students and eight instructors. Our investigation revealed that undergraduate CS students experience distress due to intellectual challenges, as well as social and emotional. Students expressed a desire for more opportunities for collaboration as a way of addressing their distress, and though faculty acknowledged the potential benefits, there are concerns that must first be addressed.

The Impact of Social Media in Learning and Teaching: A Bibliometric-based Citation Analysis

Ramla Al-Maamari, Abdul K. Shaikh, Saqib Ali
Department of Information Systems, Sultan Qaboos University, Muscat, Oman

Abstract: This paper presents the results of a systematic review of the literature on the impact of social media in learning and teaching through bibliometric-based Citation analysis. The objective of the review was to map the evolution of the current literature and identify the leading sources of knowledge in terms of the most influential journals, authors, and articles. From a total of 50 top most relevant articles, selected from the Scopus database, the detailed citation analysis was conducted. The study explored the overall theoretical foundation of social media research involving in learning and studying; and identified the leading sources of knowledge in terms of the most influential journals, authors, and papers; and revealed research trends over the last four years by citation analysis. The analysis of citation data showed that "International Journal of Management Education" is the leading journal in social media in learning and teaching research. Author "Abdullah Z" was found to be the leading author in this field in terms of a total number of publications, total citations, and h-index, while the most cited article was authored by Baaran S and by Bapitha L. The contribution of this study is to clearly outline the current state of knowledge regarding social media in learning and teaching services in the literature.

Perspectives on the Pedagogical Practice of High School Teachers in the State of Amazonas and its Implications

Aldemir M. de Oliveira, Floriano Augusto Veiga Viseu, Ana Paula Florencio Aires, Maria Teresa Bixirao Neto
Amazonas State Media Center - SEDUC AM/CEMEAM, Brazil;
Department of Integrated Studies in Literacy, Didactics and Supervision, University of Minho, CIED, Portugal;
Department of Mathematics, University of Tras-os-Montes and Alto Douro, CIDTFF, Portugal;
Department of Education and Psychology, University of Aveiro, CIDTFF, Portugal

Abstract: Teaching mathematics depends on several factors, especially on training. This assumption led us to investigate the relationship between mathematics teachers' perspectives on pedagogical practice and their training. We adopted a qualitative and interpretative approach, through interviews, in an attempt to understand the meanings that forty high school mathematics teachers in the state of Amazonas give to their pedagogical actions and their training. The results show the centrality of the teacher in the dynamic of classroom activities, which can be noticed by the resolution of exercises and the dependence on the textbook. The relationship between training and teaching practice shows the influence of initial training instead of continued training, presenting consequences for pedagogical practice.

Blockchain in Education: Design and Proposal of a Secure Management Model

Wilson Rojas, Victor Gayoso Martinez, Araceli Queiruga-Dios
University Of Salamanca, Spain;
Institute of Physical & IT (ITEFI), Spanish National Research Council (CSIC), Madrid, Spain

Abstract: Introduction of new technologies has been accepted and incorporated into our lives. This is how our way of working and communicating with each other change with the pervasive use of computers. Our way of acting, communicating, and working has also changed with the emergence of the Internet. Companies have made use of all these emerging technologies to improve their business models and to be more efficient and competitive. With the invention of Blockchain Technology, a similar journey to that of personal computers, mobile phones, internet and other technologies is expected. The financial sector has been the leading user of this technology which has impacted different fields such as services, industry, and economy, among others; whereas in the Education system its applicability is incipient. For this reason, it may be necessary to study the benefits and risks of this technology to

contribute to the transformation of the Education sector. The aim of this paper is to describe the development of this research on the application of Blockchain Technology in Education.

M-Learning based Initiative to Enhance Student Motivation in Computer Science

G. Ortega, L. O. Lopez, F. Orts, C. Diaz-Carrasco, J. J. Moreno, E. M. Garzon
Informatics Department, University of Almeria, Sacramento s/n, Almeria, Spain

Abstract: Due to the rise of remote university training, it is common for theoretical master classes to result in a monologue by the professor with low student participation. These students rarely use the available tools to participate during the class: instant messaging, microphone or "raise your hand" function. This situation is even worse when the ratio of students is high because the communication with all of them is more complicated and hence the detection of whether they are learning the explained concepts. This paper describes the teaching experience of incorporating a Mobile-Learning tool called Kahoot in a subject of the Computer Engineering degree. This tool allows the teacher to propose participatory activities in the online classes to reinforce learning and increase student involvement. The aim of this work has been twofold. Firstly, to explain the methodology carried out to motivate students to increase the participation during the master class, and, secondly, to check if students' final overall grades have improved with the use of this tool compared with the previous academic year.

Introducing Big Data Modules in a Traditional Database Course

Hong Wang, Robert Langenderfer
Department of Engineering Technology, University of Toledo, Toledo, Ohio, USA

Abstract: This paper introduces a sequence of modules developed to serve as an introductory course for a University data science minor. An existing database management system course is modified to meet big data requirements. In the original course, students are introduced to database design, implementation, and SQL query statements. The big data modules are introduced in the latter part of the course to teach basic data structures, data analysis, data visualization and machine learning in Python on Jupyter notebook. This method makes it easy for students to transition from SQL to Python data analytics because of the similarity in syntax. The course modules have been well received by students demonstrated by an increased number of who students have enrolled in the course to pursue their interest in data science.

CodeAPeel: An Integrated and Layered Learning Technology for Computer Architecture Courses

A. Yavuz Oruc, A. Atmaca, Yusuf Nevzat Sengun, Ali Semi Yenimol
Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland, USA;
Department of Computer Engineering, Bilkent University, Ankara, Turkey

Abstract: In this paper, we introduce a versatile, multilayered technology to help support teaching and learning computer architecture concepts. This technology, called CodeAPeel is already implemented in one particular form to describe instruction processing in assembly and machine layers by a comprehensive simulation of fetch-decode-execute process as well as animation of the behavior of CPU registers, RAM, VRAM, STACK memories, and various control registers of a generic instruction set architecture. Unlike most educational CPU simulators, CodeAPeel does not simulate a real processor such as MIPS or RISC-V, but it is designed and implemented as a generic RISC instruction set architecture with both scalar and vector operands, making it a dual architecture, supporting Flynn's SISD and SIMD instruction set architectures.

Lessons from an Online Multidisciplinary Undergraduate Summer Research Program

Matthias K. Gobbert, Jianwu Wang
Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, Maryland, USA;
Department of Information Systems, University of Maryland, Baltimore County, Baltimore, Maryland, USA

Abstract: During 2018, 2019, and 2020, the UMBC CyberTraining initiative "Big Data + HPC + Atmospheric Sciences" created an online team-based training program for advanced graduate students and junior researchers that trained a total of 58 participants. The year 2020 included 6 undergraduate students. Based on this experience, the authors created the summer undergraduate research program Online Interdisciplinary Big Data Analytics in Science and Engineering that will conduct 8-week online team-based undergraduate research programs (bigdatareu.umbc.edu) in the summers 2021, 2022, and 2023. Given the context of many institutions still being online and potentially expanding their online instruction in the wake of the COVID-19 pandemic, we share our experiences how the successful lessons from CyberTraining transfer to a high-intensity full-time online summer undergraduate research program.

Enabling Underprivileged Non-technical Women for Technology Careers: A Case Study

Bushra Anjum, Faiza Yousuf
CodeGirls Global Advisor, Santa Margarita, California, USA;
CodeGirls Co-Founder, Karachi, Pakistan

Abstract: The underrepresentation of women in computer science, especially those with lower socioeconomic status, remains a national issue in Pakistan. CodeGirls is the first of its kind not-for-profit initiative that focuses on women from underprivileged demographics, is fully funded by donations and sponsorships, has an active postgraduation placement strategy, and teaches a curriculum developed by global technical advisors and led by local industry experts. CodeGirls was launched in 2018, and over the past 3 years, the initiative has enabled 700+ graduates, 80+ placements, 67% retention rate, and a massive network of organizations supporting the cause. In this paper, we present CodeGirls as a case study and discuss various challenges faced and achievements made along the way. We discuss the rationale for the curriculum and its content, achieving financial stability, adding placement to our list of services, and retention strategies. We also highlight some of the careers enabled by the program. We hope that the unique approach and learnings of CodeGirls will inspire other advocates of computer science education for underprivileged women to adopt our successful model.

A Custom Web Service for Analyzing Athletic Performance Data in Collegiate Sports

Nathan Alden, Daniel Haugen, Andrzej Przybyla
Department of Computer Science, University of North Georgia, Dahlonega, Georgia, USA;
Department of Physical Therapy, University of North Georgia, Dahlonega, Georgia, USA

Abstract: With the growing influence of athletic performance data in collegiate sports, well-known sports analytics services require lofty annual subscriptions that might be too expensive for some university athletic programs. This research encompasses the development of a data analytics web service for automating the process flow of analyzing and generating data visualizations for collegiate athletics. Developed as a case study through the University of North Georgia's soccer teams, this web service provides a convenient interface for allowing student-athletes, coaches, and undergraduate researchers with custom data analytics and visualizations from the performance data collected during training sessions and matches. Lastly, this paper discusses the current state of the service and its utilization for aiding university athletic programs and undergraduate research.

Low Code/No Code Application Development and CS Education

Weidong Liao, Osman Guzide
Department of CS, Mathematics & Engineering, School of STEM, Shepherd University, Shepherdstown, West Virginia, USA

Abstract: The Low Code/No-Code (LCNC) movement has gained popularity in the software development market. In addition to companies specializing in LCNC software development such as Appian, OutSystems, and Mendix, key IT and cloud service providers, such as Amazon, Microsoft, and Google, have also released their LCNC development platforms. This concept and phenomena have brought us a series of questions to ponder: what is LCNC, essentially? What is its origin? How will the LCNC movement impact higher education to produce more "citizen developers"? How should we adjust to it if this is the future of application development? This article will investigate the LCNC movement and its impact on computer science education and higher education overall.

Using Interactive In-Class Activities to Teach Dynamic Programming

Saeed Al-Haj
Department of ECE & CS, Ohio Northern University, Ada, Ohio, USA

Abstract: Dynamic Programming is considered to be one of the difficult topics for undergrad students to understand in Data Structures and Algorithm courses. Previous studies have shown promising results in terms of the concept understanding and knowledge retention when active learning methods are used in the classroom. To overcome the challenges that students face in learning the dynamic programming topic, this paper presents several interactive in-class activities. These activities are designed to answer the major questions students may have during learning and applying dynamic programming to solve optimization problems. The interactive activities cover three dynamic programming problems in an engaging multi-staged learning environment. The problems are Fibonacci sequence, rod-cutting problem, and 0/1 Knapsack problem. Each problem is presented in a real world context to hookup students and get their attention to the new topic. The results show an improvement in student performance compared to the traditional teaching method.

Getting Lazy and Pure in Code Contests by Using Haskell

Chen Huo
Software Engineering, Shippensburg University of Pennsylvania, Shippensburg, Pennsylvania, USA

Abstract: Lazy purely functional languages, like Haskell, are never the first choices for code contests or competitive programming. We studied 107 problems from online code contest platforms, and found that Haskell users do not yet have standardized solutions for common situations in code contests due to the limitations of being lazy and pure. To name some, with side-effects prohibited (pure), it is tricky to do IO and write graph algorithms under time complexity requirements. To help laziness and purity reconcile with code contests, we derive an innovative collection of template solutions inspired by both the functional programming literature and actual user solutions from online code contest platforms. The collection will serve as an entry point for functional programming learners to code contests and a showcase of Haskell usage in this domain.

Renewable Energy Research and Capacity Building for STEM Recruitment and Reinforcement

Bahram Asiabanpour
Ingram School of Engineering, Texas State University, San Marcos, Texas, USA

Abstract: Re-Energize was a renewable energy research and education partnership program among six institutions, including Texas State University, San Antonio College, Coastal Bend College, Southwest Texas Junior College, Huston-Tillotson University, and University of Houston- Downtown, to reinforce the recruitment and retention of minority students in STEM Programs through Faculty Professional Development, Students STEM Awareness, and Infrastructure Development in each of these institutions. 22 faculty were involved. A total of 346 students, including 241 minority students, were supported in this program. Six STEM labs were developed/renovated under this program, 48 student presentations/publications/poster sessions were presented or published, and students received multiple awards in national competitions. This article discusses the program, assessments, successes, and lessons learned.

A Case Study of MOOC in Flipped Class for Non-Major Students

Jaekwang Kim
Sungkyunkwan University, Korea

Abstract: Contribution: This paper presents the results of a flipped class case study using MOOC in teaching Convolutional Neural Networks for Visual Recognition to non-computer students in the first semester of 2019 in Sungkyunkwan University. Background: There are three research questions such as “How can flipped class teaching using MOOC be designed?”, “What is the development of the Flipped Class teaching method using MOOC?”, “What are the results of developing Flipped Class teaching methods using MOOC?” Intended outcomes: A total of 12 students participated in the spring season in 2019, and under the guidance of one instructor, the MOOC in Flipped Class was based on the curriculum of cs231n: Convolutional Neural Networks for Visual Recognition, an online public lecture from Stanford University. Application design: Classes consisted of three parts: 1) instructor led offline preview lectures, 2) online video lectures from cs231n in YouTube, and 3) student-led QnA based offline complementary lectures. Findings: In the computer engineering domain, at least, flipped class with MOOC contents were found to be effective through this study. In addition, in the class designed like this, the instructor confirmed that the student was a mentor and played an important role as a leader.

An Integrated Approach to Learning Analytics

Uchendu Nwachukwu, Sandy Ingram, Juan Carlos Farah, Denis Gillet
College of Engineering and Architecture of Fribourg, Switzerland;
Ecole Polytechnique Federale de Lausanne, Switzerland

Abstract: Despite the rise in services generating learning analytics, there is a lack of standard models and guidelines for data integration and aggregation to support design choices of applications supporting learning analytics. We propose a bottom-up user-driven approach enabling educators to select, match, and contextualize activity traces from several data sources used in their teaching practice to perform and visualize meaningful learning analytics. To facilitate the process, the proposed approach recommends building customized auxiliary plugins that can be shared and re-purposed. We present the implementation of a use case following this approach, supporting the import and side-by-side comparison of activity traces from multiple data sources. Implications of this approach on cross-platform learning analytics and future work is discussed.

A Women Friendly Signal-and-Systems Curriculum Design and Implementatoin Based on Blended Teaching

Lichuan Liu
Department of Electrical Engineering, Northern Illinois University, Illinois, USA

Abstract: The paper presents a female-friendly engineering education in Signals and Systems at Northern Illinois University (NIU), more generally, revising the negative value traditionally placed upon engineering, especially electrical engineering achievement for women. Important tasks are as follows: (1) Design the signals and systems curricula in a female-friendly way. (2) Developing appropriate teaching strategies based on connected learning. (3) Offer teaching supplements for the enrolled students.

Shared Intentionality in Advanced Problem Based Learning: Deep Levels of Thinking in Coherent Intelligence

Igor Val Danilov, Sandra Mihailova, Iegor Reznikoff
Academic Center for Coherent Intelligence, ACCI, Rome, Italy;
Riga Stradins University, RSU, Riga, Latvia;
Department of Philosophy, University of Paris X - Nanterre, Paris, France

Abstract: Shared intentionality contributes to a subjectively distinct perception of sudden and unexpected insights when solving problems in the learning process. This ongoing research project supposes that shared intentionality relates to deep levels of thinking and memorizing. The research problem is to define a predictable mechanism for the occurrence of shared intentionality with the teacher in learning to facilitate cognition. We conducted two case studies with a 28-month-old child with no developmental disabilities and another with a 33-month-old child with epilepsy and cognitive delay. Young children successfully bypassed the important milestone of a numerosity ability in a short course at an age younger than others. The highlight of this short research paper is interaction elements that suit to ensure shared intentionality in learning.

Teaching Robotics and Artificial Intelligence for Kids by Developing a Virtual Escape Room

Francisco A. Pujol, Sergio Gonzalez, David Tomas, Pilar Arques, Mireia Sempere
Department of Computer Science and Technology, University of Alicante, Alicante, Spain;
Department of Software and Computing systems, University of Alicante, Alicante, Spain;
Department of Computer Science & Artificial Intelligence, University of Alicante, Alicante, Spain

Abstract: This is a scientific game in which participants were introduced to the world of robotics and programming. To do this, we proposed an online Escape Room. This game was carried out in groups using the Google Meet platform and consisted of solving a series of tests related to programming and robotics to find clues and passwords that allow access to the next room, to finish the course in the shortest possible time. In each of these phases, different aspects of robotics and programming were explained, showing videos, programming small applications, explaining topics of interest in the field of technology, etc. In this way, while the participants played, they learned basic concepts of programming and robotics, and completed different tasks with robots virtually while experiencing the excitement of an Escape room in an online format.

Stackable Cybersecurity Pathway Credentials Through Digital Badging

Guillermo Francia, III, Eman El-Sheikh, Tirthankar Ghosh
Center for Cybersecurity, University of West Florida, Florida, USA

Abstract: The critical need to fill the cybersecurity workforce gap is a pressing national issue that requires immediate action. Accelerating cybersecurity workforce development through reskilling and upskilling of workers can be a pragmatic means in resolving this issue. This action, supported by verifiable and digital records through a stackable credentialing system, presents a viable model of workforce shortage solution, not only for cybersecurity but also, for other careers as well. This paper is an exposition of our on-going project on cybersecurity workforce development through upskilling and reskilling efforts enhanced with digital and stackable credentials.

Undergraduate Coaches for Young Learners with Playful Computer Science

Jinsook Park, Mira Yun, Andrew Jung, Charlie Wiseman
Department of Mathematics, University of Hartford, West Hartford, Connecticut, USA;
School of Computing and Data Science, Wentworth Institution of Technology, Boston, Massachusetts, USA;
Department of Computing Sciences, University of Hartford, West Hartford, Connecticut, USA;
Department of Computer Science, Boston College, Massachusetts, USA

Abstract: Growing interest in computer science (CS) has led to a shortage of qualified educators for K-12 level instruction. Many outof-school-time computing programs are led by well-intentioned individuals with little formal training in education such as undergraduate students pursing computing degrees. This results in environments that are often not conducive to learning for students coming from diverse cultural and educational backgrounds. This work describes an approach to address some of these challenges by using undergraduates as coaches in K-12 outreach CS activities for underrepresented students. These coaches are given essential pedagogical training with playful material that allows them to succeed as educators.

Augmenting a Course on Ethics and Professional Responsibility in Computer Science and Software Engineering via Cloud Computing and Interdisciplinary Collaboration

Ranran Feng, Vidroha Debroy
Computer Science Department, The University of Texas at Dallas, Richardson, Texas, USA;
Public Cloud and Infrastructure, AT & T, Dallas, Texas, USA

Abstract: In the ongoing age of COVID, this paper discusses efforts to augment a course focused on Ethics and Professional Responsibility in Computer Science and Software Engineering. The course traditionally relies heavily on in-class discussions and inter-personal communication, which unfortunately were reduced because the modality of instruction needed to be converted to an online format. To counter this, efforts were made to rely on Cloud Credits – leveraging the Google Cloud for Higher Education program – to construct demonstrations/experiments such that students could utilize virtualized infrastructure (as opposed to resources available on campus) and also use credits for their own learning needs. This was realized through the collaboration between Full-time and Adjunct faculty and came at no cost to the University. The experiments are sustainable through other semesters and are also fit for purpose in other educational venues.

Challenges and Opportunities of Seeking ABET Accreditation for the First Time

Patricio Zapata, Marlene Roeckel, Ximena Garcia
Chemical Engineering Department, Universidad de Concepcion, Chile

Abstract: The Chemical Engineering Department (ChED) of the University of Concepción has presented its Chemical Engineering Program for an initial ABET-accreditation. The accreditation process requires adjusting the teaching process to new forms of assessment, control and evaluation. The axis of accreditation focuses on a process of continuous improvement. This work shows the adopted and developed strategy for the SOs assessment and evaluation, the implemented Continuous Improvement Model, and the challenges and opportunities of this process. The accreditation process means installing a procedure for continuous improvement in teaching, collecting key evidence about the type of teaching that is taught, and incorporating into the improvement plans the perception of graduates and employers, which finally allows teaching according to the needs of industry and society.

Evaluating Splitting Approaches in the Context of Student Dropout Prediction

Bruno de M. Barros, Hugo A. D. do Nascimento, Raphael Guedes, Sandro E. Monsueto
Federal University of Goias, Goiania, Brazil

Abstract: The prediction of academic dropout, with the aim of preventing it, is one of the current challenges of higher education institutions. Machine learning techniques are a great ally in this task. However, attention is needed in the way that academic data are used by such methods, so that it reflects the reality of the prediction problem under study and allows achieving good results. In this paper, we study strategies for splitting and using academic data in order to create training and testing sets. Through a conceptual analysis and experiments with data from a public higher education institution, we show that a random proportional data splitting, and even a simple temporal splitting are not suitable for dropout prediction. The study indicates that a temporal splitting combined with a time-based selection of the students' incremental academic histories leads to the best strategy for the problem in question.

Research Experience for Undergraduates in Applied Computing with Regional Research Themes

Dulal C. Kar, Longzhuang Li, Ahmed Mahdy
Texas A&M University - Corpus Christi, Texas, USA

Abstract: Texas A&M University-Corpus Christi, a Hispanic Serving Institution, offered an NSF sponsored REU (Research Experience for Undergraduates) program in applied computing research in marine sensor networks for three summers. Currently, the institution is an REU site for applied computing research in unmanned aerial systems which has completed two years of the REU project. Altogether over a period of five years in two sponsored REU projects, the project team has gained very useful experiential knowledge that can benefit many present and future REU sites sponsored by NSF to plan and implement their programs for success, particularly in a minority serving institution. In this work, we demonstrate how a unified research theme with regional research significance can be an effective way to engage undergraduate students in research and inspire and prepare them for career and graduate education in research. We also show how to form a group of research mentors with diverse research backgrounds by incorporating their research projects for REU participants under the same unified theme. Effectiveness of our programs are demonstrated through the survey results based on the questionnaire developed by an NSF CISE funded project at the University of North Carolina - Charlotte.

Student Perception of Autograding and Automated Feedback

Sonya Cates
Roger Williams University, Bristol, Rhode Island, USA

Abstract: Autograding and automatically generated feedback are commonly used in undergraduate computer science courses. Previous research has investigated methods for the creation of autograding tools that provide meaningful feedback to learners and noted potentially beneficial changes in student behavior in response to autograding. However, the question of how students, as users of these systems, feel about autograding is relatively unexplored. This work presents a study using Wizard of Oz methodology to explore student perception of and preferences toward autograding and automatically generated feedback on programming assignments. We find that students strongly prefer to receive feedback from their instructor, independent of the quality of the feedback. Results also suggest that some subgroups, such as students identifying as female, may have an even stronger preference for human over automated grading.

Collaborative Culture: Analyzing Global Trends in Computing Education

Maira Marques Samary, Stephanie Lunn, Alan Peterfraud
Boston College, Chestnut Hill, Massachusetts, USA;
Florida Intenational University, Miami, Florida, USA;
Sagefox Consulting Group, Amherst, Massachusetts, USA

Abstract: Computing education (CEd), or computer science education (CSEd), research has the potential to affect not only what and how we teach, but also who is taught and where. While CEd has grown as a discipline over the past two decades, many institutions still lack formal departments or programs. Given that it is a specialized and interdisciplinary area of research, we wanted to assess the values of collaboration, and access. To develop a better understanding of the researchers and institutions working in CEd, we manually collected publication data from the Innovation and Technology in Computer Science Education (ITiCSE) and the ACM International Computing Education Research (ICER) conferences, and the ACM Transactions on Computing Education journal, between 2015 and 2020. Using a collective total of 1099 publications, we analyzed affiliation information about the authors and their institutions. Although we hoped to uncover a global presence and collaborative relationships demonstrating a "CSEd for all" mindset, instead we found that North America and Europe were over-represented relative to other continents. Additionally, collaborations remained a national or regional affair, for the most part. While many factors may contribute, from language barriers to financial obstacles, communication across country lines needs to improve to truly develop a more equitable international presence in the field. Through this research, we hope to raise awareness of where CEd research is being conducted and what level of collaboration occurs between institutions and countries. Moreover, we want to encourage researchers to seek alternative perspectives and to expand their collaborations to ensure CEd work truly encompasses a broader worldview.

Teaching During a Pandemic: Observations and Reflections

Donald R. Schwartz
School of Computer Science and Mathematics, Marist College, Poughkeepsie, New York, USA

Abstract: As was true with most colleges and universities, the onset of the Covid-19 pandemic forced higher education (and the rest of society) to tackle a wide variety of challenges all at once, with almost no warning. Our abrupt transition from in-class instruction to online-only instruction, then to a variety of hybrid options presented many challenges. Our college offered three alternatives: Blended Synchronous, Flexible Hybrid or Online-only. This paper describes the challenges and successes of how I transitioned my in-class courses to Blended Synchronous courses in which half the class attends online while half the class attends in person, each group swapping every other class meeting.

Investigating Perceptions of NUS Lecturers on the Impact of COVID-19 on the use of Online Learning in Education: Its Benefits and Challenges

Tara Patu, Ioana Chan Mow
National University of Samoa, Samoa

Abstract: This paper documents the perceptions of lecturers from the National University of Samoa (NUS) with regards to the impact of COVID-19 on the use of online technologies in education in Samoa. The impact of the pandemic worldwide included effects on teaching and learning in Higher Education in Samoa. The research explored the perceptions of teachers about the benefits and challenges of using technology in teaching during COVID-19. The study's findings revealed that lecturers' perceptions were highly positive with high levels of motivation and satisfaction. Also evident was increased usage of technology to teach and deliver courses on online platforms. The findings also indicated both benefits and challenges of online learning as well as support a strong recommendation to continue using technology to provide learning in the event of further lockdown.

Toward a Digital University: Unicorn Learning Management System

George Feuerlicht, Marek Beranek, Vladimir Kovar
Unicorn University, Prague, Czech Republic

Abstract: Over the last decade most universities around the world have implemented some form of online learning delivery to supplement face-to-face teaching. The sudden lockdown in response to the COVID-19 pandemic acted as a catalyst forcing educational institutions to transfer all learning activities online. While online learning has well-documented advantages, it is also evident that this rapid transition is associated with challenges caused by poorly developed learning materials and a plethora of poorly integrated applications, tools and utilities that deliver a suboptimal environment for students, academic staff, administrators and management. In this paper we review the benefits and challenges of online learning and discuss the trend towards digitizing university environments. We argue that a critical pre-requisite for a successful transition to Digital University environment is a fully integrated technology platform that covers all aspects of university's operation. We describe the Unicorn Learning Management System (ULMS) platforms, developed at the Unicorn University in Prague and designed to provide uniform and integrated access for online learning and for all functions that support the day-to-day operation of the university.

Assessing the Class Performance of Online Programming Courses using Interactive Learning Platform with a Built-in IDE vs. using Adaptive Learning Platform with an External IDE

Yanzhen Qu, Richard Cai, Mazen Haj-Hussein, Howard Evans
Colorado Technical University, Colorado Springs, Colorado, USA

Abstract: In recent years, along with the fast development of online Computer Science (CS) and Information Technology (IT) degree programs, the online learning platforms used by online CS and IT courses have also experienced progressive changes. Due to the nature of CS & IT degree programs, three of the critical requirements for any online learning tool for online CS & IT degree programs are: an ability to adapt and customize the learning materials for each individual student; an ability to leverage a built-in Integrated Development Environment (IDE), which is a required tool for software development that enables students to effectively exercise hands-on programming skills; and an ability to allow students to discuss academic subjects in groups which promotes teamwork engagement. Among these three requirements, the third requirement is easily supported through an online "Discussion Board". While the first requirement is usually supported through an adaptive learning platform, and the second requirement can be realized through a newer generation of interactive learning technology. We have conducted a study in the online modality to compare the class performance of four programming courses between the classes using an adaptive learning platform with an external IDE and the classes using an interactive learning platform with a built-n IDE. The initial results of our project have consistently shown that the classes using an interactive learning platform with a built-

in IDE have yielded superior class performance over the classes using an adaptive learning platform with an external IDE.

JavaScript Frameworks

Jonathan Fuqua, Luay A. Wahsheh
Department of CSE, University of Tennessee at Chattanooga, Chattanooga, Tennessee, USA

Abstract: With modern web development becoming more complex over recent years, the number of libraries, frameworks, and tools seems to be increasing greatly, and as a web developer it can often be daunting to find a solution that suits the needs of a given project. In this paper, we identify the most useful tools that are available to the modern web developer and provide the scenarios, advantages, and disadvantages to using each given solution. We perform analysis on four popular front-end JavaScript frameworks: Angular, React, EmberJS, and Vue.js. Our analysis is introduced in this paper to assist future developers and prepare university graduates for industry to learn which frameworks are appropriate for any given situation or scenario.

On Leveraging Computing Connection in STEM Domains

Ray Kresman
Department of Computer Science, Bowling Green State University, Bowling Green, Ohio, USA

Abstract: Students in the STEM domain take classes in the major, but also take a few classes from other STEM disciplines as part of the graduation requirements. Assessment activities that cross discipline boundaries are desirable to help them prepare for the workforce or post-baccalaureate studies, In this paper we describe a variety of assessment activities suitable for an inter-disciplinary senior and/or graduate level course in the STEM domain. The activities are selected from different STEM disciplines to help students connect with ideas from other classes, but also build much needed (python-based) computational skillsets along the way. Our approach is somewhat unique for a couple of reasons: allow the student to step outside of their comfort zone and really see the problem from the client's/domain expert's perspective; add new experiences to the student's skillset, and at the same time help improve an area of expertise the student already has.

Incorporating the ARM Raspberry Pi into a Computer Architecture Course

C. Gerousis, C. Baker
Department of Physics, CSE, Christopher Newport University, Newport News, Virginia, USA

Abstract: The study of computer architecture, which focuses on the processor blueprints, hardware-software interface, balancing the efficiency, performance, cost, and reliability of a computer system has historically made the teaching of computer architecture to be mostly theoretical. The arrival of the ARM Raspberry Pi (RaPi) nearly a decade ago, is making it possible for students to supplement their computing skill and acquire a hands-on experience using a single-board computer. RaPi is a credit card-sized computer powered by the Broadcom BCM2835 system-on-a-chip (SoC), which contains many on-board ready to use features. One advantage of the RaPi is that its OS is open source and runs a suite of open source software. In addition, there is a plethora of Raspberry Pi's schematics that are regularly released as online documentation. In this paper, we discuss how the RaPi was introduced into the computer architecture course at CNU in conjunction with a description of the RaPi-laptop setup and connectivity with examples of student assignment projects. The purpose of projects is to enhance student learning of the ARM-based system architecture in a practical way.

Utilizing Modern Pedagogy for Teaching an Undergraduate OS Course

Jessy Ayala
Tandon School of Engineering, New York University, Brooklyn, New York, USA

Abstract: This paper explores a developed approach to teaching an undergraduate Operating Systems (OS) course based on modern findings of effective pedagogical techniques. More specifcally, how one would go about designing an OS course relying on approaches proven to engage students with material historically rooted in theory and typically approached in a stand-and-deliver method of instruction. The existing methods tackling this topic include utilizing hands-on assignments, project-based learning, and constructivism as a theory for teaching. This approach focuses on combining the aforementioned domains and acts as a survey for educators interested in creating an enriching experience for students enrolled in an OS course.

The 17th International Conference on Grid, Cloud, & Cluster Computing (GCC 2021)
https://www.american-cse.org/csce2021/conferences-GCC

Public, Private, and Hybrid Cloud Computing

David Yoon
CIS Department, University of Michigan - Dearborn, Dearborn, Michigan, USA

Abstract: We will present three types of cloud computing systems: Public, Private, and Hybrid using examples. E-commerce, home-monitoring system and data-deduplication system in cloud are presented as respective examples of the three types of cloud. The access keys will differentiate the three types.

Multi-Factor Authentication Framework for Cloud Computing

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

Abstract: Cloud Computing is an evolving technology which provide convenience for users and portability in which information can be exchanged in various platform as long as internet connection is available. The main issue concerning cloud is security of data. Security is word to refer to guarantee, preserve and restore the protection of data against several threats and attacks. Data security and privacy is a significant property of cloud security while it affects the organization trust and brand. Cloud security can be accomplished through authentication, authorization and access control which could provide an enhanced security for the cloud if they employed properly. The research work discuss the various researchers work regarding cloud security and their countermeasures. Moreover, to come up with an improved framework, the different attacks in the cloud system are discussed in order to provide data security. The proposed research work combines cryptography, authorization and authentication mechanisms in order to ensure access control.

Cloud Computing and Blockchain Adoption: A Systematic Review

Jacques Bou Abdo, Rayane El Sibai, Ahmed Bounfour
University of Nebraska at Kearney, Kearney, Nebraska, USA;
Al Maaref University, Beirut, Lebanon;
RITM, Universite Paris-Saclay, Sceaux, France

Abstract: The assessment and adoption of disruptive technologies is a complicated and risky task that requires the efforts of the technical and management teams at an organization. In this paper, we systematically survey the literature concerning the adoption of cloud computing and blockchain. We also compare the adoption trends and draw conclusions for other disruptive technologies.

Investigations of Micro-Benchmarks for Performance Profiling in Multi-Tenant Clouds

Hamidreza Moradi, Wei Wang, Dakai Zhu
University of Mississippi Medical Center, Mississippi, USA;
University of Texas at San Antonio, Texas, USA

Abstract: With the significant increase in the adoption of cloud computing, modeling and predicting the performance of the applications executing in clouds have gained major interests from cloud researchers and users. However, shared resource contention caused by multi-tenancy makes the performance modeling and prediction a challenging task. In this paper, we designed and evaluated a comprehensive set of micro-benchmarks to probe and profile the impacts of contention of key resources. Various machine-learning models have been exploited to model and predict cloud application performance with the profiling information of these micro-benchmarks. The results of our extensive experiments in both a private and the Chameleon clouds showed that the micro-benchmarks can effectively probe the contention levels of the resources, which enables the building of accurate performance prediction models for cloud applications with only 6.8% error on average. Moreover, to reduce profiling overhead, the results showed that the duration of 0.4 seconds execution for each micro-benchmark can achieve relatively accurate prediction even with only the CPU micro-benchmark, where the prediction error is about 8.5% on average.

Analysis of Machine Learning as a Service

Gokay Saldamli, Nishit A. Doshi, Vishal Gadapa, Jainish J. Parikh, Mihir M. Patel, Levent Ertaul
Computer Engineering Department, San Jose State University, San Jose, California, USA;
Department of Computer Science, California State University - East Bay, California, USA

Abstract: Machine Learning as a Service is a set of services that allows the Machine Learning models to run on the cloud using ready-made, easily configurable tools. The demand for MLaaS is increasing day by day as Machine Learning revolves around data processing, algorithms, and computational power, which further requires a highly skilled workforce. The market value of the industry is likely to grow from USD 1.0b to USD 8.48b within the next five years according to the study conducted by Modor Intelligence. Machine Learning as a Service offers services for data transformation, predictive analysis, data visualization, and advanced machine learning algorithms for the same. Cost-effectiveness and faster product delivery make MLaaS one of the most in-demand cloud-based services.MLaaS being arelatively new technology requires many factors to be considered such as the accuracy of the ML model, availability and cost of the service before choosing any cloud provider. Users get overwhelmed while selecting one cloud service provider over another because of the high volume of the vendors. Accordingly, there is a need for research to find the most suitable provider based on the requirements.In this paper, we propose a comparison of MLaaS provided by different cloud vendors to identify the most efficacious one. Our approach includes a thorough analysis on Natural language Processing APIs to draw conclusions on cost, time, accuracy, and ease of use of these service providers. Theoutcome will be better and cost-effective decision making forusers in a lesser amount of time.

Consideration of the Need for Quantum Grid Computing

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

Abstract: Quantum computing is poised to dramatically change the computational landscape, worldwide. Quantum computers can solve complex problems that are, at least in some cases, beyond the ability of even advanced future classical-style computers. In addition to being able to solve these classical computer-unsolvable problems, quantum computers have demonstrated a capability to solve some problems (such as prime factoring) much more efficiently than classical computing. This will create problems for encryption techniques, which depend on the difficulty of factoring for their security. Security, scientific, and other applications will require access to quantum computing resources to access their unique capabilities, speed and economic (aggregate computing time cost) benefits. Many scientific applications, as well as numerous other ones, use grid computing to provide benefits such as scalability and resource access. As these applications may benefit from quantum capabilities – and some future applications may require quantum capabilities – identifying how to integrate quantum computing systems into grid computing environments is critical. This paper discusses the benefits of grid-connected quantum computers and what is required to achieve this.

The 7th International Conference on Health Informatics & Medical Systems (HIMS 2021)
https://www.american-cse.org/csce2021/conferences-HIMS

Data Visualization for Knowledge Discovery and Sense-Making in Coronary Heart Diseases

Prajwal Koirala, Abeer Alsadoon, Mokhtar Mohammadi, Tarik A. Rashid, P.W.C. Prasad,
Thair Al-Dala'in , Nectar Costadopoulos, Angelika Maag
School of Computing and Mathematics, Charles Sturt University, Sydney, Australia;
Information Technology Department, Asia Pacific International College (APIC), Sydney, Australia;
School of Computing Engineering and Mathematics, Western Sydney University, Sydney City Campus, Australia

Abstract: The use of virtual tools has been less common in terms of knowledge discovery and making sense of data due to the limitations in the CART (classification and regression tree) algorithmic process and data collection. This is important to the field, particularly in the case of coronary heart disease. This paper focuses on the improvement of accuracy and processing time when using virtualization. The proposed system introduces Entropy and Decision Forest by penalizing attributes (EDFbPA). This decreases the impurity and improves the accuracy of the CART algorithm. It also improves the deep prediction by introducing the log function of Entropy and increasing the classifier's amount in the CART algorithm. The evaluation of the proposed hybrid model demonstrates that the new technique improves heart disease prediction accuracy and classification by 3 to 4 %. Furthermore, the algorithm's processing time was also improved compared to the ROC curve of the current best solution. The proposed algorithm focuses on providing better visualisation to improve identification of Coronary Heart Disease in patients.

Workflow Analysis for Supporting Breast Cancer Therapy Using Process Mining

Marie Bommersheim, Patrick Philipp, Jurgen Beyerer, Dirk Hempel
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany; Steinbeis-Transfer-Institut, Klinische Hamatoonkologie, Donauworth, Germany

Abstract: Modern medicine offers a variety of treatment approaches for breast cancer patients. Medical office software supports doctors in documenting patient-specific examination values and medication prescriptions. The aim of this work is to investigate how process mining can be used to automatically extract the course of therapy processes in breast cancer from existing data. For this purpose, existing patient data were adapted and process mining methods were applied and evaluated for homogeneous patient groups. Heuristic mining algorithms allow a clear presentation of complex therapy processes. In addition, the results of process discovery show that therapies carried out correspond to treatment structures described in medical guidelines: Patients with advanced breast cancer change their therapies several times during the course of treatment.

Optimizing EUD Model Parameters in Radiotherapy Planning

S. Puertas-Martin, J. J. Moreno, J. L. Redondo, P. M. Ortigosa, E. M. Garzon
Informatics Department, University of Almeria, Carretera Sacramento s/n, Almeria, Spain

Abstract: Intensity Modulated Radiotherapy (IMRT) is an effective cancer treatment for dose deposition in tumour tissues, while preserving surrounding healthy organs. Models based on the Estimated Uniform Dose (EUD) are able to generate feasible plans with excellent PTV coverage. However, manual tuning of the EUD parameters is required to find plans that fulfill the constraints specified by clinical professionals. In this work we propose a bi-objective optimization problem, solved with a genetic algorithm, with the objective of finding the set of parameters that generate the best plans for different criteria, using an HPC-accelerated EUD model for IMRT plan optimization.

Validating the Design of a Cloud-based Telehealth Service Cost Effective Solution Management System through a Quantitative Study to Use Cases

Ravi Vuppalapu, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract: Telehealth has been recognized as a possible and practical solution to address issues of rural area patients who do not have total access to a provider or primary care facility. However, Telehealth service systems are still rarely available due to high development cost and needed technical expertise of software design and development. To address these two challenges, we have conducted a research project in which we have demonstrated that in contrast with the traditional software engineering methodologies which have always conducted both verification test and validation test of the software system on the real implemented software, we have conducted the validation test against the system design of a Cloud based Telehealth Service Cost Effective Solution Management System (TSCESMS) at the design stage through applying data analytics against all the use cases of TSCESMS. In this paper, we have reported the details of how to identify all the use cases, and to utilize them to create validation tests and then applying the quantitative method to determine correlations between some key predicators in the use cases and the design success rate of TSCESMS. The result of our research project has provided a solid reference to a practical approach to develop a Telehealth software system costeffectively.

Effect of Progressively Available Features in the Clinical Timeline on Predicting 3-Days Readmission in an Emergency Department with Machine Learning Methods

Yi-Kuan Liu, Jia-Da Li, Tsung-Hung Hsieh, Cheng-Chung Fang, Chu-Song Chen
MOST Joint Research Center for AI Tech. & All Vista Healthcare, National Taiwan University, Taipei, Taiwan;
Institute of Information Science, Academia Sinica, Taipei, Taiwan;
Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan;
Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

Abstract: Readmission is a frequently used indicator for evaluating medical services quality, reflecting whether patients undergo appropriate clinical care while discharging from an emergency department. Previous works focus on readmission prediction close to or after discharge to decide further intervention and prevent unplanned revisit. However, such prediction could also assist discharge decision-making in the earlier stage for conservative clinicians and facilitate throughput. To this end, the predictive model should consider the character that features in the electric health records available progressively. Besides, a thorough feature importance analysis is essential for understanding benefits as features increasing gradually. This study adopted four powerful machine learning methods and performed sequential predictions with more and more input features to mimic practical scenarios. With Shapley Value for the feature importance analysis, we found features collected from triage and the last period before discharge may provide helpful information for the prediction task. In addition to clinical features, patient-specific non-clinical factors also make a considerable contribution to the prediction, which may be responsible for making this task more difficult. Finally, we suggest that the clinicians be aware of a trade-off between the prediction performance and the early usage, and recommend the prediction after triage for a better performance.

Blockchain Technology for Electronic Health Record (EHR) Management and Monitoring in Hospitals - an Insightful Perspective

Shaik Asif Hussain, Ahlam Salim Abdullah Al-Balushi
Middle East College, Muscat, Oman

Abstract: Background: Ever since the launch of Bitcoin in late 2009, blockchain technology has gained rapid popularity. The implementation of blockchain does not require a centralized authority but rather requires a distributed ledger among the blockchain architecture which adds a layer of security to the system. Objectives: The scope of blockchain is tremendous and huge. Through this research paper an attempt for systematic analysis of the relevant literature studied for implementation of Blockchain platform in the healthcare industry. The theme of the work is aligned to the existing literature on implementation of Electronic Health Record Management and monitoring in hospitals. The research work and their findings are discussed in brief to evaluate the recent blockchain technology platforms. Design: The literature was chosen from high ranked scientific conferences and journals. The glossary search terms included "Blockchain", "Implementation" and "Healthcare". The articles obtained from the databases considered with the following requirements: (1) Article aims to point out ways to implement blockchain. (2) Research Data on Hospitals and Clinics. The analysis results listed out the Benefits and Challenges of using blockchain technology in Healthcare. Methods: Over the last decades, several research papers have suggested ways of implementing blockchain in industries such as agriculture, E-currency, Cybersecurity, Banking etc. But in health care society there are very few papers which specifically speaks about Electronic Health record Management as it plays a prominent role to support patient, hospitals, and Health insurance companies. The literature findings have been carried out from different academic library databases such as EBSCO, ProQuest central and Omren Masader. The articles referred and studied are completely based on the empirical data. To further establish the quality criteria with the selected articles the following factors are considered as merit such as full text, Article structure and direction, influential, criticism and evaluation, Suitable evidence, and implications.
Analysis: In total 426 articles were identified and out of which 26 articles are chosen as they were complete with considered factors. Similarly, 22 articles are presenting the complete framework of Electronic health record management such as comprehensive classification of EHR, implementation architectural framework, research gap, exploratory analysis in future. Conclusion: Furthermore, to establish the feasibility in implementation of blockchain

the ratio of advantages to disadvantages of Blockchain in healthcare is conducted. The EHR systems strategies were evaluated and its performance factors were considered in its design and implementation.

Sleep Quality Analysis Based On Wi-Fi Received Signal Strength

Ho-Wai Wong, Po-Tung Zhang
Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR

Abstract: Sleep is a state of reduced activity and energy consumption, which helps restoration and memory processing. Our health and work productivity hinge on the quality of our sleep. Thus, the quality of sleep needs to be monitored. Professional sleep-tracking often involves special equipment, such as electroencephalography and electromyography sensors, which are costly and not generally accessible. There are cheaper wearable electronic devices for monitoring sleep, but they will involve contact with the user, consequently hindering their sleeping quality during the monitoring process. Therefore, we are investigating the possibility of using a nonintrusive, simple, and commonly accessible sleep quality monitoring system in the home. Sleep quality can be determined through the degree and frequency of body movement during our sleep. In this project, as Wi-Fi waves will be interfered or obstructed when passing through objects, Wi-Fi Received Signal Strength Information (RSSI) is used to analyze body movements during sleep. Changes in signal strength indicate the movement of the user. The signal is further processed by a sleep quality score system, which subsequently displays the result of sleeping quality. The monitoring system includes a Wi-Fi access point and a laptop with Wi-Fi. The access point will send radio waves, and the laptop will receive radio waves as well as signal strength information of the waves. By analyzing the signal strength information and using the above sleep quality scoring system, the corresponding sleep quality can be identified.

Personalized Question Answering on the Web using an Ontology

Amal Aljohani
Department of Computer Science, University of Idaho, Idaho, USA

Abstract: Leveraging the available semantic web technologies with common Knowledge Graphs such as DBpedia and Wikidata have provided great value in improving the quality of web search answers. Semantic web search improves the accuracy of the search and provides relevant results. Most often, people search their symptoms on the internet due to the availability of general health information on the web. Thus, there is a need of improving the quality of question answering over the web by personalizing user's search and leveraging the Personal Health Library (PHL). We are imagining a Personalized Ontological system that can answer patients' queries over the web in a human like manner within the context of the PHLs. Our goal is enabling users to obtain better answers that might positively affect their healthcare decision-making process. The proposed system consists of five stages that will be briefly described in this paper.

Application of Machine Learning in Early Recommendation of Cardiac Resynchronization Therapy

Brendan E. Odigwe, Francis G. Spinale, Homayoun Valafar
Computer Science & Engineering, University of South Carolina, Columbia, South Carolina, USA;
School of Medicine, University of South Carolina, Columbia, South Carolina, USA

Abstract: Heart failure (HF) is a leading cause of morbidity, mortality, and health care costs. Prolonged conduction through the myocardium can occur with HF, and a device-driven approach, termed cardiac resynchronization therapy (CRT), can improve left ventricular (LV) myocardial conduction patterns. While a functional benefit of CRT has been demonstrated, a large proportion of HF patients (30-50%) receiving CRT do not show sufficient improvement. Moreover, identifying HF patients that would benefit from CRT prospectively remains a clinical challenge. Accordingly, strategies to effectively predict those HF patients that would derive a functional benefit from CRT holds great medical and socioeconomic importance. Thus, we used machine learning methods of classifying HF patients,

namely Cluster Analysis, Decision Trees, and Artificial neural networks, to develop predictive models of individual outcomes following CRT. Clinical, functional, and biomarker data were collected in HF patients before and following CRT. A prospective 6-month endpoint of a reduction in LV volume was defined as a CRT response. Using this approach (418 responders, 412 non-responders), each with 56 parameters, we could classify HF patients based on their response to CRT with more than 95% success. We have demonstrated that using machine learning approaches can identify HF patients with a high probability of a positive CRT response (95% accuracy), and of equal importance, identify those HF patients that would not derive a functional benefit from CRT. Developing this approach into a clinical algorithm to assist in clinical decision-making regarding the use of CRT in HF patients would potentially improve outcomes and reduce health care costs.

Making Treatment Knowledge Available via 3D-models

Thomas Driebe
Department of Business and Economics, Philipps-University of Marburg, Germany

Abstract: This paper describes the development of an application to support surgical procedures and workforce development by utilizing WebGL and 3D-equipment. Improvement is obtained by codifying the knowledge of experienced clinicians through the use of 3D-models. Soft tissue filler treatments of the face serve as an example for applying this knowledge through the use of 3D-templates. The application aims to help novices to plan a treatment, validate a treatment model and improve the training and professional development of future clinicians.

Construction of Knowledge Graph of High Incidence Time, Area and People of Diseases

He Wang, Li Yao, Shaowei Cai, Honghai Feng
School of Computer and Information Engineering, Henan University, Kaifeng, P. R. China

Abstract: Information of high incidence time, area and people of diseases can be used in disease diagnosis. There are a large number of descriptions of the relationship between the high incidence information of diseases in the medical literature documents. However, manual handling them is time-consuming and laborious because of thousands of relational data. Using the method of relation extraction in NLP can quickly and accurately extract millions of relation data. This paper proposed an unsupervised algorithm for extracting high incidence information of diseases based on Chinese semantic rules, and establishes corresponding knowledge graphs. Through learning of 24 million web pages, information on the high incidence information of diseases of nearly 44237 diseases has been obtained.

Artificial Intelligence Use in Social Determinants of Health: A Comparative Study

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

Abstract: Social Determinants of Health analyzes the members environmental conditions based on geographical data, members economic conditions (income), type of work, age, access to care and type of care etc. to determine the overall life circumstances of the member. In this work, multiple publicly available datasets associated with Medicare enrollees, educational trends among students and U.S. chronic disease indicators were analyzed to understand the various dimensions of Social Determinants of Health. Based on the data, a linear regression model was implemented to predict the number of Medicare enrollees on a year-by-year basis through a regression line as well as a future county wise prediction model based on the past data. Finally, a recommendation for future work is discussed.

Single-cell RNAseq Analysis Reveals Diverse Expression Patterns of Cancer Survival Genes in Immune Cells

Wenjuan Zhang, John Talburt, Mary Qu Yang
MidSouth Bioinformatics Center & Joint Bioinformatics Graduate Program of University of Arkansas at Little Rock, USA;
University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA;
Department of Information Science, University of Arkansas at Little Rock, Little Rock, Arkansas, USA

Abstract: Immunotherapy has been used for treating certain types of cancer and improving patient survival. Breast cancer was initially considered as a non-immunogenic disease. Recently, several studies demonstrate the efficacy of immunotherapy in breast cancer treatment. Here, we combined breast cancer bulk and single-cell RNA sequencing data for investigating the expression alterations of survival-related genes in various immune cell types. We obtained RNA sequencing expression profiles of 100 normal and 934 breast cancer tissue samples from the Cancer Genome Atlas data portal. The patients were separated into two groups based on the expression change of a gene in a tumor sample compared to its mean expression level in the normal samples. We performed univariant survival analyses and then ranked the genes according to the resulting P-value. The top 266 genes were chosen from the sorted gene list as a survivalrelated gene set whereas the bottom 258 genes were a control set. We compiled expression profiles for naïve and central memory CD4+ T cells as well as naïve and effector memory CD8+ T cells from high-throughput single-cell RNA-seq datasets of matched normal and breast tumor samples. After filtering low quality cells and not-expressed genes, 49 out of 266 and 59 out of 258 genes remained for further analysis. Multiple survival-related genes were simultaneously differentially expressed in the CD4+ and CD8+ T cells. Our works help us to better understanding the interactions of tumor and immune systems and provide novel molecular prognostic markers for survival prediction in breast cancer patients.

Diagnosing COVID-19 Using Convolutional Neural Networks and Deep Learning Approaches

Raouf Eskandar, J. Zhong, Y. Hu
Department of Computer Science, California State University East Bay, Hayward, California, USA;
Evomics Medical LTD., Shanghai, China

Abstract: Coronavirus Disease 2019 (COVID-19) is a disease attacking human beings all over the world, which caused a great catastrophe globally. This infectious disease has spread rapidly and has affected over 170 million population globally and caused more than 3 million total deaths [1, 2]. Many research works have been developed since 2020 using various technologies to diagnose COVID19 infections with medical images based on the technology of deep learning (DL). Deep learning plays a successful role in giving fast and accurate results by processing medical images such as chest X-rays, CT, and PET as inputs, and utilizing DL algorithms to classify these images. This paper demonstrates several methods based on deep learning and convolutional neural networks (CNN) to detect COVID-19 infections using various datasets. Tested some of the existing methods to highlight the features of each, that might help to develop more successful algorithms and functions for further improvement in the field of disease diagnosis using medical images based on CNN and deep learning approaches.

A Recent Study on Cybersecurity in Healthcare

Casey Legeret, Bradley Ciuchna, Joel Mozo, Arianna Mills, Angelo Arteche, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA

Abstract: Over the last decade, cybersecurity in healthcare has increased. Cyber attacks are concerned for the health sector because their attacks can directly threaten not just the security of systems and information but also the health and safety of patients. According to a report from 2020, about 24% of U.S. healthcare employees received cyber security training. The same study found that 93% of healthcare organizations have experienced at least one data breach

over the past 3 years, and 57% had more than 5 within the same timeframe. With the way technology is advancing these numbers could get worse if not addressed properly. The purpose of this paper is to provide an outline on what is cybersecurity, healthcare, how would you improve cybersecurity, the risks of having poor cybersecurity, and finally why cybersecurity is important.

The 23rd International Conference on Artificial Intelligence (ICAI 2021)
https://www.american-cse.org/csce2021/conferences-ICAI

A Web-based Data Analysis Platform for Collaborative Decision Assistance

Christian Diaz Caez, Charles Kim
Department of Electrical Engineering & Computer Science, Howard University, Washington, DC, USA

Abstract: Political regimes and social status are always shifting in a region. It is one of the principal reasons why its analysis is needed to decipher for military or humanitarian missions for the region. More data from heterogeneous sources and locations may improve the mission decision-making, by which more accurate and less variant prediction of social unrest threat in a local area would be possible. This paper discusses a decision assistance system with a machine reasoning algorithm built on the information entropy minimum principle,and its implementation on a web-server to accommodate data upload from collaborators. The developed web-server has the capacity to add and update a database, and the updated database is supplied to the decision assistance algorithm for decision rule generation regarding social unrest threat or regime change prediction. The web-server based decision assistance system has been evaluated using publicly available polity data on regime change.

Least-Squares Estimation of Circuit Element Values from Measured Voltage and Current Waveforms

Charles Kim
Electrical Engineering and Computer Science, Howard University, Washington, DC, USA

Abstract: This article discusses an estimation approach for the values of passive elements such as resistance and inductance for a series inductive circuit using measured voltage and current waveforms at the source side. The proposed method finds the unknown two values using the single equation of the circuit by the least-squares approach with the pre-processed data. The preprocessing of the voltage and current waveforms includes concatenation of the discrete samples of voltage and current waveforms with one sampling period shifted. The approach works for steady-state sinusoidal waveforms as well as transients. The proposed approach can find the resistance and the inductance to the transient source in as short as 1/2 cycle length of discrete sample data. The proposed method is tested with waveforms generated from LTspice for a circuit of nominal 12 kV distribution circuit and with real transient waveform from a power distribution circuit. The evaluation result in both cases is good. Its good performance with transient waveforms is particularly significant in that, in such soft fault situations in utility distribution lines as self-clearing faults, the transient lasts just about 1 or 2 cycles before the system returns to the normal state as if nothing happens.

Streamline Characterization by Autoencoders and Unsigned Distance Field

Heyang Huang, Zhibin Huang, Tianhao Fa, Guangtao Fu, Min Yu
Beijing Key Lab of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing, P. R. China;
School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, P. R. China;
China Academy of Aerospace Aerodynamics, Beijing, P. R. China

Abstract: The characterization of streamlines is of great value for researchers to understand complex flow fields, traditional streamline characterization methods (e.g., using bag-of-words model) have defects such as large amount of calculation and nonend-to-end learning. Autoencoder is a neural network that can accurately extract feature information from high-dimensional data, which can achieve end-to-end and fast extraction of streamline features in complex flow field. However, existing works based on binary voxel can't extract feature information that has similarity which is not restricted by the streamline's original spatial location in the flow field. For this defect we propose a convolutional autoencoder based on unsigned distance field for streamline characterization.

A Spatio-Temporal Memory System for Multi-Modal Sensor Fusion

James A. Crowder, Michael J. Hirsch
Colorado Engineering Inc., Colorado, USA;
President & CTO, ISEA TEK LLC, USA

Abstract: Intelligence, surveillance and reconnaissance (ISR) systems require the ingest, correlation, and analysis of a variety of heterogeneous sensor data. Multi-sensor systems require advanced techniques to enhance their accuracy and efficacy, including artificial intelligence (AI) architectures that can process, analyze, and provide actionable intelligence, providing understanding and reporting on actions and events focused on Objects of Interest (EOI). This includes storing, correlating, and retrieving spatial and temporal tags and information associated with sensor readings. Here we present the initial design and testing of a Spatio-Temporal database system capable of storing and correlating complex spatiotemporal information that allows inferences to be made across both time and space (geography) to provide situational awareness to sensor processing analysts.

Forecasting the Extinction Risk for the Isle Royale Wolf Population using Discrete Event Simulation, Statistics and Neural Networks

Charles Eugene Knadler, Jr.
Independent Researcher, Alpine, Utah, USA

Abstract: In 2016 the Isle Royale National Park wolf population was reduced to a single nonbreeding pair. As a result, the United States Park Service made the decision to capture wolves in the United States and Canada and to release then on Isle Royale to recreate a viable wolf population, which would then maintain a healthy predator/prey relationship with the island's moose population. This study developed and evaluated discrete event simulations of the Park's ecology in order to forecast the extinction risk for the Isle Royal National Park wolf population, for each of four possible wolf reintroduction programs. Wolf extinction risk was found to be most sensitive to wolf pup initial survival probability and that four or more years of wolf reintroductions reduced extinction risk compared to three years of reintroduction.

Deep Learning-based Real-time Boat Detection and Localization

Ju Wang, Wookjin Choi, Jose Diaz, Curtrell Trott
Computer Science, Virginia State University, Virginia, USA

Abstract: This work investigates an ML object localization and tracking system that detects a boat and estimates its absolute geographic location (GPS coordinates) based on the RGB pixel values. Our processing pipeline utilizes a YOLOv5 neural network for target detection. A filtering module improves detection and tracking accuracy. We train a multimodal model to predict the world frame target location based on the 2D image-domain label. The filtering process removes false positive labels to improve tracking accuracy. The proposed multimodal takes advantage of the co-plane assumption of boats to calculate the 2D-3D mapping. A less than 10-meter error is observed for 90% of near-range and mid-range cases.

Avoiding Catastrophic Forgetting with Short-Term Memory During Continual Learning

James A. Reggia, Garrett E. Katz, Gregory P. Davis, Rodolphe Gentili
Department of Computer Science, University of Maryland, College Park, Maryland, USA;
Department of Kinesiology, University of Maryland, College Park, Maryland, USA;
Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, New York, USA

Abstract: When neural networks are used for continual (life-long) learning, catastrophic forgetting often occurs. Most recent past work on avoiding catastrophic forgetting has focused on long term memory and gradient descent learning. In contrast, here we consider short-term working memory and one-step training. In such models, catastrophic forgetting is especially severe, erasing not only old stored memory patterns but also causing a near-complete inability to learn new ones. We systematically examine the extent to which a recently proposed store-erase learning rule can ameliorate catastrophic forgetting over long time horizons during one step learning with models of short-term working memory. Our computational experiments reveal that the combined Hebbian and anti-Hebbian components of the store-erase learning rule, especially when combined with weight decay, is more effective than basic Hebbian learning alone or with weight decay.

Random Subspace Mixture Models for Interpretable Anomaly Detection

Cetin Savkli, Catherine Schwartz
Applied Physics Lab, Johns Hopkins University, Laurel, Maryland, USA;
Department of Mathematics, University of Maryland, College Park, Maryland, USA

Abstract: We present a new subspace-based method to construct probabilistic models for high-dimensional data and highlight its use in anomaly detection. The approach is based on a statistical estimation of probability density using densities of random subspaces with equal representation of each attribute combined with geometric averaging. Gaussian mixture models (GMMs) are used to obtain the probability densities for each subspace with techniques included to mitigate singularities allowing for the ability to handle both numerical and categorial attributes. The number of components for each GMM is determined automatically through Bayesian information criterion to prevent overfitting. The proposed algorithm obtains competitive AUC scores compared with prominent algorithms against benchmark anomaly detection datasets with the added benefits of being simple, scalable, and interpretable.

Under the Hood of Neural Networks: Characterizing Learned Representations by Functional Neuron Populations and Network Ablations

Richard Meyes, Constantin Waubert de Puiseau, Andres Posada-Moreno, Tobias Meisen
Institute of Information Management in Mechanical Engineering, Aachen, Germany;
Institute of Technologies and Management of Digital Transformation, Wuppertal, Germany

Abstract: The need for more transparency of the decisionmaking processes in artificial neural networks steadily increases driven by their applications in safety critical and ethically challenging domains such as autonomous driving or medical diagnostics. We address today's lack of transparency of neural networks and shed light on the roles of single neurons and groups of neurons within the network fulfilling a learned task. Inspired by research in the field of neuroscience, we characterize the learned representations by activation patterns and network ablations, revealing functional neuron populations that a) act jointly in response to specific stimuli or b) have similar impact on the network's performance after being ablated. We find that neither a neuron's magnitude or selectivity of activation, nor its impact on network performance are sufficient stand-alone indicators for its importance for the overall task. We argue that such indicators are essential for future advances in transfer learning and modern neuroscience.

How Do You Act? An Empirical Study to Understand Behavior of Deep Reinforcement Learning Agents

Richard Meyes, Moritz Schneider, Tobias Meisen
Technologies and Management of the Digital Transformation, Bergische Universitat Wupeprtal, Wuppertal, Germany

Abstract: The demand for more transparency of decision-making processes of deep reinforcement learning agents is greater than ever, due to their increased use in safety critical and ethically challenging domains such as autonomous driving. In this empirical study, we address this lack of transparency following an idea that is inspired by research in the field of neuroscience. We characterize the learned representations of an agent's policy network through its activation space and perform partial network ablations to compare the representations of the healthy and the intentionally damaged networks. We show that the healthy agent's behavior is characterized by a distinct correlation pattern between the network's layer activation and the performed actions during an episode and that network ablations, which cause a strong change of this pattern, lead to the agent failing its trained control task. Furthermore, the learned representation of the healthy agent is characterized by a distinct pattern in its activation space reflecting its different behavioral stages during an episode, which again, when distorted by network ablations, leads to the agent failing its trained control task. Concludingly, we argue in favor of a new perspective on artificial neural networks as objects of empirical investigations, just as biological neural systems in neuroscientific studies, paving the way towards a new standard of scientific falsifiability with respect to research on transparency and interpretability of artificial neural networks.

Integrated System for Large-scale Knowledge Graph Inferencing using Spark DataFrame

Min-Ho Lee, Min-Sung Kim, Wan-Gon Lee, Young-Tack Park
Computer Science Department, Soongsil University, Seoul, South Korea

Abstract: Studies on big data mass ontology inferencing methods have become commonplace. However, inference performance and processing speeds decrease with the increasing number of data. This study proposes a two-step integrated system for performing effective inferencing using the Spark dataframe in a cloud-computing environment. Axiom rules of the Horst web ontology language are inferred using an inference engine based on the Spark dataframe. Then, user-defined rules are inferred via the semantic web-rule language inference engine. An experiment is conducted using the Lehigh University Benchmark (LUBM) dataset to evaluate the proposed system's performance. The

developed system was ~6.3× faster than the existing inference system based on MapReduce and ~1.9× faster based on the Spark resilient distributed dataset for LUBM 1500 data.

Machine Learning for Surrogate Modeling in Ground Vehicle Mobility Analysis

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

Abstract: High-fidelity physics-based modeling and simulation tools are commonly used to perform analysis tasks in a variety of contexts. Because these tools are often computationally expensive, surrogate models can be employed to approximate metric values of interest without the computational burden of a full-fidelity simulation. In this paper, we examine the use of machine learning techniques to construct surrogate models, specifically within the context of US Army ground vehicle mobility analysis. We compare several machine learning algorithms and investigate the error rates of the resulting surrogate models for each. The results were very positive, with resulting average meansquared errors of under 3%.

Weighted Graph Nodes Clustering via Gumbel Softmax

Deepak Bhaskar Acharya, Huaming Zhang
Computer Science Department, The University of Alabama in Huntsville, Alabama, USA

Abstract: Graph is a ubiquitous data structure in data science widely applied in social networks, knowledge representation graphs, recommendation systems. When given a graph dataset consisting of one graph or more graphs, where the graphs are weighted in general, the first step is often to find clusters in the graphs. This paper presents some ongoing research results on graph clustering algorithms for clustering weighted graph datasets, which we name as Weighted Graph Node Clustering via Gumbel Softmax (WGCGS for short). We apply WGCGS on the Karate club weighted network, Madrid train bombing, Graph drawing contest, 08Blocks datasets. Our experiments demonstrate that WGCGS can efficiently and effectively find clusters in all the datasets used in our experiments. Our algorithm's effectiveness is shown by (1) comparing the clustering result obtained from our algorithm and the given labels of the dataset; and (2) comparing various metrics between our clustering algorithm and other state-of-the-art graph clustering algorithms.

Brain Dominance Level Prediction Using Hybrid Convolutional Neural Network-Long Short Term Memory

Zheng You Lim, Kok Swee Sim, Shing Chiang Tan
Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia;
Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia

Abstract: Brain dominance theory is often applied in the education field to discover the brain potential of a student. However, the current approaches used to determine brain dominance level are questionnaire-based assessments, and are not referred to biological information. Hence, we demonstrate a deep learning neural network namely the hybrid convolutional neural network-long short-term memory (CNN-LSTM) model to predict brain dominance level (in terms of percentage) from the resting-state electroencephalogram (EEG) signals. With the implementation of variational autoencoder (VAE) to generate additional training datasets, the CNN-LSTM outperforms the other benchmark neural network model: convolutional neural network, bidirectional recurrent neural network, bidirectional gated recurrent unit, and bidirectional long short term memory. The proposed CNN-LSTM achieves mean squared error as low as 0.0044 and means absolute error as low as 0.06. Besides, the proposed model also achieves the R-Squared value up to 0.92, which indicates that the predicted values are very close to the actual values. Hence, the results demonstrate this CNN-LSTM is reliable and has made a huge breakthrough contribution in the education field.

Generalization Over Different Cellular Automata Rules Learned by a Deep Feed-Forward Neural Network

Marcel Aach, Jens Henrik Goebbert, Jenia Jitsev
Juelich Supercomputing Centre, Forschungszentrum Juelich GmbH, Wilhelm-Johnen-Strasse, Juelich, Germany;
University of Cologne, Mathematical Institute, Cologne, Germany;
Helmholtz AI, Forschungszentrum Juelich GmbH, Juelich, Germany

Abstract: To test generalization ability of a class of deep neural networks, we randomly generate a large number of different rule sets for 2-D cellular automata (CA), based on John Conway's Game of Life. Using these rules, we compute several trajectories for each CA instance. A deep convolutional encoder-decoder network with short and long range skip connections is trained on various generated CA trajectories to predict the next CA state given its previous states. Results show that the network is able to learn the rules of various, complex cellular automata and generalize to unseen configurations. To some extent, the network shows generalization to rule sets and neighborhood sizes that were not seen during the training at all.

Airbnb Price Prediction with Sentiment Classification

Peilu Liu, Robert Chun
Computer Science Department, San Jose State University, San Jose, California, USA

Abstract: Airbnb is an online platform that provides arrangements for short-term local home renting services. It is a challenging task for the house owner to price a rental home and attract customers. Customers also need to evaluate the price based on the listing details. This paper demonstrates existing Airbnb price prediction models using machine learning and external data to make price predictions. The goal of this paper is to build a prediction model using machine learning and sentiment analysis techniques to help hosts and customers to adjust and evaluate the offered price. We classify sentiments into three types: positive, neutral, and negative. We observe that using classified sentiments with Regression Tree provides accurate prediction results compared with using numerical sentiment scores.

A Power Control Scheme for Artificial Intelligence and IoT (AIoT) ProSe-enabled Sensors

Weston Mwashita, Marcel O Odhiambo
Vaal University of Technology, Johannesburg, South Africa;
Mangosuthu University of Technology, Durban, South Africa

Abstract: The Internet of Things (IoT) and Artificial Intelligence (AI) are influential technologies. When Artificial Intelligence (AI) and Internet of Things (IoT) are combined, the result is AIoT, or Artificial Intelligence of Things. Artificial intelligence can be thought of as the brain of a system, while IoT devices can be thought of as the digital nervous system. The combination of AI and IoT allows systems to be predictive, prescriptive and autonomous. IoT sensors on one hand, allow for the collection of massive quantity of data, while AI on the other hand aids in the development of smarter applications for a smarter world. AIoT sensors are crucial in the Internet of Things. A power control approach involving two power control mechanisms is presented in this study: an open loop power control (OLPC) mechanism that can be used by a ProSe-enabled sensor to establish communication with a base station (BS) and a closed loop power control (CLPC) mechanism that can be used by a ProSe-Enabled sensor to establish communication with a base station (BS). Several researchers have proposed power control schemes to mitigate interference in mobile networks, but none has attempted to deal with the interference caused by ProSe-enabled sensors communicating with smart phones and 5G BSs. Simulation results showed that the proposed strategy performed well when compared to an optimal solution when it came to evaluating it. The proposed approach does not have a negative impact on the quality of service (QoS) of a 5G mobile network, according to a simulator created using the MATLAB high-level programming language.

Tagged Documents Co-Clustering

Gaelle Candel, David Naccache
Departement d'informatique de l'ENS, ENS, CNRS, PSL University Worldline Labs, Paris, France

Abstract: Tags are words or short sequences associated with multimedia, such as music, image, video, movie, blog post, patent, shoping item. They can be used to describe the content of a document. These tags are useful for human and machine information retrieval systems to access quickly a document or use recommender systems to suggest similar items. This short description allows it to guess if the resource would be relevant for a user. Due to the description shortness of a document, the match between the user query and the document's tags is often low. For example, to improve matching, tags can be clustered into conceptual groups containing synonyms and related concept to refine the query. In this paper, we propose a methodology to cluster tags into conceptual groups. As keywords follow Zipf's law distribution, data are preprocessed to remove power-law effects and enhance the context of low-frequency words. Then, a hierarchical agglomerative co-clustering algorithm is proposed to group together the most related tags into clusters. The capabilities were evaluated on a sparse synthetic dataset and a real-world tag collection associated with scientific papers. The task being unsupervised, we propose some ending criterion for an optimal partitioning.

Evaluating Chat-Type Conversation Systems: Knowledge-Based vs Machine Learning Hybrid

Kenta Nakamura, Seiji Tsuchiya, Hirokazu Watabe
Intelligent Information Engin., Graduate School of Science and Engineering, Doshisha University, Kyo-Tanabe in Kyoto, Japan;
Intelligent Information Engin., Faculty of Science and Engineering, Doshisha University, Kyo-Tanabe in Kyoto, Japan

Abstract: We compared and evaluated a chat-type conversation system that incorporates machine learning and a chat-type conversation system that does not incorporate machine learning. Recently, hybrid chat-type conversation systems that incorporate some machine learning into a knowledge-based chat-type conversation system has begun to be developed. Therefore, we evaluated the questionnaire after letting the participants use the knowledge-based chat system and the hybrid conversation system. The concepts to be measured were the appropriateness, the comfort, the naturalness, and the humanlike-ness of the conversation. As a result, it was evaluated that the hybrid chat type conversation system had more humanlike conversation than the knowledgebased conversation system. However, it was unclear how the incorporation of machine learning would improve the conversation system.

Dialogue Architecture for Personalized Carputer

Masato Tokuhisa, Shuhei Kimura
Tottori University, Tottori, Japan

Abstract: This paper addresses a personalized carputer that collects user's experience via dialogue to provide user-adapted services. Since user sometimes does not say user's experience to the carputer, the carputer needs to take dialogue initiative. Our proposed dialogue architecture contains user initiative process and system initiative one. The former precisely analyzes user's intention, fills slots and calls carputer's services. The latter arouses system's desire to launch dialogue. The carputer is an offline and low-spec computer, so we provide light algorithms for domain specific pos/vocabulary tagging, k-NN intention analysis, slot-filling by reinforce learning, and graph-based dialogue management with desire and emotional features. As the results of our experiments, we confirmed the system interviewed the user in the ideal way.

The Construction Method of Image Concept-Base and Verification of the Number of Attributes used in DoA

Misako Imono, Seiji Tsuchiya, Hirokazu Watabe
Department of Information Systems, Daido University, Nagoya, Japan;
Department of Intelligent Information Engineering and Science, Doshisha University, Kyotanabe, Kyoto, Japan

Abstract: In a Concept-Base, concepts are assigned attributes and weightings that express their importance. Concepts and attributes are defined based solely on the relationships that can be associated with each other. Using such definitions, a CB aims at including various meanings that human beings understand automatically based on words used, not simply definitions as described in dictionaries. This paper describes the construction method of Image Concept-Base defined by visual information called Image-Concept and Image-Attribute. In addition, the relationship between a image and a word called Degree of association (DoA) is calculated using the created Concept-Base and verify the number of attributes used when calculating DoA.

Assessing the Role of Demand Response Program in the Performance of Residential Energy Hub

H. A. Shayanfar, H. Shayeghi, L. Bagherzadeh
College of Technical & Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran;
Energy Management Research Center, University of Mohaghegh Ardabili, Ardabil, Iran

Abstract: This paper presents assessing the role of demand response program (DRP) on optimal energy management system of an extended residential energy hub (EH) containing solar photovoltaic (PV) generation, solar heat exchanger (SHE), and a battery energy storage system (BESS). Thus, a mathematical model is formulated based on daily load demand (such as electricity, heat, and cooling) and time-of-use (TOU) energy prices to improve energy efficiency and reduce operational cost. The results are in the form of four case studies to evaluate the effect of solar energy applications, BESS and DRP on the operational efficiency. The desired model has been solved using CPLEX solver in the GAMS program. The end results show that the proposed model leads to a significant reduction in energy costs. They also demonstrate that the model is appropriate for the characteristics of residential areas load.

Deep Learning for Uniformity and Texture Feature Extraction in Oral Cancer: Enhanced Scalar Probability of Number of Images

Ashok Bhatta, Abeer Alsadoon, P.W.C. Prasad, Indra Seher, Tarik A. Rashid,
Ahmed Dawoud, Mokhtar Mohammadi,
Sarkhel H.Taher Karim, Azheen Ghafour Mohammed, Omar Hisham Alsadoon, Oday D. Jerew
School of Computing and Mathematics, Charles Sturt University (CSU), Wagga Wagga, Australia;
School of Computer Data and Mathematical Sciences, University of Western Sydney (UWS), Sydney, Australia;
School of Information Technology, Southern Cross University (SCU), Sydney, Australia;
Crown Institute of Higher Education, Information Technology Department, Sydney, Australia;
Computer Science and Engineering, University of Kurdistan Hewler, Erbil, KRG, Iraq;
Department of Information Technology, Lebanese French University, Erbil, KRG, Iraq;
Computer Department, College of Science, University of Halabja, Halabja, KRG, Iraq

Abstract: Oral Squamous Cell Carcinoma (OSCC) is a common type of cancer of soft tissue in oral epithelium. However, deep learning has not been successfully implemented in the detection of OSCC due to the possibility of high classification errors. This research aims to increase the classification accuracy by automating the detection of OSCC in the early stages. The proposed system consists of a Convolutional Neural Network (CNN) to enhance the overall accuracy of classification for the detection by using Local Binary Pattern (LBP) and Grey Level Co-occurrence Matrix (GLCM). Confocal Laser Endomicroscopy (CLE) images are scaled down and then the noise is reduced up to 50% by half-size down-sampling. Then images are divided into patches with fifty percent overlap that enhances the

accuracy of classification. Patches are 2-fold augmented for training and testing of the neural network. This then feeds CNN for feature extraction through convolution layers followed by pooling layers followed by LPB and GLCM. The extracted features are passed to the Support Vector Machine (SVM). SVM classifier is implemented using Python, with a 2-fold cross-validation approach. Finally, the selected features are passed to classifiers. These features are combined during the fusion of patches using patch activity map and patch count map for classification of CLE Images. Accuracy is measured using a probability score, and the processing time is measured by measuring the runtime. The accuracy is found to be varying for different sources of CLE images. Based on the results, the proposed approach was found to provide a better solution to the problem. The proposed solution increased the average classification accuracy from 89.56 to 93.38 and decreased the average processing time from 438 milliseconds (ms) to 435 ms.

Finding Phishing and Fraudulent Websites using Logo Detection

Abhilash Garimella, Matin Hosseini, Abhishek Dubey, Shashi Prakash
Department of Research, Bolster, Inc., California, USA

Abstract: Finding phishing and fraudulent websites are one of the most critical problems on the Internet these days. In the covid situation, the number of phishing websites increased. In this paper, we proposed the use of logo detection to find the website's brand. This paper used dif erent versions of Yolo for logo detection algorithms to find the phishing and fraudulent websites. The current work is applied at checkphish.ai[1]. In this paper, we used our logo detection dataset containing 198 logos. We used all the recent versions of the Yolo, including yolov3, yolov4, and yolov5, and compared their performance in our customized logo detection dataset.

Quantum Circuits for Quantum Convolutions: A Quantum Convolutional Autoencoder

Javier Orduz-Ducuara, Pablo Rivas, Erich Baker
School of Engineering & Computer Science, Department of Computer Science, Baylor University, Texas, USA

Abstract: Quantum machine learning deals with leveraging quantum theory with classic machine learning algorithms. Current research efforts study the advantages, if any, of using quantum mechanics or quantum information theory to accelerate learning time, or convergence. Other efforts study data transformations in the quantum information space to evaluate robustness and performance boosts. This paper focuses on processing input data using randomized quantum circuits that act as quantum convolutions producing new representations that can be used in a convolutional network. Experimental results suggest that the performance is comparable to classic convolutional neural networks and in some instances, using quantum convolutions can accelerate convergence.

Intentions Prediction at Road Intersections using Ensemble Hidden Markov Models

Jian Yang, Christian Poellabauer, Pramita Mitra, Arpita Chand, Abhishek Sharma, Cynthia Neubecker
Department of Computer Science & Engineering, University of Notre Dame, Notre Dame, Indiana, USA;
Research and Advanced Engineering, Ford Motor Company, Dearborn, Michigan, USA

Abstract: Modeling the mobility of traffic participants is a crucial and challenging problem in urban planning, and predicting the movement intentions of traffic participants at road intersections has drawn increasing attention in recent years. In this paper, we propose an ensemble model for predicting such intentions, where, in order to accommodate different types of contextual information, we design an ensemble model that integrates three Hidden Markov models (HMMs) from three different aspects. The prediction of each HMM is integrated via linear combination. We further train and validate the proposed model with both synthetic and realworld high-resolution trajectory data, and our model shows an inference accuracy as high as 0.92. Further, the feature importance of each individual HMM has been tested and we provide insights for fine-tuning the model in different use cases.

Exploring Behavior via Neural Network Activations in Deep Reinforcement Learning Agents

Melanie Lu, Richard Meyes, Andres-Felipe Posada-Moreno, Tobias Meisen
Chair of Technologies and Management of Digital Transformation, University of Wuppertal, Germany;
Chair of Information Management in Mechanical Engineering, RWTH Aachen University, Aachen, Germany

Abstract: Transparency of learned behavior of deep reinforcement learning (DRL) agents is crucial for their applications in safety critical or ethically challenging domains such as autonomous driving, medical diagnostics, or industrial production scenarios. Especially the role of individual neurons for the successful execution of a given task is largely neglected in state-of-the-art applications. We propose an analytical approach to characterize learned behavior of DRL agents trained in a variety of continuous motor control environments and explore how individual neurons or groups of neurons of the trained networks contribute to the formation of key behavioral components, i.e., specific movements that are repeatedly performed during task execution. We show that groups of neurons contribute to specific behavioral components via specific activations during task execution, a phenomenon that has been well studied in motor cortex neurons of the mammalian brain. We demonstrate the universality of our results for different network topologies, weight initializations and control domains ultimately aiming to promote empirical study designs inspired by neuroscientific studies of the brain to facilitate transparency of learned representations in deep neural networks.

Data Implantation - Creating Minority Samples for Extremely Imbalanced Datasets

Francis Onodueze, Josyula Darsana, Josef Schaff
Computer Science Department, Bowie State University, Bowie, Maryland, USA;
Naval Air Systems Command (NAVAIR), USA

Abstract: Creating accurate machine learning models that can distinguish between normal and abnormal data on extremely imbalanced datasets that contain only a few or no minority instances is challenging. One-class classifiers that fit a model on the data that belongs to a single class and predicts whether new data belongs to that class or not are generally used in such scenarios. In this work, we present a novel method called Data Implantation for heavily imbalanced datasets or one-class problems. This paper introduces the concept of Data Implantation which creates synthetic minority samples from a general knowledge of the entire structure of the training set and what a deviation from a normal sample might look like. This novel approach opens up new horizons for algorithms that fail when tasked with classifying datasets that lack minority samples and for real life problems where there is very little or no minority samples available to learn from. This approach has shown to increase performance significantly in many algorithms.

Estimate and Predict Fire Weather Index Using ANFIS and ANN Algorithms

Ying Bai, Dali Wang
Johnson C. Smith University, Charlotte, North Carolina, USA;
Christopher Newport University, Newport News, Virginia, USA

Abstract: Due to the huge and tremendous damages made by forest fires and wild fires in recent years, to correct and accurate predict the possibility or rate of generation of forest fires becomes a crystal research topic. In this study, some artificial intelligent (AI) methods, such as an adaptive neuro fuzzy inference system (ANFIS) and an artificial neural network (ANN) algorithm are developed and used to correctly predict and estimate the possibility of generation of forest fires based on the fire weather index (FWI). These methods can be adopted by some decision makers to effectively estimate the dangerous degree of possible forest fires, and furthermore to issue some commands and processing steps to protect national and personal properties from being damaged or to reduce the degree of those damages.

Machine Learning in Geochemistry: A Bibliometric Analysis

Jodi D. Garcia, Adriana L. Garcia Garza, German H. Alferez, Benjamin L. Clausen, Ana Maria Martinez Ardila
School of Engineering and Technology, Universidad de Montemorelos, Mexico;
Geoscience Research Institute, USA;
Department of Earth and Biological Sciences, Loma Linda University, California, USA

Abstract: The use of machine learning in geochemistry has grown in the last 70 years. Although there are several publications written on machine learning applied to geochemistry, there is no existence of the analysis of the state of the art of these publications. Our contribution is the bibliometric analysis of 1,432 publications on the application of machine learning to geochemistry. The analysis was performed with Bibliometrix, an R-based tool, using information from Elsevier's bibliographic database Scopus. The results obtained were grouped into four sections based on scientific production, trending words and keywords, conceptual structure, and geographic results. Findings indicate a great interest in machine learning methods in geochemistry and also a lack of scientific production in some areas of the world.

Human-Centered Computing Based on Shared Intentionality in Human Cognition Model: A New Approach to Contactless Human-Computer System

Igor Val Danilov
Academic Center for Coherent Intelligence, ACCI, Rome, Italy

Abstract: Recent findings in physics and neuroscience challenge our knowledge about human cognition models and shed light on how social interaction shapes the mind. Shared intentionality is an essential quality of all organisms that contributes to cognition. This extended abstract presents neurobiological grounds for shared intentionality. It proposes several general arguments showing the quality of shared intentionality in different species. The abstract also notes recent findings in physics, concluding that living beings tend to quantum entangled state under specific conditions. This publication poses a research problem on whether a contactless interface between the brain and computer is possible without introducing artificial elements into the nervous system. Can a quantum harmonic oscillator induce quantum entanglement between neurons and a computer interface to provide the contactless management of the computer. Whether this cooperation provides a synergy of computation of the same task simultaneously. A valuable outcome of this further research about possible contactless brain-computer interaction may be the direct or indirect evidence of (i) contactless braincomputer interaction; (ii) synergy of this interaction. Notable, both sides of this interaction should not achieve this synergy outcome separately.

A Fog-Assisted Architecture for Urban Surveillance Using AI and Data Fusion

Arslan Munir, Jisu Kwon, Joonho Kong, Erik Blasch, Alexander Aved, Khan Muhammad
Kansas State University, Manhattan, Kansas, USA;
Kyungpook National University, Daegu, South Korea;
Air Force Office of Scientific Research (AFOSR), Arlington Virginia, USA;
Air Force Research Laboratory (AFRL), Rome, New York, USA;
Sejong University, Seoul, South Korea

Abstract: Urban surveillance provides situational awareness (SA) and timely response to emergencies, and safeguards national security. This article proposes a fog-assisted and artificial intelligence (AI)-empowered surveillance architecture that integrates Internet of things (IoT) devices with cloud servers via intermediary fog nodes for enabling realtime surveillance and monitoring. The article also proposes an AI-fusion framework for the surveillance architecture that alleviates the computational burden of the cloud by incorporating information/data fusion, and distributing AI processing, analytics and decision-making among all three levels of the proposed surveillance architecture, viz., IoT, fog, and the cloud. This article then experimentally evaluates the effectiveness of combining

data fusion with AI for surveillance by providing a latency and energy efficiency comparison between different deep learning models' processing with and without data fusion. Results indicate that combining AI with data fusion can provide a speedup of 9.7× over AI without data fusion while reducing energy consumption up to 88.5% over AI without data fusion.

Incorporating Abstract Behavioral Constraints in the Performance of Agent Tasks

Robert E. Wray, John E. Laird
Center for Integrated Cognition/IQMRI, Ann Arbor, Michigan, USA;
University of Michigan, Ann Arbor, Michigan, USA

Abstract: The behavior of autonomous agents needs to be conditioned by considerations that go beyond the functional requirements of executing a task. These considerations, which we term abstract behavioral constraints, range from operational and contextual constraints to conformance with norms (obligation) to ethical consequences of actions. These constraints are "abstract" because it is typically challenging to specify, at the level of task execution knowledge, how all of the constraints that bear on an agent should be applied in specific cases ("operationalization problem"). The paper reviews several categories of past approaches to encoding abstract constraints and discusses their strengths and limitations. The primary contribution of the paper is the identification of the "operationalization problem" and the specification of functional requirements that must be met by novel approaches to resolve it. Solutions to the operationalization problem will more readily allow autonomous agents to incorporate these constraints into their task knowledge to enable both efficient and reliable task execution.

Effectiveness of Proximal Policy Optimization Methods in Training a Neural Virtual Machine

Runxing Lin, Garrett E. Katz, James A. Reggia
Computer Science Department, University of Maryland, College Park, Maryland, USA;
Department of EECS, Syracuse University, Syracuse, New York, USA

Abstract: The Neural Virtual Machine (NVM) is a novel neurocomputational architecture designed to emulate the functionality of a traditional computer. A version of the NVM, called NVM-RL, supports reinforcement learning based on vanilla policy gradient methods as a mechanism for performing neural program induction. In this work, we trained NVM-RL using one of the most popular reinforcement learning algorithms, proximal policy optimization (PPO). Surprisingly, using PPO with the existing all-or-nothing reward function did not improve its effectiveness. However, we found that PPO did improve the performance of the existing NVM-RL if one instead used a reward function that grants partial credit for partially correct outputs. We conclude that, in some situations, PPO can improve the performance of reinforcement learning during program induction, but that this improvement is dependent on the quality of the reward function that is used.

A Deep Neural Network for Network Intrusion Detection System Based on VAE and RNN

Elie Fute Tagne, Audrey Tsotsop Nangue, Emmanuel Tonye
University of Buea, Buea, Cameroon; University of Dschang, Dschang, Cameroon;
University of Yaounde, Cameroon

Abstract: The emergence of a multitude of computer networks connected to the Internet has led to an exponential growth of the traffic circulating on the network over the last decades. This gigantic traffic attracts the interest of malicious attackers who constantly launch attacks (resulting in an abnormal flow circulating in the network) in order to damage these networks. Given the observed success of deep learning in various domains, we propose in this paper,

in response to the multiple attacks perpetrated in computer networks, a Network Intrusion Detection System (NIDS) using deep neural networks to distinguish normal and abnormal activities. To make our NIDS robust and efficient, we combine two neural networks: variational auto-encoder and recurrent neural networks. Our proposed approach is analyzed, tested, and then compared with some existing traditional and recent approaches on the publicly available NSL-KDD dataset for intrusion detection in anomalous networks. Evaluation metrics such as accuracy, false positive rate have been used to illustrate the satisfactory results obtained by our NIDS.

A Model-Agnostic SAT-based Approach for Symbolic Explanation Enumeration

Ryma Boumazouza, Fahima Cheikh-Alili, Bertrand Mazure, Karim Tabia
CRIL, Univ. Artois and CNRS, Lens, France

Abstract: In the literature dealing with explainable AI, we find either numerical approaches which quantify the contribution of each feature in a prediction or approaches providing certain forms of symbolic explanations such as Counterfactuals. In this work, we propose a generic agnostic approach allowing to generate different and complementary types of symbolic explanations. Our approach is based on the encoding of the model to be explained in an equivalent symbolic representation, this latter serves to generate in particular two types of symbolic explanations which are Sufficient Reasons and Counterfactuals. Our experimental results show the feasibility of the proposed approach and its effectiveness in providing Sufficient Reasons and Counterfactual explanations.

Source-Code Plagiarism Detection in C/C++ Language Programs using Neural Networks

Farhan Ijaz, Muhammad Arslan, Wasi Haider Butt, Ali Saleh Aziz Shah
National University of Sciences and Technology (NUST), Islamabad, Pakistan;
National University of Computer and Emerging Sciences (FAST), Islamabad, Pakistan

Abstract: Source code plagiarism is defined as an act of stealing/copying a whole or a part of someone else's code and presenting it as one's own. This dishonestly is mostly seen in academics where students copy their home assignments from other students or the internet to get good grades. Copying code from another source without understanding it is harmful. By considering this problem, code plagiarism detection appears as an active area of research. There exist many tools and techniques to detect source code plagiarism in different programming languages, but most of them lack accuracy or time efficiency while testing on a real-world standard dataset. Also, most of the techniques discussed in the literature are unable to detect the difficult type of plagiarism "logical changes" or type-5 plagiarism. This research presents a neural-network-based hybrid approach to overcome these issues by detecting similarity as well as plagiarism in C and C++ source code using intermediate representation (IR). The proposed framework uses an intermediate representation to generate static code features which are further use in a pairwise manner to train supervised-learning based Multilayer Perceptron classifier. The devised approach is compared to a recent state-of-the-art deep-learning-based tool called FC-detector over the OJClones labelled dataset which includes C and C++ program files. Our approach shows better results with a 7x faster prediction time and 97% f1-score. Further tested on plagiarism dataset and results show that our system is capable of identifying multiple types of plagiarism.

Entropy-Based Algorithm of Transfer Learning and Genetic Algorithm

Tami Alghamdi, Robert B. Heckendorn, Terence Soule
Department of Computer Science, University of Idaho, Moscow, Idaho, USA

Abstract: Transfer Learning (TL) is a Machine Learning strategy that employs previously gained knowledge in solving one problem to solve another similar problem or task. We created a model of TL using Genetic Algorithms (GA) to better understand TL. A Genetic Algorithm is a stochastic optimization method that searches and evolves candidates in a known environment looking for the best candidates. We proposed an algorithm that used the concept

of entropy (H) to discover, maintain, and transfer known knowledge of a solved problem to solving another problem. The results demonstrate that our proposed algorithm is more efficient than some other TL strategies.

Toward Contextual AI/ML

Robert Pilgrim, Andrew Bevilacqua
Bevilacqua Research Corporation, Huntsville, Alabama, USA

Abstract: This paper presents a road-tested, pragmatic approach to verifiable Artificial Intelligence (AI) and Machine Learning (ML) based on a consensus among multiple agents. It addresses the need for more precise monitoring of information exchange at the boundaries between the real-world and the digital domain. It suggests the importance for a willingness to share and discuss shortfalls in order for developers to ensure the deployment of safe autonomous AI/ML applications.

Sentiment Analysis and Text Summarization Across Different Domains of Data

Poornima Narendra Sapkal, Hans-Peter Bischof
Department of Computer Science, Rochester Institute of Technology, Rochester, New York, USA

Abstract: This paper is about Natural Language Processing techniques like Sentiment Analysis and Text Summarization. It sheds some light on the important concepts surrounding these topics while evaluating the existing methodologies. The Random Forest and Na¨ıve Bayes algorithms for Sentiment Analysis are explored to understand how the performance of the model is affected when they're subjected to different feature selection techniques and hyper-parameters. The Grid Search and Random Grid Search algorithms are used to tune the hyper-parameters for the Random Forest model, and feature selection techniques like TF-IDF, Bag of Words, and Ngrams are applied. For Text Summarization, the LexRank, LSA and TextRank algorithms are subjected to the same text corpus and their resulting summaries are evaluated using metrics like ROUGE.

Opinion Dynamics on a Dual Network of Neighbor Relations and Society as a Whole using the Trust-Distrust Model

Nozomi Okano, Akira Ishii
Tottori University, Koyama, Tottori, Japan;
Center for Computational Social Science, Kobe University, Kobe, Japan

Abstract: In society, people form networks through daily connections and SNS connections. In this study, we propose to divide the network into a network where the entire society is connected through SNS, etc., and a network of dense connections within a narrow range such as daily life, workplace, and friendships. We use the Trust-Distrust model, which deals with both trust and mistrust among people, to consider opinion dynamics in this dual network structure.

The Evaluation of Rating Systems in Team-based Battle Royale Games

Arman Dehpanah, Muheeb Faizan Ghori, Jonathan Gemmell, Bamshad Mobasher
School of Computing, DePaul University, Chicago, Illinois, USA

Abstract: Online competitive games have become a mainstream entertainment platform. To create a fair and exciting experience, these games use rating systems to match players with similar skills. While there has been an increasing amount of research on improving the performance of these systems, less attention has been paid to how their performance is evaluated. In this paper, we explore the utility of several metrics for evaluating three popular rating systems on a real-world dataset of over 25,000 team battle royale matches. Our results suggest considerable differences

in their evaluation patterns. Some metrics were highly impacted by the inclusion of new players. Many could not capture the real differences between certain groups of players. Among all metrics studied, normalized discounted cumulative gain (NDCG) demonstrated more reliable performance and more flexibility. It alleviated most of the challenges faced by the other metrics while adding the freedom to adjust the focus of the evaluations on different groups of players.

Automated Materials Spectroscopy Analysis Using Genetic Algorithms

Miu Lun Lau, Min Long, Jeff Terry
Department of Computer Science, Boise State University, Boise, Idaho, USA;
Department of Physics, Illinois Institute of Technology, Chicago, Illinois, USA

Abstract: We introduce a Genetic Algorithm (GA) based, open-source software to solve a multi-objective optimization problem to automate various types of analysis of materials characterization data including EXAFS, XPS and nanoindentation. The modular design makes the software extensible for other applications. This automation of the analysis is crucial in the era when instrumentation acquires data orders of magnitude more rapidly than it can be analyzed by hand. We provided in the software multiple crossover and mutation methods. Our results demonstrated good fitness scores with minimal human intervention.

A Review on Machine Learning Algorithms for Dust Aerosol Detection using Satellite Data

Nurul Rafi, Pablo Rivas
School of Engineering and Computer Science, Department of Computer Science, Baylor University, Texas, USA

Abstract: Dust storms are associated with certain respiratory illnesses across different areas in the world. Researchers have devoted time and resources to study the elements surrounding dust storm phenomena. This paper reviews the efforts of those who have investigated dust aerosols using sensors onboard of satellites using machine learning-based approaches. We have reviewed the most common issues revolving dust aerosol modeling using different datasets and different sensors from a historical perspective. Our findings suggest that multi-spectral approaches based on linear and non-linear combinations of spectral bands are some of the most successful for visualization and quantitative analysis; however, when researchers have leveraged machine learning, performance has been improved and new opportunities to solve unique problems arise.

Artificial Intelligence Applied to Dairy Cattle Anomalies Detection: A Systematic Review

Antonio Jorge de P. N. Cassimiro, Guilherme Margato Vital, Diogo Oliveira, Eduardo Noronha de Andrade Freitas
Federal Institute of Goias, (IFG), Goiania-Go, Brazil;
College of Communication & Information (CCI), Florida State University, Florida, USA

Abstract: Context: Anomaly detection in dairy cattle with artificial intelligence (AI), aims to increase the efficiency of production systems in dairy farming, with many studies indicating that different AI techniques can be successful in the field. However, the best techniques and features currently used to achieve a high anomaly detection rate are not in an well organized way. Objective: The objective of this study is to evaluate and classify existing AI techniques that detect anomalies in dairy cattle. Method: We developed a search protocol on different databases, using the same keywords to perform a systematic literature review and evaluation of empirical studies using relevant criteria. In order to calibrate the efficiency of our protocol, target articles were expected in the search. Result: From 1408 articles that resulted from the search, we identified 94 articles among them that used AI techniques for anomaly detection in dairy cattle. From that, it was possible to identify the most accurate AI techniques for anomaly detection from those articles. Conclusion: Given the many possibilities for anomaly detection in dairy cattle, one can perceive that the best AI

techniques and that several techniques are still prone to further investigation, even though there are some combinations of features and techniques that are potentially better than others.

Swarm Intelligence and UAV Security

Natalia S. Martinez Ojeda, Jeffrey L. Duffany
Electrical and Computer Engineering, Ana G. Mendez University, Gurabo, Puerto Rico;
Polytechnic University of Puerto Rico, San Juan, Puerto Rico

Abstract: This investigation aims to explore swarm intelligence exhibited by both animals and plants to see if there are any behaviors that can be applied to the security of Unmanned Aerial Vehicles (UAVs). Animals and plants display intelligence in their social structures, communication methods and antipredation strategies. Ants form colonies with complex division of labor strategies to find nourishment. Fish form schools to confuse predators. The anti-predation strategies found in redcrowned and common cranes, wild guinea pigs, lizards, and wind-dispersed trees are security mechanisms that each species has evolved to improve their odds of survival and it is possible that some of these strategies may be adaptable to UAV security.

Feature Optimization for Machine Learning

Jeffrey L. Duffany, Carla A. Varela-Rosa
Electrical and Computer Engineering, Ana G. Mendez University, Gurabo, Puerto Rico;
Polytechnic University of Puerto Rico, San Juan, Puerto Rico

Abstract: In the past decade machine learning has given us self-driving cars, practical speech recognition, effective web search, email spam filters and many other things. Machine learning has evolved due to increasing availability of a variety of different types of sensor data. Examples of sensors include smoke detectors, motion sensors and contact sensors. Hardware is becoming smaller and sensors are getting cheaper, making sensor data available for a variety of applications ranging from predictive maintenance to behavior monitoring. One of the most difficult challenges is how to choose an optimal set of features from the sensor data. We investigate the topic of optimal feature selection using a Human Activity Recognition dataset and various classifiers such as the support vector and the k-nearest neighbors.

Visual Response to Emotional State of User Interaction

Nina Marhamati, Sena Clara Creston
Computer Science Department, Sonoma State University, Rohnert Park, California, USA;
Department of Art, Sonoma State University, Rohnert Park, California, USA

Abstract: This work proposes an interactive art installation “Mood spRing” designed to reflect the mood of the environment. Through interpretation of language and tone, Mood spRing controls an immersive 3D animation of the seasons. An AI program interprets the language and tone of the users: if perceives them as pleasant, the animation progresses through idealized renditions of seasons. Otherwise, the animation slips into unpleasant weather and natural disasters of the season. To interpret the language and tone of the user interaction, hybrid state-of-the-art emotion detection methods are applied to user audio and text inputs. The emotional states detected separately from tone and language are fused by a novel approach that focuses on minimizing the possible model disparity across diverse demographic groups.

Detection of Heart Rate Variability based on Restricted Boltzmann Machines

Shijun Tang
Department of Science and Mathematics, Alvernia University, Reading, Pennsylvania, USA

Abstract: In this paper, a novel ECG detection algorithm based on Restricted Boltzmann Machines (RBM) and feature extraction was proposed to detect normal and abnormal ECG signals. RR interval, half width of R peak as well as their distributions have been taken as the features of ECG signals. Then, the RBM network with feature extraction are trained to classify the imbalanced ECG signals. This research shows how to detect heart rate variability and classify electrocardiogram (ECG) data from the PhysioNet using Restricted Boltzmann Machines. The experimental results demonstrate that the proposed method achieves a good classification performance. The proposed method can be applied to assist cardiologists into more accurately and objectively diagnosing ECG signals.

Improving Data Center Peak-Shaving with Deep Reinforcement Learning

Aashish Adhikari, Lizhong Chen
School of EECS, Oregon State University, Oregon, USA

Abstract: Increasing reliance on cloud services has led to a rapid growth of data center usage. As a consequence, data centers pay a high peak demand cost in millions of US $. Several techniques like task migration, server request cancellation, dynamic programming for energy arbitration between sources, etc. have been proposed to reduce total energy consumption cost that do not address peak-shaving. Prior works on peak shaving at data centers, grids, and smart homes also suffer from weak assumptions and cost formulations - they assume knowledge of future demands or make an attempt to predict them, use a workload that does not resemble a data center, optimize an unrealistic smooth and dense reward function, or learn non-generalizable closed-form solution. We use Google's data center's cluster resource usage trace to introduce a Markov Decision Process formulation of a data center. We use deep reinforcement learning to learn a dynamic policy that works to optimize a non-smooth and sparse but practical objective of a data center in realtime. We also explain the challenges in learning a policy in presence of a non-smooth and sparse reward function that is not present in prior literature for peak-shaving in data centers, grids, and smart homes.

A Data-Driven Approach to Efficient Library Management: Predicting Library Checkouts using Machine Learning

Nikesh Kumar Pahuja, Maria H. Rivero
Harrisburg University School of Science and Technology, Harrisburg, PA, USA

Abstract: Library circulation facilitates the lending activities, including checkouts and renewals by patrons. This service consumes more extensive Library resources and is an area that could be improved by applying Machine Learning techniques. This study uses the Integrated Library Systems dataset from the Office of Chief Data Office City and County of San Francisco. This study builds a Machine Learning model to predict the total checkouts by patrons in public libraries in San Francisco, California, USA. The study used fourteen predictors to train several Machine Learning models to predict the checkouts. The study trained and evaluated several Machine Learning models using 10-fold Cross-Validation (10-fold CV): Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), Lasso Regression (LSR), Neural Network (NN), Support Vector Regression (SVR) and XGBoost (XGB). The model performance was evaluated using the R Square metric. The best predictive algorithm was the XGBoost model with an R Square score of 0.72 and std dev of 0.002. This study and the predictive model will allow libraries to predict the number of checkouts and renewals a patron will do in the future. It will allow libraries to manage their resources efficiently. The methods and models developed in this study can also be applied by other libraries on their data to manage their operations efficiently.

Examining the Gap Between Pseudo-Real and Real Images with CycleGAN

David Schweitzer, Jing-Ru C. Cheng, Justin T. Carrillo, Luis De Jesus-Diaz
Information Technology Lab. (ITL), Engineer Research and Development Center (ERDC), U.S. Army Corps of Engineers, Vicksburg, Mississippi, USA;
ERDC Geotechnical and Structures Lab., U.S. Army Corps of Engineers, Vicksburg, Mississippi, USA

Abstract: It is well known that object detectors trained on synthetic data do not perform well on real data. However, in many situations, real data may not exist or may be too small to train effective models on. Generative Adversarial Networks (GANs) are rapidly gaining traction in image analytics for the purposes of translating images from one domain to another, such as synthetic to real. However, image similarity and GAN effectiveness are largely subjective or qualitative approaches performed by an end user, introducing biases by simply viewing the results and comparing them to the desired outcomes. In trying to generate datasets where data is very small or nonexistent, there is a critical need to be able to objectively evaluate GANs' efficacies, especially if datasets are meant for object detection. Eventually, the question to answer is then can models trained on translated "pseudo-real" images perform well on real images provided the gap between pseudo-real and real is small? To begin answering such a question, this work considers some metrics that are used to compare image similarities in order to evaluate the overall effectiveness of CycleGAN, the GAN used for this effort. The gap between synthetic images and translated pseudo-real images is measured to demonstrate that CycleGAN is making significant changes in the image domain translation from synthetic to pseudo-real, and the gap between pseudo-real and real images is measured to demonstrate that CycleGAN is creating images that closely approximate real images.

Mind Readers and Mind Users: The Utility of Sharing Architectural Components Across Multiple Robots

Matthias Scheutz
Tufts University, Medford, Massachusetts, USA

Abstract: We introduce the concept of architectural "component-sharing" as the basis for "knowledge sharing" in hive minds, e.g., a system multiple robots connecting by shared components in their control architectures. We discuss the architectural requirements and demonstrate the utility for multi-robot instruction and automatic reasoning across robotic platforms with two examples of natural language human-robot interactions with mind-sharing robots. Finally, we discuss some of the challenges of making instructing hive minds intuitive for humans and point to questions that need to be addressed in the future.

Testing the Performance of Multi-class IDS Public Dataset using Supervised Machine Learning Algorithms

Vusumuzi Malele, Topside E. Mathonsi
Department of Information Technology, Tshwane University of Technology Pretoria, South Africa

Abstract: Machine learning, statistical-based, and knowledge-based methods are often used to implement an Anomaly-based Intrusion Detection System which is software that helps in detecting malicious and undesired activities in the network primarily through the Internet. Machine learning comprises Supervised, Semi-Supervised, and Unsupervised Learning algorithms. Supervised machine learning uses a trained label dataset. This paper uses four supervised learning algorithms Random Forest, XGBoost, K-Nearest Neighbours, and Artificial Neural Network to test the performance of the public dataset. Based on the prediction accuracy rate, the results show that Random Forest performs better on multi-class Intrusion Detection System, followed by XGBoost, K-Nearest Neighbours respective, provided prediction accuracy is taken into perspective. Otherwise, K-Nearest Neighbours was the best performer considering the time of training as the metric. It concludes that Random Forest is the best-supervised machine learning for Intrusion Detection System.

A Distributed Scheme for Accelerating and Scaling PSPNet

Ajay Krishna, Xiang Gu, Andrew Larkin, Perry Hystad, Lizhong Chen
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon, USA;
College of Public Health and Human Sciences, Oregon State University, Corvallis, Oregon, USA

Abstract: Advances in deep learning based image processing techniques have led to their adoption for a wide range of applications, and in tow with these developments is a dramatic increase in the availability of high quality datasets. With this comes the need to accelerate and scale deep learning applications in order to keep up with the size of data and best use available hardware resources. In this paper, we propose a useful scheme to enable accelerated and distributed deployment of PSPNet inferences, which allows scaling across disparate compute clusters with recourse for system instability and heterogeneity of resources. The systemic incorporation of multiprocessing paradigms and pipelining also allows the scheme to mitigate the time penalties of multiscale inference and significantly improves total inference times when used on extremely large datasets. The use of this distributed scheme to process close to 150,000 Google Street View images as part of a public health study has demonstrated excellent promise in addressing issues of deployment, maintenance, crash recovery, and dataset management at scale.

Artificial Intelligence, Data and Decisions in the Pandemic

John N. Celona
Decision Analysis Associates, LLC, San Carlos, California, USA

Abstract: Although "artificial intelligence" ("AI") systems have made mindboggling strides over the last seventy years, these are based on advances in processing and data storage rather than on conceptually new techniques. These techniques lack a fundamental capability of organic intelligence: the ability to recognize when something is wrong and the codified rules should not apply. This is because AI expresses only a portion of the processing methods organic intelligence is capable of, a model for which is presented. To enable AI to contribute better to novel, complex, and uncertain situations like the Covid pandemic, we need to rethink the machine-human system design and interface. A possible starting model for doing so is presented.

CAGE: Classification with Convolutional Neural Networks for Age and Gender Estimation

Shahrullohon Lutfillohonov, Joonho Kwon
School of Computer Science and Engineering, Pusan National University, Busan, South Korea

Abstract: Automatic classifying age and gender on human face images has been hot topic recently. Although there have been several machine learning and deep learning models offered to solve the given task, there is still lacking performance on real world cases due to the large intra-class variety of face images (lighting variation, low-resolution, posing, scale, expressions, occlusions). To tackle the mentioned issues, we propose CAGE multitask deep learning architecture which obtained high accuracy rate in performance by learning representations. To further increase representation and classification capability of the proposed framework, we benefited from the data pre-processing, self-normalizing to the input and suitable parameter initialization methods. The model is trained on the Adience benchmark for age and gender prediction to label the input images into eight class ranges and two classes of gender. With the new proposed CAGE framework and pre-processing steps, we approach improved accuracy in both age and gender classification, reaching state of the art performance in gender recognition.

Reinforcement Learning for Auto-Reconnaissance and Decision Support

Joshua Haley, Jonathan Tucker, Benjamin Woodward, Jonathan Takahashi, Mark Taipan
Elbit Systems of America, Advanced Capabilities Center, Orlando, Florida, USA;
CVision AI, Medford, Massachusetts, USA

Abstract: Applied machine learning innovations have lead to an increase in automation in surveillance and reconnaissance systems. However, this increase in automation has lead to an increase in operator cognitive load. Operators are being asked to interpret more data, understand more system characteristics and in some instances control complex technologies. The volume of data can lead to cognitive overload and reduce efficacy if new sensors and algorithm characteristics need to be understood for operation. This paper describes an approach to characterize system behavior implicitly and only bring operator attention to triggers of high confidence after they have been verified by an automated reconnaissance system.

Dark Spot Identification using Clustering Techniques for Analysis and Classification of Cases of Molestation Against Women in Pune-India

Anjum A. Patel, Prasad Gokhale
Department of Computer Science, Faculty of Science and Technology, Vishwakarma University, Maharashtra, India

Abstract: Worldwide, women and children safety is a major issue as cases of violence against them has seen multifold increase. In India, crime against women and minor has increased 277% in past five years [1]. The rate of increase in crime depends upon many factors which need to be analyzed to identify the predominant factors. These predominant factors can then be used by law enforcement agencies, NGOs and others in an appropriate way to curb the crime rate. The crime rate can be reduced, if we can identify the locations where the crime rate is above threshold limit, which is defined as 'Dark Spot'. In order to decrease in molestation against women and minor at dark spot, we need accurate and timely information such as location, age group of accused and victim, relation of accused with victim etc to take action against crime. Hence, methodological approach for analyzing and identifying type and pattern of crime is required. This paper is divided into three sections: In first section; overview of problem defined with literature review. Second part elaborates methodologies applied for better results and Third section represents the analysis using statistical methods and regression analysis. The factors affecting to this analysis are age group, location and type of crime. Data mining is one of the best approaches to find out previously unknown, unstructured data and useful information by using different algorithms like regression, k-means, and apriori algorithm to find out trends, pattern and association of crimes at dark spot.

Designing an Efficient Machine Learning Algorithm to Offer Advices to Students based on Their Multiple Learning Attributes

Yanzhen Qu, Mikayla Cohen
Colorado Technical University, Colorado Springs, Colorado, USA

Abstract: Over the last ten years, AI technology has been gradually integrated into many modern learning supporting tools such as the adaptive learning platform, e-tutoring system, and education data mining software, etc. A common goal of all AI based learning support software is to identify gaps in a student's knowledge and enhance learning effectiveness through providing students with relevant advice. For example, the adaptive learning platform software based on a student's existing knowledge on a particular subject can determine whether certain learning materials will be offered to student. The student's existing knowledge on a subject is only one learning attribute, however, and cannot reflect all the aspects impactful to a student's learning effectiveness. A natural solution to overcoming this weakness is to improve the adaptive learning decision-making algorithm by making it capable of efficiently processing the dataset of a student's multiple learning attributes. This paper presents a design of a machine learning algorithm to efficiently process the dataset of a student's multiple learning attributes. The main enabling foundation for this new algorithm is a data structure called "student learning attributes index" which represents every learning attribute as a

tuple of three elements: the "learning-attribute-ID", the "weight" of the learning attribute among all the learning attributes, and the "efficient factor" of the contribution to the student's learning effectiveness made by the learning attribute. This study has applied unequal weight to each of the learning attributes, more accurately reflecting that different learning attributes will have different impacts on a student's learning effectiveness. This new algorithm enables various learning support applications to bring more practical factors into the decision-making process so that the advice presented to students becomes more practical and personalized.

Determination of Exploration Errors in Bernoulli Bandits Under the Epsilon-First Strategy

Clement Leung, Longjun Hao, Kristen Li
School of Science and Engineering, Chinese University of Hong Kong, Shenzhen, China;
School of Data Science, Chinese University of Hong Kong, Shenzhen, China;
Department of Computer Science, Hong Kong Baptist University, Hong Kong, China

Abstract: Bandit models are pervasive and occurin a wide variety of contexts such as patient treatment options investigating the effects of different experimental treatments, adaptive routing efforts for minimizing delays in a network, and financial investment portfolio design. Very often bandit exploration problems do not occur in isolation but in groups where several machines need to be explored. Here, we study Bernoulli multi-armed bandit machines where there are binary outcomes within the framework of the epsilon first strategy. Prior to exploitation, we explore and learn which machines among M bandit machines should be shortlisted for subsequent exploitation and which ones should be excluded. In the course of the learning process, we establish two classes of bandit machines, an exploitable class and a nonexploitable class, and cast the problem as a binary classification problem, where the ground truth class is determined by the majority of play outcomes. Here, we model the bandit plays as a random walk process, and predict the occurrence of erroneous classification. We obtain closed-form solutions for the probability of classification inaccuracies and derive useful error bounds as a function of the finite exploration capital. Experiments are also performed which show good agreement with the theoretical predictions.

Artificial Intelligence Applications to Develop the Skills of Soccer Players

Abdelfatah A. Tamimi, Hussein Ali Mohammed
Department of Software Engineering, Faculty of Science & IT, Al-Zaytoonah University of Jordan, Amman, Jordan

Abstract: In this paper, we will talk about the importance of artificial intelligence and its applications used in football, as the world is witnessing the development of technology, especially artificial intelligence and its use in all fields, and how it can be used to improve the level of football in particular and its impact on the level of players by observing and reading their own data of speed, strength and all What is related to the player's condition and how the data can be used for the purpose of analyzing it and obtaining the best strategies for matches in summary.

Examining the Applications of Artificial Intelligence and Machine Learning in the Automotive Industry

Behrouz Aslani, Ahemen Lucia Gege, Sandip Rakshit, Narasimha Rao Vajjhala
La Verne University, La Verne, California, USA; American University of Nigeria, Yola, Nigeria;
University of New York Tirana, Tirana, Albania

Abstract: Artificial intelligence, data science, and machine learning are key terms for improving almost every economic sector today. This paper defines these terms and how they relate to producing reliable processes that would change the future of the automotive industry. It includes an explanation of some of the areas of artificial intelligence and data science that come together to produce automated vehicles. It uses examples to show how some companies today are already taking advantage of these techniques to improve the production and performance of their automobiles. The paper highlights some of the benefits, disadvantages, and challenges of this massive transition from

manual cars to AI-driven cars and possible future implementations. Artificial intelligence and machine learning methods are presented along with a review of the state-of-the-art applications in the context of the automotive industry.

Next Frame Video Prediction using Hirachical Peredictive Coding Model

Matin Hosseini, Antoni Maida, Majid Hosseini, Raju Gottumukkala
University of Louisiana at Lafayette, Louisiana, USA

Abstract: In this paper, we proposed a new hierarchical deep learning architecture for video frame prediction we used the Prednet [4] similar architecture and changed the architecture based on the HER (Hierarchical Error Representation) model. We tested our model with two different size datasets, KITTI and UCF11. The results show a promising results for next frame video prediction.

Evaluating the Effectiveness of Chatbots as Experiential Learning Tools in an Introductory "Doing Business With A.I." Course

Thomas Menkhoff, Lydia Teo Ying Qian
Organisational Behavior and Human Resources (Education), Singapore Management University (SMU), Singapore;
Lee Kong Chian School of Business, Singapore Management University (SMU), Singapore

Abstract: In this qualitative-exploratory paper, we reflect about the design of an introductory chatbot workshop, a core element in an undergraduate business elective module “Doing Business with A.I.” conducted by the authors at Singapore Management University (SMU). Our research interest is to examine the effectiveness of using chatbots as a learning tool in introductory A.I. courses. We identify relevant learning theories and models that relate to the study of technology-mediated learning (TML) in general and chatbot-mediated learning in particular, examining how that may relate to students’ dispositions at the input level, the learning process (featuring a chatbot exercise) and learning outcomes. In the paper, we share our lesson plan (based on our current pedagogical workshop approach) and discuss how it integrates key ideas from conceptual frameworks and models. Overall, our ongoing study (formal data collection will begin in AY 21-22) intends to shed light on some of the critical factors that affect the usefulness and success of using Chatbots in higher education courses to support students’ learning outcomes in relation to the increasingly widespread practice of machine intelligence in business management with particular reference to NLP.

Segmentation of Timeseries Manufacturing Data Using CNN-LSTM

Sung-Hoon Kim, Ga-Ae Ryu, Hyoseok Oh, Aziz Nasridinov, Kwan-Hee Yoo
Department of Computer Science, Chungbuk National University, South Korea;
Department of Big Data, Chungbuk National University, South Korea

Abstract: Most of the time series data in the manufacturing industry repeats with a specific pattern. In this case, the CNN-based model has the potential to classify states by extracting features from the data shape regardless of noise. In this paper, we propose a hybrid model of CNN and LSTM to perform data segmentation on time series-sensor data of manufacturing industry that had undergone minimal preprocessing.

Anomaly Detection Algorithm for Acoustics Phenomena

Thomas Barguil, Johann Laurent, Nicolas Bohelay, Dominique Heller
CNRS UMR6285 - Lab-STICC, Univ. Bretagne Sud, Lorient, France;
Wavetel, F-56260, Larmor-Plage, France

Abstract: The evolution of technologies in call centers towards communications via ethernet is at the origin of a certain number of perturbations. These perturbations can take different forms but the most important one is the acoustic

phenomena. In this paper, we present an anomaly detection algorithm based on the One-Class Support Vector Machines (OC-SVM), for the detection of these acoustic phenomena. We are exploring different feature functions and seeking to find the best pairing with the OC-SVM to most effectively detect those acoustic problems that may pose a risk to consultants. Our experimental results show a good detection rate for amplitude levels equal or higher than -15 dB.

Detecting Phishing and Fraud Sites at Scale Using Deep Learning During Covid

Matin Hosseini, Abhilash Garimella, Abhishek Dubey, Shashi Prakash
Research Department, Bolster Inc., Los Altos, California, USA

Abstract: In this paper, we introduce a technology to detect phishing and fraud websites in real-time using deep learning, natural language processing, and computer vision. We also discuss the impact of COVID-19 on online fraud and its contribution towards record-breaking cybercriminal activity in the first half of 2020. We also introduce a free community tool checkPhish.ai that leverages this technology to detect fraudulent websites

An Artificial Neural Network and Cellular Automata Vertical Urban Growth Model Using Major Socioeconomic and Geographical Factors

Luis Antonio Lopez-Rivera, Marcelo Romero-Huertas, Eduardo Jimenez-Lopez
Universidad Autonoma del Estado de Mexico, Mexico;
El Colegio Mexiquense A.C., Mexico

Abstract: Vertical urban growth is a dynamic and continuous process that arises as a consequence of geographical, social, and economic factors. Traditionally, vertical urban models are not automated which implies complexity of analysis, probability of errors in calculations, spatial data incompatibility and a greater investment of time to generate the results. On the other hand, the use of artificial intelligence in vertical growth models is limited and presents several areas of opportunity, therefore this article presents a multidisciplinary model that works with automated mathematical rules, spacial analysis methods and raster image processing that allow project and simulate different vertical urban growth scenarios. Using neural networks and cellular automata integrated into the free Geographic Information System QGIS through the Python programming language. The model is tested in the metropolitan areas of: Mexico City, Guadalajara and Monterrey during the years 2015-2020, we obtained results in a range of 72% to 76%, validated by: i) Average number of projected skyscrapers, ii) Position using the Kappa index and, iii) Value in the raster image using the Jaccard index. Analysing different vertical urban growth scenarios, by modifying growth factors or cellular automata rules, allow better-informed decisions for urban planning and anticipating new infrastructure needs, projections, and regulations.

Know-linking: When Profile Techniques Meet Documents Classification to Generate a Documents Sharing Strategy based on Knowledge Requirement

Elamin Abderrahim, Nada Matta, Hassan Atifi, Vincent Maugis
University of Technology of Troyes, Troyes, France;
National Radioactive Waste Management Agency, Chatenay-Malabry, France

Abstract: Large companies represent a large frame of collaboration between several actors working on different fields. During their working process, actors require high knowledge sharing between each other, as well as re-usability of their past experience. Although, classical solutions couldn't deal with the challenge of managing knowledge flow which consists of identifying, extracting, capitalizing, keeping track and sharing knowledge, as big companies expected in some contexts. we proposed previously [28] an approach called "know-linking" organising actors in a company into profiles based on their knowledge requirement. Our first job of profiling was based on a data lake

collecting data from different sources, we complete our approach by two different type of analysis: operational tools and organisational tools to build a more preferment profile structures and we go through discovering semantic links existing between those profiles expressed deeply in documents, which will solve documents retrieval and access problems.

Deep Learning Techniques for Constructing Unmanned Store with Low Cost

Shi-Jinn Horng, Ken-Hao Wang, Ping-Chen Tsai
Department of CS & Info. Eng., National Taiwan University of Science and Technology, Taiwan, ROC

Abstract: Reducing operating costs and increasing profits are the future trend of the retail stores. In this paper, we use the deep learning techniques to implement the unmanned store. Without the checkout platform, we achieve the goal, "Get and Go" for the unmanned store. We propose a construction of the environment of the unmanned store by integrating product recognition, face recognition, shopping judgment algorithms, sensors and hardware devices to build a store environment. Compared to the existing approaches, we have successfully reduced the usage of hardware to reduce the cost of the construction quite much, and also resolved the potential theft issues.

Predicting Human Movements Using Machine Learning

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

Abstract: Assurance of safety in a crowd, such as the density of mass gatherings in religious rituals, or sports events, represents an important matter for authorities. Chaos and crowds create a big challenge for movement prediction. Even though, knowing the pedestrians' next positions is difficult regarding their behavior and their intentions, it is useful for ensuring their security. In this paper, we predict the future positions of pedestrians using the Recurrent Neural Network (RNN) and, specifically, its extension, Long Short-Term Memory (LSTM). We use the past directions of individuals predict their next directions, which leads to their future positions. We compared our proposed LSTM method, based on previous directions (LSTM-direction), with the neural networks and genetic algorithms (NN-GA) method. We used several datasets that are based on various scenarios and several simulation types, to evaluate both methods. Our results show that the LSTM based on the sequence of past directions and its features is superior to NN-GA.

A Contribution to Crowd Control by Detecting Critical Density Spots

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

Abstract: Crowd control is a complicated process and a highly costly operation, but it is essential when large number of people gather and create herd movement, in which there is a high chance of casualties from trampling and crushing. People's safety is critical in order to enhance their experience when attending large events or religious gatherings; overall public safety must be guaranteed. A crowd control process can help organizers make decisions in order to save manpower and create time for preventing fatal accidents in a crowd without guidance or leadership. By focusing on the two main crowd types, structured and unstructured crowd behaviors, it is possible to contribute to the crowd control process with logic. Monitoring individuals' location and status in a crowd is necessary for crowd control. Our method is based on major factors such as an individual's speed,direction, and detecting the direction of other individuals. The proposed model in this paper is based on [?] an alert system that can inform authorities by flagging individuals who are in a situation of risk when a crowd is being formed and while events are taking place. The alert system scrutinizes individual's location status using fuzzy logic to identify the critical density spots that can obstruct crowd motion.

Crowd simulations were built using Netlogo to generate datasets of individuals' locations based on the flocking model and the social forces model. We compare the crowd behavior is different situations to detect the critical density spots distinguishing the levels of risks among individuals.

The Color of Music: Towards Hearing Impaired Enrichment with Cognitive AI/ML and Multi-sensory Emotion Fusion

Ryan A. Carbone, John N. Carbone
OMS-III, UIW School of Osteopathic Medicine, San Antonio, Texas, USA;
Senior Director/Chief Solution Architect, Forcepoint Global Governments & Critical Infrastructure, USA;
Baylor University, USA;
Southern Methodist University, USA

Abstract: Music, a sound-based artform, has influenced human culture for generations, however, many hearing-impaired individuals have been unable to fully experience the substantial breadth and depth of music. Inherently complex human sensing, deficient cochlear spectrum range, and narrowly focused sound-based Music Emotion Recognition (MER) research complements the challenge. Thankfully, rapid advances in Artificial Cognition, advanced sensing, data fusion research, and Machine Learning enable new approaches to significantly enhance the emotional experience of music, and ultimately provide innovative mitigations for sensory impairments. Therefore, we propose a multidisciplinary approach, fusing artificial cognition architecture (ACA), emotion processing, intricate brain and multisensory performance characteristics, and adaptive neural framework machine learning. We leverage and extend existing audio-visual learning research and utilize knowledge density equation research to algorithmically map fine-grained emotions and sensing characteristics to brain sensing performance and explore intersensory relationship characteristics. Thereby, infusing emotion context for enhanced individualized visual and tactile listening and potentially novel enhancements for sensing-centric applications, and for maximizing normal and hearing-impaired music emotion perception, enjoyment, learning and performance.

SESSION:
XXI - Applications of Advanced AI Techniques to Information Management for Solving Company-Related Problems

Co-Chairs: Dr. David de la Fuente and Dr. Jose A. Olivas***
**University of Oviedo, Spain*
***University of Castilla - La Mancha, Spain*

Operations Management through the Prism of Distributed Ledger Technology

Fernandez-Vazquez S., Rosillo R., Gomez A., Priore P., Parreno J.
Escuela Politecnica de Ingenieria de Gijon, Campus de Viesques s/n, Gijon (Asturias), Spain

Abstract: Blockchain has become increasingly one of the hot topics in the financial world. Due to its characteristics, it is applicable to almost any kind of industry. In this paper we explore the relation between blockchain and operations management, discussing the main advantages and challenges of this new technology. Although more and more companies are using blockchain in a day-to-day basis, it may still be too soon to see the implementation of this technology across all industries. Through an in-depth analysis of blockchain in operations management, its main features are presented.

The Learning Machinery in Organisations: The Role of Problem-Solving

Costas J., Pastor R., Puche J., Fernandez I., Pino R.
Department of Applied Economics, Faculty of Economics and Business, University of Burgos, Burgos, Spain

Abstract: An extremely important feature of any industrial company to survive is its ability to solve problems. This work shapes a framework in which each one of the company processes is devised through three interconnected pipelines that together care to generate value for external or internal customers, react to all kind on non-conformities (accidents, downtimes and defects), and implement solutions to either adapt to external changes or to improve current performance. Connecting those three pipelines, either within or among processes, is achieved by means of a ticketing system. This pipeline architecture is inspired upon the BDI paradigm for artificial intelligent agents, and addresses some of the challenges that organizations face to grow up their talent, which in the end is a basic need for survival.

Using Deepfake Techniques to Emulate Deceased People: The 'Lola Flores' Case and its Legal Implications

Elisa Gutierrez, Cristina Puente, Pilar Cousido, Ana Laguna, Jose A. Olivas
Advanced Technical Faculty of Engineering - ICAI, Pontificia Comillas University, Madrid, Spain

Abstract: N/A

Evaluating Advance Purchases of COVID19 Vaccine Candidates under Fuzzy Conditions

Puente J., Alonso Bonis S., Gascon F., Ponte B., De la Fuente D.
Department of Business Administration, Faculty of Economics and Business, University of Oviedo, Oviedo, Asturias, Spain

Abstract: Many countries signed contracts with biotechnology and pharmaceutical companies to purchase millions of doses of several COVID-19 vaccine candidates before proving their efficacy or being approved by the relevant bodies (such as the FDA or the EMA). Indeed, the total number of doses that were purchased in advance exceeded in many cases the doses needed to vaccinate the whole population of these countries. Nevertheless, there is not a straightforward relationship between the vaccination rates of the countries and the number of doses they prepurchased. From this perspective, this short research paper investigates three key issues that affected the way countries built in advance portfolios of COVID-19 vaccine candidates and models their accelerating effects on vaccination rates. Specifically, we consider three dimensions: the leaders' decision-making authority, the efforts to guarantee a continuous flow of vaccines, and the compromise to donate vaccines. To meet our goals, we employ fuzzy techniques, which allow us to accommodate the intrinsic uncertainty of the problem under consideration. In this sense, the methodology that we develop may be applied to investigate other problems of similar nature.

Multilevel Representation of Key Factors for the Prediction of the Madrid'21 Elections Results using Case Based Reasoning

Antonio Lorenzo, Jose A. Olivas
Department of Business Intelligence, Castilla-La Mancha Government, Toledo, Spain

Abstract: Usually statistical techniques are used to make election predictions, but some machine learning techniques are also used. The application of both sets of techniques are usually based on historical data results without considering important key factors that influence the outcome. We propose new approaches based on Case Based Reasoning (CBR) to improve the accuracy of the predictions. A multi-level representation of the main characteristics that could influence

the electoral process is proposed, relating key factors to election results and trying to convert correlation into causality. This allows us to compare different electoral processes to know their similarity to the current case. Once the most similar elections to the ones to be predicted have been retrieved and patters are discovered, the knowledge about current specific key factors is applied to make predictions.

SESSION:
Advancements in Mathematical Models for Adversarial Machine Learning and Defenses

Chair: Prof. Young-Ae Jung
Division of Information Technology Education, Sunmoon University, Asan, Korea

Robust Industrial Control Systems based on Network Anomaly Detection

HyunJin Kim, Taeshik Shon
University of Ajou, Yeongtong-gu, Suwon-si, Gyeonggi-do, Korea

Abstract: Industrial control systems (ICS) are systems for controlling and managing the devices and systems of national infrastructure. The existing ICS domains are isolated environments from the outside world for protecting sites and systems. But the following the trend of the 4th Industrial Revolution, the ICS also increase connectivity of various domains and the convergence of various technologies. These led to security concerns at ICS domains from Operation Technology (OT) and Information Technology (IT) perspectives. However, many exiting security studies in the ICS domain focus on polling network mechanism systems. But, in ICS sites that require fast or real-time, the polling method is used in the initial setting stage and the environment change setting whereas the reporting method is used for other operations. So, this paper proposes an anomaly detection method for the non-polling or reporting network operation of ICS. The proposed method uses time information and pattern of time intervals and classified messages. So, it does not require pre-knowledge of the packet content. It means that the proposed method applies the various ICS protocols and encrypted contents.

Cybersecurity Threat Analysis for Smart Manufacturing Systems

SungJin Kim, Taeshik Shon
University of Ajou, Yeongtong-gu, Suwon-si, Gyeonggi-do, Korea

Abstract: Smart factories using various technologies such as cloud service and big data analysis have improved product production efficiency in all areas of the industry. Although the smart factory has these advantages, the risk of cyber-security incidents has increased due to increased external connections. Various attacks targeting the industrial area have been reported, so countermeasures are essential. Many studies, guidelines, and standards have been published to cope with this problem, but smart factories in a transition stage, which uses both legacy equipment by targeting the latest smart factories, are still exposed to risk. In this paper, to solve this problem, we analyze the expected structure of a smart factory in a transition stage and discuss potential risks and affected devices/systems due to differences between the legacy systems and smart factories. The legacy system transforming into a smart factory has seven points of change. We discuss the threats posed by these changes. And it analyzed which areas of the factory these threats affect. The mappings for changes, threats, and affected points are provided. As a result, it is judged that all the threats caused by the seven changes can affect the OT area.

Real-Time Evaluation of Player Experience in Games

Shin Jin Kang, Soo Kyun Kim
Department of Computer Engineering, Jeju National University, Jeju, South Korea

Abstract: This paper introduces an automatic detection system of the emotional status of game users via a multi-modal interface. Our system can detect emotional pleasure and arousal, which are useful in user behavior analysis, without obtrusive contact. By using a web visualization method, our system can summarize time-serious user's behaviors in an intuitive way. Experiment results show that the system can classify five emotional statuses with a 76% success rate in a laboratory condition. Our results can be applied to the game industry as game user experience analysis middleware for evaluating the excitement of a game quantitatively.

Research on Security Risk Detection Model for Manufacturing Plant Business Resilience

Yanghoon Kim, Hangbae Chang
Department of Cyber Drone Bot Military Engineering, Shinhan University, Gyeonggi-do, Korea

Abstract: Manufacturing plants in early information days had a system that low computing power equipment simply and repeatedly produced focusing on machinery. These information plants account for a large portion of the manufacturing industry. The plants built in early days couldn't interrupt business to check information devices. In addition, it was difficult to develop and install new high-performance equipment due to insufficient investment cost. These problems pose a significant problem in securing business resilience in events of security incidents. Accordingly, in this study, we proposed a methodology to solve problems in production reliability due to security problems in information manufacturing plants. Specifically, we have devised a model and measurement that could detect a problem situation in a production process due to security problems.

Virtual Characters based Personalized Intelligent Agent Service using Face Recognition

Bumsoo Kim, Mikyung Bae, Taemin Lee, Sanghyun Seo
Chung-Ang University, Anseong-si, Gyeonggi-do, Republic of Korea

Abstract: This paper introduces an expandable system structure for Personalized Intelligent Agent (PIA) Services based on Virtual Characters using Face Recognition. Our PIA is integrated in Unity3D Engine, it can visually provide a variety of services. The services are divided into General Services and Personalized Services by classification whether it is a trained classes or unknown. The services data is updated and pre-processing each hour, and it has an adaptive system structure that can easily annotate additional class dataset for face recognition. Embodied virtual characters help intuitive information a communication as virtual assistant included 3D Animation, Speech using Digital Information Display. Our PIA with personalized face recognition can be utilized as various domain such as education, institutions, companies, and public facilities in the future.

Face Mask Detection System using Generative Adversarial Networks

M. Kaliappan, S. Vimal, K. Vijayalakshmi, C. Jegatha, V. Vaishnavi, Sanghyun Seo
Chung-Ang University, Anseong-si, Gyeonggi-do, Republic of Korea

Abstract: Recent global pandemic of Covid-19, there is a vital need for security mechanisms, with face masks being one of the most significant. The main purpose of this project is to detect face masks on human faces in both live streaming video and photographs. Our face detector model was created using deep learning.Face recognition has recently been successfully implemented using deep convolutional neural networks. Despite significant progress, most

current detection methods only use a bounding box to locate each face, making it impossible to segment each face from the background image at the same time.To address this shortcoming, we present Generative Adversarial Networks -Mask, face detection and segmentation approach based on an improved Mask CNN that combines face detection and segmentation into a single framework with the goal of obtaining more fine-grained face information. This project deals with the following problems: • Identifies whether the person is wearing a mask or not. • Detects multiple faces at a time. GANs achieved good performance and results in all the aspect directions. It forces the generated samples to be indistinguishable from the target distribution by introducing an adversarial discriminator. The proposed schemeemploys embedding information for facial to generate portrait images.

A Survey on Byzantine Fault Tolerant Consensus Algorithms

Mingyu Jo, Jaehwan Lee, Sanghyuck Nam, Sangoh Park
Department of Computer Science & Engineering, Chung-Ang University, Seoul, Republic of Korea

Abstract: A lot of research is being done on algorithms for reaching a consensus on the same value among distributed nodes. Paxos is a representative algorithm that achieves consensus among distributed nodes. However, Paxos is a crash fault tolerance algorithm. It means, if a node maliciously forges a message, consensus may not occur or consensus may occur with an incorrect value. A Byzantine fault-tolerant (BFT) consensus algorithm was devised to prevent malicious message forgery of arbitrary nodes. In this paper, we survey BFT consensus algorithms and analyze their advantages and disadvantages.

Micro-Macro Network Approach for Explainable and Reusable Reasoning

Yujung Cho, Namyeon Lee, Ohbyung Kwon
Department of Big Data Analytics, Kyung Hee University, Seoul, Republic of Korea;
Department of IT Management, Hanshin University, Osan, Republic of Korea;
School of Management, Kyung Hee University, Seoul, Republic of Korea

Abstract: The purpose of this study is to develop a method of learning problem solving strategies for new tasks by combining micro-networks as small deep neural networks that have been learned to perform small-scale functions to solve the limitations of Neural Architecture Search (NAS). To this end, when a new task is given, we compared the task solved by other agents and referred to a solution method for similar tasks, and developed a technology to quickly resolve the problem through the combination of micro-networks. The proposed method is useful for the developers to build a reusable deep learning inference system. In addition, sustainable learning is possible through the expansion of network banks, and a platform can be built through sharing and practical use of learned models.

Estimation of Respiratory Rate Using the Embedded Microphone System on the Mask

Chhayly Lim, Yunyoung Nam
Department of ICT Convergence Rehabilitation Engineering, Soonchunhyang University, Cheonan-si, South Korea

Abstract: Respiratory rate (RR) is an important sign to predicts various serious clinical events. However, this vital sign is often neglected in the clinical records due to lack of reliable, available, and cost-effective. Therefore, in this paper, we present respiratory rate estimation using nasal breath sound recording by a low-cost microphone device embedded on mask. Experiments involved the participants using the proposed device to breathe in ranging from 12 to 24 breaths per minute. The results demonstrate that the proposed device and method can estimate RR effectively from both breaths through the nose and the mouth. The results yielded median and interquartile range (IQR) errors as low as 0% by using a Welch periodogram method to estimate the power spectral density of the breath audio signal envelope.

Classification of Yoga Pose using Transfer Learning

Chhaihuoy Long, Yunyoung Nam
Department of ICT Convergence Rehabilitation Engineering, Soonchunhyang University, Cheonan-si, South Korea

Abstract: Yoga practice is a healthy exercise from India that focuses on physical, mental, and spiritual. In order to release stress or depression, yoga asana is a popularly well benefit for health. In brisk technology advancements, the challenge in artificial intelligence, machine learning, and deep learning techniques for utilization of image recognition like the yoga pose domain is interesting. The system that classifies a yoga pose is proposed in this work based on transfer learning of deep neural network architectures. The yoga poses data were collected in RGB images for the training model to evaluate the classification accuracy. Data augmentation technique was also applied to the training set. The classification performance shows possibility of using for yoga self-training tool.

Service Platform Model for Contactless Care of the Elderly Living Alone in Aging Era

Yang Sun Lee, Ji-Hee Kim, Ki-Hong Park
Division of Computer Engineering, Hanshin University, Kyunggi-do, Republic of Korea

Abstract: In this paper, we propose a service model for contactless care services for the elderly living alone. The proposed service model aims to check the life patterns of the elderly living alone, who spend most of their daily lives at home, and to effectively check their personal safety. The main core of the proposed service model is the companion toy robot. The companion toy robot has built-in IR sensor, push sensor, vibration sensor, microphone and speaker, and acts as an interactive emotional device by activating and interacting with the elderly living alone. In order to evaluate the possibility of the proposed service model, AI analysis of the collected data and visualization of the monitoring results were performed. As a result of the experiment, the possibility of the proposed contactless care service model was confirmed.

Current Big Data Security Issues for Wearable Devices

Jaeseok Choi, Yanghoon Kim, Mucheol Kim
Department of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea;
Department of Cyber Drone Bot Military Engineering, Shinhan University, Uijeongbu, Republic of Korea

Abstract: The number of the elderly is increasing due to the influence of the super-aged society. Most of these elderly people live alone due to the urbanization. For the elderly's health, the appropriate management measures are required. Wearable devices are being studied as the devices have a function to monitor the elderly's healthcare. Wearable technology is used to recognize and collect the data in real time. However, the collected data include personal information. The information is sensitive for privacy. Therefore, this paper examines the literature related to wearable devices, wearable information, security issues, and information security technologies.

Deep Learning Based Automatic Questionnaire Generator

M. Kaliappan, S. Vimal, K. Vijayalakshmi, L Ganesan, P. Rajarajeswari, K. Iyappan, Mucheol Kim
Department of Computer Science and Engineering, Ramco Institute of Technology, Tamilnadu, India;
Department of Electronics and Communication Engineering, Ramco Institute of Technology, Tamilnadu, India;
Oracle Functional Consultant, IBM, Bangalore, India; Heukseok-ro, Dongjak-gu, Seoul, Korea

Abstract: Manual Question generation is a cumbersome task. It consumes a lot of time and takes alot of effort and evaluation of the assessment becomes tedious also. When the content of the assessment is too long to read and frame questions, the automatic question generator becomes helpful. In view of the evolving practice of digital learning platforms, it is essential to develop a system that generates questions from the course content for the assessment

automatically. Manual question generation is burden some and evaluation takes time and sometimes inaccurate. To overcome these issues, the question generator will be used as a solution. There have been many question generators in the existing literature and they use complex workflows. The aim of our proposed work is to generate an automatic questionnaire using Deep learning models. The paper represents our work on using Transformers for generating questions. The system uses answer aware model and pretrained T5 for the processes of answer extraction and question generation.

An Evolutionary Algorithm Based Approach for Efficient Cricket Highlight Generation

Momina Bukhari, Muazzam Maqsood, Sadaf Yasmin, Irfan Mehmood, Seungmin Rho, Mucheol Kim
Heukseok-ro, Dongjak-gu, Seoul, Korea

Abstract: Significant events in a cricket match are often missed because of longer match durations, broadcasting time considerations, and a vast amount of accessible multimedia material. With this in mind, this study proposes an automated system for key event identification and highlights production for cricket video. In this work, an efficient method for video summarization is proposed based on textual features to detect key events from different input videos. Moreover, the score caption is analyzed to identify the significant changes in the score, and the wickets counter is used to detect the boundary, six, and wicket events. This paper offers a modular platform for cricket content highlights with genetic algorithms (GA). A cricket match highlight can include a wide range of activities, cover the whole match, and include events of comparatively high significance. A multi-objective optimization problem is formed by these three variables which is solved through evolutionary algorithm. The findings of our cricket match experiment have been quite promising. As a result, cricket videos can be effectively transmitted over low-bandwidth networks with time-constrained delivery.

An Efficient Method for Emotion Recognition Using Machine Learning Techniques

Anam Moin, Wajiha Ilyas, Muazzam Maqsood, Irfan Mehmood, Dongwann Kang
Department of CSE, Seoul National University of Science and Technology, Nowon-gu, Seoul, Republic of Korea

Abstract: Emotion recognition is widely gaining importance in research areas nowadays. Electroencephalography (EEG) signals are a new emerging trend currently being used for the recognition of emotions. Good emotion recognition using EEG greatly depends on the selection of significant electrodes and features extracted. In this proposed work, a dataset named DEAP is used. The main purpose of our research work is to make a model capable of subject-independent emotion recognition using deep learning. We classified emotions on the Valence-Arousal 2D scale. Our proposed work achieved 63.72% accuracy for valence and 55% accuracy for arousal. Although the accuracy is somewhat low as compared to other work this is because testing is performed on two subjects' data completely unseen by the trained model so, it is significant accuracy for cross-subject emotion recognition.

Five Pixel Adversarial Attack for Gait Recognition Based Biometric System

Maryam Bukhari, Muazzam Maqsood, Sadaf Yasmin, Irfan Mehmood, Seungmin Rho
Department of Industrial Security, Chung-Ang University, Dongjak-gu, Seoul, Republic of Korea

Abstract: Gait is a unique and behavioral characteristic of human beings, which is widely used for their identification. Its acquisition does not require human cooperation. It is still advantageous even if the subjects are at a longer distance from the camera. Recently, deep learning-based methods are extensively used in gait recognition-based biometric systems. However, due to connatural vulnerability of Convolutional neural networks (CNNs), there exists some potential risks to all these social security systems. In this research, we propose a method in which we perturbed the human gait features by adding noise to only five random pixels of an image using the Fast gradient sign method (FGSM). This small perturbation is invisible to humans and generates adversarial images, which are more natural and

very close to original images. Moreover, this research investigates the performance of CNNs towards human gait recognition. The results demonstrate that CNNs based gait recognition methods are highly vulnerable to these attacks even when the perturbation is much smaller.

Dynamic Search for Disaster Relief Applications

Faisal Saeed, Anand Paul, Seungmin Rho
Department of Industrial Security, Chung-Ang University, Dongjak-gu, Seoul, Republic of Korea

Abstract: Searching techniques along with smart agents for intelligent searching are beneficial in disaster management. We projected a navigational model for smart agents intended for forest observation from fire dangers. Our principal intention is on fire detection then proactively to explore the zone which are additional probable to have directions on the way to the target. We interpreted this delinquent problem into an optimization query or problem over a state space, which can be resolved proficiently by a greedy searching technique. The proposed approach is compared mutually with uninformed and informed searching techniques.

Adversarial Networks for Blockchain Attack Classification

Junaid Gul Malik, Urfa Gul Malik, Anand Paul, Seungmin Rho
Department of Industrial Security, Chung-Ang University, Dongjak-gu, Seoul, Republic of Korea

Abstract: The time series model for forecasting and classification for blockchain attacks requires more studies with neural networks as compared to statistical methods. Time-series predictions are made after analyzing the historic time-series patterns. Fewer studies are made with neural networks but still require more exploration in the field of the blockchain. GANs (Generative Adversarial Networks) have proven to be a highly effective system for training models to generate realisticlooking results. For our study, we have gathered hash rate data from miners. Metrics for comparison are MAPE and RMSE. We observe lower values for the error functions with GAN in comparison.

Research on Development of Security Level Evaluation Model for Manufacturing Industry Partner Companies

Yurim Choi, Hangbae Chang
Department of Industrial Security, Chung-Ang University, Dongjak-gu, Seoul, Republic of Korea

Abstract: In the era of the Fourth Industrial Revolution, in the manufacturing industry, collaboration with other companies that have different technologies has become an indispensable element. However, for collaboration, companies need to share data and technology between the parent company and the cooperating ones. In this process, cooperating companies with a high proportion of SMEs with relatively few security resources including personnel or management systems often incurs technology leak accidents. In this research, to solve these problems, we propose a model that can objectively and rationally evaluates the security level required of cooperating companies to share data and technology securely from the perspective of the parent company of the manufacturing industry. We analyze related precedent research and design a reference model for security level evaluation that reflect the characteristics of cooperating companies that need to be managed from parent company's view. Through this research, we expect not only manufacturing parent company could collaborate with cooperating companies securely but also cooperating companies could manage themselves utilizing the model we propose in this research.

A Novel Hybrid CNN based Smart Detection and Classification for Banana Plant Disease Using Leaf Image

K. Lakshmi Narayanan, R. Santhana Krishnan, Y. Harold Robinson, E. Golden Julie, S. Vimal, Mi Young Lee
Department of Software, Sejong University, Seoul, South korea

Abstract: Agriculture is a backbone of India of that Banana cultivation is a important factor as banana fruit has its market all over the world.The major issue in the cultivation of banana plants is the crop getting affected by numerous diseases and pest symptoms very often so proper monitoring system is required to identify the infections in the banana plant early to avoid financial losses to the farmer who cultivated the crop.In order to avoid these problems we propose a hybrid Convolution Neural Network (CNN) and Fusion Support vector machine (SVM) based banana disease detection and classification in the earlier stages of its disease infection and also suggest the farmer with possible fertilizers to be used to get rid of the disease in its earlier stages.With this proposed system we achieve accuracy above 99% and system was developed using python programming language.

SESSION: Social Robots
CHAIR: Dr. Patrick C. K. Hung
University of Ontario Institute of Technology, Canada

A Conceptual Model for a Parental Control Tool for Smart Toys

Otavio de P. Albuquerque, Marcelo Fantinato, Sarajane M. Peres, Farkhund Iqbal, Patrick C. K. Hung
School of Arts, Sciences and Humanities, University of Sao Paulo, Sao Paulo - SP, Brazil;
College of Technological Innovation, Zayed University, United Arab Emirates, UAE;
Faculty of Business and IT, Ontario Tech University, Oshawa - ON, Canada

Abstract: The use of smart toys by children raises new concerns for parents and researchers. Children are more likely to share sensitive data and are unaware or rarely care about online risks. Parents play a relevant role in protecting the children, and parental control tools are needed to take control and adequately manage their child's data according to their preferences. To our best knowledge, current parental control tools neither meet parental needs nor compliant with a standard for toy makers. This paper presents a conceptual model for the development of a parental control tool for smart toys. This conceptual model relies on a set of requirements based on literature and related privacy standards.

Only Words? Examining the Effect of a Social Robot's Use of the Bajan Dialect's Lexicon on User Attitude and Perception

Curtis L. Gittens, Damian Garnes
The University of the West Indies, Cave Hill Campus, Bridgetown, Barbados

Abstract: N/A

BDI Agents for Adaptive Cyber Defense of IoT Systems

Laura Rafferty, Patrick C. K. Hung, Eleanna Kafeza
Faculty of Business and IT, Ontario Tech University, Canada;
College of Technological Innovation, Zayed University, Dubai, United Arab Emirates, UAE

Abstract: As cybersecurity threats to Internet of Things (IoT) devices, including social robots, are becoming increasingly prevalent, the characteristics introduce a unique challenge to the security industry, which can benefit from the autonomous response. The Beliefs-Desires-Intentions (BDI) model provides a foundation that can be extended to model rational agent behavior, which can be applied to adversarial cybersecurity scenarios. This paper defines the base model for a cyber control system and presents an architecture for adaptive multi-agent defense policy generation and deployment, network visibility, and access control in a smart home IoT environment. Finally, we present a demonstration of the functional components for the practical implementation of an agent for a domain-specific use case.

Sliding Window Trust based Implicit Authentication for Service Robots

Pengming Zhang, Xiaomei Zhang, Yajie Wang, Xin Zeng, Zhicai Shi, Chi-hung Chi
School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China;
Commonwealth Scientific and Industrial Research Organization (CSIRO), Sandy Bay, Tasmania, Australia

Abstract: With advances in networks, Arti_cial Intelligence (AI), and the internet-of-things, humanoid robots are rising in many areas, including elderly care, companion, education, and services in public sectors. Given their sensing and communication functionality, information leakage and unauthorized access will be of big concerns. Very often, authentication techniques for service robots, especially those related to behavioral identifcation, have been developed, in which behavior models are created using raw data from sensors. However, behavioral-based authentication and re-authentication is still an open area for research, including cold start problems, accuracy, and uncertainty. This paper proposes a hierarchical implicit authentication system by joint built-in sensors and trust evaluation, coined sAuth, which exploits sensor data based sliding window trust model to identify the service robot and its expected users. In order to mitigate the uctuations of identification results in the real world environment, the trust evaluation is computed via combining the weighted intermediate identi_cation probability of various small sliding windows. The performance of sAuth is evaluated under di_erent scenarios where we show that (i) approximately 5%-7% higher accuracy and 2%-18% lower equal error rate can be achieved by our method compared to other works; and (ii) the hierarchical scheme with joint sensors and trust sliding windows improves the authentication accuracy significantly by comparing it with only sensor-based authentication.

A Case Study of Facial Expression and Color for Social Robot's Emotion Expression

Pei-Chun Lin, Patrick C. K. Hung
Department of Information Engineering and Computer Science, Feng Chia University, Taiwan, R.O.C.

Abstract: N/A

SESSION: Intelligent Linguistic Technologies (ILINTEC'21)

Chair: Dr. Elena B. Kozerenko
Institute of Informatics Problems of the Federal Research Center;
Computer Science and Control of the Russian Academy of Sciences, Moscow, Russia

Cognitive Aspects of Mathematical Education

Konstantin Kozerenko, Elena B. Kozerenko
Lyceum "The Second School", Moscow, Russia;
Institute of Informatics Problems of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia

Abstract: N/A

Analytical Textology in Establishing the Explicit and Implicit Semantic Features of Discourse

Elena B. Kozerenko, Mikhael Yu. Mikheev, Lev I. Erlikh, Konstantin I. Kuznetsov, Nikolai V. Somin
Institute of Informatics Problems of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia;
Moscow State University, Research Computer Center, Moscow, Russia

Abstract: The paper presents a new direction of research at the intersection of linguistics, informatics, philology using logical and statistical methods for the analysis of unstructured data in the form of natural language texts in order to solve a number of problems of extracting explicit and implicit knowledge from texts using a semantically oriented linguistic processor, the formation of lexical and statistical representations of texts, the construction of analytical conclusions, determination of the author's idiostyle and textual similarity of written speech based on the analysis of service words and other micro elements of the text; revealing the emotional coloring of texts, building a holistic profile of the author's text based on the superposition of techniques. An example of discourse profiling is based on the "Blue Book" of the "Petersburg Diary" by Z.N. Gippius.

Computer-Assisted Analysis of the Contemporary American Political Discourse from the Perspective of Cognitive Linguistics

Lidia Grigoryeva, Diana Kasimova
Institute of Informatics Problems of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia;
Institute of Foreign Languages, Peoples' Friendship University of Russia, Moscow, Russia

Abstract: With reference to the recently completed elections in the USA that highly attracted attention of the world media and international community we exmined the modern political discourse in terms of cognitive linguistics complemented with computer assisted approach. We collected valid corpora of Joe Biden’s and Donald Trump’s speeches and scrutinized average sentence length, distribution of stop words, number of clauses in each particular sentence within the individual corpora and analyzed cognitive models that were considerably resorted to by the politicians. Thorough computational analysis revealed a significant difference in complexity and information redundancy of the political speeches. The comparative analysis confirms the premise that superfluous speech with semantic redundancy and irrelevant concepts hampers the speech perception and may lead to lower political competitiveness. On the basis of G. Lakoff’s cognitive framework we verify the assumption on the possible ongoing shift of national American values and beliefs. The possible current shift can be presumably associated with the impact of the global pandemic.

The 17th International Conference on Data Science (ICDATA 2021) https://icdata.org https://www.american-cse.org/csce2021/conferences-ICDATA

Solid Log Chain: Novel Blockchain Driven Audit Log Management System for GDPR Compliance

Ulas Aslan, Baha Sen
Ankara Yildirim Beyazit University, Turkey

Abstract: Audit log files contain personal information, protecting the data is crucial for GDPR compliance. As stated in GDPR, data subjects have the right to be forgotten, so personal data must be removed from the log records upon user demand. In this paper, we present a GDPR compliant immutable log storage system called Solid Log Chain (SLC). The goal of the SLC is to store audit logs and assure immutability of log data while meeting the GDPR. The presented system is intended to be an alternative to expensive hardware-based solutions by combining and innovatively using existing technologies. We describe the design concept, architecture, and system flow of the SLC and evaluate its performance in terms of latency and payload size.

A Deep Neural Network Based Approach to Building Budget-Constrained Models for Big Data Analysis

Rui Ming, Haiping Xu, Shannon Gibbs, Donghui Yan, Ming Shao
University of Massachusetts Dartmouth, USA

Abstract: Deep learning approaches require collection of data on many different input features or variables for accurate model training and prediction. Since data collection on input features could be costly, it is crucial to reduce the cost by selecting a subset of features and developing a budget-constrained model (BCM). In this paper, we introduce an approach to eliminating less important features for big data analysis using Deep Neural Networks (DNNs). Once a DNN model has been developed, we identify the weak links and weak neurons, and remove some input features to bring the model cost within a given budget. The experimental results show our approach is feasible and supports user selection of a suitable BCM within a given budget.

Recommending Teammates with Complementary Skills via Matrix and Tensor Decompositions

Wookey Lee, Jafar Afshar, Arousha Roudsari
Inha University, Korea

Abstract: N/A

The Winning Mining Formula for Football Competitions

Ana Clarissa Miranda Pena, Francisco J. Cantu
Tecnologico de Monterrey, Mexico

Abstract: Football is practiced in almost every country, each league having its playing style. This research proposes a Football Framework Machine Learning Pipeline (Football MLP) which predicts soccer results by understanding teams performance based on statistics. Current research studies have a common weakness, this is due to limited access to seasons and league data, those issues have been resolved by Football MLP's capability to retrieve updated statistics from several countries. Football MLP's winning formula consists of a pipeline that ranks teams based on Hierarchical Clustering, uses match games statistics to rate offensive and defensive teams playing styles, and generates a prediction with an Ensemble Machine Learning Model. FootballMLP was proved using data from seven European leagues on seasons 2016/2017, 2017/2018, 2018/2019, 2019/2020, and 2020/2021, including cost-benefit analysis.

Tyro: A First Step Towards Automatically Generating Parallel Programs from Sequential Programs

Greg Speegle, Arun Sanjel
Baylor University, USA

Abstract: For the past several years, parallel programming has been exploding in popularity. Fueled by the development of tools for automatically executing code across large data clusters and the demand for analysis of enormous data sets, parallel computation has become a standard. Unfortunately, current parallel programming languages often limit the possible parallelism by only supporting limited operations. As a result, the "natural" way to write a program sequentially is not possible within the parallel languages. Recently, there have been several efforts to automatically generate parallel programs from sequential programs (see e.g., [1]–[3]). In this paper, we present Tyro, a new tool that automatically translates a sequential Python program into a parallel PySpark program. Tyro identifies potential code fragments where parallelism can be done and translates them. It uses Abstract Syntax Trees (AST) for fragment detection and gradual program synthesis to convert the Python operations into PySpark operations. Tyro verifies the generated code against given user test cases. We evaluated Tyro by automatically converting different sequential Python programs into PySpark programs. The resulting PySpark programs performed up to 9x faster (on 9 parallel machines) compared to the original. These promising results indicate Tyro is another option for automatic code parallelization

Towards Containing COVID-19 Pandemic by Mining Knowledge from Scientific Literature and Social Media

Syeda Jannatus Saba, Biddut Sarker Bijoy, Souvika Sarkar, Md Saiful Islam, Sheikh Rabiul Islam, Ruhul Amin
University of Hartford, USA

Abstract: The ongoing pandemic, COVID-19, has been sweeping the world, affecting millions of people from 221 countries and territories, claiming 3.01 million deaths around the world in an unprecedented way. There is a pressing demand to elicit the unknown facts to contain and eradicate the virus globally. In this work, we apply a semi-supervised text-mining technique on the social media data and research paper archive (i.e., CORD-19 dataset), to uncover new information about the COVID-19 virus transmission. The discovered novel information, demonstrated by an enhanced Chain of Infection—an epidemiological term describing the infection process, provides valuable insights about the spread of the virus, which in turn can aid in developing improved policies towards containing the virus.

Effects of Alpha, Beta and Patch Size on Performance of Layer-wise Relevance Propagation

Lijun Sun, Douglas Talbert
Tennessee Technological University, Tennessee, USA

Abstract: N/A

Learning with Noisy Inconsistent Data

Eman Allogmani, Darsana Josyula
Bowie State University, USA

Abstract: Deep learning algorithms learn best from large amounts of data. The quality of data plays an important role in the performance of deep learning models. Mislabeled data instances existing in the training data often degenerate the performance of the classification models. Identifying and eliminating mislabeled training instances can noticeably improve the performance of the trained classifiers. We compare and contrast three different approaches for improving classifier performance in the presence of inconsistently labeled data. The approaches are based on detecting mislabeling and filtering them based on cross-validation losses during the training process. The prediction performance of classifiers trained using the three approaches on the MNIST dataset injected with different percentages of mislabeling – 0, 5, 10, 15, 20, and 25 – are evaluated against baseline CNN models. The results indicate that the majority-majority (MM) method performs better than the consensus-majority (CM) and the majority-consensus (MC) methods.

A Comprehensive Analysis of Air Pollution and Equity During COVID-19 in Los Angeles County

Kevin Marlis, Justine West, Dawn Comer, Irene Burga, Jeremy Taub, Chisato Fukuda Calvert, Jeanne Holm, Mohammad Pourhomayoun
California State University, Los Angeles, California, USA; City of Los Angeles, California, USA; openAQ, USA

Abstract: The COVID-19 global pandemic's impact on Los Angeles County is not evenly distributed across the region, where the socioeconomic characteristics of a neighborhood can inform the spread and deadliness of the virus and the rate of vaccination. This research consists of a data analysis of the neighborhoods within Los Angeles County by looking at socioeconomic characteristics and COVID-19 case, death, and vaccination rates, culminating in a COVID-19 equity hardship index that can identify the neighborhoods that are most socioeconomically disadvantaged, hit hardest by the COVID-19 pandemic, and slowest to receive vaccinations. This comprehensive research will contribute to understanding the consequences of COVID-19 on low-income communities where health and healthcare are at a premium.

An Approach to Bridge the Gap Between State-of-the-art Predictive AI Algorithms and Real-world Activity Forecasting Constraints

Anna Nesvijevskaia, Edoardo Aliprandi, Catherine Lesperance, Sophie Ouillade, Valentin Masdeu, Cyril Esnault, Jean-Daniel Zucker
Quinten, France

Abstract: The need to adapt human resources to the activity workload is omnipresent in most sectors, especially when the activity variability depends on exogenous factors. Artificial intelligence seems to offer unprecedented opportunities for improving forecasting. However, its application to real-world business contexts still reveals critical

breaches. In this paper, we review the state-of-the-art traditional and recent Machine Learning methods for predicting time series and highlight the main gaps between their capabilities and business constraints, illustrated through car claims forecasting for a French leading assistance call center. We then propose an original approach called HYLAF with proven performance and operationality, combining business expertise and a statistical graphical stagewise-inspired approach. It reminds that overperforming algorithms should not overshadow the operational purposes of a field solution.

Handling Concept Drift using LSTM based Autoencoders

Bander Allogmany, Darsana Josyula, Tagrid Alshalali
Bowie State University, USA

Abstract: We live in a dynamic world full of changes. Hence, static models will lose their effectiveness over time. Machine learning models suitable for such non-stationary environments must be adapted to deal with the changes. The changes in the underlying data distribution or data generation process are termed concept drift. In this paper, we propose the use of an integrated model that merges the Long-Short Term Memory (LSTM) network and autoencoder technique to better handle concept drift. We used three popular synthetic datasets that are under various types of concept drift to evaluate our model performance in dealing with concept drift. Experimental results show that our method is effective for handling concept drift in streams of data.

Detecting Informed Traders in the Financial Markets

Damian Etchevest, Sebastian Calzadilla, Louis Alvarez, Sean Mondesire
St. Thomas University, USA

Abstract: N/A

Using Social Signals Predict Bitcoin Price Fluctuations

Tanzhou Liu, Wei Zhang, Yilmazcan Ayurt
Nanjing University, China

Abstract: Bitcoin is a cryptocurrency with large price uctuations. In our analysis, we add social signals related to public attention (Google Trends data) and sentiments from the word of mouth (Tweets) to test whether they have e_ect on Bitcoin price uctuations. The sentiments of tweets are from people who have many followers and actively tweet about Bitcoin, called Bitcoin inuencers. In addition, we also present the empirical _ndings of network characteristics of inuencers. Our analysis reveals that public attention has slight and insigni_cant impact on price volatility and anticipation emotion from inuencers has the greatest impact reaching 10% of the price change. Hence, we conclude that the sentiment constructed from these inuencers provides a more robust predictor of Bitcoin price uctuations than the general public attention.

Data Mining Product Cross-Selling Association Rules from Jewelry Sales Order Data

Sherrene Bogle, Robert Airth
Humboldt State University, USA

Abstract: This paper outlines the process of mining a large set of product sales data to create association rules for cross-selling. These association rules could then be used to provide insight into customer purchase patterns. Once purchase patterns are established, the association rules can provide statistical insight into which products would likely

be purchased along with a set of items. The results showed that support, lift, conviction, and confidence of the {Earrings, Bracelets} ⇒ {Necklaces} rule stayed consistently high throughout each mining attempt, thus revealing a good cross-selling opportunity.

Predicting the Progression in Interstitial Lung Disease Using Computer Tomographic Scans with Transfer Learning

Frank Liu, Olajide Salawu, Bing Zhou (Qingzhong Liu)
Sam Houston State University, USA

Abstract: The Interstitial lung disease (ILD) can cause stiffness in the lungs which makes it difficult to breathe and to get oxygen to the bloodstream. Lung damage from ILDs is often irreversible and gets worse over time. In this paper, we develop a method to predict the progression in ILD by using computer tomographic scans with fine-tuning transfer learning approach. The experiment shows the effectiveness of proposed approach, and it would also improve the detection of severity which would also help with treatment trial designs and the development of novel treatment.

Tumor Detection and Segmentation using Deep Learning Models

Frank Liu, Emeka Kanikwu (Qingzhong Liu)
Sam Houston State University, USA

Abstract: The detection and segmentation of brain tumor is critical for accurately diagnosing and effectively treating patients. To automate and standardize this task, we apply multiple convolutional neural networks (CNN) to the Multimodal Brain Tumor Segmentation Challenge (BraTS) 2020 training dataset with a comparison study. We utilized four CNN models for detection and segmentation. The models include the 3D U-Net, V-Net, spatial-channel Attention Gate (scAG), and a modified 3D U-NET Convolutional Neural Network. Meanwhile, different hyperparameters, optimization techniques are compared. Our study shows that CNN-based models are effective and very promising to realize the detection and segmentation of brain tumor.

Global Production Database Ready for the Large ATLAS Particle Physics Project

Marek Beranek, Vladimir Kovar, Vaclav Vacek
Unicorn University, Prague, Czech Republic

Abstract: N/A

Detection of Seam-carving Forgery with Transfer Learning

Frank Liu, Alberto C. Delgado (Qingzhong Liu)
Sam Houston State University, USA

Abstract: In image forensics, the forgery detection on JPEG images is a hot spot in the field. Seam carving was originally designed for content aware image resizing. It is also being used for forgery manipulation. It is still very challenging to effectively identify the seam carving forgery in JPEG images since the original manipulation traces were compressed and compromised. This research investigates the use of convolution neural network models for seam-carving forgery detection in JPEG images. The models evaluated in this research are EfficientNet-B0, E_cientNet-B1, and EfficientNet-B2. The models accuracy, precision, and F1 score were recorded during training and validation to evaluate their performance. Experimental results show that transfer learning is effective and promising to discriminate manipulated JPEG images after seam-carving from the untouched.

Exploring Deep Learning to Improve Compton Camera Based Prompt Gamma Image Reconstruction for Proton Radiotherapy

Gerson Kroiz, Carlos Barajas, Matthias Gobbert, Jerimy Polf
University of Maryland, Baltimore County, Maryland, USA

Abstract: Proton beam radiotherapy is a cancer treatment method that uses proton beams to irradiate cancerous tissue while simultaneously sparing doses to healthy tissue. In order to optimize radiational doses to the tumor and ensure that healthy tissue is spared, many researchers have suggested verifying the treatment delivery through real-time imaging. One promising method of real-time imaging is through a Compton camera, which can image prompt gamma rays emitted along the beam's path through the patient. However, the images reconstructed with modern reconstruction algorithms are often noisy and unusable for verifying proton treatment delivery due to limitations with the camera. This paper demonstrates the ability of deep learning for removing false prompt gamma couplings and correcting the improperly ordered gamma interactions within the data for the case of Double events.

Novel Semi-parametric Tobit Additive Regression Models

Hailin Huang
George Washington University, USA

Abstract: N/A

Data Science in Procurement - a case study of a wind turbine manufacturer

Manuel Lutz, Niklas Stepanek
Technical University of Munich, DE, Germany

Abstract: The digital transformation within makes it possible to uncover previously undiscovered potential in procurement and to tap into it. In this context, the application of Data Science approaches has gained increased attention. While such approaches can lead to significant cost reductions on the part of both suppliers and buyers, their use is anything but trivial. This paper presents a case study of a wind turbine manufacturer determines the application possibilities in order to achieve procurement goals like transparency of the own commodities and cost savings. Management recommendations are to be derived by bringing together scientific as well as practical approaches in industry.

Optimization of Programmatic Advertising Supply Chains using Evolutionary Algorithms

Tanusha Goswami
MiQ, India

Abstract: Supply Path Optimization (SPO) – the process of selectively choosing the intermediaries involved in a real time programmatic auction for an online advertisement slot to improve campaign performance – has mostly been a heuristic one. It is based on traditional relationships between buyers and sellers, fragmented datasets, and industry biases. Extensive research in this area leveraging big data and implementing machine learning algorithms is scarce. With multiple interactive components, this becomes an Optimization problem where evolutionary algorithms seem appropriate to apply. Hence, this paper proposes the use of the Genetic Algorithm to employ a data driven approach towards SPO using the IAB's SupplyChain Object and historical advertising campaign data. This method results in consolidated supply partners (quality control) and a 60% reduction in current costs (cost efficiency).

An Analysis of the Relationship Between Household Income and School Prestige on Student Income

Christina Sherpa, Jiselle Ramirez, Naomi Giancola
Smith College, USA

Abstract: Many studies in the past have set out to examine the relationship between household income and college turnout; however, not many studies have looked at the impact of household income and tier of college on student income post-graduation. In this study, using the data collected by Opportunity Insight on college income segregation and social mobility, we examined the relationship between mean parent income, tier/prestige of the school, and student income. Our dataset has information on US-accredited colleges and universities(n = 2,199). We hypothesized that the students whose families have a higher income and students who attend schools of higher prestige will have a higher post-graduation income on average. Our analysis showed that both parents' income rank and college tier are significant predictors of post-graduation student income. Additionally, we found that the model with both predictors was able to explain the most variability (had the highest Rˆ2), meaning this model is the best at predicting the post-graduation income. Based on our findings, we found evidence to conclude that there is a positive linear correlation between mean parents income, prestige of school and student income.

Imitation Learning for VM Placement Problem Using Demonstration Data Generated by Heuristics

Sangyoon Oh, Seungjun Lee, Daegun Yoon
Ajou University, Korea

Abstract: Data centers are key components of cloud computing to run virtual machines. For saving the cost to operate data centers, it is important to decide how to allocate each virtual machine to a certain physical machine. Because the virtual machine placement problem is NP Hard, there are many heuristics to obtain near-optimal solutions as quickly as possible. The reinforcement learning technique can be applied for virtual machine placement problem. However, if the problem size gets bigger, the convergence speed of reinforcement learning gets slower. The possible solution is that the agent imitates the behavior of given demonstration, called imitation learning. In this paper, we propose a method combining reinforcement learning with imitation learning. In our proposed approach, demonstration data is generated by simple heuristics not human experts.

Making Easy to Replicate Dishes from Free Format Descriptions on Recipe Sharing Sites

Hiroshi Uehara, Takayasu Yamaguchi, Miu Shibata
Rissho University, Japan

Abstract: This study aims to acquire complementary information for repli-cating dishes based on cooking instructions on recipe sharing sites. Because the descriptive styles of the cooking instructions are considerably variated depending on each individual posting of a recipe, partial omissions of the in-structions often happen so that users cannot correctly replicate the dishes. To address the issue, we try to detect essential structures among the variated in-structions describing the same cooking step, then, collect descriptions that semantically complement those structures. Acquired structures with the com-plementary descriptions for each cooking step are found to provide efficient information for complementing the partial omissions in cooking instructions.

Evaluating a Semi-Supervised Intrusion Detection Algorithm Through Benford's Law

Innocent Mbona, Jan H. P. Eloff
University of Pretoria, South Africa

Abstract: The CIC-IDS2017 and CSE-CIC-IDS2018 are state-of-the-art network intrusion data sets. They are typified by high dimensions and high volumes of benign network traffic and low volumes of network intrusion attacks, including sophisticated brute-force attacks. The imbalance and high-dimensional nature of such big data sets challenge machine learning algorithms used for network intrusion detection, particularly for feature selection. This paper adopts Benford's Law as a feature selection method, since current research shows it can detect network intrusion attacks. The method was chosen based on observations about the frequency distribution of leading digits of network traffic data records. The paper investigates whether features identified by Benford's Law significantly impact the performance of a popular semi-supervised anomaly detection Gaussian Mixture Model.

SESSION:
DATA SCIENCE, BIG DATA ANALYTICS AND APPLICATIONS

Chair: Prof. Hamid R. Arabnia
Department of Computer Science, University of Georgia, USA

Patterns of Information in Financial Dynamical Systems

Wenzhi Cai, Amir Assadi
University of Wisconsin, Madison, Wisconsin, USA

Abstract: Analysis of time-series in financial markets poses numerous challenges and many possibilities. Viewing a time-series as part of a dynamical system naturally lead us to explore the influence of various parameters and their relative significance in driving the dynamics. In the following article, we construct a (smooth) dynamical system from a discrete time series, and investigate its dynamics. We construct a quantitative measure for implicitely encoded patterns of information dynamics for the foreign currency exchange rates time series [Japanese/U.S. exchange rates]. Further, we demonstrate its usefulness for better understanding the influence of various parameters and their relative significance in driving the system. An extension of Principal Components Analysis (PCA) helps us identify tumultuous week in the market for Japanese/U.S. exchange rates. Our method facilitates analysis of the dynamics before and after the said tumultuous week. Analysis of information dynamics shows regular behavior in the market before the .catastrophic. week and greater uncertainty afterwards. In summary, our research presents a new scheme to evaluate the behavior of international currency markets and the significance of .catastrophic. events on these markets.

Optimization of Smoking Classification by Applying Neural Network with Variable Importance Using Cytokine Biomarkers

Seema Singh Saharan, Pankaj Nagar, Kate Townsend Creasy, Eveline O. Stock, James Feng, Mary J. Malloy, John P. Kane
Department of Statistics, University of Rajasthan, Jaipur, India; Voluntary Data Scientist UCSF Kane Lab, San Francisco, California, USA; University of California Berkeley Extension, California, USA; Cardiovascular Research Institute, Department of Medicine, University of California, San Francisco. California, USA; Cardiovascular Research Institute, Departments of Medicine and Pediatrics, University of California, San Francisco. California, USA; Cardiovascular Research Institute, Department of Medicine, Department of Biochemistry and Biophysics, University of California, San Francisco, California, USA

Abstract: N/A

Estimation of Average Annual Daily Bicycle Count Using Bike-Share GPS Data and Bike Counter Data for an Urban Active Transportation Network

Marzi Rafieenia, Liza Wood, Mohsen Zardadi, Scott Fazackerley, Ramon Lawrence
Department of Computer Science, University of British Columbia, Kelowna, BC, Canada

Abstract: In 2018, the City of Kelowna entered into a license agreement with Dropbike to operate a dockless bike-share pilot in and around the downtown core. The bikes were tracked by the user's cell phone GPS through the Dropbike app. The City's Active Transportation team recognized that this GPS data could help understand the routes used by cyclists which would then inform decision-making for infrastructure improvements. Using OSMnx and NetworkX, the map of Kelowna was converted into a graph network to map inaccurate, infrequent GPS points to the nearest street intersection, calculate the potential paths taken by cyclists and count the number of trips by street segment. Combined with the data from four counters around downtown, a mixed effects statistical model and a least squares optimization were used to estimate a relationship between the different traffic patterns of the bike-share and counter data. The relationship was used to estimate and visualize the annual daily bicycle volume in downtown Kelowna. The visualization helped to understand how the bike network was being used, including using laneways and crossing highways.

Prediction of Days-On-Market for Single-Family Homes in the Housing Market of Southeast Georgia

Keagan Galbraith, Ray R. Hashemi, Omid M. Ardakani, Jason S. Beck
Department of Computer Science, Honors College, Georgia Southern University, Statesboro, Georgia, USA;
Department of Computer Science, Georgia Southern University-Armstrong Campus, Savannah, Georgia, USA;
Department of Economics, Georgia Southern University-Armstrong Campus, Savannah, Georgia, USA

Abstract: The number of days that a home stays on the housing market (Days-On-Market—DOM) provides crucial information at both microlevel (behavior associated with the buyer's/seller's decision) and macro level (risk associated with real estate investments and also housing bubbles' identification). Housing data has a mixture of simple and complex attributes. A complex attribute in contrast with a simple attribute, has an array of values for a real estate property, which creates a major challenge in prediction of DOM. The goal of this research effort was: (a) Providing for complex attributes in DOM's prediction, (b) Analyzing, designing, and implementing a DOM prediction's package using Naïve Bayesian and Linear Regression separately, and (c) Establishing the superiority and robustness of the underline models.

Predicting Traffic Accident Severity with Deep Neural Networks

Meghan Bibb, Pablo Rivas
School of Engineering and Computer Science, Department of Computer Science, Baylor University, Texas, USA

Abstract: Traffic accidents can be studied to mitigate the risk of further events. Recent advances in machine learning have provided an alternative way to study data associated with traffic accidents. New models achieve good generalization and high predictive power over imbalanced data. In this research, we study neural network-based models on data related to traffic accidents. We begin analyzing relative feature colinearity and unsupervised dimensionality reduction through autoencoders, followed by a dense network. The features are related to traffic accident data and the target is to classify accident severity. Our experiments show cross-validated results of up to 92% accuracy when classifying accident severity using the proposed deep neural network.

Modern Cybersecurity Measures for Organizations

Fadi Al-Ayed
Independent Researcher, Culpeper, Virginia, USA

Abstract: To enhance current cybersecurity models, the recently most popular MITRE ATT&CK which is reached out to provide better accuracy for organizations. The adjustments are repackaged into what is alluded to as the Mightier Attack Model. This model not just presents the particular reaction steps of the business in averting away hacker's step-by-step move, but also gives an estimation of likely cost to the organization. The inclusion of these highlights presents expanded practicality and relevance to business concerns as its assts focus on preventing hacker's mischievous exploits. The cost estimation of the hacker's attempts is intended to give businesses a means to determine a cost-benefit analysis of its cybersecurity systems.

Shap-Based Study for Heart Disease Prediction

Casey Legeret, Corey Dutkiewicz, Camron Grant, Devin Yang, Darius Banks,
Dale Kauffman, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA

Abstract: To understand how Machine Learning is used to gives us a prediction on heart disease, one must first understand what the data is, how the variables within the data are important and how they are used to create results. We have developed a step-by-step guide, where we show what the dataset looks like, explain each variable including the relevance each has with heart disease. We show the data using a tree diagram which shows important relationships, along with using a confusion matrix to validate the diagram. After these initial steps, and your understanding of the data is clear, we will use Machine Learning with SHAP to show and explain the impact individual variables have on different genders.

Applying Different Machine Learning Algorithms to Heart Disease Data

Jonathan Esho, George Feliciano, Jackie Gates, Jessica Lafraugh, Fabian Isho,
Dylan Idzkowski, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA

Abstract: This research aims to predict the risk of heart disease in a person using Machine Learning algorithms. We will predict such a disease using the four following algorithms: KNN, CatBoost, Linear Regression and Decision Tree Classifier. With these four algorithms, we can make an accurate prediction on how heart disease correlates between a person's age and gender. We have concluded that the most accurate algorithm is the KNN algorithm. This algorithm has a high accuracy of 88% when presented with a reasonable sized dataset such as the heart disease dataset we are working with. We concluded that any of these algorithms would suffice when working with any dataset.

A Study of Outliers Detection Techniques in Heart Disease Data

Paul Albrecht III, Kiratdeep Kahlon, Kristian Simovski, Micahel Reed, Patrik Lekocaj,
Lucas Schleben, Jordan Zavich, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA

Abstract: In this paper, we have discussed what an outlier is and the causes of outliers, as well as the multiple methods used for finding outliers including Hypothesis Testing (Grubbs Test), Z-Score, Robust Z-Score [11][12], I.Q.R, Winsorization (Percentile Capping) [3], DBSCAN [13] and Isolation Forest methods. We will then compare each of these methods to determine which is best for finding outliers [7][8] in the dataset [16], leading to better predictions from Machine Learning. Out of all the methods to finding the outliers [7][8], the best method is the Winsorization

Method (Percentile Capping) [3]. You will learn more about these methods and why the Percentile capping method is the best method to use for our predictions within the next few pages.

Using Deep Learning to Optimize Software Define Radio in Smart City and Health Care

Khalid Amen, Mohamed Zohdy, Mohammed Mahmoud
Department of Electrical and Computer Engineering, Oakland University, Rochester, Michigan, USA

Abstract: Software Defined Radio is a single radio that can be by simply reconfiguring the radio with different software in Internet of Things Devices in Smart City. In this paper, we will present how Deep Learning is used to optimize SDR in terms of flexibility, analytics and real-time data. We will also describe a method for uniquely identifying a specific radio, among nominally similar to IOT devices, using a combination of SDR sensing capability and Machine Learning techniques. The key benefit of this approach is that Machine Learning operates on raw I/Q samples and distinguishes devices using only the transmitter hardware-induced signal modifications, that serve as a unique signature to identify large number of devices at distances of 100-200 ft.

Decision Tree Classifier Based Model for Disease Prediction

Andrew Alhaj, Arianna Fuoco, Adam Komeshak, Camron Farida, Caleb White, Mohammed Mahmoud
Department of Computer Science and Engineering, Oakland University, Rochester, Michigan, USA

Abstract: According to the Centers for Disease Control and Prevention (CDC), diseases such as malaria, diarrheal infections, meningitis and HIV/AIDS heavily affect developing countries. Research conducted previously used many different approaches; some relying on multiple Machine Learning models, while others relied on using a single model. In this research, we analyzed our dataset using the Decision Tree Classifier model in Python. Our goal is to see if we could find any correlation between age and diseases. We wanted to know which diseases had the biggest effect on each age group. With our findings, we were successfully able to identify that disease outbreaks were very prevalent in children ages 10 and below.

Quantifying the Performance of Interval Merging Binary Trees using a Matrix Representation

Istvan Finta, Lorant Farkas, Sandor Szenasi
Nokia Bell Labs, Budapest, Hungary;
Obuda University, Budapest, Hungary

Abstract: Interval Merging Binary Tree is a binary tree where each node maintains the endpoints of an interval, rather than a single value. It shows certain advantages in distributed/networked environments with packet loss and duplication, when packets have incremental sequence numbers. Assessing the applicability of the data structure to detect losses and duplications is a complex task since typical performance metrics depend not only on the total number of arrived/managed keys, but also on the statistical distributions of the keys of arrived packets in the space of sequence numbers. In this contribution the matrix representation of the IMBT is introducted in order to overcome this complexity and be able to assess its performance in comparison to other well known data structures and formulate a sufficient condition on the dynamic equilibrium of IMBT.

Epidemic Analysis of Covid-19 in California

Marisabel Chang-Chan, Lan Yang
Computer Science Department, California State Polytechnic University, Pomona, California, USA

Abstract: The outbreak of novel coronavirus-caused pneumonia (COVID-19) has been impacted worldwide specially United States. By April 2021, it has been more than one year since COVID-19 pandemic appeared in the US. During this period of time, good sizes of datasets have been built up by recording infection cases, and many researchers have performed diverse COVID-19 data analyses to try to predict and explorer epidemiological data. In this paper, we explore the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, to perform a detailed analysis of COVID-19 cases from January 26th, 2020 to January 25th, 2021 in California. We observe the impact of the outbreak during holiday season (Thanksgiving, Christmas, and New Year), utilize a SEIR model to observe the impact of the social distancing and stricter mitigation measures on death and infected cases, and demonstrate that if the control rate were higher the COVID-19 death and infected cases would have been under better control.

Metadata Provenance for Analysis Applied to Ecological Niche Modelling

Agnei C. Silva, Pedro L. P. Correa, Cleverton Borba
University of Sao Paulo, Brazil;
Andrews University, USA

Abstract: The search for knowledge causes the scientific community to look not only for new ways and strategies of improving the quality of scientific experiments, but also to reduce time and costs required for its execution using computational distributed environments. The management of provenance descriptors collected throughout scientific experiments life cycle represents a preponderant facet in this new scenario. This paper aims to present a computational strategy to support the discovery and reuse of data in the area of biodiversity. It is supported by a computational architecture for the management of provenance metadata throughout the data life cycle. A case study about an experiment for ecological niche modeling is presented. The results show that this approach is effective in collecting and storing metadata and allows the scientific community to query the provenance database in order to enable the understanding of the experiment, the data and the data products (results) generated by the experiment

Link Prediction in Unipartite and Bipartite Networks Influenced by Six Degrees of Separation

Purushottam Kumar, Dolly Sharma
Department of Computer Science and Engineering, Shiv Nadar University, Delhi NCR, India

Abstract: Recently, network science has gained popularity in the research community and the reason for its popularity is its application in various domains. Many real-world complex systems can be represented using networks and prediction of links in the network holds significant importance in network analysis. In this paper, we present a novel similarity metric for link prediction. The proposed algorithm uses the concept of six degrees of separation and the total number of possible walks between node pairs. We did exhaustive experiments on 13 disparate real-world network datasets. To test the performance of the proposed algorithm, we compared the proposed algorithm with 14 baseline algorithms. Comparison results show that the proposed algorithm outperforms the baseline algorithms based on four given metrics.

Clustering Automatic Identification System (AIS) Data using Density-based Spatial Clustering of Large and Complex Datasets on Pairwise Distance Matrix

Cheronika Manyfield-Donald, Tor A. Kwembe, Jing-Ru Cheng
Jackson State University, Jackson, Mississippi, USA;
Engineer Research and Development Center, U.S. Army Corps of Engineers, Vicksburg, Mississippi, USA

Abstract: In this paper, we introduce a 2-D points data collection system developed using Automatic Identification System (AIS) data that will aid in monitoring maritime traffic and directly assist in adverting accidents, tracking vessels, and support in avoidance of dangerous environments. Density-based spatial clustering of Applications (DBSCAN) using a pairwise distance matrix and the Haversine distance function are methods we used to find core points in relation to a vessel and its outliers. This will show AIS equipped vessels/objects in an inputted radius to a given Latitude/Longitude coordinate pair and identifies anomalies or what could possibly be other AIS equipped vessels/object with erratic behavior.

Multi-Objective Quantum-Behaved Particle Swarm Optimization Training Feed-Forward Neural Network

Pei-Yau Weng, Gilbert S. Young
Department of Computer Science, California State Polytechnic University, Pomona, California, USA

Abstract: Multi-objective Quantum behaved Particle Swarm Optimization (MOQPSO), a kind of swarm optimization algorithm, directly trains Feed-forward Neural Network without using Backpropagation. MOQPSO does not have gradient error problems, such as gradient vanishment and gradient explosion and get rids of overfitting and training error by objective functions. MOQPSO is composed of the double-well Dirac delta function, Pareto-optimality, and the mesh refinement Compared to Backpropagation, MOQPSO applies on the parallel computing and is less likely to trap in the local minimum. Finally, the result performs that MOQPSO has higher bitch size and faster convergence than Backpropagation on Iris classifier.

Smoking Classification using Novel Plasma Cytokines by Implementing Machine Learning and Statistical Methods

Seema Singh Saharan, Pankaj Nagar, Kate Townsend Creasy, Eveline O. Stock,
James Feng, Mary J. Malloy, John P. Kane
Department of Statistics, University of Rajasthan, Jaipur, India;
UCSF Kane Lab, San Francisco, California, USA;
University of California Berkeley Extension, California, USA;
Cardiovascular Research Institute, Department of Medicine,
University of California, San Francisco, California, USA;
Cardiovascular Research Institute, Departments of Medicine and Pediatrics, University of California, San Francisco, California, USA; Cardiovascular Research Institute, Department of Medicine, Department of Biochemistry and Biophysics, University of California, San Francisco, California, USA

Abstract: Smoking is a major cause of premature and preventable death. It has a detrimental effect on many human organs. Illnesses such as COPD, arteriosclerotic disease, cancer, diabetes etc. are linked to tobacco exposure. Cytokines are inflammatory biomarkers that are mechanistically associated with smoking. Machine Learning algorithms allow for the quantitative assessment of the contributions of individual cytokines to tobacco-related disease. The mapping of cytokines to disease can facilitate and direct treatment modalities. By the application of k-NN and Random Forest Machine learning algorithms on 46 plasma cytokines we will demonstrate the classification of smoking induced disease. To ensure optimal results, performance improvement techniques like bootstrap sampling

and hyper parameter tuning are employed. Efficiency of separability achieved by the models is evaluated using the Area Under the Receiver Operating Characteristic (AUROC) metric. The most significant cytokines that enabled prediction are highlighted. Using two sample independent t test, we compare the cytokine biomarkers that are statistically significantly different for smokers versus nonsmokers. The discovery and transference of novel biomarkers such as cytokines from the platform of molecular investigation to clinical practice, can facilitate precision medicine-based therapeutic interventions.

Logistic Regression and Statistical Regularization Techniques for Risk Classification of Cardiovascular Disease using Cytokines Transported by High Density Lipoproteins (HDL)

Seema Singh Saharan, Pankaj Nagar, Kate Townsend Creasy, Eveline O. Stock, James Feng, Mary J. Malloy, John P. Kane
Department of Statistics, University of Rajasthan, Jaipur, India;
UCSF Kane Lab, San Francisco, California, USA;
University of California Berkeley Extension, California, USA;
Cardiovascular Research Institute, Department of Medicine, University of California, San Francisco, California, USA; Cardiovascular Research Institute, Departments of Medicine and Pediatrics, University of California, San Francisco, California, USA; Cardiovascular Research Institute, Department of Medicine, Department of Biochemistry and Biophysics, University of California, San Francisco, California, USA

Abstract: Cardiovascular disease (CVD) is a leading cause of mortality in the world. It is paramount to be able to detect the severity of the disease proactively, using novel biomarkers like cytokines that are indicators of inflammation. Atherosclerosis, the primary cause of CVD, is an inflammatory disease involving cytokines. Identifying impactful cytokine species can advance diagnosis and personalized treatment. It has been hypothesized that HDL transports cytokines. Therefore, it is vital to explore the roles of HDL-borne cytokines in vascular inflammation. Machine Learning algorithms are harnessing pioneering research from the standpoint of precision medicine. Their technological prowess can materially enable the translation of scientific research to clinical practice. In this research we implemented logistic regression and the derived regularized techniques using multidimensional cytokine biomarkers with the objective of identification of individuals "at risk" for CVD. These techniques were further empowered by k-fold cross validation and hyper parameter tuning. The identification and quantification of cytokines transported by HDL provide novel mechanistic insights that can inform the assessment of risk and therapeutic intervention in atherosclerotic vascular disease.

Trends and Challenges of Clustering Techniques in Complex Networks Analysis

Matin Pirouz
Department of Computer Science, California State University, Fresno, California, USA

Abstract: Complex network analysis has revolutionized various discplines including how businesses perceive customer habits and how social networks form from the ground up. In this study, we review the state-of-the-art literature in clustering techniques. In particular, we propose a nodel taxonomy by dividing the research in this field into five classes of modularity models, label propagation models, clique models, stochastic block models, and neighbor influence models. This enables us to provide a comparative analysis of their performances and limitations with respect to time and space complexities. We conclude the research with the limitation faced by the existing methods and how future research can improve upon such limitations.

Quantitatively Examining the Relationship between Social Media Messages and the Risk Management at Financial Institutions

Issayas Haile, Yanzhen Qu
School of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA

Abstract: This study addresses the relationship between social media text themes and the outcome of the bank risk rate. The study demonstrates how financial institutions can use opensource intelligence (OSINT) to manage their risks with the help of text analytics. A statistical information is used to categorize different types of risks from twitter data that affects financial institutions. The twitter data collected were 2019 data that came from different sources, mainly news outlets and U.S. Government agencies that focused on major U.S. financial institutions. These data collected from twitter were analyzed with the use of IBM SPSS software to find correlation between the themes of social media text and bank's risk rate. Social media text themes are the dependent variable with three discrete independent variables: competitive risk, performance risk, and compliance risk. This research indicates data collected from twitter shows how opensource intelligence can be utilized to manage risk in financial institutions with the help of text analytics. Keywords– Text Analytics, OSINT, Big Data, Social Media, Risk Management, and Bank.

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

Damian Martinez Diaz, Francisco Javier Luna Rosas
Departamento de Sistemas y, Computacion Instituto Tecnologico de Aguascalientes, (ITA), Aguascalientes, Mexico

Abstract: Depression is on the rise globally according to the World Health Organization (WHO) and has affected more than 322 million people around the world, in the most severe cases, if left untreated it can lead people to suicide. Experts say that one of the best ways to prevent depression is to listen to the people who are close to them, social networks such as Twitter or Facebook are in a unique position to help these people to connect in real time in difficult situations, but also represents a potential risk to receive information that could later be harmful. In this research, we propose a model to optimize global time processing in detecting depression-related patterns on the social network Twitter. The results show that the proposed model can be a good alternative when it comes to optimizing response time in this type of problem.

Random Database Creation and Use in a Context-Free Grammar Data Generator

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

Abstract: The Data Generation Language (DGL) is (or was) a popular tool for generating program test data. It has also been used as a simulation tool for verifying the correctness of VLSI simulators and for testing VLSI circuits. We are currently in the process of upgrading DGL to make it more useful in modern programming environments. One of the many new features is the ability to obtain test data from database tables. In addition, it is now possible to create randomly generated databases that combine random data with solid structure to test specific program features. This upgrade should prove useful for testing a wide variety of modern software.

A Priori Unification vs A Posteriori Composition for Better Data Science

Ravishankar Nair
Chief Consultant, PassionBytes LLC, Jacksonville, Florida, USA

Abstract: In the past where the data was considerably manageable and data sources were minimal, stand-alone applications for machine learning or AI was just enough to handle them effectively. From monolithic data sources of past, we are now in the age of polyglot persistence. The advancements in the world of machine learning and artificial intelligence demanded the use of more and more data thus generating a myriad of architecture and software applications for distributed computing like map-reduce, Spark and Flink etc. However, computing is different from data processing and fetching data from multiple data sources into these distributed applications poses a challenge – the need to join data from polyglot sources causes shuffling and huge data transfer between the nodes where these distributed systems are running. This causes the distributed systems degradation in performance and memory exceptions. The paper presents an innovative way of creating a virtualization layer for data unification prior to the consumption by the compute nodes termed "a priori unification".

Prediction of Number of Personnel to Deploy for Wildfire Containment

Olivia Samples, Todd Gary, Qingguo Wang
College of Computing and Technology, Lipscomb University, Nashville, Tennessee, USA;
School of Applied Computational Sciences, Meharry Medical College, Nashville, Tennessee, USA

Abstract: Wildfire size, frequency, severity, and associated fatalities have surged at an alarming rate in recent years, particularly throughout the western United States, resulting in steep budget increasesfor fire suppression. According to the U.S. Forest Service, a subdivision of the U.S. Department of Agriculture, the annual budget allocated to wildfire suppression has more than tripled, increasing from 16% to 52% between 1995 and 2015, exceeding $2 billion in 2017. The ability to assign personnel, equipment, and other resources to a fire, in a timely manner, is vital in suppressing fire, reducing costs, and saving lives. This is especially true when national, state, and local budgets and capacity constraints are considered. In this paper, we use gradient boosting decision trees to predict the number of personnel needed to effectively fight a typical wildfire. By combining the U.S. wildland fire incident data and weather data, our model obtained a coefficient of determination (R2) of 77.78%, which marks a significant improvement over existing solution. Our model can potentially be used as a rapid-response resource by those units who coordinate and address wildfires.

Data Coverage Analysis and Improvement of Real-Time Twitter Streaming

Min-Seon Kim, Hyuk-Yoon Kwon
Department of Industrial Engineering, Seoul National University of Science and Technology, Seoul, Korea;
Department of Industrial Engineering and Research Center for Electrical and Information Technology, Seoul National University of Science and Technology, Seoul, Korea

Abstract: Recently, as real-time data gathering becomes attention, the importance of new architecture for large-scale real-time data collection, ingestion, and storing has been emphasized. Due to the nature of streaming data, which are generated with a fast speed, the data coverage to handle a large portion of the streaming data without loss is a challenging issue. In this study, we deal with a problem to resolve the coverage limitation for collecting real-time Twitter streaming data. For this, we propose a framework working with Apache Kafka and Apache Spark Streaming to maximize the data coverage of real-time twitter streaming.

Search Algorithms for Automated Hyper-Parameter Tuning

Leila Zahedi, Farid Ghareh Mohammadi, Shabnam Rezapour, Matthew W. Ohland, M. Hadi Amini
Knight Foundation School of Computing & Information Sciences, Florida International University, Miami, USA; Sustainability, Optimization, and Learning for InterDependent Networks Laboratory (solid lab), Florida International University, Miami, USA; Department of CS, University of Georgia, Athens, Georgia, USA; Enterprise and Logistics Engineering, Florida International University, Miami, Florida, USA; School of Engineering Education, Purdue University, West Lafayette, Indiana, USA

Abstract: Machine learning is a powerful method for modeling in different fields such as education. Its capability to accurately predict students' success makes it an ideal tool for decision-making tasks related to higher education. The accuracy of machine learning models depends on selecting the proper hyper-parameters. However, it is not an easy task because it requires time and expertise to tune the hyper-parameters to fit the machine learning model. In this paper, we examine the effectiveness of automated hyper-parameter tuning techniques to the realm of students' success. Therefore, we develop two automated Hyper-Parameter Optimization methods, namely grid search and random search, to assess and improve a previous study's performance. The experiment results show that applying random search and grid search on machine learning algorithms improves accuracy. We empirically show automated methods' superiority on a real-world educational data (MIDFIELD) for tuning HPs of conventional machine learning classifiers. This work emphasizes the effectiveness of automated hyper-parameter optimization while applying machine learning in the education field to aid faculties, directors', or non-expert users' decisions to improve students' success.

Practical Graph Refinement for Fast Enumeration of

Size-Bounded Maximal Cliques

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

Abstract: In this paper, we are concerned with a task of enumerating maximal cliques, for a given graph, with size no less than a designated threshold. For fast enumeration, we present a practical method which refines the original graph into a shrunken graph from which the same maximal cliques can still be extracted, where the refinement is carried out as a preprocess for the enumeration task. Although the method is simply based on fundamental properties of cliques, it can often bring us drastic reductions in computation times for our enumerations. Our experimental results for SNAP benchmark graphs show actual effectiveness of the proposed method.

The 22nd International Conference on Internet Computing & IoT (ICOMP 2021)
https://www.american-cse.org/csce2021/conferences-ICOMP

Fog Computing Strategy for Assembly Using Reconfigurable Hardware

Victor Lomas-Barrie, Mario Pena-Cabrera, Tamara Alcantara-Concepcion
Instituto de Investigaciones en Matematicas Aplicadas y en Sistemas, Universidad Nacional Autonoma de Mexico, Mexico; Direccion General Computo y de Tecnolgias de Informacion y Comunicacion, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico

Abstract: A strategy based on fog computation is presented for the assembly of manufacturing pieces using pattern recognition algorithms implemented in re configurable hardware. The strategy is based on the fact that each part of the digital machine of the recognizer can execute the tasks completely or independently. Thus, each task can be run on different FPGA cards to balance work across multiple collaborative manufacturing cells in case communication is disrupted.

Internet of Things Cyberthreats Assessment

Ivan Ortiz-Garces, Aaron Echeverria
Universidad de Las Americas, Quito, Ecuador

Abstract: One of the emerging technologies is the Internet of Things, this technology does not have an adequate level of security, so they have been exposed to various attack vectors, vulnerabilities, and high risks. This paper will perform an assessment of cyber threats surrounding Internet of Things (IoT) ecosystems. An experimental environment will be conducted with IoT devices where the risk of vulnerabilities detailed in the OWASP IoT Top 10 project and various cyberattacks will be analyzed. In the IoT ecosystem, the secure communication protocols will be applied to reduce the risks of the IoT ecosystem. Again, the risks will be evaluated to know their weighting and if the IoT ecosystem is suitable to work in operational environments.

Sentiment Analysis of Long-term Social Data during the COVID-19 Pandemic

Sophanna Ek, Marco Curci, Xiaokun Yang, Beiyu Lin, Hailu Xu
Department of Computer Engineering & Computer Science, California State University, Long Beach, CA, USA; College of Science and Engineering, University of Houston, Clear Lake, Texas, USA; Department of Computer Science, University of Texas, Rio Grande Valley, Texas, USA

Abstract: The COVID-19 pandemic has bringing the "infodemic" in the social media worlds. Various social platforms play a significant role in instantly acquiring the latest updates of the pandemic. Social media such as Twitter and Facebook produce vast amounts of posts related to the virus, vaccines, economics, and politics. In order to figure out how public opinion and sentiments are expressed during the pandemic, this work analyzes the long-term social posts from social media and conducts sentiment analysis on tweets within 12 months. Our findings show the trend topics of long-term social communities during the pandemic and express people's attitudes towards progress of major actions during the pandemic. We explore the main topics during the prolonged pandemic, including information surrounding economics, vaccines, and politics. Besides, we show the differences in gender-based attitudes and propose future research questions refer to the "infodemic". We believe that our work contributes to attracting public attention to the "infodemic" of the social crisis.

A Scheme to Reduce MVCC Conflict by Adjusting Block Size in Hyperledger Fabric

Hyeonwoo Shin, Taerim Lee, Ho-Se Moon, Juwook Jang
Department of Electronics Engineering, Sogang University, Seoul, Korea

Abstract: Hyperledger Fabric is one of the blockchain development projects hosted by the Linux Foundation. Hyperledger Fabric can build an optimized network by selecting appropriate endorsement policy, identity management, organizations, channels and so on according to various business purposes. In addition, data throughput is improved by parallel processing in which only some nodes execute smart contracts. However, in Fabric, because transactions are processed in parallel, transaction conflicts may occur. Among them, a typical collision is MVCC conflict. MVCC conflict is a collision that occurs when more than one transaction with the same key are registered in the same block, causing read after write(RAW) hazard. When MVCC conflict occurs, the transactions are treated as failed, thus lowering the success rate of the transactions and reducing the TPS. In this paper, we propose a method to reduce MVCC error and improve network performance by adjusting block sizes in Hyperledger Fabric.

Computational Perspective of the Fog Node

Joao Bachiega Jr, Breno Costa, Aleteia P. F. Araujo
Department of Computer Science, University of Brasilia (UnB) - Brasilia - DF – Brazil

Abstract: Fog computing is a recent computational paradigm that was proposed to solve some weaknesses in cloud-based systems. For this reason, this technology has been extensively studied by several technology areas. It is still in a maturing stage, so there is no consensus in academia about its concepts and definitions, and each area adopts the ones that are convenient for each use case. From an analysis of some current research on the subject, this article proposes a definition and a classification for the "fog node", which is a fundamental element in a fog computing environment, relying on a computational perspective. In addition, the main challenges related to the fog node are also presented.

Network Emulator Approach to Testing Internet of Everything

Keita Yoshida, Akihito Nakamura
Graduate School of CSE, Computer Science Division, University of Aizu, Aizu-Wakamatsu, Fukushima, Japan

Abstract: The Internet of Things (IoT) has become one of the most popular topics in the IT over the last few years. However, reliability and security of IoT systems become a serious problem. It is necessary to reduce those risks by supporting developers and system integrators at the development and system construction phases, respectively. This paper presents a new technical approach to the problem: network emulator. The network emulator works almost like a software router. It relays or alters communication between two parties. Exceptional situations like network failure or cyberattack can be reproduced by manipulating packets at the network layer in a transparent way. Therefore, IoT systems can be tested against such situations without any modification to the devices and their software. Naturally, our system can be used for traditional networked IT system testing. The evaluation results revealed that the system has high throughput and accuracy.

Security of IT/OT Convergence: Design and Implementation Challenges

Bassam Zahran, Adamu Hussaini, Aisha Ali-Gombe
Computer and Information Sciences, Towson University, Towson, Maryland, USA

Abstract: IoT is undoubtedly considered the future of the Internet. Many sectors are moving towards the use of these devices to aid better monitoring, controlling of the surrounding environment, and manufacturing processes. The Industrial Internet of things is a sub-domain of IoT and serves as enablers of the industry. IIoT is providing valuable services to Industrial Control Systems such as logistics, manufacturing, healthcare, industrial surveillance, and others. Although IIoT service-offering to ICS is tempting, it comes with greater risk. ICS systems are protected by isolation and creating an air-gap to separate their network from the outside world. While IIoT by definition is a device that has connection ability. This creates multiple points of entry to a closed system. In this study, we examine the first automated risk assessment system designed specifically to deal with the automated risk assessment and defining potential threats associated with IT/OT convergence based on OCTAVE AllegroISO/IEC 27030 Frameworks.

Leveraging IoT and Optical Tracking to Monitor Bee Hive Health

Paul R. Raupach III, Varick L. Erickson
Department of Computer Science, California State University, East Bay, California, USA

Abstract: Honey bees and their pollinating activity are essential to supporting global agriculture and food production. Effectively monitoring bee hive health is crucial to industries which depend on bees as pollinators. However, monitoring bee hive health can be incredibly complex and inefficient. In the present work, we built a "smart" beehive which collects realtime data to monitor hive health. This hive is low-cost, robust, able to be remotely monitored, and easily integrated with an industry standard Langstroth hive. We leverage multiple sensors including weight sensors, temperature sensors, and cameras with computer vision tools to continuously monitor hive health. Using these sensors, we applied various image processing techniques to develop a hive activity metric. Using these tools, we successfully monitored an actual beehive's health and activity over a 6 month period. This smart beehive system could be used in development of remote, data-driven tools to help beekeepers better support and monitor hive health.

Model-Based Testing of Smart Home Systems

Afnan Albahli, Anneliese Andrews
Department of Computer Science, University of Denver, Denver, colorado, USA;
Department of Information Technology, Princess Nourah Bint Abdul Rahman University, Riyadh, Saudi Arabia

Abstract: A Smart Home System is one of many Internet of Things (IoT) applications. The increase in the number of devices connected to the Internet cause a challenge for software testers to find a suitable approach to test systems using IoT. This paper provides a model-based testing technique (MBT) to test the functionality of Smart Home systems that use IoT. MBT employs Extended Finite State Machines (EFSMs) and a Communicating Extended Finite State Machines (CEFSMs) to model the smart home system. Our approach generates test paths for individual components of the smart home system through device level testing. We apply our approach to two devices in the smart home system.

An Improved Novel Low Power and Low Cost IoT Wireless Sensor Node for Air Quality Monitoring

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

Abstract: Battery autonomy is critical in wireless devices especially in large scale wireless sensor node deployments. In this paper fragmented dynamic power management is used to achieve an improved SOA low power and low cost

air quality monitoring IoT wireless sensor node. The improved novel device is designed to fit in a ventilated plastic enclosure and is equipped with low cost air quality sensors and a LoRa transciever. The current draw of the sensor node in sleep mode is 570 nA, the average active current demand is 44 mA and the average power consumption per hour is 379 µAh. The improved device can be operated for up to 20 months using a 6000 mAh battery.

Macro Benchmarking Edge Devices Using Enhanced Super-Resolution Generative Adversarial Networks (ESRGANs)

Corwin Stanford, Anthony Lam, Warren Williams, Jing-Ru C. Cheng
Computational and Data-Enabled Science and Engineering, Jackson State University, Jackson, Mississippi, USA;
Engineer and Research Development Center, U.S. Army Corps of Engineers, Vicksburg, Mississippi, USA

Abstract: In standard machine learning implementations, training and inference takes place on servers located remotely from where data is gathered. With the advent of the Internet of Things (IoT), the groundwork is laid to shift half of that computing burden (inferencing) closer to where data is gathered. This paradigm shift to edge computing can significantly decrease the latency and cost of these tasks. Many small, powerful devices have been developed in recent years with the potential to fulfill that goal. In this paper we analyze two such devices, the NVidia Jetson AGX Xavier and the Microsoft Azure Stack Edge Pro—2 GPUs. For comparison, the performance of these devices is compared to more common inferencing devices, including a laptop, desktop, and high performance computing (HPC) device. The inferencing model used for testing is the Enhanced Super-Resolution Generative Adversarial Networks (ESRGANs), which was developed using techniques borrowed primarily from other GAN designs, most notably SRGANs and Relativistic average GANs (RaGANs), along with some novel techniques. Metrics chosen for benchmarking were inferencing time, GPU power consumption, and GPU temperature. We found that inferencing using ESRGANs was approximately 10 to 20 times slower on the Jetson edge device, but used approximately 100 to 300 times less power, and was approximately 2 times cooler than any of the other devices tested. The inferencing using ESRGANs performed very similarly on the Azure device as on the more traditional methods. The Azure device performed with slightly slower speeds and equivalent temperatures to the other devices, but with slightly less power consumption.

Predicting Air Quality Index Using Deep Neural Networks

Sean McGrath, Eoin Garrigan, Liaoyuan Zeng
Department of Electronic and Computer Engineering, University of Limerick, Limerick, Ireland;
University of Electronic Science and Technology of China, Chengdu, China

Abstract: Air Quality Index is a measure of how unhealthy the air is, along with associated health consequences of exposure, and is composed of a mix of atmospheric pollutants. Traditionally, this index has been created from a suite of expensive monitoring sensors that evaluate maximum levels of these pollutants. This paper pursues an alternative approach to calculating the AQI. Utilising Machine Learning, an Artificial Neural Network method is employed to enable specific sensor data to be substituted or replaced, while still obtaining robust results.

Disruptive Computing Technologies and the Influence of 5G, AI/ML on Edge to Core to Cloud

Sean McGrath, Liam Quinn, Colin Flanagan, Liaoyuan Zeng
Department of Electronic and Computer Engineering, University of Limerick, Limerick, Ireland;
Dell Technologies, Austin, Texas, USA;
University of Electronic Science and Technology of China, Chengdu, China

Abstract: This paper focuses on the changes in computing and network architectures driven by the emergence of 5G and artificialintelligence/machine-learning (AI/ML), and how these disruptive technologies will radically reshape computing and networking, leading to a new sort of hybrid "compute/net" of the future. The way users develop, access

and consume content and data will be reshaped by the impact of disruptive and emerging technologies such as 5G, Edge, Cloud, IoT, and AI/ML. The application of accelerators along with AI/ML on the explosive growth of data sets will be essential to optimize, manage, control, and support compute elements and network infrastructures. It is argued that these trends will generate revolutionary rather than evolutionary change in the landscape of networking and computing.

Threat Intelligence Detection and Response Time Modelling

Choong-Hee Han, Soon-Tai Park, ChangHee Han
Korea Power Exchange, Korea;
Korea Internet Security Agency, Korea;
Korea Military Academy, Korea

Abstract: Threat Intelligence Detection & Response Time (TIDRT) is the sum of IIDRT and EIDRT. With this TIDRT formula, the activity time of a Security Operation Center (SOC) can be measured in accordance with three variables; the number of security systems to be enrolled (S), the amount of internal cyber threat intelligence (X), and the amount of external cyber threat intelligence (Y). This activity modelling method will be very useful to determine how many people must work against cyber threats in a SOC. Furthermore, this modelling method can be used to measure the improvement of a SOC's situation after adding a security investment.

Privacy for IoT Communication in a Smart City Environment

Matteo Anedda, Mauro Fadda, Roberto Girau, Vlad Popescu, Mariella Sole, Daniele Giusto
Department of Electrical & Electronic Engineering - DIEE/UdR CNIT,
University of Cagliari, Piazza, Cagliari, Italy;
University of Sassari, Italy;
Department of Informatics, University of Bologna "Alma Mater Studiorum", Cesena, Italy;
Department of Elektronics and Computers, Transilvania University of Brasov, Romania

Abstract: The smart city concept is based, among others, on the interaction and cooperation of various smart devices that regularly exchange large amounts of data. One of the biggest communication problems among such devices is the data treatment in order to avoid exposure to data tampering and unlawful manipulation. These risks can be mitigated with an adequate knowledge of the existing faults and the use of various protection methods, such as anonymization and pseudonymization of data. In this paper we analyze the stateof-the-art on the concepts of security and privacy for smart devices, including the specific anonymization algorithms. This analysis will be focused on one fundamental aspect of data security, privacy, which is investigated through the light of the European GDPR Regulation adopted in 2016 and with respect to the practical techniques involved. Upon an analysis of the advantages and disadvantages of the current algorithms, we propose a new solution called MISSION and we evaluate its behavior in an urban mobility scenario in terms of QoS parameters such as delay, latency, packet loss and prediction error.

Websites Evaluation Questionnaire for some Saudi Higher Education Institutions (HEIs)

Fahd Almutari, Sung-Chul Hong
Department of Computer and Information Sciences, Towson University, Towson, Maryland, USA

Abstract: Website evaluation questionnaires measure the quality of the website depending on the selected measurement criteria and methods. Many higher educational websites are currently assessed with questionnaires that concentrate on some aspects to assess the quality of universities' websites. However, the lack of knowledge in the methods available to measure website information quality creates a gap in assessments. As a result, website quality is not empirically validated. This paper presents improving the Website Evaluation Questionnaire (WEQ) and a generic approach to evaluating higher educational websites using domain ontology and questionnaires that measure the distinct quality of a website. The website evaluation ontology is developed based on the literature on website evaluation

methods and questionnaires. The proposed approach was verified by assessing 12 Saudi Higher Educational Institutions (HEIs) websites, resulting in a reliable measure of specific quality dimension measure of Saudi higher educational websites.

The 20th International Conference on Wireless Networks (ICWN 2021) https://www.american-cse.org/csce2021/conferences-ICWN

VoIP University Solution: VoIP UEMA

Carlos Henrique Rodrigues de Oliveira, Luiz Ricardo Souza Ripardo, Magno Castro Moraes, Rogerio Moreira Lima Silva, Leonardo Henrique Gonsioroski Furtado da Silva, Luis Carlos Costa Fonseca, Ivanildo Silva Abreu
Computer Engineering - Technological Sciences Center (CCT), State University of Maranhao (UEMA), Sao Luis, Brazil

Abstract: VoIP University is a solution that enables the development of communication projects based on the IP protocol. This paper presents a low-cost deployment called VoIP UEMA, the first project of the VoIP University solution. Technical feasibility to deploy this project was checked in a proof of concept. The concept was proven showing it is possible to create user extensions with information as name, ID, job role, course, center, and campus getting information of all registered users in the academic and administration system server and make voice and video calls through VoIP UEMA softphones, showing this project is viable technically. The low-cost solution was justified by a financial analysis showing the annual phone bill expense dropped, representing savings of over 97%.

A User-Based Authentication and DoS Mitigation Scheme for Wearable Wireless Body Sensor Networks

Nombulelo Zulu, Deon P. du Plessis, Topside E. Mathonsi, Tshimangadzo Tshilongamulenzhe
Department of Information Technology, Tshwane University of Technology, Pretoria, South Africa

Abstract: Wireless Body Sensor Networks (WBSNs) is one of the greatest growing technology for sensing and performing various tasks. The information transmitted in the WBSNs is vulnerable to cyber-attacks, therefore security is very important. Denial of Service (DoS) attacks are considered as one of the major threats against WBSNs security. In DoS attacks, an adversary targets to degrade and shut down the efficient use of the network and disrupt the services in the network causing them inaccessible to its intended users. If sensitive information of patients in WBSNs, such as the medical history is accessed by unauthorized users, the patient may suffer much more than the disease itself, it may result in loss of life. This paper proposes a user-based authentication scheme to mitigate DoS attacks in WBSNs. A five-phase user-based authentication DoS mitigation scheme for WBSNs is designed by integrating Elliptic Curve Cryptography (ECC) with Rivest Cipher 4 (RC4) to ensure a strong authentication process that will only allow authorized users to access nodes on WBSNs.

An Efficient Channel Estimation Technique for MIMO-OFDM Systems, Under the Effects of Partial-Band Jamming and Doppler Spread, Using Two-Sided Complementary Codes (TSCC-CE)

Said El. Khamy, Noha Korany, Hossam Hassan
Faculty of Engineering, Alexandria University, Alexandria, Egypt

Abstract: MIMO-OFDM channel estimation techniques have received a lot of interest recently. Channel Estimation (CE) is a vital step in the design of channel equalizers. Many CE techniques have been proposed including the use of orthogonal codes and orthogonal space-time block codes (OSTBC). Complementary codes (CC) were also considered for MIMOOFDM CE. CC are characterized by zero side lobe levels in their autocorrelation function and were originally introduced as optimum phase coding signals in radar systems. They require the use of two independent channels for the transmission of the twofold pilot code components. Thus, a two-sided CE (TSCE) approach is proposed, in this paper, where the two pilot CC codes are placed in different subcarrier locations. In harsh environment, the system is confronted by interference and jamming which can highly affect the accuracy of channel estimation and correspondingly causes equalization errors. This particularly happens in the presence of harsh Doppler spread due to the deterioration of the OFDM subcarriers' orthogonality. In this paper, we investigate the performance of the proposed MIMO-OFDM TSCE technique, based on complementary codes, namely TSCC-CE, under the effects of partial band jamming and Doppler spread. The proposed TSCC-CE technique is proven to highly withstand the effect of partial band jamming on the accuracy of MIMO-OFDM CE even in the presences of moderate values of Doppler spread and the partial band jamming ratio. The performance of TSCC-CE is shown to significantly outperforms other MIMOOFDM CE techniques.

The Transferable Belief Model for Network Performance Reliability Analysis in Wireless Sensor Networks

Elie Tagne Fute, Hugues M. Kamdjou, Adnen El Amraoui, Armand Nzeukou
Department of Mathematics and CS, Faculty of Science, University of Dschang, Dschang, Cameroon;
Department of Computer Engineering, Faculty of Engineering and Tec., University of Buea, Buea, Cameroon;
LGI2A (Lab. de Genie Informatique et d'Automatique de l'Artois), University of Artois, UR, Bethune, France;
Department of Industrial Systems and Environmental Engineering, Fotso Victor University Institute of Technology, University of Dschang, Cameroon, Faculty of Science, University of Dschang, Bandjoun, Cameroon

Abstract: In view of the prospects offered by the Wireless Sensor Networks (WSNs), some of the main challenges in recent years concern access to more reliable information (event detection) for an efficient decision making. Moreover, WSNs are one of the most speedily developing information technologies and promise to have a variety of applications in Next Generation Networks (NGN), Internet of Things (IoT), and for critical and safety relevant applications. Reliability is one of the most important attribute of such systems. Thus, in the field of artificial intelligence, the development of intelligent systems for processing imprecise and/or uncertain information is proposed. Truth be told, reliable event detection at the sink node depends on collective information given by source nodes and not just on any individual report. Subsequently, the WSNs world view requires a collective event to the sink reliability notion instead of the conventional end to end notion. This paper analyses the reliability of WSNs by adopting the theory of belief functions. The proposed Transferable Belief Model (TBM) approach aims to analyse reliability in WSNs according to the network topology. It includes a reliability of coverage, reliability of packet delivery, reliability of secure data exchange, reliability of network availability and reliability in terms of network latency

Analyzing the Weighted Polling on WBAN

Tatiana Annoni Pazeto
Federal University of Mato Grosso - Rondonopolis Campus, Graduation in Info Systems, Mato Grosso, Brazil

Abstract: This article analyzes the scheduler Weighted Polling in an application for Wireless Body Area Networks (WBAN), because the IEEE 802.15.6 has not yet adopted the protocol Medium Access Control (MAC) for these applications. The aim is to check their performance with regard to packet loss and queue time on the sensor nodes. For this a simulation platform encompassing various sensors, the proposed scheduler, that sends packets from the sensors to the sink node, which includes in its output link the algorithm First In, First Out (FIFO), was developed in MATLAB. Through of the results obtained can be seen that the Weighted Polling has a behavior similar of the Limited Polling.

Security in Distributed Systems by Verifiable Location-Based Identities

Simon Tschirner, Katharina Zeuch, Sascha Kaven, Lorenz Bornholdt, Volker Skwarek
RTC Digital Business Processes, Faculty of Life Sci., Hamburg University of Applied Sciences, Hamburg, Germany

Abstract: Proof-of-Location (PoL) is a lightweight security concept for Internet-of-Things (IoT) networks, focusing on the sensor nodes as the least performant and most vulnerable parts of IoT networks. PoL builds on the identification of network participants based on their physical location. It introduces a secondary message type to exchange location information. Via these messages, the nodes can verify the integrity of other network participants and reach a consensus to identify potential attackers and prevent malicious information from spreading. The paper presents the concretization of the concept to allow implementation on real hardware. The evaluation based on this implementation demonstrates the feasibility of PoL and enables identifying further steps to develop a deployable protocol.

A New Compact Size Dual-Bands 5G Patch Antenna Array

Chemseddine Benkalfate, Achour Ouslimani, Abed-Elhak Kasbari, Feham Mohammed
Quartz Laboratory, ENSEA, France;
STIC Laboratory, University of Tlemcen, Algeria

Abstract: This paper presents the design of a new compact size 5G patch antenna array covering both 3.5 GHz and 28 GHz bands. Its designed-on Teflon glass substrate. With a size of 15 x 30 mm² which makes it well suited for IOT applications. The main idea of array design is to couple a dipole antenna with several elementary wire antennas superposed on the two arms of the dipole. The proposed antenna is adapted to the [3 - 4] GHz and [26 - 29.3] GHz bands with resonance frequencies of 3.5 GHz and 27.65 GHz respectively. The simulated return loss is of -35 dB and -44 dB for the two resonance frequencies respectively. The antenna is pseudo-omnidirectional for the [3.4 - 3.8] GHz band with a directivity of 3 dBi, and directive for the [26 - 29.3] GHz band with a directivity of 6.5 dBi.

Intelligent Throughput-based Sleep Control Algorithm for the 5G Dense Heterogeneous Cellular Networks

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

Abstract: In the recent past, many mobile/telecom operators have seen a continuously growing demand for ubiquitous high-speed wireless access and an unprecedented increase in connected wireless devices. As a result, we have seen explosive growth in traffic volumes and a wide range of QoS requirements. The Fifth generation (5G) heterogeneous cellular networks (HetNets) have been developed by different mobile operators to achieve the growing mass data capacity and to reconnoiter the energy efficiency guaranteed trade-off between throughput QoS requirements and

latency performance. However, existing energy efficiency algorithms do not satisfy the throughput QoS requirements such as reduced latency and packet loss, longer battery lifetime, reliability, and high data rates with regards to the three components of energy consumption of the 5G radio access network (RANs) that dominate the overall mobile communication networks. In addition, real-time traffic types such as voice and video require high computational load at the terminal side which have an undesirable impact on energy/battery lifetime which further affects the throughput QoS performance such as reduced packet loss, longer battery lifetime, reliability, and high data rates. As a result, this paper proposed an Intelligent Throughput-based Sleep Control (ITSC) algorithm for throughput QoS and energy efficiency enhancement in 5G dense HetNets. In the proposed ITSC algorithm, a deep neural network (DNN) was used to determine the cell capacity ratio for the small base stations (SBSs). Hence, the SBSs cell capacity ratio was employed as decision criteria to put the SBSs into a sleep state. Furthermore, transferable payoff coalitional game theory was used in order to ensure real-time applications have a higher priority over non-real time applications. Numerous Network Simulator 2 (NS-2) results confirmed that the proposed ITSC algorithm reduced packet loss and produced better QoS. Moreover, the ITSC algorithm provided a longer battery lifetime, reliability, and high data rates for real-time traffic. The network throughput was improved as a result.

Smart Grid Communication Technologies and their Applications for Smart Cities

Deok Hee Nam
Wilberforce University, Wilberforce, Ohio, USA

Abstract: Smart grid and its technologies are introduced to address the relationship between smart grid and smart city. The smart grid is an essential technology for the smart city to modernize the infrastructure of an electric power grid that use high power converters, automation control, modernized communication architecture, smart metering technologies, modernized energy management techniques, network, and energy availability to enhance the reliability and efficiency of the electric power grid. Along with the smart grid technologies, communication architecture and correct information are two main components of the current electric power systems, but the smart grid requires much larger and more complex communication architecture for power systems. This survey paper is an excerpt of more comprehensive studies on a smart grid (SG) and the concepts of smart grid technologies to carry out the feasibility studies for a smart city with the introduced smart grid technologies.

Receiver-Initiated RTS/CTS Control for Even Transmission Opportunities of Data Messages among Neighbor Wireless Nodes

Momoka Hara, Hiroaki Higaki
Department of Robotics and Mechatronics, Tokyo Denki University, Tokyo, Japan

Abstract: In a wireless ad-hoc network, collisions of control and data messages are avoided by CSMA/CA and RTS/CTS control of wireless LAN protocols. Random backoff timers for avoidance of collisions among RTS control messages provide equal opportunities to transmit data messages to neighbor wireless nodes since the value of the backoff timer monotonically decreases. In wireless ad-hoc networks, wireless nodes are not equally distributed and frequency of transmission requests in wireless nodes is also not the same. Thus, especially in a region with high density of transmission and receipt requests for data messages, it is not always possible for a wireless node with a transmission request to receive a response CTS control message even though it has an opportunity to transmit an RTS control message. Hence, the equal opportunities to transmit an RTS control message is not enough to realize the equal opportunities to transmit a data message. In order to solve this problem, this paper proposes an extended receiver-initiated RTS/CTS control to equally provide opportunities to transmit data messages whose receiver node is hard to send back a CTS control message in response to an RTS control message. Here, a transmission of a CTS control message precedes a transmission of an RTS control message after failed transmissions of a CTS control message. Simulation experiment results show the proposed method improves equality of transmission opportunities of data messages even among densely distributed wireless nodes.

Emerging and Deadly Botnet on Mobile Network

H.D.N.S Priyankara, Sureswaran Ramadass
Institute of Postgraduate Studies, Malaysia University of Science and Technology, Selangor, Malaysia;
National Advanced IPv6 Center of Excellence (NAv6), Universiti Sains Malaysia, Penang, Malaysia;
The Open University of Sri Lanka, Sri Lanka

Abstract: Mobile Botnet is a collection of inter-connected mobile devices, which is a serious cybersecurity threat. Mobile Botnets have been used in various cyberattacks, including identity fraud, phishing, stealing bank credentials, and distributed denial-of-service (DDoS). Although there are surveys on the Botnet, there is no comprehensive study on mobile Botnet in the last decade. This paper presents emerging botnets evolution from 2016 to 2021, including Mandrake, Cerberus, EventBot, PreAMo, Triada, Rouge, and Redaman malware. The study covers Botnets' classification, architecture, deploying mechanism, malware functionality, algorithms, encryption methods used in the C&C server, and different malware versions. Further paper present how these malware escape from intrusion detection mechanism. Finally conclude the research areas could be carried further.

The 20th International Conference on Information & Knowledge Engineering (IKE 2021)
https://www.american-cse.org/csce2021/conferences-IKE

A Comparative Performance Analysis of Estimation Techniques in Community Detection

Matin Pirouz
Department of Computer Science, California State University, Fresno, California, USA

Abstract: Community structures and relation patterns, and ranking them for social networks provide us with great knowledge about the networks. A community is defined by determining a lower density of relations between groups comparing to higher density among every group. Such knowledge can be utilized for grouping similar, yet distinct, nodes with applications in health, marketing, and many more. The ever-growing variety of social networks necessitates detection of minute and scattered communities, which are important problems across different research fields including biology, social studies, physics, etc. As a result, analyzing complex networks has become very popular among researchers in academia and the industry. Interactions and inter-individual relations are captured and depicted as social networks graphs. The aforementioned structure analysis helps researchers determine the model, the type, and the degree of relationships. Besides, it can provide guides to predict the future behavior of social networks. In this study, we take a comparative approach to analyze the performance of the existing estimation methods for community detection.

Performance of Distributed Join Algorithms on Large Data Sets

Felicia Imbriano, Paul Bergstein
Department of Computer and Information Science, University of Massachusetts Dartmouth, Massachusetts, USA

Abstract: In relational database management systems, computing joins is a common and expensive operation in query processing. When the tables to be joined are large and reside on different network nodes, performance becomes especially critical. In the simplest strategy for computing distributed joins, the smaller table is first copied across the

network and then a standard join algorithm is employed. The Semi-Join algorithm is a well-known alternative that attempts to improve the overall performance of distributed joins by reducing the amount of data transferred across the network at the expensive of additional processing time. CNF-Join is a new algorithm that takes a different approach to reducing network traffic, and CNF-SemiJoin combines the Semi-Join and CNF-Join techniques. In this paper, we empirically compare the performance of these algorithms under a variety of conditions.

A Novel Method for Calculating Query Hashes for Improved Query Grouping in Relational Database Management Systems

Derek Colley, Md Asaduzzaman
Department of Computing, School of Digital, Technologies and Arts, Staffordshire University, Stoke-on-Trent, UK;
Department of Engineering, School of Digital, Technologies and Arts, Staffordshire University, Stoke-on-Trent, UK

Abstract: Database queries are stored and compared by relational database management systems as hashes, or short unique representations, of the original query text. This leads to cache misses and increased resource consumption by database engines when queries differing only in non-syntactic detail, or queries which are relationally equivalent, are presented to the query parser. We propose a new method of structural query decomposition, transforming database queries into multidimensional adjacency cubes (MACs), allowing the codification of queries by structure rather than content as currently implemented. We build and test our solution, demonstrating superior query hash grouping to that currently offered by a leading relational database platform, and consider the applications of this new technique.

Spider Monkey Algorithm based on Differential Adaptive

Xiaoping Xu, Feng Wang, Jingjing Rong, Long Liu
School of Sciences, Xi'an University of Technology, Xi'an, P. R. China;
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, P. R. China;
School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, P. R. China

Abstract: The spider monkey algorithm based on differential adaptive is proposed in this paper. On the basis of the basic spider monkey algorithm, three parts of the algorithm are improved: phase factor is used in the local leader stage to realize the iterative updating from coarse to fine; adaptive learning factor is adopted in the global leader stage to balance the global exploration and local mining capacity; the mutation operation of differential evolution algorithm is applied in the local leader decision-making stage to find the potential solution in the local region of the current solution and improve the search efficiency and solution accuracy of the algorithm. Finally, multi-dimensional and multi-modal test functions are calculated using the improved spider monkey algorithm, and compared with other algorithms, the results show that the proposed algorithm in stability, accuracy and other performance has been significantly improved

A Clustering Based Data Quality Consistency Assessment Architecture for Categorical Data

Mahmood Mohammed, John R. Talburt, Onais Mohammed
Information Science, University of Arkansas at Little Rock, Arkansas, USA;
Hyderabad, TS, IND

Abstract: Companies and institutions dependent on data for business decisions and operations recognize the need for robust data quality management. One of the major data quality challenges is maintaining consistent representation of the same information stored in different systems across the organization. We design a machine learning based data quality assessment architecture for categorical data to identify inconsistencies. Our work overcomes some of the challenges with the traditional rules-based approach of data quality assessment. Additionally, we demonstrate and compare various approaches for applying unsupervised learning to data quality assessment. We discuss our decisions on designing the resultant architecture and present an experimental evaluation on datasets.

Classification with a Decision Tree Algorithm and Neural Networks

John Tsiligaridis
Math and Computer Science Department, Heritage University, Toppenish, Washington, USA

Abstract: Decision Trees (DT) and Neural Networks (NN) have wide applications in classification. However, NN needs enormous memory space and training time to build the model. In this project the Projection Algorithm (PA) a sort of Decision Tree based on the purity criterion, and Neural Networks are used for classification. The DTNN is created by mapping the DT into a NN. It has hyperparameters, from the PA, considering one or more hidden layers. Next an Incremental Neural Network (INN) is created from dataset instances. The INN consists of a set of rules for the structural and weight adaptation. Comparisons show that DTNN outperforms NN. PA is slightly better than ID3. Simulation results are provided with various datasets.

A Knowledge Perspective on Co-Opetition - Supporting Development of Innovative Business Co-Operations

Jan Aidemark
Linnaeus University, Vaxjo, Sweden

Abstract: IS/IT solutions are important for learning and knowledge processes within and between organizations. This paper explores the topic of co-opetition from a knowledge perspective. A planning framework for support for co-opetition based on knowledge management theory is proposed. The framework is intended as support for planning of support systems and inter-business learning processes.

Implementation of Full-Automatic Stock Trading System to Prevent Investor Intervention

Sojin Lee, Jihyeon Lee, Yoojeong Song, Jongwoo Lee
Department of IT Engineering, Sookmyung Women's University, Cheongpa-ro, Yongsan-gu, Seoul, Korea

Abstract: In a modern capitalist society, people want to make a big profit by trading stocks. There are various stock price analysis and investment strategies for successful stock investment, but it is still difficult to prove its effectiveness. This is because, despite the fact that investors use an effective investment strategy, the change in their mind fails to carry it out to the end. Also, investors have to focus a lot of time on trading in stocks, but in reality, it is difficult to invest all of an individual's time into trading. As stock prices change in real time, investors are highly stressed and may become addicted to stocks. In this paper, we developed a full-automatic stock trading system. The full-automatic stock trading system sell stocks according to the rules directed by a recommendation file, and automatically runs at the stock market opening time. The system proposed in this paper is expected to enable investors who lack stock trading knowledge to earn returns with less time and effort. We are confident that this system will be a good solution for investors and stock addicts who have spent a lot of time trading stocks.

A Novel Deep Belief Network and Enhanced Root Mean Square Error for Stock Exchange Price Prediction

Ananda Prasad Jaisy, Abeer Alsadoon, Omar Hisham Alsadoon, Tarik A. Rashid, P.W.C. Prasad, Rasha S. Ali, Oday D. Jerew, Dara Sherwani, Mokhtar Mohammadi, Shima Rashidi
School of Computing and Mathematics, Charles Sturt University (CSU), Wagga, Australia;
School of Computer Data and Mathematical Sciences, University of Western Sydney (UWS), Sydney, Australia;
School of Information Technology, Southern Cross University (SCU), Sydney, Australia;
Crown Institute of Higher Education, Information Technology Department, Sydney, Australia;
Department of Islamic Sciences, Al Iraqia University, Baghdad, Iraq;
Computer Science and Engineering, University of Kurdistan Hewler, Erbil, KRG, Iraq;
Department of Computer Techniques Engineering, AL Nisour University College, Baghdad, Iraq

Abstract: Due to the dynamic, complex, and chaotic of the stock market, its prediction is one of the most challenging tasks to be performed. Meanwhile, the rapid development in machine learning and deep learning has contributed to several parts of image processing and voice recognition. Deep learning-based classification and prediction have not been successfully implemented in stock exchange prediction due to the limitation in the root mean square error (RMSE) and total return (TR). This paper aimed to improve the classification and prediction accuracy of the stock exchange perform on the multiple organization stock dataset. The proposed system consists of a two-dimension principal component analysis (2D), the purpose of using PCA is for dimension reduction, and deep belief network (DBN) algorithms have been proposed to improve the classification performance and prediction accuracy of stock exchange price by minimizing the root mean square error (RMSE) and total return (TR). The Number of iteration (N), the enhanced root mean square error (ERMSE), and the total return (TR) are considered as the evaluation matric for accuracy and performance. The deep belief network (DBN) must reduce the number of iterations by 1:2 ratios. DBN algorithm must perform well on enhance root mean square error to DNN, as it is 4.1% more accurate than DNN and total return 7.3 % better than DNN. The proposed system concentrates on minimizing the root mean square error (RMSE) and total return (TR) value. Thus, this study provides a better result on accuracy and processing speed for stock exchange value.

REST Interface Generator (RESTInG)

Kenneth Higa, Daniela Ushizima
Energy Storage and Distributed Resources Division, Energy Technologies Area,
Lawrence Berkeley National Laboratory, Berkeley, California, USA;
Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA;
Berkeley Institute for Data Science, University of California, Berkeley, California, USA

Abstract: Many computing projects require distributed read and/or write access to relational databases such as PostgreSQL. Building these capabilities for a project can involve significant time and effort. Here we describe a package which allows one to easily build simple project-specific deployments of PostgreSQL databases with web interfaces using the Python-based Django REST Framework. These interfaces can be further customized as needed. This tool is presently intended for creation of services hosted on the "cloud-like" NERSC Spin platform, but future development might include extension to standalone servers.

Event Detection in Edge Streams of Time-Evolving Graphs

Monika Bansal, Dolly Sharma
Department of Computer Science & Engineering, Shiv Nadar University, India

Abstract: Event detection refers to the problem of detecting anomalous time-points at which the properties of data changes significantly in the time-series data. In this work, we introduce a fast, scalable and unsupervised method for event detection (or anomaly detection) in large time-evolving1 graphs. We define an anomaly as a sudden change in the features of nodes causing from arriving groups of multiple similar edges. Given a stream of multi-dimensional

edges of a time-evolving graph, the proposed algorithm exploits its local structure to discover (1) anomalous time-points at which many nodes deviate from their normal behavior and (2) those nodes, features and edges that majorly contribute to the change. The algorithm is online such that it processes each edge in a constant time using constant memory. Moreover, we have also provided the theoretical guarantee on false positive probability of the algorithm.

Token Correction for the Data Washing Machine: Types and Comparisons

Fumiko Kobayashi, John R. Talburt, Awaad K. Al Sarkhi
Donaghey College of Science, Technology, Engineering, and Mathematics, The University of Arkansas at Little Rock, Arkansas, USA

Abstract: Many researchers and practitioners are working to design an unsupervised system accepting any type of data as an input and generates a fully cleaned, and ready to use data product. This type of system is known conceptually as a Data Washing Machine (DWM). To build a DWM, the system must abstract away the concept of structured data and leverage metadata products derived from the data itself to assist in the cleansing of the data. This paper describes the results of using an unsupervised process for correcting inconsistent representations of data. The correction process is based on an analysis of the frequency of the tokens (strings) used to express data attribute values. Token correction can be performed at various points in DWM. This paper explores two (2) of these, global (file) and (ER) block level and compares the performance of each. Based on F-measure, our unsupervised approach to global spelling corrections improved results by 75%, the block-level corrections by 53%, and in combination improved results by 83%.

Strategies To Counter Artificial Intelligence in Law Enforcement: Cross-Country Comparison of Citizens In Greece, Italy And Spain

P. S. Bayerl, B. Akhgar, E. La Mattina, B. Pirillo, I. Cotoi, D. Ariu, M. Mauri, J. Garcia, D. Kavallieros, A. Kardara, K. Karagiorgou
CENTRIC, Sheffield Hallam University, Sheffield, United Kingdom, UK;
Engineering Ingegneria Informatica, Palermo, Italy;
PLURIBUS ONE S.r.l., Cagliari, Italy; Vicomtech, Donostia-San Sebastian, Spain;
Center for Security Studies (KEMEA), Athens, Greece;
University of the Peloponnese, Athens, Greece

Abstract: This paper investigates citizens' counter-strategies to the use of Artificial Intelligence (AI) by law enforcement agencies (LEAs). Based on information from three countries (Greece, Italy and Spain) we demonstrate disparities in the likelihood of ten specific counter-strategies. We further identified factors that increase the propensity for counter-strategies. Our study provides an important new perspective to societal impacts of security-focused AI applications by illustrating the conscious, strategic choices by citizens when confronted with AI capabilities for LEAs.

DeepQ: A System to Peek Inside the eCommerce Deep Web

Amal Aljohani
Department of Computer Science, University of Idaho, Moscow, Idaho, USA

Abstract: Contemporary search engines such as Google do not allow peeking into a database behind the firewall, or the so called Deep Web. These engines can only find information that are surfaced to the Shallow Web in some manner and archived for later use. That is usually the reason items displayed on Google search often not found when clicked on the vendor site, or are not the same item, giving an impression of baitand-switch. This discrepancy or the inability for the search engines to query the deep web databases is fundamentally related to their inability to actually simulate a search of database contents that a human can do. In this paper, we present a system called DeepQ to search the database contents behind the firewall inside the deep web. We leverage a recently proposed deep web query language called the DQL and show that its implementation is feasible.

A Case Study on Record Matching of Individuals in Historical Archives of Indigenous Databases

Matthew Currie, Ramon Lawrence
Department of Computer Science, University of British Columbia, Kelowna, BC, Canada

Abstract: Digitization of historical records has produced a significant amount of data for analysis and interpretation. A critical challenge is the ability to relate historical information across different archives to allow for the data to be framed in the appropriate historical context. This paper presents a real-world case study on historical information integration and record matching with the goal to improve the historical value of archives containing data in the period 1800 to 1920. The archives contain unique information about Metis and Indigenous people ´ in Canada and interactions with European settlers. The archives contain thousands of records that have increased relevance when relationships and interconnections are discovered. The contribution is a record linking approach suitable for historical archives and an evaluation of its effectiveness. Experimental results demonstrate potential for discovering historical linkage with high precision enabling new historical discoveries.

Domain Interaction in Scientific Academic Communities

Rares Nesa, Niculescu Virginia, Adrian Sterca, Darius Bufnea
Faculty of Mathematics and Computer Science, Babes-Bolyai University, Cluj-Napoca, Romania

Abstract: Co-authorship networks are used to provide collaborative insights into various research communities, from university-level to country-level and to global level. They can be used in order to generate macro level research strategies for the whole community (i.e. university) or to suggest future collaborations for individual researches. In this paper we build a co-authorship network from the scientific output of a Romanian university with multiple research domains. We then analyse this network from two perspectives: a domain-level study of collaborations and a social network analysis of the co-authorship network. This dual approach analysis of the network offers us important insights that would otherwise be missed if we only choose one single approach.

Automated Construction of Skills Knowledge Linked Data Cloud

Edward Schwalb
Schwalb Consulting, USA

Abstract: To remain competitive, organizations need to understand the skills and competencies of their human resources. Ontologies and knowledge graphs are two distinct approaches known for their powerful representation of knowledge enabling such understanding. On the one extreme, ontologies tend to be manually constructed, are often associated with closed world assumptions which enable formal consistency assurance and inference, and are thus trustworthy but not scalable, inflexible, difficult to harmonize and evolve over time. On the other extreme, knowledge graphs tend to be automatically constructed, are often associated with open world assumptions which render formal consistency and inference difficult, and are thus not trustworthy but scalable, flexible, automatically harmonize and evolve over time. We leverage machine learning methods to develop a middle ground between the two extremes, which is trustworthy, salable, automatically harmonizes and evolve over time. Our approach stochasticaly lifts a baseline Boolean knowledge graph by associating each triplet with a belief ranging from 0 to 1. We assemble the graph from a cllection of rigid manually constructed ontologies together with big-data sources of job postings and resume data. Our frameworks can support knowledge which is incomplete, ambiguous and dissonant, and yet extract meaningful signal and actionable insights. As an example, we are able to extract insight about skills which serve as scaffolding for other skills, and leverage such insight to identify emerging trends and make insightful recommendations about closing skill gaps.

Adaptive Information Architecture: Information Element Identification and Organisation Patterns for Digital Ecosystem

Asif Qumer Gill
University of Technology Sydney, Ultimo, NSW, Australia

Abstract: Data-intensive digital ecosystem is a complex actor network, where information flows beyond the boundary of a single enterprise. There are several actors and types of information in the digital ecosystem. Often information identification approach is ad-hoc. The challenge is: where to begin, and how to identify and organize information elements for designing adaptive information architecture. This research aims to address this important challenge and presents the information element identification and organisation patterns. The applicability of these patterns is demonstrated with the help of an industry case study example. These patterns will help researchers and practitioners who are interested in identifying, classifying, managing, and securing information elements for digital ecosystems. The results of this research indicate that organism-interaction centric approach seems useful for identify the information elements in digital ecosystem.

A Text Mining Approach to Identify IT Research Analyzing the NSF Grants

Said Varlioglu, Hazem Said, Murat Ozer, Nelly Elsayed
School of Information Technology, University of Cincinnati, Ohio, USA

Abstract: Information Technology (IT) is recognized as an independent and unique research field. However, there has been ambiguity and difficulty in identifying and differentiating IT research from other close variations. Given this context, we analyzed 40,521 NSF-CISE awards and abstracts to find keywords that might help to understand core concepts of IT research. Using text mining techniques, we analyzed the similarities and differences between IT research and other related research domains. Results show that IT differentiates itself from other close variations by focusing more on the needs of users, organizations, and societies.

ILE, a Linked Database for Improving the Reliability, Efficiency, and Security of the Electric Power Grid

Vitit Kantabutra, Natalia Castaneda
Department of Electrical and Computer Engineering, Idaho State University, Pocatello, Idaho, USA

Abstract: Electric power grids are frequently threatened by cyberattacks, solar flares, failing equipment, and more. Because of these threats, and because of the need to operate and maintain electric grids well in real time, including load balancing, we need a secure, efficient, and well-designed computer database for representing the nodes of a grid, their attributes, and their interconnections. We argue that the ILE (Intentionally-Linked Entities) database system is ideally suited for this task. More than other database systems, ILE allows grid nodes to be represented hierarchically and connected if, and only if, they are supposed to be connected. ILE represents nodes and connections in a rich, flexible, accurate, and organized fashion, thereby enabling essential improvements in grid operation, maintenance, and security.

Identifying and Using Phases of Flight for Unmanned Aircraft Systems through Machine Learning

Li Yu, Brent T. Langhals, Ryan D. L. Engle, Michael R. Grimaila, Douglas D. Hodson
US Air Force Institute of Technology (AFIT), USA

Abstract: Machine learning techniques provide the means to leverage large quantities of data to aid in operational decision making and autonomy. The proposed research applies machine learning to unmanned aircraft systems (UAS)

telemetry and system data to unlock unique trends and patterns in the data based on UAS phases of flight. Specifically, critical parameters such as exceeding systems thresholds which may indicate potential maintenance concerns or operational limitations are identified. Finally, accurately understanding phase of flight based on telemetry and system data alone allows efficient post-flight analysis of observed anomalies during operations.

Performance Characteristics and Techniques of Discrete Database Schemas

Carter Grove, Brent T. Langhals, Ryan D. L. Engle, Michael R. Grimaila, Douglas D. Hodson
US Air Force Institute of Technology (AFIT), USA

Abstract: With the advent of the Big Data era, an arms race developed to design and field advances in data storage and retrieval. Traditional databases, while still and important element in many business, academic and government environments are simply unable to hand the rigors of the Big Data world. The NoSQL database which emerged since 2000 have generally been built to suit a particular purpose of a given business application, with many designs claiming superiority. Research into efficacy of each database design generally compared two or more instantiation of databases. This research proposes a meta-analysis of the body of literature since 2010 in order to determine if common performance characteristics and/or techniques can be derived from the individual studies.

Exploring News Articles on Bitcoin in New York City - Case Study

Anastasia Sagalovitch
City College, City University of New York, New York, USA

Abstract: A case study approach was utilized to explore news articles published by a popular daily newspaper in New York City - the New York Post, via the Natural Language Tool Kit in Python on the topic of bitcoin to assess any potential trends in its reporting. The larger goal of this project would be to compare reporting across cities and publications, across time, to determine if there were/are differences in how different publications and by extension cities or countries, report on the emergence of digital currency. Currently, the legislation around bitcoin is fragmented as Federal, State and City agencies may create their own respective rules. As an emerging financial technology, empowered by a technology backbone centered around cryptography, distributed computing, that is to date "untied" to anything else in the market, the aim of the project was to investigate in what terms has the reading public been exposed to the topic and in what ways might this differ across locales and publications. The project was amended to reflect reviewer feedback to condense/limit the scope for this initial exploration.

Recycling 4.0 - Solving the Information Puzzle of the Circular Economy

Lars Kintscher, Sebastian Lawrenz, Hendrik Poschmann, Priyanka Sharma
Institute of Distributed Systems, Ostfalia University of Applied Sciences, Wolfenbuttel, Germany;
Institue for Software and Systems Engineering, Clausthal University of Technology, Clausthal-Zellerfeld, Germany;
Institute of Production Technology, Ostfalia University of Applied Sciences, Wolfenbuttel, Germany

Abstract: Recycling of products and the establishment of the Circular Economy are some of the highly regarded topics in the 21st century. The industry is growing, more and more products are entering the world market. Additionally, the complexity of products is increasing. But how can these products be recycled efficiently? Our Recycling 4.0 framework targets this question by using Industry 4.0 technologies as a solution. A framework for a Recycling 4.0 was developed and data exchange between stakeholders of a product's lifecycle was identified to play a key role in future recycling processes – the Circular Economy is expanded to a DataDriven Circular Economy. This paper evaluates the Recycling 4.0 framework by using the concrete example of electric vehicles and their batteries. We present our framework to solve the information puzzle with the concept of a shared digital twin in OPC UA and summarize the research of our previous work.

MetaScribe: Data Provenance in Scholarly Publications

Maria Joseph Israel, Ahmed Amer
Santa Clara University, Santa Clara, California, USA

Abstract: MetaScribe aims to broaden provenance application beyond scientific workflow management and database systems by developing a general provenance framework to capture richer and extensible metadata in unstructured textual data sources such as literary texts, commentaries, translations, and digital humanities. Specifically, it intends to demonstrate the feasibility of capturing and representing expressive provenance metadata including an author's intentions and contexts of citing scholarly works at the time of developing his/her research content for publication, while also supporting subsequent tagging for such richer metadata by human or automated third party. While achieving this goal, we also ensure that MetaScribe is not simply a bibliographic citation application but an architecture for a data storage system suitable for this data model as it requires features like immutability and flexible authentication.

The 25th International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV 2021)
https://www.american-cse.org/csce2021/conferences-IPCV

A Revision Spatial CNN Steganalysis Based on Yedroudj-Net

Hsien-Wen Tseng, Hui-Shih Leng
Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan;
Department of Mathematics, National Changhua University of Education, Changhua, Taiwan

Abstract: Steganography hides secret data into a cover image so that it cannot be detected by unintended recipients, whereas steganalysis attempts to discover hidden data in suspected images. In recent years, much progress has been made in developing steganalysis because of deep learning. The performance of steganalysis has been improved using the Convolution Neural Network (CNN). Yedroudj et al. proposed the Yedroudj-Net CNN that composed of 30 pre-processing filters block, 5 convolutional blocks, and a fully connected block. Yedroudj-Net is an efficient CNN and achieves performance improvement. In this paper, we proposed a revision method which selects 10 out of the 30 filters in the pre-processing block. The proposed method not only achieves close performance, but also saves half of the training time. The performance is evaluated against WOW and S-UNIWARD embedding algorithms.

Confidential Deep Inference with the Aid of Fully Homomorphic Encryption

Wun-Ting Lin, Ja-Ling Wu
Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

Abstract: In this work, we study mechanisms for running deep inference with only addition and multiplication operators based on the approximate number encryption techniques and the numerical approximation algorithms. These mechanisms allow us to securely infer deep neural network models without eliminating or replacing any activation layer. To justify the effectiveness of the mechanisms mentioned above, we conducted a series of experiments. In our experiments, the encrypted inference accuracy compared to its nonencrypted counterpart is identical in the benchmarking of MNIST classification problem. This investigation implies that the deep confidential inference with homomorphic encryption aid is workable and of great potential.

A Quantitative Interpretable Convolutional Neural Network to Diagnose Diabetic Retinopathy

Wei Xiang Lim, Zhiyuan Chen, Amr Ahmed
Faculty of Science and Engineering, School of Computer Science, University of Nottingham Malaysia, Malaysia

Abstract: Deep Learning, a subset of machine learning has been adopted in various fields, especially in computer vision. However, due to the "black box" design architecture of Deep Learning models, the deep network-based computer-aided diagnosis (CADx) systems hinder end-users from understanding how the models work and ultimately impedes model adoption. The crux of the problem is that these models could not explain what representations that have been learned or why the models generate such decision. In this paper, we propose a quantitative interpretable Diabetic Retinopathy (DR) detector that produce qualitative visual interpretations of vital features and quantitatively evaluate the visual interpretation by measuring the degree of alignment between the produced attribution values and the pixel-level annotated lesion ground-truth. Our work focused on quantifying the generated visual interpretation of learned representations by adopting a quantitative evaluation metric called Intersection over Union (IoU) score. Our results highlight that quantifying the visual interpretation to explain and strengthen trust of end-users which pixels were important and accurate for the classification.

Deep Learning for White Blood Cells Classification: Enhanced Rectified Linear Unit (ReLU)

Ali Haider, Abeer Alsadoon, Omar Hisham Alsadoon, Wafaa M. Salih, Tarik A. Rashid, Mokhtar Mohammadi, P.W.C. Prasad, Ghossoon Alsadoon, Shima Rashidi, Mohammed Kamal Majeed
School of Computing and Mathematics, Charles Sturt University (CSU), Australia;
School of Computer Data and Mathematical Sciences, University of Western Sydney (UWS), Sydney, Australia;
School of Information Technology, Southern Cross University (SCU), Sydney, Australia;
Crown Institute of Higher Education, Information Technology Department, Sydney, Australia;
City University College of Ajman, General Education, Ajman, UAE;
Computer Science and Engineering, University of Kurdistan Hewler, Erbil, KRG, Iraq;
Department of Information Technology, Lebanese French University, Erbil, Kurdistan Region, Iraq;
Computer Department, College of Science, University of Halabja, Halabja, Iraq;
Business Informatics Department, AMA International University Bahrain (AMAIUB), Bahrain

Abstract: In deep learning, the recent use of convolution neural layers are inefficient, as they are unable to find a long-term dependent relationship between the label and key features. The proposed system consists of an improved convolutional neural network with ReLU activation function and color detection, to obtain the key features and introduce the color feature of white blood cells. Accuracy is measured using a probability score and processing time is measured using the total execution time. Different value of accuracy and processing time is achieved while using the different sample image group of WBC images. In addition to this, the proposed system was found to have improved the classification accuracy that was enhanced to 98% from a 96% average and reduced the processing time by 0.311 seconds against the original 0.581 seconds and this is approximately 30-60 seconds on average without damaging the other features. Moreover, the proposed system had the color and volume features added. In the state of art, white representation of WBC having no color as compared to the proposed system, which would be the original WBC color. The proposed system is focused on improving the accuracy and processing time of the WBC. It consists of feature extraction from the convolutional neural network with a ReLU activation function. In addition to this, the proposed system has added color, which does not exist in state-of-art. Finally, this study enhances accuracy. However, it will not negatively impact the other features.

Acquiring Semi-automated Annotations of Bounding Boxes based on Only One Hand Written Bounding Box per Class

Hiroshi Uehara, Kaito Ishizaki, Kasuki Saruta
Department of Data Science, Rissho University, Saitama, Japan;
Department of Management Science and Engineering, Akita Prefectural University, Akita, Japan;
Department of Information and Computer Science, Akita Prefectural University, Akita, Japan

Abstract: Annotating bounding boxes for object detections is one of the most time-consuming work among annotations for supervised machine learnings. This study aims to reduce the manual work concerning annotations of bounding boxes. Semi-automated annotations in related work requires manual tasks in proportion to the number of training data. Meanwhile, our proposal requires only one hand written bounding box per class to acquire automated the bounding boxes in a variety of training images. We applied our proposal to sample datasets, and evaluate the performance by computing IoUs of automated bounding boxes against precisely hand written bounding boxes. The average IoU was 0.85, which supports the efficiency of our proposal.

3D Face Re-Construction from Front and Profile Image

Sankarshan Dasgupta, Mehdi R. Zargham
Department of Computer Science, University of Dayton, Ohio, USA

Abstract: In this fast-paced, unpredictable world, the difference we can make in someone's 3D view experience is significant, and our objective is to create an automated procedure to construct an avatar of an individual from his/her front and profile image. 3D re-construction from a limited number of images has been a difficult problem to overcome but machine learning based approach is a state-of-the-art approach for such kind of challenges to build a UV-texture and mesh. Haar-Cascade classifier, a machine learning approach for effective face detection is used to generate our UV-textures. Convolution Neural Network model is used with our own dataset and Cosine similarity for obtaining nearest mesh for 3D head models to resemble the corresponding individuals. However, more work is required for identical copy

Analysis of Post-Disaster Damage Detection using Aerial Footage from UWF Campus after Hurricane Sally

Rafael de Sa Lowande, Andrew Clevenger, Arash Mahyari, Hakki Erhan Sevil
University of West Florida (UWF), Florida, USA;
Institute for Human and Machine Cognition (IHMC), USA

Abstract: In this study, we investigate the feasibility of detecting post-disaster damages through camera images obtained onboard an Unmanned Aerial Vehicle (UAV). Aerial footage from the University of West Florida (UWF) after being hit by hurricane Sally in 2020 is used in our study. Our goal is to automatically locate and identify all the roof damages caused by Sally on the university campus and compare two methods of detection. The first is a Convolutional Neural Network (CNN) based approach and the second is a cascade of classifiers model. We utilize cascading classifiers from the OpenCV Python library and a TensorFlow Object Detection API model both retrained on images hand annotated by our team to demonstrate the damage detection capabilities of these models. The aim of this study is to analyze feasibility and compare results between CNN and cascade classifier model for post-disaster damage detection to aid the effort of damage recovery after hurricanes.

Automated Sorting of X-ray Scattering Patterns with Convolutional Neural Networks

Daniela Ushizima, Flavio Araujo, Romuere Silva, Harinarayan Krishnan, Eric Roberts, Alexander Hexemer
Lawrence Berkeley National Laboratory, Berkeley, California, USA;
University of California Berkeley, Berkeley, California, USA;
Federal University of Piaui, Picos PI, Brazil

Abstract: X-ray scattering is an experimental technique that generates patterns conveying material structural information. This technique presents four main modes: small and wide-angle X-ray scattering (SAXS and WAXS) and their respective surface-sensitive variation due to grazing-incidence known as GISAXS and GIWAXS. During a single, high-throughput experiment, these modes can be used interchangeably, therefore sorting is challenging. This paper proposes machine learning tools to sort datasets with thousands of patterns from experiments performed at a synchrotron-light beamline. Each pattern undergoes featurization with convnet, then the extracted features serve as input to AutoML, an automated paradigm for classification and performance evaluation. Our analysis shows that different convnet architectures lead to sorting schemes with high accuracy rates, achieving accuracy of 99.11%.

A Review of Pulse-Coupled Neural Network Applications in Computer Vision and Image Processing

Nurul Rafi, Pablo Rivas
School of Engineering and Computer Science, Department of Computer Science, Baylor University, Texas, USA

Abstract: Research in neural models inspired by mammal's visual cortex has led to many spiking neural networks such as pulse-coupled neural networks (PCNNs). These models are oscillating, spatio-temporal models stimulated with images to produce several time-based responses. This paper reviews PCNN's state of the art, covering its mathematical formulation, variants, and other simplifications found in the literature. We present several applications in which PCNN architectures have successfully addressed some fundamental image processing and computer vision challenges, including image segmentation, edge detection, medical imaging, image fusion, image compression, object recognition, and remote sensing. Results achieved in these applications suggest that the PCNN architecture generates useful perceptual information relevant to a wide variety of computer vision tasks.

A Proposal for Semi-Supervised Learning with GBDT in Image Recognition

Toshiki Hatano, Toi Tsuneda, Yuta Suzuki, Satoshi Yamane
EE & CS, Kanazawa University, Kakumamachi, Kanazawa Shi, Ishikawa Ken, Japan

Abstract: In recent years, deep learning technology has made remarkable progress in the field of image processing [1], and has been applied to a variety of problems. However, since a large number of correctly labeled images must be prepared for learning, there are many difficult situations in which it can be used in the real world. Therefore, in this study, we aimed to learn using a small number of labeled images. As a result, we were able to achieve higher accuracy with a smaller number of images than the previous method and with fewer images than the previous method.

Palm Vein Recognition using SVM and CNN: A Comparative Performance Investigation

Felix Marattukalam, Waleed H. Abdulla, Akshya Swain, Chathura Wanigasekara, Jesin James
The University of Auckland, New Zealand

Abstract: This study carries out a comparative investigation on the classification performance of Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) for palm vein recognition. A relatively small dataset was

used to reduce the computational complexity associated with SVM. The investigation is carried out considering the HK PolyU Multispectral palm vein database using samples from 25 people, each with 12 images. Many studies on palm vein recognition use CNN and SVM as approaches for recognition. But studies till date have not done a comparative evaluation of the two approaches. The results of the study demonstrate that the performance of CNN is superior compared to SVM where the CNN yields 91.7% recognition accuracy compared to 82.4% with SVM. The average precision and recall value for best case of SVM is 0.82 whereas for CNN is 0.92. It is seen that the precision and recall values often reach the ideal value of unity in case of CNN. The results, therefore, suggest that it is possible to get better classification accuracy with CNN from small size of training data and hence more suitable to be applied in palm vein recognition systems.

Early Prediction of Sepsis Utilizing CNN, LSTM and Fuzzy Twin SVM

Chien-Feng Kung, Chia-Te Kung, Chih-Min Su, Pei-Yi Hao
Department of Emergency Medicine, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung, Taiwan, R.O.C.;
Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan, R.O.C.

Abstract: Sepsis is a clinical complication with high mortality and morbidity. The most serious complication of sepsis is septic shock, the mortality rate is as high as 50%. This study proposes a novel technique that combines long short-term memory (LSTM), convolutional neural network (CNN), fully connected neural network and fuzzy twin support vector machine to achieve early prediction of sepsis. We used data from the electronic health records (EHRs) of all patients above 17 years old admitted to the medical ICU at Chang-Gung Memorial Hospital between January 2010 and December 2018. The extracted data contained sets of static features and temporal features. In this study, we proposes a hybrid deep neural network framework which incorporates two additional components with the aim of improving LSTM to automatically extract important features. The first component, a CNN, is added before LSTM to obtain local characteristics of EHRs. The second component, a fully connected neural network, introduces static information (e.g., age) to LSTM. Finally, a LSTM is applied to handle dynamic information (e.g., lab result). The features learned by hybrid deep learning model are fed to a novel fuzzy twin support vector to predict sepsis onset in patients admitted to an intensive care unit. Using EHRs data, sepsis onset can be predicted up to 28 d in advance. Our findings will offer an early solution for mitigating the risk of sepsis onset.

SetGAN: Improving the Stability and Diversity of Generative Models through a Permutation Invariant Architecture

Alessandro Ferrero, Ross Whitaker, Shireen Elhabian
Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, Utah, USA

Abstract: Generative adversarial networks (GANs) have proven effective in modeling high-dimensional distributions. However, their training instability is a well-known hindrance to convergence, which results in practical challenges in their applications. Furthermore, even when convergence is reached, GANs can be affected by mode collapse, a phenomenon for which the generator learns to model only a small part of the target distribution, disregarding the vast majority of the data manifold or distribution. This paper addresses these challenges by introducing SetGAN, an adversarial architecture that processes sets of generated and real samples, and discriminates between the origins of these sets (i.e., training versus generated data) in a flexible, permutation invariant manner. The advantages of this approach are studied theoretically while the state-of-the-art evaluation methods show that the proposed architecture, when compared with GAN variants stemming from similar strategies, produces more accurate models of the input data with lower sensitivity to hyper-parameter settings.

New Efficient Visual OILU Marker

Youssef Chahir, Messaoud Mostefai, Hamza Saidat
Normandie University, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France;
MSE Laboratory, University of Bordj Bou Arreridj, Algeria

Abstract: This paper presents the development of a new real time visual marker, derived from a recent patented work in the field of numbering systems. Proposed visual marker uses a group of projective invariant straight-line patterns, easily detectable and remotely recognizable. Based on efficient data coding scheme, developed marker enables producing a large panel of unique real time identifiers with highly distinguishable patterns. The simplicity and the robustness of the proposed marker makes it well suited for the major augmented reality (AR) and UAV's navigation applications. Extensive experiments tests validate the robustness of the marker against acquisition and geometric distortions.

On The Uphill Battle of Image Frequency Analysis

Nader Bazyari, Hedieh Sajedi
Department of Computer Science, School of Mathematics, Statistics and Computer Science,
University of Tehran, Tehran, Iran

Abstract: This work is a follow up on the newly proposed clustering algorithm called The Inverse Square Mean Shift Algorithm. In this paper a special case of algorithm for dealing with non-homogenous data is formulated and the three dimensional Fast Fourier Transform of images are investigated for hidden patterns.

Skin Cancer Classification Utilizing Atrous Convolutions in a Fine-Tuned Network

Ranpreet Kaur, Hamid GholamHosseini, Roopak Sinha
School of Engineering, Computational and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand

Abstract: Skin cancer is the most common problem all over the world, where some forms of skin cancer are not as aggressive as melanoma. It is vital to identify the type of skin cancer whether benign or malignant for providing timely treatment to the patients to increase the survival rate. The proposed work aims to address this task by proposing a convolutional neural network (CNN) model by leveraging the learning of a pre-trained network called EfficientNetB0. The network is fine-tuned with the optimized selection of hyperparameters and modified network layers at the end to make it suitable for the given dataset. Moreover, the atrous dilated convolution rate is added in some of the feature extraction layers of the existing network. The outcome of the network is analyzed using the locally interpretable model-agnostic explanation (LIME) technique to verify whether the proposed network learned suitable features from the desired portion of lesion region in different skin cancer images. The proposed model employed three datasets of skin cancer; International Skin Imaging Challenge (ISIC), PH2 , and MEDNODE. It is concluded from the experimental results that the adopted deep neural model with the proposed modification is effective in classifying forms of cancer.

A Neural Network Model for Multi-class Image Classification Using Divide and Conquer Method and Bootstrap Sampling

Seunghui Jang, Yanggon Kim
Department of Computer and Information Sciences, Towson University, Towson, Maryland, USA

Abstract: As research on deep neural networks (DNNs) becomes more active, research in the field of computer visions is more sophisticated and enable classifications that are as accurate as humans or more accurately. In order to learn faster and more accurately, many models use transfer learning method, and a lot of research is being conducted on a lot of good pre-trained models. However, many pre-trained models are trained through a large amount of genetic dataset, so in order to use transfer learning to solve a specific-purpose problem, network modification or additional data preprocessing is required. In this paper, we propose a way to improve performance without data preprocessing or network changes of pre-trained model based on the idea of divide and conquer. The proposed method allows more precise feature data to be extracted by subdividing many problem points, and as a result, all problems in the original dataset can be solved quickly with higher accuracy by combining the extracted data into one classifier. We experimented using two CIFAR datasets, divided into two methods, and compared and demonstrated the performance of our approach with the original way

Human Position Estimation with a Mobile Robot Using a Single 360 Degree Camera

Martin Reber, Thomas Kronfeld, Guido Brunnett
Chemnitz Univeristy of Technology, Chemnitz, Germany

Abstract: In the field of human-robot interaction, the accurate recognition and location of persons is of utmost importance. Existing methods consider these two aspects separately. Recognition is based on image analysis methods, whereas depth sensors are often used to calculate the spatial position. A complete coverage of the environment of a mobile robot requires the fusion of several depth sensors due to the limited field of view. In order to avoid this time-consuming calculation step, we present a method that uses a single 360-degree dual fisheye camera to detect people and estimate their positions relative to the camera. We use state-of-the-art methods of human recognition in the image space and use the properties of equirectangular projection to determine the 3D positions of the detected people. Our method does not need a change of the panorama projection and does not require additional training of the underlying neural network.

Varied Few-Shot Relation Network for Non-Uniformly Distributed Source

Yufang Bao
Department of Mathematics and Computer Science, Fayetteville State University, North Carolina, USA

Abstract: In this paper, we propose a new algorithm, called varied fewshot learning, for training the network weights to classify objects by taking into account the distribution of the source when selecting images (shots) from various classes for classification. This method extends the existing few-shot RN network that uses constant number of shots. We allow varied selection of few shots that is in proportion to the source distribution. This new method allows the network weights to be learned in favor of the categories that contains classes that naturally have more objects in the source than others. Our method has significantly higher accuracy rate when compared to the few-shot algorithm in the case when testing objects are distributed nonuniformly, namely, each category carries varied number of objects.

Imagering Hurricane Forecasting based on Artificial Intelligence: A Systematic Review

Raquel Pereira de Alcantara, Adauto Tertulino de Freitas Neto, Diogo Nunes de Oliveira, Eduardo Noronha de Andrade Freitas
Federal Institute of Goias, (IFG), Goiania-Go, Brazil;
College of Communication & Information (CCI), Florida State University, Florida, USA

Abstract: Context: Hurricanes have promoted catastrophes and arduous consequences in many places around the world. Many studies are applied to address the challenge of forecasting them in order to create contingential plans, more especificaly using artificial intelligence techniques. Objective: The goal of this study is to investigate hurricane forecasting techniques whose explores the use of images and prediction models based on artificial intelligence. Method: We develop a search protocol based on database search to perform the systematic literature review and evaluate studies by applying rigor and relevance criteria. Result: We identified 33 studies presenting hurricane forecasting techniques from 153 retrieved studies. Conclusion: We conclude that the lack of information about the implementation of the techniques, indicates that the results of the existing approaches have not been adequately tested and replicated, which requires more rigorous studies on the subject.

A Novel Approach for Parking Space Occupancy Detection Using Deep Learning and Fusion Algorithm

Hertog Nugroho, Ginanjar Suwasono Adi, Ferry Satria, M. Yusuf Fadhlan
Department of Electrical Engineering, Politeknik Negeri Bandung, Bandung, Indonesia

Abstract: Since the rapid economic growth in some countries, the use of private cars is also increase with the same rate. But this phenomenon raises new problems in big cities, especially in finding empty parking spaces. This work proposed an intelligent parking space occupancy detection as a solution to the problem. We adopted YOLO4 classifier to identify cars and parking slots and proposed 3 algorithms, which we call Intersection over Reference (IoR), Euclidean Distance (ED), and a fusion of IoR and ED to detect vacant and occupied parking areas. Based on our experiment, our approach has achieved 99.176% accuracy for IoR algorithm, 99.091% accuracy for ED algorithm, and 99.171% a fusion of IoR and ED algorithm, which is comparable with other works.

An Application of Deep Learning for Sweet Cherry Phenotyping using YOLO Object Detection

Ritayu Nagpal, Sam Long, Shahid Jahagirdar, Weiwei Liu, Scott Fazackerley, Ramon Lawrence, Amritpal Singh
Department of Computer Science, University of British Columbia, Kelowna, BC, Canada;
Agriculture and AgriFood Canada, Summerland, BC, Canada

Abstract: Tree fruit breeding is a long-term activity involving repeated measurements of various fruit quality traits on a large number of samples. Traits such as larger size and specific colors in sweet cherries are of significant commercial value to the growers and consequently the germplasm evaluation process also necessitates evaluation of these traits. Traits are measured manually by counting fruits, weighing and sizing, and classifying them subjectively into different color categories. This process is slow, tedious, expensive and subject to evaluators' bias and fatigue. Recent advancements in the field of artificial intelligence, specifically deep learning, can help automate this process and overcome several limitations of manual measurement. Objective data can be generated for consistent characterization of germplasm, with greater speed and higher accuracy. A method to count the number of sweet cherry fruits in a camera's field of view in real time was developed to automate the process of fruit counting using YOLOv3. A system capable of analyzing the image data for other traits such as size and color was also developed using Python. The YOLO models obtained close to 99% accuracy in object detection and counting of cherries and 90% on the IoU metric for object localization when extracting size and colour information. The model surpassed human performance

and detected manual errors on previous years' image data set, offering a significant improvement compared to manual counting.

Texture Image Retrieval using a Classification and Contourlet-based Features

Asal Rouhafzay, Nadia Baaziz, Mohand Said Allili
Departements d'informatique et d'ingenierie, Universite du Quebec en Outaouais, Gatineau, Canada

Abstract: In this paper, we propose a new framework for improving Content Based Image Retrieval (CBIR) for texture images. This is achieved by using a new image representation based on the RCT-Plus transform which is a novel variant of the Contourlet transform (previously defined in [1]) that extracts a richer directional information in the image. Moreover, the process of image search is improved through a learning-based approach where the images of the database are classified using an adapted similarity metric to the statistical modeling of the RCT-Plus transform. A query is then first classified to select the best texture class after which the retained class images are ranked to select top ones. By this, we have achieved significant improvements in the retrieval rates compared to previous CBIR schemes.

Deep Learning for Infrastructure Defect Monitoring and Forecasting

Erika D. Estrada Medina, Mehran Mazari, Mohammad Pourhomayoun
Department of Computer Science, California State University, Los Angeles, California, USA;
Department of Civil Engineering, California State University, Los Angeles, California, USA

Abstract: Concrete structures are prone to develop surface cracking. Assessing these defects is important to prevent further deterioration and potential failure. This study proposes a deep learning method to predict the growth pattern of cracks in the concrete surface. The method consists of analyzing video of the propagation of individual cracks to generate a series of frames that predict the continuation of the growth sequence over time. Due to their ability to process spatiotemporal data, the suggested models are Convolutional Long Short-Term Memory and Convolutional Gated Recurrent Unit. Transfer learning and data augmentation are implemented to deal with a limited dataset. The results demonstrate that the models are comparably capable of learning the behavior of cracks over multiple time steps.

Multi-Class, Multi-Object Tracking on Multiple Domains

David Schweitzer, Jing-Ru C. Cheng, Benjamin S. Parson, Justin T. Carrillo
Information Technology Lab. (ITL), Engineer Research and Development Center (ERDC), U.S. Army Corps of Engineers, Vicksburg, Mississippi, USA;
ERDC Geotechnical and Structures Lab., U.S. Army Corps of Engineers, Vicksburg, Mississippi, USA

Abstract: Multiple object tracking (MOT), essentially, is an extension of object detection wherein a target in an image needs to be both classified and localized (typically through a bounding box) cross-referenced with some groundtruth. Tracking extends this idea by seeking to both predict where a target will be in a future frame of a video as well as ensure that the target in the future frame is the same target in the current frame; these efforts can impact the Department of Defense's (DOD) missions in the case of (for example) tracking a convoy of vehicles moving through hostile territory. However, most work is done on videos where targets of interest are clear. This work considers the MOT problem on unclear videos where the video may be low in quality or targets (vehicles and persons) may be moving too quickly and the camera may be very unsteady as well as synthetic videos. Such situations are important for the DOD's efforts because hostile territories have a variety of unknowable variables that can affect the quality of information. We examined SORT and Deep SORT as trackers with YOLOv2 (DarkFlow) as a detector and demonstrated the feasibility of such efforts, finding that although effort has obstacles to overcome, it can be improved through more training where shortcomings include a lack of noisy data. We also demonstrate acceptable accuracies for the Deep SORT plus DarkFlow implementation that adheres to the MOT Challenge benchmark metrics.

3D Features Recovery from Images of a Multi-Camera System

Marcela A. Guerrero, Juan M. Santos, Juliana Gambini
Departamento de Ingenieria Informatica - ITBA, Argentina

Abstract: The objective of this work is to automatically find features in 2D images which can be used as markers that allow 3D reconstruction without user intervention. In estimations by using markers, the reconstruction is based on highlight points that the user strategically places on an object of interest or over the suit of a person, designed for this purpose. The idea of this work is to detect feature points in two images taken from different points of view. With this objective, the Scale Invariant Feature Transform (SIFT) method is used. The obtained markers will be utilized to estimate three-dimensional structures such as pose, skeleton, shape or concavity, among others.

Mitigating Algorithmic Bias on Facial Expression Recognition

Glauco A. Amigo Galan, Pablo Rivas, Robert J. Marks
Department of Computer Science, Baylor University, Waco, Texas, USA;
Department of Electrical and Computer Engineering, Baylor University, Waco, Texas, USA

Abstract: Biased datasets are ubiquitous and they present a challenge for machine learning. When a number of categories are equally important but some of them are sparse on the dataset and others are common, the learning algorithms will favor the ones with more presence. This problem is specially sensitive when we deal with people and minorities. How can we, from biased data, generate algorithms that treat every person equally? This work explores one way to mitigate it using a debiasing variational autoencoder with experiments on facial expression recognition.

Tracking of Shorelines Using Unmanned Aerial Systems

Adnan Yunus, Garrett Shake, Dulal Kar
University Of California Santa Cruz, Santa Cruz, California, USA;
Texas A&M University - Corpus Christi, Texas, USA

Abstract: Existing ground surveying techniques for shoreline mapping and detection are difficult, expensive, time-consuming as well as hazardous due to presence of physical obstacles [3]. The measurements obtained using ground surveying techniques can be inaccurate due to unavoidable approximations needed to accommodate inaccessible regions in the measurement. In this work, using unmanned aerial systems (UAS), we propose an augmented method that can overcome some difficulties and reduce inaccuracies as mentioned above. Using computer vision and image processing algorithms, the proposed methodology can detect shorelines by analyzing the characteristics of the video frames obtained using unmanned aerial vehicles (UAV). Specifically, the methodology is based on a machine learning technique applied on filtered video frames of an ocean's shoreline. The results corroborate that the proposed method based on the Support Vector Machine (SVM) learning model for shoreline detection is effective and accurate.

Deep Morphological Network-based Artifact Suppression for Limited-angle Tomography

Ishak Ayad, Cecilia Tarpau, Mai K. Nguyen, Ngoc Son Vu
Equipes Traitement de l'Information et Syst., CY Cergy Paris Universite, ENSEA, CNRS, Cergy-Pontoise, France;
Laboratoire de Physique Theorique et Modelisation, CY Cergy Paris Universite, CNRS Cergy-Pontoise, France;
Laboratoire de Mathematiques de Versailles, Universite de Versailles Saint Quentin, CNRS, Versailles, France

Abstract: Computed tomography has been widely used in biomedical, industrial, and other applications. However, images reconstructed from an insufficient angular range of projection data suffer from important artifacts. Limited-angle tomography is a challenging issue. In this situation the well-known filtered back projection (FBF) reconstruction is inefficient. In order to suppress these artifacts, we developed a new deep learning approach based on the U-net

including a morphological operation in the up-sampling path. Mathematical morphology provides useful non-linear properties for various treatments (segmentation, specific feature extraction). Our method provides good results for angular ranges of 170, 150, 130, even 110 degrees. To the best of our knowledge, it is the first time a limited-angle artifact suppression method works with 110 projections.

Efficient Data Augmentation within Deep Learning Framework to Improve Cross-Dataset Facial Emotion Recognition

Taif Anjum, Steven Lawrence, Amir Shabani
School of Computing, University of the Fraser Valley, B, Canada

Abstract: Facial emotion recognition (FER) is a key component for affective computing in social companion robotics. Current FER datasets are age-biased predominantly excluding seniors. This reduces the performance of deep learning methods in cross-dataset experimentations and hence presents a challenge in our application of social companion robots in long-term care home facilities. Collecting data among seniors is particularly challenging due to privacy concerns and restrictions under pandemic situations (e.g., COVID-19). We address the age-bias in crossdataset experiment using face aging augmentation to include synthetic sample data of the target age group. Our extensive experiments showed that including synthetically aged images during training improves the ability of deep learning classifiers (i.e., MobileNet and DCNN) to detect emotions among seniors. Furthermore, it increases the age diversity in training data and appears as a viable generalization technique.

Regression Networks for Automatic Pain Intensity Recognition in Video using Facial Expression on the X-ITE Pain Database

Ehsan Othman, Philipp Werner, Frerk Saxen, Ayoub Al-Hamadi, Steffen Walter
Department of Neuro-Information Technology, Otto-von-Guericke University, Magdeburg, Germany;
Department of Medical Psychology, Ulm University, Ulm, Germany

Abstract: Facial expression is one of the most prominent responses to pain, and it is very informative for pain recognition in healthcare and medical field. Most works on automatic pain intensity recognition focus on facial features that are extracted independently for every frame of a given video. This approach fails to represent movement dynamics, such as speed, tendency, or overall variation. To overcome this problem, we propose three distinct real-time regression methods to recognize continuous pain intensity automatically based on classifying descriptors of Action Unit (AU) time series and using sliding-window strategy to obtain fixed-length input samples. The first proposed method is a Random Forest regression (RFr) baseline method; the second is Long-Short Term Memory Network (LSTM); and the third is LSTM using sample weighting method (called LSTM-SW). Experimental results (1) report the first results of continuous pain intensity recognition in video from the X-ITE pain database based on facial expression analysis, (2) show that RFr as baseline method is significantly better than guessing to recognize pain intensity, (3) investigate the most efficient loss function with LSTM method by comparing the results achieved, and (4) show that LSTM and LSTM-SW perform better than RFr in recognition pain intensity from video sequences. The findings of this study are the baseline results for future research related to real-time pain intensity recognition in video from X-ITE Pain database.

The Hummingbird Drone: Using Visual Coordination to Intersect Flying Objects

Andrew Sumsion, Dah-Jye Lee
Department of Electrical and Computer Engineering, Brigham Young University, Provo, Utah, USA

Abstract: We present a demonstration of an automated drone's ability to do complicated tasks, such as chasing birds away from dangerous locations, restricted areas, or fragile assets. While this ability applies to multiple industries, it

has particular interest to surveillance systems regarding birds or in cases where birds are considered a pest. In order to demonstrate this complicated feat, we present an autonomous drone that is capable of recognizing a ball that is thrown overhead, calculating the ball's trajectory, and then moving to intersect the ball in the air. This supports the concept of automated drones doing more complicated tasks in the future.

Decorrelated Age-invariant Face Feature for Cross-age Face Recognition

Suli Li, Hyo Jong Lee
Division of Computer Science and Engineering, Jeonbuk National University, South Korea

Abstract: Deep-learning-based face recognition has improved significantly in recent years, even surpassing human capability. However, these methods may not be robust to identify faces across a wide range of ages. In this paper, we propose a new multi-task network architecture based on the feature decompose module for age-invariant face recognition. We use the Resnet101 pre-trained on the VGGFace2 dataset as the pre-trained CNN to capture face structure or the most prominent features for identity discrimination. The features decompose module to obtain the age-invariant features by subtracting age-specific factors from the representations with the help of the age estimation task. Experimental results on the cross-age dataset CACD_VS and CALFW indicate that the proposed method obtains superior performance.

Skin Lesion Classification Using Transfer Learning and DenseNet-201

Juan Pablo Villa-Pulgarin, Anderson Alberto Ruales-Torres, Daniel Arias-Garzon, Mario Alejandro Bravo-Ortiz, Harold Brayan Arteaga-Arteaga, Alejandro Mora-Rubio, Jesus Alejandro Alzate-Grisales, Esteban Mercado-Ruiz, M. Hassaballah, Simon Orozco-Arias, Oscar Cardona-Morales, Pablo Guillen-Rondon, Reinel Tabares-Soto
Department of Electronics and Automation, Universidad Autonoma, de Manizales, Manizales, Colombia;
SEDMATEC, Corporacion Universitaria Autonoma de Narino, Pasto, Colombia;
Department of Computer Science, Faculty of Computers and Information, South Valley University, Qena, Egypt;
Department of Computer Science, Universidad Autonoma de Manizales, Manizales, Colombia;
Department of Systems and informatics, Universidad de Caldas, Manizales, Colombia;
Department of Computer Science, University of Houston Downtown, Houston, Texas, USA

Abstract: Skin cancer is one of the most severe diseases, and medical imaging is among the main tools for cancer diagnosis. This paper focuses on the classification of skin lesion images. Here we report a novel strategy using transfer learning and a new convolutional neural network (CNN) model based on DenseNet-201 that combines recent advancements in the field. Our strategy based on deep learning technique achieve remarkably high skin lesion diagnostic accuracy: 1) We build a CNN with transfer learning taking advantage of ImageNet weights, and various stages of the workflow, including data augmentation and fine-tuning optimization, a detection accuracy of 98% is obtained by using the HAM10000 dataset; 2) This model is tested on the ISIC 2019 dataset obtaining an accuracy of 93%. The satisfactory results obtained with this computational approach could help dermatologists and improve their ability to diagnose skin cancer.

On Two-Person Mutual Action Recognition by Deep Learning

Ting-Hui Chiang, Liang Chia Jung, Yu-Chee Tseng
Department of Information Engineering and Computer Science, Feng Chia University, Taiwan;
Department of CS, College of Computer Science, National Yang Ming Chiao Tung University, Taiwan

Abstract: Human activity recognition is an active research topic in computer vision. While most existing researches focus on single-person activity recognition, understanding the interactions between two persons has its social implications. Recent approaches in this topic mainly focus on using CNNs, LSTMs, and/or GCNs built upon Euclidean distance relations. We observe that using only Euclidean distances is deficient in capturing spatial information in many cases. In this paper, we propose to construct intra- and inter-skeleton relationships based on Vector Matrix (VM). Intra- and inter-VMs are fed into Inception CNNs in parallel with different patch sizes (1×1, 2×1, 3×1) for

extracting critical spatial information. The results are then sent to a LSTM network for analyzing temporal relation of the target interactive activity. The proposed model is validated to achieve better performance compared to the distance-based models on smaller-scale SBU and larger-scale NTU RGB+D and NTU RGB+D 120 datasets.

Preliminary Study on Using Facial Motions for Identify Verification

Zheng Sun, Dah-Jye Lee, Dong Zhang, Weicong Chen, Xiao Li
Brigham Young University, Provo, Utah, USA;
Sun Yat-sen University, Guangzhou, Guangdong, China

Abstract: Many functions in our daily life require identity verification. Biometric-based identity verification techniques are popular and more secure than traditional methods but cannot prevent passive verification or provide liveness assurance. In this paper, an alternative identity verification technique is presented to address these challenges. This unique approach, called Facial Motion Aided IDentification (FMA-ID), analyzes spatial facial features and temporal facial movements concurrently to verify identity. In this study, a metric learning model using recurrent neural network (RNN) is implemented to demonstrate its feasibility. The experiment result of our proof-of-concept study using this network shows a very encouraging 85.2% average precision. Development of more sophisticated networks specifically for this unique application is underway

Semi-Automatic Spatial Markov Chains

Edgar Jardon, Marcelo Romero
Autonomous University of the State of Mexico, Mexico

Abstract: This paper presents a novel methodology to project the urban growth. This proposal begins by computing a binary image to which it is applied our Spatial Markov Chains approach. Contrarily to related literature, we are including a semi-automatic image pre-processing, which traditionally has been done manually. Furthermore, we are introducing our Spatial Markov Chains technique, which is able not only to estimate the urban expansion probability, but also, the expansion position within the map. Six Mexican cities are taken as study case: Acapulco, Puebla, Queretaro, Tampico, Tijuana and Toluca; by training with data from the year 2003 to project that metropolitan urban sprawl in years 2017 and 2031. In this research we are experimenting with state-ofthe-art Landsat 8 satellite images. For evaluation performance, four goodness-of-fit metrics are used, namely Kappa index, Jaccard index, Shannon's entropy and fractal dimension. Then, using such indices, it is possible to measure the concentration degree, geometry level and similarity degree on created projections. We are achieving encouraging results showing a high precision among our projections, making this a suitable option to estimate the urban expansion.

Vision-based Path Planning for Exploration in an Unmanned Aerial Vehicle using OpenCV

Alan Garduno, Mehrube Mehrubeoglu, Pablo Rangel
Electrical Engineering Program, Department of Engineering, Texas A&M University - Corpus Christi, Texas, USA

Abstract: The tag detection concept has been an important area of implementation when coordinating multiple robots. For instance, live camera feed on an autonomous robot needs to be processed in a timely manner. We present the use of OpenCV algorithms and the Augmented Reality Tags for efficient detection of objects and to coordinate a robot's path plan. This visual marker system uses a tag that allows six degrees-offreedom localization for faster and robust visual detection. OpenCV algorithms and functions have been used to support the Augmented Reality approach in camera calibration, tag detection, and path-planning for exploration in a UAV that is extendable to multi-robotic systems. This research was aimed to develop an improved approach for detecting images, where robustness, speed, and flexibility are critical in the high-level controls of autonomous systems.

No-reference Screen Content Image Quality Assessment Using Deep Neural Networks

Shih-Hsien Chuang, Jin-Jang Leou
Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, ROC

Abstract: In this study, a no-reference (NR) screen content image (SCI) quality assessment using deep neural networks is proposed. The proposed approach consists of two main stages, namely, feature extraction and feature regression. In the feature extraction stage, the feature maps of each SCI are extracted in a nonuniform fashion via multiple convolutional layers. The extracted maps are transformed into uniform column feature vectors via a spatial pyramid pooling (SPP) layer, and stacked into one column feature vector. In the feature regression stage, the column feature vector is processed by two fully connected layers with Euclidean normalization, which is used to train a regression model. The predictive image quality assessment (IQA) score of each SCI is assessed via the trained regression model. Based on the experimental results obtained in this study, the performance of the proposed approach is better than those of most comparison NR-SCI-IQA approaches.

The 18th International Conference on Modeling, Simulation & Visualization Methods (MSV 2021) https://www.american-cse.org/csce2021/conferences-MSV

Identification of Complex System using Hybrid Method

Guangjun Liu, Xiaoping Xu, Feng Wang, Long Liu
School of Sciences, Xi'an University of Technology, Xi'an, P. R. China;
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, P. R. China;
School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, P. R. China

Abstract: System identification is a powerful technique for building models of complex systems in communications, signal processing, control, and other engineering disciplines. This paper discusses an identification approach for a class of complex system with unknown structure. The basic idea is as follows. Firstly, a system model composed with classical models is adopted to transform the system structure identification problem into a combinatorial optimization problem. Then, a modified artificial fish swarm algorithm with strong convergence is applied to synchronously implement the identification on the system's structure and parameters. Finally, in simulation experiments, compared with the existing method, simulation results show the feasibility of the presented approach.

Active Non-Terminals in Grammar-Based Data Generation

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

Abstract: Active non-terminals change the way data is selected from productions in grammar-based data generation. Productions become active devices that can respond to a variety of commands instead of being passive, descriptive entities. Active non-terminals have been introduced into the context free grammar-based data generation tool DGL. This change is the enabling technology for several new features of DGL, such as external file management and database access.

Revisiting the Out-of-Order Superscalar Processor Core

Jongbok Lee
Department of Electronics Engineering, Hansung University, Seoul, South Korea

Abstract: As multi-core processors, GPUs, SoCs, NPUs are widely used today, the importance of out-of-order superscalar core processor, which is an essential basic unit or the component for all these architectures cannot be under estimated. Most of the out-of-order superscalar core processor exploits Tomasulo algorithm where the reservation stations and register status units are used for the register renaming. First, this paper proposes a modified register status unit for Tomasulo algorithm which effectively renames registers in various situations such as the case when the destination register and the source register are identical. Secondly, for the description of the hardware, an elegant way of expressing the reservation station, the reorder buffer, and the register status as VHDL record is introduced. Thirdly, for maintaining the consistency of these data structures, a round-robin connection technique of the major components is suggested. The proposing out-of-order superscalar core processor has been verified by GHDL and GTKwave.

Virtual Simulation of 3D Modeling

Juanita L. Benjamin
Texas Tech University, Texas, USA

Abstract: Education plays an integral role in the lives of students. The way they learn has been taught the same way for many years. Now that we are entering a new age with immersive media being used in all sectors, including retail, gaming, mobile development, etc. Therefore, it would be beneficial to use this technology within the education division. Students nowadays play many interactive games like Minecraft or Fortnite, which they can be playing for hours without cease. To increase students' engaged participation, motivation, and willingness, immersive media should be incorporated into their learning activities. This new addition to the curriculum can capture the student's or users' attention and compel them to be attentive, and heighten their curiosity in their education. They will see this as a fun way of learning science or math. In this extended abstract, we introduce an application for Virtual Reality with 3D modeling that encourages learning through exploration.

Automated Graphing for Modeling and Simulation

Sara Boyle
US Army - DEVCOM GVSC - Analytical Development and Data Science, Warren, Michigan, USA

Abstract: High fidelity multi-body physics modeling and simulation is used to evaluate vehicle models in specific scenarios that represent real world applications. It is a useful tool for identifying the best vehicle model designs but it can be challenging to match the physical test and the simulation runs can generate a lot of data. For verification and validation purposes it is often important to compare the simulation time series data rather than just a single output metric to identify how closely the modeling and simulation tool represents the physical world. Comparing time series

data is a challenging task and a lot of factors need to be considered to ensure that the comparison is one-to-one between the modeling and simulation data and the physical test data.

The Frequency Effect of Information State Update on Consensus in Multi-Agent Systems

Azad Azadmanesh, Hassan Farhat, Rui Yang, Satish Modugu
University of Nebraska-Omaha, Nebraska, USA

Abstract: A fundamental problem in a multi-agent system (MAS) is the cooperative control of agents in a distributed and autonomous manner to reach consensus, so that every agent has a consistent view of the global information state. This study illustrates an aspect of the consensus problem through simple algebraic and matrix theory, showing how consensus can be achieved through rounds consisting of the three simple steps of collecting, voting, and broadcasting the information state to the neighbors. It is shown how these steps can gradually achieve consensus through discrete steps and the speed at which consensus can be achieved through continuous-time with discrete time updates. The results show the limitation of continuous-time updates and the effect of the communication topology that can be important factors in the design of MASs for which consensus is a major objective. The results are applied to a simulation model in Robotic Operating System (ROS). Through simulation, the effect of frequency updates on information states is illustrated.

Periodic Steady-state Computation in Time-domain of Microgrids with PV Systems based on Companion-circuit Analysis

Julio Cesar Godinez-Delgado, Aurelio Medina-Rios
Edificio de la Division del Posgrado de la Facultad de Ingenieria Electrica, Universidad Michoacana de San Nicolas de Hidalgo, Ciudad Universitaria, Michoacan, Mexico

Abstract: This article details the application of a methodology in the time-domain (TD) for the analysis of the periodic steady-state operation of microgrids supplied by photovoltaic (PV) generation sources. The applied methodology is based on the combination of companion-circuit analysis (CCA) and two Newton extrapolation techniques to the limit cycle (LC), based on the conventional and enhanced numerical differentiation processes (ND and END), respectively, i.e. CCA-ND and CCAEND. The CCA method uses the applying integration methods, e.g. forward Euler (FE) and trapezoidal rule (TR), for the representation of electric components by Norton equivalents. The results associated with the case study are given in terms of total harmonic distortion (THD), as well as the exchange measurement of maximum power transfer at the point of common coupling (PCC) between the microgrid and the PV system. The results are successfully validated through direct comparison with the PSCAD/EMTDC R simulator.

The 27th International Conference on Parallel & Distributed Processing Techniques & Applications (PDPTA 2021) https://www.american-cse.org/csce2021/conferences-PDPTA

A Comparison of Vivado HLS, SDSoC C++ and OpenCL for Porting a Matrix-vector-based Climate Model Mini-app to FPGAs

Moteb Alghamdi, Graham D. Riley, Mike Ashworth
Department of Computer Science, University Of Manchester, Manchester, United Kingdom, UK

Abstract: The High-Performance Computing (HPC) community's interest in FPGAs as accelerators has been renewed due to the introduction of the High-Level synthesis tools (HLS). HLS tools hide the complexity of FPGA programming through raising the abstraction level for programmers. They offer environments where traditional HPC programmers can use highlevel languages such as C/C++ and OpenCL to implement HPC application kernels on FPGAs. For the traditional HPC scientific programmer, the appropriate HLS environment choice implies trade-offs between the achievable performance and programmer effort. This is related to the level of detailed knowledge of FPGA hardware and performance improvement techniques assumed by the HLS environments. This paper presents a comparative study between three HLS programming methodologies, Xilinx Vivado HLS and Xilinx SDSoC using both OpenCL and C++, all targeting the Xilinx Zynq UltraScale+ MPSoC ZCU102. Based on an existing low-level Vivado HLS design of a finite-element kernel in the context of a realistic LFRic Climate model miniapp from the Weather and Climate domain we compare the programming techniques, effort and resulting performance of implementations developed using the higher level of abstraction provided by SDSoC C/C++ and OpenCL. We provide a comparative analysis of insights that lead to the different design choices, scaling behaviour and peak performances obtained. We find that Vivado HLS provides the highest performance due to the programmer's ability to exploit low-level FPGA features in the manual construction of the hardware system design, but near equivalent solutions can be obtained with OpenCL and C++ with a higher-level view of the FPGA hardware, including automatic generations of the system's design, and, hence, less programmer effort.

Novel Prime Finding Algorithm

Yujin Yoshimura, Eduardo Colmenares
Department of Computer Science, Midwestern State University, Wichita Falls, Texas, USA

Abstract: Prime numbers are commonly used today, such as the cryptosystem RSA (Rivest-Shamir-Adleman) is guaranteed secure by large prime numbers. Prime numbers had been studied and fascinated humanity since ancient times. This research presents a novel prime finding algorithm based on geometry, which sieves more efficiently than existing algorithms. It also presents and evaluates the results of the proposed approach from three different perspectives: (1) typical sequential approach, (2) multicore shared memory approach, and (3) parallel many-core approach, that the proposed algorithm can exploit the power of parallel computation, with its ease of parallelization.

A Shared-Memory Parallel Implementation of the Minimax Algorithm

Lewis Sharpe, Hans-Wolfgang Loidl
Department of Computer Science, Heriot-Watt University, Edinburgh, Scotland, United Kingdom

Abstract: This paper investigates the process of creating a parallel implementation of the Minimax with Alpha-Beta pruning algorithm for game tree search. The implementation is written in C# and .NET's Task Parallel Library (TPL) is used to create the shared-memory parallelism within the search. The test game cases used to test the performance of the algorithm included TicTacToe, Treblecross and GoMoku, albeit on modified board sizes. The suggested parallel computational model utilises task parallelism near the root of the game tree and employs branch-and-bound to prune the game tree. The parallel variant achieves a modest but noticeable absolute speed-up over the sequential base when selecting a first move. We give evidence that the flattening out of speedups beyond 32 cores is caused by increased memory consumption.

Optimized 4-digits Quantum Carry Lookahead Adders

F. Orts, G. Ortega, E. Filatovas, E. M. Garzon
Informatics Department, University of Almeria, Spain;
Institute of Data Science and Digital Technologies, Vilnius University, Lithuania

Abstract: Quantum computing has been postulated as one of the most important Post-Moore technologies in recent years. To make progress in Quantum Computing, the development of efficient quantum circuits is of great interest. Adders are the most remarkable arithmetic circuits in quantum computing systems. For instance, they are a fundamental part of Shor's algorithm, the most famous quantum algorithm. As quantum circuits are very sensitive to noise, noise reduction and fault tolerance are two of this disruptive paradigm's most important objectives. This work presents the design of two novel reversible fault-tolerant carry lookahead adders. Moreover, a comprehensive analysis has been carried out to compare the proposed adders to the state-of-the-art circuits. The obtained results have demonstrated that our proposals outperform the best adders available in the literature

Computing Extreme Eigenvalues of General Complex Matrices on the GPU

Basileal Imana, Peter Yoon
Department of Computer Science, Trinity College, Hartford, Connecticut, USA

Abstract: The Jacobi method for computing eigenvalues for general dense matrices, though introduced in 1846, has not become popular as they were computationally expensive for sequential computing. With the advent of parallel computing, however, it has become feasible to efficiently implement such algorithms on various types of multiprocessors. In addition, the Jacobi method has been shown to compute very small eigenvalues with high accuracy compared to the conventional methods. In this paper, we present a novel parallel implementation of Jacobi method for eigendecomposition of general complex matrices on the GPU. Our preliminary results show a significant improvement over those on the CPU, running up to 94 times faster for general dense complex matrices of moderate size.

Clustered Visualization of Large Scenes using Actors

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

Abstract: This work extends on previous work done to visualize large scenes using an actor based system by distributing it across a cluster and adding photometric rendering. The cluster distribution of ray-tracing is shown to scale well. The addition of photometry required few updates to our previous actor architecture, indicating that the previous work was a good generalized framework of casting rays through large, distributed sets of geometry. We also find that it is well worth the effort required to change to a non-standard serializer.

A System for Popularity-aware and Energy-efficient Offloading Decision with Load Balancing in Fog Computing

Chiu-Shun Hsiang, Yung-Ting Chuang
Department of Institute of Information Management, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Abstract: As more and more applications with high computing requirements appear on mobile devices, considering the limited computing resources and energy of those devices, more and more devices will offload this type of task to the surrounding cloud or fog. By offloading tasks to the server, mobile devices can save execution time and energy. In addition, the energy consumption of the overall system and stability are also factors that cannot be ignored. Therefore, we propose an offloading model that considers the load balance of the server and its popularity to achieve the goal. Our goals are as follows: 1) Provide a more stable and robust system; 2) Minimize the energy consumption of the overall system; 3) Maximize the execution efficiency of the mobile device.

Distributed Photometry of Large Scenes Using Spark

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

Abstract: This work builds on previous work using Spark to do large, distributed renderings of scenes with more geometry than will fit in the memory of a single machine. Here we look at photometric rendering to see how performance scales with geometry size and photon counts. We find both scalings to be sub-linear on a small cluster of nine machines. The time scaling with photon counts is remarkably sublinear. We currently have no explanation for what is causing that. In general, we find that Spark works well as a general distributed computation engine and allows us to set up largescale computations with remarkably little code.

An Analytical Model of Virtual Cut-Through Routing in a Multidimensional Toroidal Network

Lev B. Levitin, Yelena Rykalova
Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts, USA;
Computer Science Department, University of Massachussetts, Lowell, Massachusetts, USA

Abstract: A theoretical model of an n-dimensional toroidal interconnection network with the virtual cut-through routing is developed. The network performance is characterized as a function of the message load for various values of the network and communication parameters. An exact analytical expression for the saturation point has been obtained. The latency as a function of message generation rate under the mean field theory approximation has been derived. The saturation point has been found proportional to the number of the torus dimensions and inversely proportional to the message length and to the length of the path from the source to the destination. It was found that the transition to the saturation state is a second-order (continuous) phase transition with a critical exponent equal to 1.

Partitioning based Incremental Travelling Salesman Algorithm on Time Evolving Graphs

Shalini Sharma, Jerry Chou
Institute of Information System and Applications, National Tsing Hua University, Taiwan;
Department of Computer Science, National Tsing Hua University, Taiwan

Abstract: Travelling salesman problem (TSP) has been applied in different application domains, such as transportation and logistics. However, due to NP hardness of TSP, it is vital to reduce the computational cost of TSP algorithm. Reducing the complexity of TSP becomes a even greater challenge, when we apply the problem on time

evolving graph(TEG), where the topology of graph changes over time intervals because of update events. Our recent work [21] made one of the earliest attempt to address this issue by proposing incremental TSP algorithms. In this paper, we further improve that result by using divide and conquer paradigm to reduce the problem size and proposed Partitioning algorithms P-TSP and Pg-TSP. From our experimental results, we have observed partitioning TSP algorithms (P-TSP and Pg-TSP) are faster than the incremental algorithms (ITSP and Ig-TSP), and the computation time can be reduced to at least half with more number of partitions.

Set-to-Set Disjoint Paths in a Folded Hypercube

Hiroyuki Ichida, Keiichi Kaneko
Graduate School of Engineering, Tokyo University of Agriculture and Technology, Koganei, Tokyo, Japan;
Institute of Engineering, Tokyo University of Agriculture and Technology, Koganei, Tokyo, Japan

Abstract: The hypercube is a very popular topology for the interconnection networks of parallel systems, and the folded hypercube is a variant topology of the hypercube. The folded hypercube attains much higher performance by introducing one additional link to each processing element. Therefore, there are many research activities regarding the folded hypercube. We focus on this topology and address an unresolved problem, that is, the set-toset disjoint paths problem in it. In this paper, we show an algorithm that solves the problem in a folded hypercube. We show that the time complexity of the algorithm is $O(n\ 3\log\ n)$ and the maximum length of the paths is $2n$ + 2 if the algorithm is applied to an n-dimensional folded hypercube.

Dimension-Balanced Pancyclicity on T_{m,n} When Both m and n are Odd

Justie Su-Tzu Juan, Wen-Fan Peng, Zong-You Lai
Department of Computer Science and Information Engineering, National Chi Nan University, Puli, Nantou, Taiwan

Abstract: The dimension-balanced cycle problem is a quiet new topic of graph theorem. Given a graph G = (V, E) and a cycle C on G, the edge set E is partition into k dimensions for a positive integer k. The set of all i-dimensional edge of C, a subset of E(C), is denoted as Ei(C) for $1 \le i \le k$. If $||Ei(C)| - |Ej(C)|| \le 1$ for $1 \le i < j \le k$, C is called a dimension-balanced cycle. For toroidal mesh graph Tm, n, the dimension-balanced cycle problem when at least one of m, $n \ge 3$ is even have been discussed. In this paper, we study the problem for both m, $n \ge 3$ are odd, and prove that when both m, n > 3 are odd, Tm, n contains a dimension-balanced cycle whose length is 4k for any integer $1 \le k \le \lfloor (mn - 3) / 4 \rfloor$.

Parallel Frequency Detection and Modelling of Large Time Series Based on Block Recurrent Pseudoinverse Matrices

Christos K. Filelis - Papadopoulos, Panagiotis E. Kyziropoulos, John P. Morrison, Philip O'Reilly
Department of Computer Science, University College Cork, Cork, Ireland;
Cork University Business School, University College Cork, Cork, Ireland

Abstract: Time series modelling and forecasting techniques have a wide spectrum of applications in several fields including economics, finance, engineering and computer science. Recently, a novel modelling and forecasting technique based on recursive pseudoinverse matrices for general time-series has been presented. Herewith, a parallel block variant based on harmonics for large general time-series is proposed and analysed. The frequencies in the initial signal are estimated using a parallel adaptive approach that searches potential frequencies based on initial guess obtained by the Fast Fourier Transform. The novel block technique is combined with an optimization procedure to determine the fitting tolerance that leads to better generalization in order to improve forecasting accuracy. The implementation of the proposed technique is performed using Python and parallelization for shared memory parallel systems is performed using the Message Passing Interface, in order to avoid the Global Interpreter Lock. Numerical results depicting the efficiency and applicability of the proposed technique are also given.

Novel Micro Road Pricing Method Based on Spatio-Temporal Grid Reservation for Collecting Gasoline Tax Revenue from Zero-Emission Vehicle

Habibullah Babakarkhail, Kenny Sato
Department of Info. Science and Tech., Graduate School of Engineering, Doshisha University, Osaka, Japan;
Network Information System Laboratory, Doshisha University, Osaka, Japan

Abstract: This work proposes a novel micro road pricing alternative method to replace the current gasoline tax collection method by providing a software architecture platform. In this platform, the automated vehicle reserves (virtual) spatio-temporal sections on the road in real-time for road pricing. The time and space are divided into grids and assigned a designated charge to a grid managed on a dynamic map. Automated vehicle reserves the planned travel route, time in advance, and mediates based on the reservation information. The performance evaluation results show that the system reserves the requested grid, route, and collects the gasoline tax charges with minimum communication time with no data package loss. Furthermore, the simulation results demonstrate that the proposed method succeeded in compensating and sustaining the annual revenue with a constant rate and an increase in revenue estimated as 21.4% in 2025. As a result, the proposed method is promising and efficient for future applications.

Selfish-Stabilizing Spanning Tree for Two-Colored Graphs

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

Abstract: In today's self-stabilizing distributed systems, each process or a node or each administrative domain may want to optimize its own incentive. They might have selfish motives in addition to the global convergence property. The nodes or the domains do not require to give up their stabilization property, but they could make moves governed by different rules, depending on their types to maximize/minimize own payoffs. Optimizing individual payoffs without sacrificing the stabilization property is a relatively new trend compared to traditional self-stabilization and this characteristic of the system is known as selfish stabilization. In this paper, we discuss a selfish stabilizing algorithm for constructing spanning tree for two-colored graphs, where the colors represent different types of processes/nodes.

GPU Based ODE System Solving Framework

Ma Lin, Zhang Xuesong, Lei Xinli
Computer Science and Technology, JiLin University, ChangChun, China

Abstract: In complex simulation scenario, there usually includes large number of differential equations which cost too much computing time, leading to low simulation efficiency due to the insufficient utilization of computing resource including CPU and GPU. For this reason, we propose a pipeline ODE solving framework. In this framework, a high efficiency CPU side dynamic solver including traditional numerical solver and the RIDC solver is implemented temporarily, and the preliminary experiment result shows the desired efficiency and accuracy.

Identity: In, By, and, Of, Blockchains

Steven A. Wright
College of Law, Georgia State University, Atlanta, Georgia, USA

Abstract: Permissionless blockchains provide a parallel computing environment over a heterogeneous peer-peer network infrastructure. Many of the proposed applications use blockchains to control the disposition of property as digital assets via smart contracts. These transactions are between pseudonymous identities based on cryptographic keys. Identities are thus key to the initiation and effect of blockchain operations in the real world. This paper explores notions of identity in, by, and, of, blockchains in order to identify barriers to commercial deployment as areas for further research.

SESSION: Blockchain Technology (BT 2021)

Co-Chairs: Dr. Amjad Gawanmeh and Dr. Kashif Saleem***
** University of Dubai, UAE*
***King Saud University, Saudi Arabia*

Big Data Analytics and IoT for Personalised Healthcare: Opportunities and Challenges

Sitalakshmi Venkatraman, Sazia Parvin, Wathiq Mansoor, Amjad Gawanmeh
Melbourne Polytechnic, Australia;
Curtin University of Technology, Australia;
Universithy of Dubai, United Arab Emirates, UAE;
University of Dubai, United Arab Emirates, UAE

Abstract: The increased availability of health data in digital platforms provides huge potential for the use of Big Data in personalized healthcare. While there has been some advancement in the adoption of information and communication technology (ICT) in this context, there is huge amount of under-utilized health data. With the prevalence of Internet of Things (IoT) based point-of-care (POC) devices used for personal health monitoring and diagnosis, Big Data can help in healthcare. Big Data analytics can provide useful insights in monitoring, diagnosis, and self-management of health issues for a better personalized healthcare. In this paper, we identify the current trends that motivate us to propose a big data analytics framework for personalized healthcare. We provide the baseline architecture in adapting three popular machine learning algorithms to leverage the current trends in IoT, Big Data infrastructures and data analytics for advancing personalized healthcare of the future.

BCT-Voting: A Blockchain Technology Based Voting System

Deepali Raikar, Avimanyou K. Vatsa
Fairleigh Dickinson University, Teaneck, New Jersey, USA

Abstract: In order to support the nation's security of a modern democratic country, the security of the election polling system has a major role. It includes high-level security and immutable feature of votes cast by voters. The identification, authentication, and detection of fraud voters are also major concerns in democracy. Furthermore, the background check of electoral candidates, detection of fraud voters and donors, and detection and prevention of malicious software (Malware) or cyber-attacks on e-voting systems are challenging issues. Moreover, the effort, time, and cost of an election process need to be minimized. Therefore, there is a need to model, design, implement a trusted, transparent, integrated, and secure voting system that is based on blockchain technology, called BCT-Voting. Blockchain technology works for an untrusted ecosystem in a decentralized environment. Additionally, blockchain technology provides privacy oftraceable and verifiable features of a voting process such that fraud activities may be detected and prevented. This paper has contributed four models - registration and login, vote casting andcounting, result declaration, and donation. These models are designed and implemented using a blockchain technology framework enabled with smart contracts and cryptocurrency wallets. These proposed secure and immutable voting system (BCT-Voting) solves five major issues of the electoral process – nomination of candidates, secure voter Identity, vote tampering, a donation to electoral parties or candidates, and counting of votes and winner announcement. These processes will be validated and verified by a consensus mechanism. Also, a user-friendly graphical user interface (GUI) - Dapp - is proposed.

Users and Security Elicited Requirements for System (USERS)

Natchapol Worakitpreeda, Monvorath Phongpaibul
Thammasat University, Thailand

Abstract: This framework illustrates a proposal for eliciting security requirements from the earliest approach of design specifications to improve the security levels of software since the most common cause of a data breach or system failure comes from insufficient identity security levels. This problem happens at the beginning of the first stage of software development which is the requirement engineering process, especially elicited non-functional security requirements. Perhaps not surprisingly, many people are unaware of security, and business owners, who are called users, do not have a vital role in this resulting in a lack or excess of security, so we attend to design a new framework for resolving sufficiently efficient security. The main contribution of this framework is user involvement, and the levels of each system. The result of this framework derives pragmatic evaluation from security specialists and security standards. Furthermore, we assess the practical framework with statistics of user satisfaction rates.

WORKSHOP/SESSION: Mathematical Modeling and Problem Solving (MPS)

Co-Chairs: Prof. Masahito Ohue, Prof. Masakazu Sekijima*, Prof. Kazuki Joe***
** Tokyo Institute of Technology, Japan*
***chair of Life Computing and Communication Department, Nara Women's University, Japan*

NARU: Natural-gradient AutoRegressive Unlossy Audio Compressor

Taiyo Mineo, Hayaru Shouno
The University of Electro-Communications, Chofu, Tokyo, Japan

Abstract: In lossless audio compression, it is essential for predictive residuals to remain sparse when applying entropy codings. The sign algorithm (SA) is a conventional method for minimizing the magnitude of residuals; however, it exhibits poor convergence performance compared with the other adaptive algorithms. Although the natural gradient sign algorithm (NGSA) exhibited better convergence performance than SA, its practical applications were not provided yet. This paper proposes a novel lossless audio codec based on the NGSA, called Natural-gradient AutoRegressive Unlossy Audio Compressor (NARU). Its implementation was written by C and open-source. We compared the NARU with existed well-known codecs and showed better compression performance.

Generating Diverse Solutions for Nurse Scheduling

Naoaki Katoh, Wei Wu, Atsuko Ikegami
Seikei University, Japan

Abstract: Nurse scheduling involves creating a schedule that not only has the required number of nurses in each shift each day, but also satisfies many constraints considering the health and social life of each nurse. In addition, due to implicit constraints and unstated criteria, simple optimization techniques that find a single optimal solution do not suffice for obtaining a practical solution. In hospitals, schedules are created with attention given to the work shifts of skilled nurses. In this study, we focus on decision variables concerning skilled nurses which frequently appear in the formulation of optimization models, and we treat the degree of difference in the values of these variables as the "difference" between solutions. We propose a model for simultaneously obtaining multiple solutions that differ greatly from each other. We then observe the differences between multiple solutions given by our model and how these differences affect the quality of each schedule.

An Improved Branch-and-Bound MCT Algorithm for Finding a Maximum Clique

Etsuji Tomita, Jiro Yanagisawa, Kengo Katayama, Kazuho Kanahara, Takanori Toda, Hiro Ito, Mitsuo Wakatsuki, Tetsuro Nishino
The University of Electro-Communications, Chofu, Tokyo, Japan

Abstract: We improve a branch-and-bound algorithm called MCT (Tomita et al., FAW 2016, LNCS 9711, pp.215{226, 2016) for finding a maximum clique. First, we devise a new efficient approximation algorithm for finding a maximum clique. Second, we employ MIS vertex ordering with an appropriate precondition. Third, we employ a combination of Re-NUMBER and Infra-chromatic bound. Finally, we devise an adaptive change of stages of the search tree. The finally improved MCT algorithm is named MCT*. It is shown that MCT* algorithm is significantly faster than MCT by extensive computational experiments. In addition, it is shown that MCT* algorithm is faster than the state-of-the-art IncMC2 algorithm (Li et al., INFORMS J. Computing, 30, pp. 137{153, 2018) for many instances.

OptWedge: Cognitive Optimized Guidance toward Off-screen POIs

Shoki Miyagawa
Mitsubishi Electric Co., Kanagawa, Japan

Abstract: Guiding off-screen points of interest (POIs) is a practical way of providing additional information to users of small-screen devices, such as smart devices and head-mounted displays. Popular previous methods involve displaying a primitive figure referred to as Wedge on the screen for users to estimate off-screen POI on the invisible vertex. Because they utilize a cognitive process referred to as amodal completion, where users can imagine the entire figure even when a part of it is occluded, localization accuracy is influenced by bias and individual differences. To improve the accuracy, we propose to optimize the figure using a cognitive cost that considers the influence. We also design two types of optimization with difference parameters: unbiased OptWedge (UOW) and biased OptWedge (BOW). Experimental results indicate that OptWedge achieves more accurate guidance for a close distance compared to heuristics approach.

Antisense Oligonucleotide Activity Analysis based on Opening and Binding Energies to Targets

Kazuya Isawa, Keisuke Yanagisawa, Masahito Ohue, Yutaka Akiyama
School of Computing, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo, Japan

Abstract: In the development of antisense oligonucleotide (ASO) drugs, a large number of ASOs that are reverse-complementary to the target transcripts are designed with 13–25 bases. The experimental determination of the binding affinity of ASOs is time-consuming and costly; thus, the computational prediction of desired ASOs is in high demand. In this study, we analyzed the relationship between the affinity of ASO and the inter- and intra-hybridization energies of ASO and/or target mRNA. As a result, the binding energy between ASO and mRNA was most correlated with the inhibition rate of gene expression ($r = -0.448$). Additionally, the inhibition rates tend to be low when the binding site on the target mRNA forms a strong secondary structure regardless of the highly stable complementarity between ASO and the target. These findings will be useful for building better computational prediction models in the future.

Face Impression Classification in Cosmetic Counseling Using Deep Convolutional Neural Network

Masaharu Kurosawa, Hayaru Shouno
Graduate School of Informatics and Engineering, University of Electro-Communications, Japan

Abstract: The evaluation of facial impressions by beauty experts plays an important role in cosmetic counseling. However, the opportunities to avail counseling from beauty experts are limited by location and time. Therefore, a prediction model is required for the evaluation of facial images by beauty experts. However, as an appropriate image dataset required for the evaluation by beauty experts does not exist, the evaluation is considered "tacit knowledge," making feature engineering difficult. Thus, we have developed a new dataset and used a deep convolutional neural network (DCNN) model to perform end-to-end training, from feature extraction to the classification, to develop a prediction model with the skill of a beauty expert. The dataset was developed using facial images of 377 Japanese women, labeled with a unique impression based on the sensitivity and experience of beauty experts (skin power dataset: SP dataset). We applied the transfer learning method using pre-trained DCNN models on a large facial image dataset to achieve effective features of beauty experts' evaluation with a small amount of data. Furthermore, we visualized the facial areas that contributed to the prediction of the DCNN model. Thus, although the model's accuracy needs to be improved for practical use, we conclude that the DCNN can represent the unique evaluation of face impressions in cosmetic counseling, as provided by beauty experts.

Proposal of an Essential Oil Blending Method using Q-learning

Kyoka Nakano, Masami Takata
Nara Women's University, Nara, Japan

Abstract: Essential oils are aromatic substances that are extracted from plants. Customized scents can be created by blending several essential oils. However, because there are many types of essential oils, it is difficult to select the appropriate oil. In this study, we propose a method for automatically selecting essential oil blends that are specific to individual preferences, with the aim of enabling users to easily blend essential oils. In the proposed method, Q–learning, a type of reinforcement learning, is adopted. The experimental results confirm that the proposed method can learn the preferences for different scents and environmental changes.

Implementation of Singular Value Decomposition for Complex Matrices Using the Two-sided Jacobi Method

Takahiro Miyamae, Masami Takata, Kinji Kimura, Yoshimasa Nakamura
Nara Women's University, Nara, Japan

Abstract: The time reversal method requires a singular value decomposition (SVD) for complex matrices. First, each element of the time–reversal operator obtained from the time–series data is transformed into the frequency domain. The decomposition must then be conducted within the frequency domain. Thus, all elements of the target matrices of the SVD are complex. The Jacobi method used to apply a singular value decomposition can compute all singular values and singular vectors with high accuracy. However, the computation cost of the Jacobi method is higher than that of a computational method combining the QR method with bidiagonalization using a Householder transformation. Notably, the computational cost is insignificant for considerably small matrices. Based on the Jacobi method, one– and two–sided Jacobi methods have been proposed. The one–sided method was implemented in LAPACK. In this study, we improved the two–sided Jacobi method for complex matrices. We experimentally confirmed that, for small matrices, the two–sided Jacobi method has a shorter computational time and higher accuracy than the one–sided Jacobi method.

MEGADOCK-Web-Mito: Human Mitochondrial Protein-Protein Interaction Prediction Database

Masahito Ohue, Hiroki Watanabe, Yutaka Akiyama
Department of Computer Science, School of Computing, Tokyo Institute of Technology, Nagatsutacho, Midori-ku, Yokohama City, Kanagawa, Japan

Abstract: Mitochondrial diseases are largely caused by dysfunction in mitochondrial proteins. However, annotations of human mitochondrial proteins are scattered across various public databases and individual studies. To facilitate research aimed at elucidating mitochondrial functions, we constructed the MEGADOCK-Web-Mito database as a protein–protein interaction (PPI) prediction data archive, including prediction results for exhaustive protein pairs of 654 mitochondria-related human proteins. MEGADOCK-Web-Mito enables users to search for all PPI prediction results efficiently and comprehensively. In particular, we linked functional annotations to each human mitochondrial protein. The comprehensive and specialized human mitochondrial PPI prediction results and searching function of MEGADOCK-Web-Mito will support further research on mitochondria and mitochondrial diseases.

Timing to Start to Feel Anxious while Moving with an Autonomous Personal Mobility Vehicle

Kotoyu Sasayama, Tomoya Uefune
Information Technology R & D Center, Mitsubishi Electric Co., Kamakura, Kanagawa, Japan

Abstract: In recent times, the demand for autonomous personal mobility vehicles (APMVs) has been growing. People who ride APMVs, such as those who ride autonomous vehicles, experience anxiety in many situations. One way to relieve anxiety is to inform the occupants regarding the behavior of APMVs. However, it is unclear when the APMV behavior should be notified to reduce occupants' anxiety. Therefore, to investigate when anxiety occurs in APMVs, we experimented with scenes in approaching anxiety factors. It was proposed that the occupants in APMVs became anxious 0.72–1.96 s before they were closest to anxiety factors. In addition, because there are individual differences in how stimuli are received, the relationship between individual characteristics and the timing of anxiety were investigated. The results showed a weak positive correlation between the intensity of anxiety when approaching the anxiety factor and the time it took to reach the anxiety factor after an individual starts feeling anxious.

MEGADOCK-GUI: A GUI-based Complete Cross-docking Tool for Exploring Protein-Protein Interactions

Masahito Ohue, Yutaka Akiyama
Department of CS, School of Computing, Tokyo Institute of Technology, Midori-ku, Yokohama City, Kanagawa, Japan

Abstract: Information on protein-protein interactions (PPIs) not only advances our understanding of molecular biology but also provides important clues for target selection in drug discovery and the design of PPI inhibitors. One of the techniques used for computational prediction of PPIs is protein-protein docking calculations, and a variety of software has been developed. However, a friendly interface for users who are not sufficiently familiar with the command line interface has not been developed so far. In this study, we have developed a graphical user interface, MEGADOCK-GUI, which enables users to easily predict PPIs and protein complex structures. In addition to the original 3-D molecular viewer and input file preparation functions, MEGADOCK-GUI is software that can automatically perform complete cross-docking of M vs. N proteins. With MEGADOCK-GUI, various applications related to the prediction of PPIs, such as ensemble docking that handles multiple conformations of proteins and screening of binding partner proteins that bind to specific proteins, can now be easily performed.

Prediction of G4(MP2)-level Quantum Chemical Properties from DFT-level Molecular Structures by Delta Machine Learning

Wan Mingda, Nobuaki Yasuo, Masakazu Sekijima
School of Computing, Tokyo Institute of Technology, Nagatsuta-Cho Midori-Ku, Yokohama, Japan

Abstract: Predicting the chemical properties of molecules is one of the main purposes of quantum chemical calculations. Density functional theory (DFT) is a faster calculation method than more complex methods such as G4(MP2); however, the calculation accuracy is insufficient. In this study, the difference in atomization energies calculated by DFT and G4(MP2) was corrected using deep-learning-based prediction. A deep tensor neural network on a Coulomb matrix was trained, where the input structures were optimized by DFT, and the difference in atomization energies between DFT and G4(MP2) was learned as the label. Experiments using the QM9 dataset showed that 10,000 training data enabled an average absolute error of less than 1 kcal.mol^-1 for the test data, which is comparable to the G4(MP2)-level calculations. The results suggest that the combination of DFT and machine-learning-based correction may be able to replace higher-precision methods.

Estimating Concrete Interior using a 3D-CNN

Miho Chiyonobu, Sayaka Iida, Yu Ishikawa, Masaharu Nishikino, Ryuta Kawakami, Kazuki Joe
Nara Women's University, Nara, Japan

Abstract: In recent years, the aging of concrete infrastructure constructed during the period of rapid economic growth has become a serious problem in Japan, and urgent maintenance work is required. The existing tunnel inspection method, called the percussion method, is performed manually by inspectors, but it is necessary to ensure safety and efficiency in order to perform frequent inspections. One of the alternatives to the percussion method is the laser remote sensing method, which enables efficient inspection by high-speed remote control. In this paper, we propose a method to automatically estimate the interior of concrete by combining the laser sensing method and a neural network. As a first step, we generate a small amount of training data by using scientific simulation, and construct a 3D-CNN to estimate the 3D shape.

Generating an Estimated Font Set for Early-Modern Japanese Printed Character Recognition

Yuki Takemoto, Yu Ishikawa, Masami Takata, Kazuki Joe
Nara Women's University, Nara, Japan

Abstract: A lack of training data is a problem for early-modern Japanese printed character recognition, which is difficult to convert printed images into text using modern OCR technology. In order to solve this problem, we aim to automatically generate character images with the characteristics of fonts used in early-modern Japanese printed books. In this paper, we use StarGAN v2 to generate fonts of early-modern books, aiming at more accurate original character images. From the generated images, we examine the types of modern fonts suitable for the source images and the performance of StarGAN v2 in generating fonts of early-modern books.

The 20th International Conference on Security & Management (SAM 2021)
http://sam.udmercy.edu/sam21/
https://www.american-cse.org/csce2021/

Secure End-to-End Communications with Lightweight Cryptographic Algorithm

Augustine Ukpebor, James Addy, Kamal Ali, Ali Abu-El Humos
Jackson State University, Jackson, Mississippi, USA

Abstract: Lightweight cryptography has been an active area of research in the past decades because conventional cryptographic technologies are difficult to implement in constrained environments. This research presents the use of hardware platform, ESP32 microcontroller to implement lightweight cryptographic algorithm. We use the smallest block cipher of the KATAN family, KATAN32 with 80-bit key and 32-bit blocks. The algorithm is computationally less intensive as it uses key of 80 unsigned 64-bit integers to encrypt and decrypt data. The data array is passed into the encryption function with a key where a buffer is filled with an encrypted array during encryption. The decryption function fills a buffer with an array of the original data in the form of 32 unsigned 64-bit integers during decryption. This implementation demonstrates that data can be securely end to end transmitted with good throughput and low energy consumption.

A Study of Secure Algorithms for Vertical Federated Learning: Take Secure Logistic Regression as an Example

Huan-Chih Wang, Ja-Ling Wu
Department of CSIE, National Taiwan University, Taipei, Taiwan

Abstract: After entering the era of big data, more and more companies build services with machine learning techniques. However, it is costly for companies to collect data and extract helpful handcraft features on their own. Although it is a way to combine with other companies' data for boosting the model's performance, this approach may be prohibited by laws. In other words, finding the balance between sharing data with others and keeping data from privacy leakage is a crucial topic worthy of close attention. This paper focuses on distributed data and conducts secure modeltraining tasks on a vertical federated learning scheme. Here, secure implies that the whole process is executed in the encrypted domain; therefore, the privacy concern is released.

Sandy Toolbox: A Framework for Dynamic Malware Analysis and Model Development

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

Abstract: With the emergence of zero-day threats and the constant change to the threat landscape, their must be a necessary change in the way in which we identify Malware and its variants. Legacy techniques makes use of static analysis to analyze Malware composition, but the research trend is steering towards dynamic analysis as it is less susceptible to Malware obfuscation. This work introduces a novel sandbox coined Sandy which implements many of the same functionality as existing sandboxes, but includes a framework for feature extraction and model development within the toolbox. We demonstrate the inner workings of Sandy, as well as illustrate some preliminary dynamic analysis results retrieved from execution of the Malware variant Emotet.

On the Use of Semantic Web Technologies for Forest Fire Emergencies in Resilient Societies

Panagiota Masa, Georgios Meditskos, nastasios Karakostas, Stefanos Vrochidis, Ioannis Kompatsiaris
CERTH / Information Technologies Institute, Centre for Research & Tecnhology Hellas, Greece

Abstract: This short paper presents work in progress for the semantic representation and integration of information in the SAFERS framework. More specifically, SAFERS is an H2020 project that aims to support intelligent decision-making of wildfire scenarios, intervening at critical points in the emergency management cycle. Semantic Web technologies, such as OWL 2, are extensively employed in the framework to semantically represent the available information and best practices in crisis management, fostering advanced linking and interpretation mechanisms for decision-making solutions to support human operators and authorities.

Blockchain-based Decentralized Attack Pattern Repository

R. K. N. Sai Krishna, Chandrasekhar Tekur, Ravi Mukkamala
Teradata India Pvt Ltd., Hyderabad, India;
Old Domnion University, Norfolk, Virginia, USA

Abstract: A publicly available repository of attack patterns with appropriate classifications (e.g., Common Attack Pattern Enumeration and Classification, known as CAPEC) assists software developers, software and system testers, and system administrators in building and maintaining highly secure software and systems. However, when such a repository relies on a centralized server, it may suffer from issues such as single point of failure, lack of trust and possible data tempering due to a compromised server. To solve these problems, we propose a blockchain-based decentralized attack pattern repository replacing the centralized server system. This approach relies on multiple-node repositories, where each node contributes to and uses the attack patterns. The security aspects of the blockchain architecture ensure that the attack patterns are not tampered with and that the attack patterns are authorized/approved by a subset of participant nodes. The decentralization aspect of the architecture avoids single point of failure and ensures high availability. The proposed approach also ensures scalability in terms of the number of users and the number of queries.

Cryptanalysis of a Lightweight Authentication and Key Agreement Scheme for Internet of Drones

Chun-Ta Li, Cheng-Chi Lee, Chin-Ling Chen, Chi-Yao Weng
Department of Information Management, Tainan University of Technology, Tainan City, Taiwan;
Department of Library and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan;
Department of Photonics and Communication Engineering, Asia University, Taichung City, Taiwan;
Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung City, Taiwan;
Department of Computer Science, National Pingtung University, Pingtung City, Taiwan

Abstract: Internet of drones (IoDs) has been widely used in various fields and it aims to offer a live-stream, real-time information from flying drones belonging to a particular flying zone. Since the communication between the drones are wireless in nature, numerous security and privacy concerns are arise in IoD environments. In 2020, Zhang et al. proposed a lightweight authentication and key agreement scheme for IoDs. Authors utilize the lightweight online authentication design for assuring the authentication efficiency when deploying on resource-constrained drones and they claimed that their authentication scheme achieves better security and resists various known attacks. However, in this paper, we demonstrated that the anonymity of their authentication scheme cannot be ensured from passive attacks during authentication phase. In addition, it fails to withstand drone impersonation and session key exposure attacks and is not easily reparable.

Automatic DDoS Attack Detection on SDNs

Sabrina Corsetti, Avi Purkayastha, Adarsh Hasandka, Michael Samon
University of Michigan, Ann Arbor, Michigan, USA;
National Renewable Energy Laboratory, Golden, Colorado, USA;
University of Colorado, Boulder, Colorado, USA

Abstract: Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks pose a serious threat to computing networks – especially to critical systems within the U.S. electrical grid. As attack mechanisms have increased in complexity and variety, more sophisticated detection mechanisms have become necessary to ensure network security. This paper explores the use of artificial intelligence to automate the process of detection and mitigation of DoS and DDoS attacks within the framework of Software-Defined Networking (SDN), to a high degree. Machine learning algorithms are trained to recognize DoS and DDoS attacks and are deployed in real-time to mitigate malicious network traffic. The results show a well-tuned gradient-boosted decision tree detecting DoS and DDoS attacks, as well as initial successful mitigation of attacks within an SDN framework.

Toward High Security and Energy Saving of Internet of Things Devices in Edge Computing Using Heterogeneous Architecture

Mais Nijim, Sandhya Rani Kadavakuti
Department of Electrical Engineering & Computer Science, Texas A&M University-Kingsville, Kingsville, Texas, USA

Abstract: The rapid development of the Internet of Things (IoT) and machine learning techniques have evoked a new computing model called Edge Computing. To Boost the Functionality, the energy Conserving, and the Safety of IoT Apparatus in Edge Computing, we propose using a heterogeneous multiprocessing architecture in edge computing that's two processing processors run concurrently. We recommend using the two processors: Graphics Processing Unit (GPU) and Digital Signal Processing (DSP). We will study the effect of using these hybrids on Edge Computing regarding the workload dimensions, energy efficiency, and electricity consumption. We also proposed a business rules engine considered an essential block in the entire system responsible for classifying and simulating different calculations to extract client request patterns and behavior. The business rules engine is also in charge of managing the GPU interactions and the DSP and allows the machine to use them simultaneously. The business rules engine helps supply us with comments that help divide the GPU and DSP workload, leading to higher performance and lower energy consumption. The advantage of this approach is that all cores will be used simultaneously. We've designed and implemented a Secure Socket Layer (SSL) proxy that opportunistically offloads cryptographic operations to GPUs and DSPs. We conducted a comparison experiment using Nvidia GeForce RTX 2060 GPU plus a Texas Instrument C6000 high-performance DSP. The experimental results imply that the key advantage of embracing both the GPU and DSP from the advantage computing technology is in two folds: 1) GPU and DSP guarantee high performance for edge computing. 2) The GPU's heterogeneous multiprocessor and the DSP include two to three times lower power consumption in edge computing. 3)The energy consumption of this GPU is 78 percent more compared to using the GPU and the DSP together. 4) The integration of the GPU and the DSP conserve energy by nearly 140%, then just utilizing the GPU chip. 5) Support for IoT software in terms of Safety from attacks on cryptographic operations.

A Secure Communication between Patient and Psychiatrist in Telepsycholog: Hyper Chaotic Advanced Encryption Standard (HCAES)

Anish Adhikari, Noha Hazzazi, Abeer Alsadoon, Ahmed Dawoud, P.W.C. Prasad, Tarik Rashid
School of Computing and Mathematics, Charles Sturt University (CSU), Wagga Wagga, Australia;
Department of Electrical Engineering and Computer Science, Howard University, Washington DC, USA;
School of Computer Data and Mathematical Sciences, University of Western Sydney (UWS), Sydney, Australia;
Computer Science and Engineering, University of Kurdistan Hewler, Erbil, Iraq;
Kent Institute Australia, Sydney, Australia;
Asia Pacific International College (APIC), Sydney, Australia

Abstract: Telepsychiatry is a remote psychological service provided by using several telecommunication technologies, including telephones, mobile devices, interactive videoconferencing, email, and other potential modes of communication between the patient and the psychiatry. The deployment of these technologies implies the transmission of the data over untrusted networks like the Internet. This work aims to propose a novel solution to enhance security during the telepsychiatry conference without compromising the video quality and the processing time. The proposed system focuses on preserving network security while maintaining the video quality during telepsychiatry counseling and behavioral observation. The presented solution deploys a Hyper Chaotic Advanced Encryption Standard (HCAES) that utilizes the generation of keys from a hyperchaos system and encryption with the Advanced Encryption Standard (AES). The results show that the encryption time was decreased by ~19%, and decryption time was decreased by ~22% compared with the best solution that uses the chaotic system. Also, the encrypted file size is reduced from ~38% to ~24% with maintaining the original file size. The proposed system aims to provide a new solution to secure telepsychiatry conferences by using an HCAES algorithm. It offers an efficient and robust system to protect the confidentiality and integrity of the real-time communication during telepsychiatry to communicate between the psychiatry and the patient on the remote side.

A Survey on the Future of Secure and Trustworthy Passwordless Authentication Environments

Riley Kiefer, Ashok Patel
Department of Computer Science, Florida Polytechnic University, Lakeland, Florida, USA

Abstract: Password-based authentication has long been the staple of modern society, but it now carries too many problems. Increasing administrations costs in attempt to counter advanced password attacks have forced the end-user to memorize longer and more sophisticated passwords. This survey paper reviews the state of passwordless environments as an alternative to memory-based authentication. Specifically, this paper will review the different types of passwordless schemes, explore end-user opinion on passwordless environments, and analyze the latest published work on passwordless authentication. Each paper that proposes a passwordless scheme is summarized and presented with advantages and disadvantages by passwordless topic. Passwordless schemes are divided into the following topics: E-Health, Single-Sign-On, biometrics, mobile devices, Internet of Things, and FIDO extensions.

Deep Learning Algorithm for Threat Detection in Hackers Forum (Deep Web)

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

Abstract: In our current society, the inter-connectivity of devices provides easy access for netizens to utilize cyberspace technology for illegal activities. The deep web platform is a consummative ecosystem shielded by boundaries of trust, information sharing, trade-off, and review systems. Domain knowledge is shared among experts in hackers forums which contain indicators of compromise that can be explored for cyberthreat intelligence. Developing tools that can be deployed for threat detection is integral in securing digital communication in cyberspace.

In this paper, we addressed the use of TOR relay nodes for anonymizing communications in deep web forums. We propose a novel approach for detecting cyberthreats using a deep learning algorithm Long Short-Term Memory (LSTM). The developed model outperformed the experimental results of other researchers in this problem domain with an accuracy of 94% and precision of 90%. Our model can be easily deployed by organizations in securing digital communications and detection of vulnerability exposure before cyberattack.

A GDPR-compliant Risk Management Approach based on Threat Modelling and ISO 27005

Denys A. Flores, Ricardo Perugachi
Escuela Politecnica Nacional, Departamento de Informatica y Ciencias de la Computacion (DICC), Ecuador

Abstract: Computer systems process, store and transfer sensitive information which makes them a valuable asset. Despite the existence of standards such as ISO 27005 for managing information risk, cyber threats are increasing, exposing such systems to security breaches, and at the same time, compromising users' privacy. However, threat modelling has also emerged as an alternative to identify and analyze them, reducing the attack landscape by discarding low-risk attack vectors, and mitigating high-risk ones. In this work, we introduce a novel threat-modelling based approach for risk management, using ISO 27005 as a baseline for integrating ISO 27001/27002 security controls with privacy regulations outlined in the European General Data Protection Regulation (GDPR). In our proposal, risk estimation and mitigation is enhanced by combining STRIDE and attack trees as a threat modelling strategy. Our approach is applied to an IoT case study, where different attacks are analyzed to determine their risk levels and potential countermeasures.

Exploring the Vulnerability of Side Channel Attack Countermeasures via Electromagnetic Analysis Experiments

Lewis Etheridge, Yanzhen Qu
School of Computer Science, Colorado Technical Univesrity, Colorado Springs, Colorado, USA

Abstract: This quantitative experimental design study has investigated electromagnetic side channel attack via the experiments of analyzing the security of FPGA's claimed secure by specification for military and commercial system implementation. Commercially, specifications assert FPGA's secure against electromagnetic side channel attack; however, research supporting these claims is limited at best. This research applies constructive experimental design to assess FPGA embedded hardware security and countermeasures to test specifications. Experiments employ simple electromagnetic analysis attacks (SEMA) to establish the baseline for differential and correlation electromagnetic analysis to defeat cryptographic security, configuration settings, and countermeasures thereby testing experimental hypotheses to prove or disprove security capabilities. The experiments have provided the micro-architectural specifications, operational data, and statistical information to develop the side channel attack (SCA) safe test tool kit providing users a means to test embedded cryptographic hardware security on most devices. The experimental outcomes have provided not only the basis for future work for validation of cryptographic hardware security, countermeasure development, and testing, but also the information needed to improve cybersecurity and information assurance through modified hardware design, micro-architectural engineering, sensor and scanning software development.

Basic Security Baseline for Mobile Applications

Chia-Chen Chang, Ching-Chin Lin, Jiunn-Shiow Lin, Chun-I Fan
Chinese Cryptology and Information Security Association, Kaohsiung, Taiwan;
Department of IT Industries Division, Industrial Development Bureau, MOEA, Taipei, Taiwan;
Department of Computer Science and Engineering and Information Security Research Center, National Sun Yat-sen University, Kaohsiung, Taiwan

Abstract: In recent years, people are highly reliant on mobile devices and numerous mobile applications to achieve their daily activities, such as news, navigating outside the home, checking stock prices, etc. However, due to inbuilt application flaws in mobile devices, privacy leakage of personal information, including location, lifestyle habits, and even financial accounts and credit card numbers, without being aware of them. To address such an issue, we have developed a Basic Security Autonomous Certification System with an essential security testing baseline that provides testing standards and procedures to enable developers to self-check whether their applications are vulnerable to user information leaks.

A Qualitative Study of Security in Application Programming Interfaces (APIs)

Farhan A. Qazi, Sondria Miller
Cybersecurity and Information Assurance, Capitol Technology University, Laurel, Maryland, USA

Abstract: Application Programming Interfaces (APIs) are the most vulnerable points of attack and many users are not aware of their insecurity. In this paper, our purpose was to explore how and why the insecurity of APIs is being overlooked, understand the vulnerabilities of APIs, and find ways to mitigate the vulnerabilities. This unique qualitative case study of the complex phenomenon of the insecurity of APIs yielded valuable results. A survey was used as the research method to collect responses from developers, engineers, managers, and users of APIs. The results showed that most of the respondents are unaware of the insecurity, lack resources and training to educate user about APIs, and depend on the overall security of the network instead of API security. Existing security tools for APIs along with ML(machine learning) techniques, if used, will mitigate the insecurity.

Fingerprint Liveness Detection using Minutiae-Independent Dense Sampling of Local Patches

Riley Kiefer, Jacob Stevens, Ashok Patel
Department of Computer Science, Florida Polytechnic University, Lakeland, Florida, USA

Abstract: Fingerprint recognition and matching is a common form of user authentication. While a fingerprint is unique to each individual, the authentication is vulnerable when an attacker can forge a copy of the fingerprint (spoof). To combat these spoofed fingerprints, spoof detection and liveness detection algorithms are currently being researched as countermeasures to this security vulnerability. This paper introduces a fingerprint anti-spoofing mechanism using machine learning. Specifically, the proposed algorithm builds upon existing methods of local patch-based fingerprint liveness detection by applying a dense and overlapping local patch sampling method and applying algorithm reduction principles to reduce algorithm complexity for embedded platforms. These minutiae-independent local patches are uniformly rotated according to the patch's intensity gradient. Once the preprocessing methods are completed, the patches are passed into a shallow Convolutional Neural Network (CNN) and an aggregate of the patch scores determines the fingerprint classification. This proposed method is tested on the public fingerprint liveness detection dataset: LivDet-2009 Biometrika, CrossMatch, and Identix as well as LivDet-2011 Biometrika, Digital, and Sagem. Intra-sensor models are generated, tested, and compared to other top algorithms created by researchers using the Average Classification Error (ACE) metric. A graphic user interface tailored to the proposed method is also presented to visualize the classifier results at the local level.

An Investigation of the Cyber-Attack Susceptibility of State and Local Government Websites

Patrick Hill, Yu-Ju Lin
Department of Computer Science, Charleston Southern University, South Carolina, USA

Abstract: Critical infrastructure is those assets, systems, and networks that underpin American society. Many cybercrime activities have gained attention by the U.S. Government due to the impact of state-sponsored hacktivism, hardware\software shipped with spyware, compelled certificates issued for Internet interception, and exploited zero-day exploits [1]. This study is to provide an analysis on United States government owned websites to gauge their overall susceptibility to common cyber threats. By using common IT administration tools, we examined over 2900 state and local government websites for the evidence of best practices, vulnerability to legacy exploits, and readiness of adopting emerging technologies. The results of this study show that there are aspects of our local government, citizen facing websites that are not as resilient to cyber-attack as they should be. While most of the vulnerabilities can be mitigated by patching these systems, nothing has been done due to lack of security awareness. There is a severe need for greater web security assistance from the US federal government in funding and education.

To the Edge: A New Model for Moving Sensitive Data from Cloud to Edge using Machine Learning Approach

Mais Nijim, Hisham Albataineh
Texas A&M University-Kingsville, Kingsville, Texas, USA

Abstract: The Internet of Things (IoT) is grabbing the attention of scientists all over the globe. The IoT applications, notably real-time IoT applications, require both high performance and trustworthiness. In IoT cases, they mainly consult with this dependability of information and apparatus, the convenience of fault tolerance, and the high end. There are lots of challenges that are taken into consideration. The most damaging facet of cloud computing is the community dependencies and privacy issues. Many security threats arise from the cloud computing environment; these dangers include but are not limited to data loss, malicious attacks, service disruption, loss of control, and denial of service attack. In any case, cloud sensors and IoT devices are far apart, slow reaction times also frequently occur. This delayed response time could lead to a larger problem for delay-real time sensitive applications. However, Edge computing significantly outperforms the traditional cloud computing systems in several aspects where security is considered the most essential factor. To protect confidential data stored in the cloud, we use machine learning algorithms to classify the highly confidential data and then move it to the edge platform.

Security Improvement on HMAC-based One-time Password (HOTP)

Shushan Zhao, Xiaobing Hou
Dept of Computer Electronics & Graphics Technology, Central Connecticut State University, New Britain, Connecticut, USA

Abstract: Authentication plays a paramount role in online services. Today many online services are still using password as single authentication method, but this is not considered secure any more. There have been many attempts to introduce multifactor authentication mechanism. Nowadays many online service providers use One-time Password (OTP) as a supplementary authentication method to verify identity of the user. There are two major methods to generate OTPs: Time-based One-time Password (TOTP) and HMAC-based One-time Password (HOTP). We notice that there are several limitations or weaknesses with both. In this paper, we first show some security vulnerabilities of TOTP and HOTP, then we present security improvement methods for HOTP. We analyze and discuss the security features of proposed solution. The solution is generic to all platforms and operating systems, and our analysis demonstrates that it addresses security vulnerabilities of HOTP.

Visualizing Static Analysis Warnings by Dynamic Trace

Young Lee
Department of Computing and Cyber Security, Texas A&M University-San Antonio, Texas, USA

Abstract: This study aims to analyze the requirements of the data flow visualization to track the data leaks. We compared various data flow diagrams and graphs that were used for internal representations or pictorial outputs of the static analysis tools. We proposed synchronized visualization techniques with case studies. The result of this study will aid in understanding how to visualize the results of the static analysis to help the programmers to understand the data flow in context- and object-sensitive.

Comparison of Semantic Web and Distributed Ledger Approaches to Decentralized Identity

August Worden, Ian McCloskey, Gregory Jones, Srujan Kotikela
SPAIRL Research Group, Texas A & M University-Commerce at Rellis, Texas, USA

Abstract: The centralized nature of Identity and Access control Management (IAM) on the web causes serious privacy concerns and robs the users control over their own identity. This has led to the rise of Decentralized Identity technology. The core objective of Decentralized Identity is to return control, privacy and security of user data back to the owner. While the objectives of various decentralized identity solutions are the same, their approaches are strikingly different. This paper analyzes two different distributed ledger approaches (Sovrin & uPort) against a Semantic Web based approach (SOLID). It evaluates and analyzes key features of each technology in order to obtain insights into a possible transition from a centralized data storage architecture to a decentralized format.

Applying Equivalent Mutant to Refactoring the Security Vulnerabilities

Chunyang Tang, Young Lee
Department of Computing and Cyber Security, Texas A&M University-San Antonio, Texas, USA

Abstract: A novel scheme for dealing with security warnings output from static analysis tools is proposed, with the help of the equivalent mutant technique. A pre-collected set of good equivalent mutants is defined by carefully examining a well-known vulnerable web app, and a scheme using this set of equivalent mutants on user-specified apps, with pre-defined critical test cases, is developed and initial steps are demonstrated.

Introducing Zero Trust by Design: Principles and Practice Beyond the Zero Trust Hype

Dwight Horne
AT&T Center for Virtualization, Southern Methodist University, Dallas, Texas, USA

Abstract: The zero trust (ZT) threat model is one in which no implicit trust exists. Companies across industries, governments, and more are at various stages of evaluation, planning, and implementation of re-architected networks that deviate from legacy perimeter-centric security models in favor of ZT approaches. A lack of open standards, or even consensus on ZT details and priorities in some cases, has potentially slowed adoption. But recent standardization attempts have taken strides toward filling that gap. ZT research and practice to date have focused predominantly on the network architecture and operations. However, to fully realize the potential benefits of the ZT model, we must expand our conception of ZT. In this paper, we introduce Zero Trust by Design (ZTBD), a set of guiding principles that extend beyond network architecture to also foster ZT software engineering and ZT protocol design, thereby maximizing synergies and potential benefits. Beyond the ZTBD principles, a set of ZT practices connect theory and guidelines with practical applications, while the introduction of ZT patterns provides reusable solutions to common challenges. Moreover, just as ZT research and practice are evolving, so too will ZTBD. We invite the community to further contribute to these efforts and support continued alignment of this growing body of knowledge with the rapid evolution and expansion of zero trust.

Experimental Validation of Structured Attack Patterns

Walter Alan Cantrell, Edwin Michael Bearss, Katia P. Maxwell, Spencer York, Mikel D. Petty, Laurel Strelzoff, C. Daniel Colvett, Marshall Rosenhoover
Lipscomb University, Nashville, Tennessee, USA;
University of Alabama in Huntsville, Huntsville, Alabama, USA;
Athens State University, Athens, Alabama, USA

Abstract: The Common Attack Pattern Enumeration and Classification (CAPEC) consists of patterns of malicious activity that have been observed in attempts to compromise information systems. This publicly available resource is designed for the purpose of strengthening the defenses employed by systems owners. The optimum benchmark for the validity of conceptual models, such as the CAPEC attack patterns, is comparison between the models and data observed from instances of the class of systems being modeled. Data from cyberattacks is sensitive so direct validation of the techniques that make up the CAPEC database can be complicated. The purpose of this research is validation of the techniques described in the CAPEC attack patterns by experimentation using virtual computers installed on live simulated networks.

A Virtual Reality Home Security System

Nathan Robinson, Leonidas Deligiannidis
School of Computing and Data Science, Wentworth Institute of Technology, Boston, Massachusetts, USA

Abstract: This paper presents a study and implementation of a Virtual Reality application developed with Unity for the Oculus Quest 2 that provides an interface to remotely view and control multiple video streams from Raspberry Pi and IPenabled cameras. Within a virtual environment, an operator can freely move around and monitor several camera streams on large virtual screens and use the provided control panels to control the playback of each stream. From the display area, the control panel is extended to provide an interface to pan and tilt the cameras or set the brightness of attached light sources. The application worked well when streaming MJPEG from Raspberry Pi cameras, achieving 15 FPS.

Zero-Trust Model of Cybersecurity: Characteristics and Future Challenges

Fadi Al-Ayed
Independent Researcher, Culpeper, Virginia, USA

Abstract: A scrutiny of the zero-trust model is conducted to examine its features and implementations. An examination of the current applications identifies the extent of security by establishing the phases of framework implementation. The zero-trust model is discovered to be a sensitive and anticipatory model for cybersecurity. However, since the capability of the attackers, what-if scenarios are drawn to investigate the potential areas of vulnerability. As a consequence, an enhanced examination of the incoming and outgoing (input-output) flows in business environments must be made as organizations which are considered open systems. While these flows are inherent in business operations, zero-trust policies require extending to these flows.

Towards an Ecological Model for Cybersecurity

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

Abstract: As rapid response is often needed in cybersecurity, finding ways to predict trends can be beneficial. A good model can reduce the time that needs to be spent on analysis to understand and respond to a situation. This paper proposes a cybersecurity model based on ecology. Models in ecology make predictions about the dynamics of an ecosystem, which can determine when intervention is needed. Inaccuracies in ecological models can assist with building hypotheses about ecological relations that aren't yet known or understood. If ecological modeling can be successfully applied to cybersecurity environments, it could be beneficial for contributing these attributes of ecological modeling to cybersecurity and the prediction of threats.

Threat Intelligence Using Machine Learning Packet Dissection

Viktor Sowinski-Mydlarz, Jun Li, Karim Ouazzane, Vassil Vassilev
London Metropolitan University Cyber Security Research Centre, London, United Kingdom, UK;
Cranfield University Centre for Computational Engineering Sciences, Cranfield, United Kingdom, UK

Abstract: In this research we compare different methods to examine network packets using supervised learning to predict possible intrusions. Although there have been many attempts to use Machine Learning for automated packet analysis, our application simplifies the process by taking any packet data source for analysis in a container ready for deploying on a private or public cloud without the need to pre-process the packet data. The packet is dissected extracting numerical data, describing the packet numbers, the time and length of the packets. Categorical variables are the source and destination IP addresses, protocol used and packet info/flag. The use of filters allows ability to recognize any type of packet (e.g., SYN, ACK, FIN, RST). Four machine learning models, i.e., Neural Networks, Support Vector Machines, Logistic Regression and Linear Regression, are applied respectively to calculate the probability of suspicious packets. Subsequently, the outcomes are compared. During the testing against trojan malware, the models can detect the suspicious packets sent to a bogus website and attempts at downloading malware by means of packet payload analysis.

Survey on Blockchain-Based Frameworks for Secure, Anonymous yet Verifiable e-Voting

Michel Verwer, Ioanna Dionysiou, Harald Gjermundrod
Department of Computer Science, School of Sciences and Engineering, University of Nicosia, Nicosia, Cyprus

Abstract: Even though the proliferation of technology has led into the digitization of a plethora of governmental services, still an area that one may claim is at its infancy is the applied electronic voting (e-Voting). There are numerous models and approaches to address aspects of e-Voting, however only a handful of countries have adopted the digital form of voting and integrated it as a service in their digital government environments (DGE). Election integrity issues such as vote non-tampering, voter anonymity, vote verifiability hinder the practical deployment of e-Voting. This paper presents a survey on existing e-Voting systems and frameworks along with a critical assessment on the security aspects of the various solutions.

Security and Privacy Challenges in the Cloud-IoT Paradigm

Angelina Gaber, Dina Mamdouh, Youmna Atef, Maggie Mashaly
German University in Cairo, Egypt

Abstract: The new developments such as Internet of Things (IoT), digital economy and digital transformation are currently at their peak. Due to massive data produced, the continuous growth of data storage drives the rapid development of the entire storage industry. A vast amount of data is generated by IoT-connected devices. Thus, the greatest influence of IoT for organizations is on how they treat big data. In the future, businesses will be forced to update their systems, technology, and safety in order to align with this need. For companies, handling the flow and storage of this information is a repetitive job. With its numerous models and deployment frameworks, cloud computing allows businesses to handle and evaluate this knowledge, improving overall IoT system efficiency and functionality. Currently; governments, enterprises, and individuals are actively migrating their data to the cloud. The increase of data volume results in increased capacity and complexity of data storage systems in the cloud, which may cause possible risks as unauthorized access, data leakage, data information disclosure and privacy disclosure. In this paper, we review data security and privacy issues and discuss the challenges and methods used by cloud computing to facilitate a secure transition of IoT applications to the cloud.

Towards Provably Secure Proxy Signature Scheme Based on Practical Lattice-Based Cryptography

Yong Xiao, Ao Liu, Mi Zhou, Jiangnan Li
Electric Power Research Institute, CSG, P. R. China;
Shenzhen Power Supply Bureau Co., Ltd, P. R. China

Abstract: The proxy signatures are important cryptosystems that are widely adopted in different applications, for example, grid computing, mobile agent applications, and mobile communications. Most of the proxy signature schemes so far are based on the hardness of integer factoring, discrete logarithm, and/or elliptic curve. However, Peter Shor showed that when a large enough quantum computer is built, the proxy signature schemes based on these problems are not sufficient since the quantum computers can solve the problem of prime factorization and discrete logarithm in polynomial time. In this paper, A proxy signature scheme based on a basic lattice-based digital signature scheme is proposed. The main advantage of the scheme in this paper over other proxy schemes should be its provable security feature of its underlying scheme and its post-quantum feature. A formal security proof is also given, which shows that our scheme can reach Existential Unforgeability under an Adaptive Chosen Message Attack with Proxy Key Exposure assuming that the underlying signature is Existential Unforgeability under an Adaptive Chosen Message Attack which was proved previously. All in all, the method to construct our scheme and the method to formally prove the securityof our scheme are very good exploration and valuable attempt in the area of lattice-based cryptosystems, since they can be similarly used in the construction of other primitives of a cryptosystem.

Secure Location-Based Contact Tracing for COVID-19

Gokay Saldamli, Sidharth Jayaprakash, Jojo Joseph, Sushant Mathur, Mitra Gunakara Nayak, Levent Ertaul
Computer Software Engineering, San Jose State University, San Jose, California, USA;
Department of Computer Science, California State University - East Bay, Hayward, California, USA

Abstract: COVID-19 became one of the most fatal outbreak in the modern history. At the time of writing, about 150 million people were reported to be infected and more than 3 million people dead worldwide. It has caused a worldwide economic slowdown, with many businesses having to close and lay off their employees. It has also exerted mental stress among individuals as they worry about their family and loved ones. In order to bring things back to normal, it is absolutely necessary to control the spread of this disease. This can be done effectively using contact tracing, testing and isolating of individuals who could be exposed to the Coronavirus. Many countries have followed this approach and have been successful in limiting the spread of infections to an extent. This also helps health authorities in planning the roll out of vaccines and healthcare resources to areas with higher infection rates. In this study, we attempt to create a contact tracing solution in a secure way. In our approach we try to inform people who have been in contact with an infected individual and have them quarantined and tested to avoid possible community spread. This is achieved through contact tracing using location data in a decentralised manner. The system informs the suitable group of people if an infection is reported but with complete anonymity. The system also provides a way to track the number of infections across various locations. The outcome will be better identification and control of community spread of this infectious disease.

Quantum Technology and Crypto Algorithms

Jeong Yang, Jesus Guerrero, Muammer Sahin, Thomas Hernandez, Gerardo Bilbatua, Sashikumari Singh
Department of Computing and Cyber Security, Texas A&M University - San Antonio, Texas, USA

Abstract: The current hype surrounding quantum technology is not totally in accordance with reality. Quantum computing has its merits but mostly as a hybrid system of classical and quantum. The fundamental qubit has a nature that disallows pure quantum technology to be used for daily tasks. In the current state, pure quantum technology is impractical, but there are many reasons to be excited about the hybrid system. Quantum cryptography may be a technology that ensures ultimate security in the future. This paper concentrates on the principles of quantum technology and cryptography, and the way quantum crypto algorithms contribute to security. The paper also outlines the important world application implementation of quantum technology and crypto algorithms, and therefore the future direction in which quantum cryptography accelerates.

Performance Evaluation of Virtual Private Network Protocols and Applications

Fadi F. Qazzazee, George Dimitoglou
Department of Computer Science and IT, Hood College, Frederick, Maryland, USA

Abstract: A Virtual Private Network (VPN) provides secure data transmission between network-accessible endpoints. Although the use of VPNs offers several security advantages, establishing a secure tunnel and encrypting the transmitted traffic come at a performance cost. Different VPN protocols offer varying levels of encryption and performance. In this study, we evaluated the availability performance of three modern and widely adopted VPN protocols: OpenVPN, IKEv2 and WireGuard, over numerous criteria, such as throughput, latency, jitter, and their impact on client system resources. The results provide insight into the behavior of these protocols in terms of both performance and the associated client system processing load overhead. It is clear that newer and emerging protocols, namely IKEv2, and WireGuard, offer significantly faster throughput speeds and lower latencies along with an overall more conservative use of system resources, as compared to older protocols, such as openVPN.

Impact of Deep Fakes and Social Engineering

Brennan Huber, Luay A. Wahsheh
Department of Computer Science and Engineering, University of Tennessee at Chattanooga, Chattanooga, Tennessee, USA

Abstract: Social engineering has proven that humans are prone to influence and can be coaxed into allowing attackers to compromise a computer system. This has led to the implementation of deep fakes' technology to impersonate co-workers allowing the social engineer to extract information that would normally not be provided to a non-employee. In this paper, we discuss a case study that involves a deep fake attack on a UK energy company where a manager believed he was talking to the CEO and initiated a transfer of over $200,000 into a fraudulent account. We also discuss methods for preventing deep fakes' attacks. While deep fakes are in their infancy, the impact of deep fakes in social engineering is present. We believe that social engineering attacks that threaten personal and organizational information can be prevented by creating a cyber security awareness culture. Increasing awareness by drawing attention to the deep fake attack case that is discussed in this paper is a step towards achieving this goal.

Examining Security and Forencisc Across DBMS

Abhishek Phadke, Stanislav Ustymenko
Texas A&M university, Corpus Christi, Texas, USA

Abstract: Database Management Systems (DBMS) have evolved from simple file systems to modern DBMS servers and Big data management spanning across complex hierarchies, and today, even the cloud. Data must be protected from unauthorized access, falsification, destruction, and malicious intent. The necessity of DBMS security has increased in importance with the integration of local and global networks, the increasing volume of data produced, and role of information in enterprise practices. The paper has a two-pronged approach. It studies most common attacks on DBMS and their countermeasures. The next section is an attempt to describe Database forensic model process to lay the groundwork for future study towards creating a vendor-neutral benchmarking model and a comprehensive forensic investigation model.

FPGA-Based Edge Detector Ring Oscillator Physical Unclonable Function for IoT Devices Security

Amin A. Zayed, Hanady H. Issa, Khaled A. Shehata, Hani Fikry Ragai
Department of Electronics and Communication Engineering, Faculty of Engineering, Ain Shams University, Cairo, Egypt;
Department of Electronics and Communications, Faculty of Engineering, Arab Academy for Science, Technology and Maritime Transport, Cairo, Egypt

Abstract: Physically unclonable function (PUF) is considered as one of the most popular hardware security tools. It extracts identical hardware implementation differences generated from uncontrolled fabrication process errors to generate a secret key. Ring Oscillator (RO) PUF is the most common design adopted to be implemented in Field Programmable Gate Array (FPGA). Internet of Things (IoT) is the essential application that gets benefits from the PUF. This paper introduces a new RO-PUF, Edge

Detector (ED) RO-PUF, with low dynamic power and fewer hardware overheads than the conventional RO-PUF. The proposed Edge Detector (ED) RO-PUF is designed and implemented on Cyclone V GX Starter Kit from Intel. Its performance is evaluated by measuring the performance metrics. The design shows a low power consumption while maintaining high-performance metrics. The dynamic power consumption of the proposed design is 40 nW. The average reliability of the proposed design is 99.78 %, and the uniqueness is 49.98 %. The proposed PUF decreases the power consumption while maintaining high-performance metrics.

Electronic Design Automation (EDA) and Integrated Circuits: Systems Security and Protection

Nicolas Sklavos, Charikleia Salaga
University of Patras, Greece

Abstract; Electronic system's complexity, has been increased significantly and we face a new era, where system attacks are becoming more hardware – centric. Hardware security is mandatory for every physical device, to ensure protection against any adversary, that wants to weakness the entire system. Thus, the step that needs to be taken in orientation of secure integrated circuits (ICs) design and composition, can be achieved with the aid of Electronic Design Automation (EDA). It is aware that EDA tools focus on performance, power, area which leaves a basic gap in security direction. In this paper, firstly we demonstrate the classic EDA functionality. We present the security absence facing threats to overcome that are being defined. Afterwards, we lead to the basic purpose, which is the introduction of EDA community to security-driven enhancements, that could overcome security threats and compromise the integrity and confidentiality of sensitive information and propose countermeasures that should be embraced. Finally, we discuss about challenges towards a holistic security protection of EDA tools.

Cyber Intelligence: A Minor Program for the Next Generation Cybersecurity Workforce

Avdesh Mishra, Ayush Goyal, David Hicks, Hisham Al-bataineh, Mais Nijim
Department of Electrical Engineering and Computer Science, Texas A&M University - Kingsville, Texas, USA;
Department of Physics and Geosciences, Texas A&M University - Kingsville, Texas, USA

Abstract: This paper describes a minor program in Cyber Intelligence for undergraduate science students to learn how to apply artificial intelligence and machine learning for cybersecurity. Such a program would allow more undergraduate students to gain cybersecurity skills to fulfill the growing need for a diverse and skilled workforce in the nation's cybersecurity and defense industries. This paper also reviews the artificial intelligence and machine learning curriculum that are useful for cybersecurity, which would need to be imparted in this minor program. The goal and objective of the artificial intelligence and machine learning program for cybersecurity would be providing students with skills in detection and prevention of cyber-threats over a network. Artificial intelligence and machine learning algorithms can be used to mine the data being sent over packets in a network to perform intrusion detection. In this way, packet data can be analyzed to detect and thwart network intrusion threats. This paper summarizes the course content and curriculum for introducing undergraduate science students to artificial intelligence and machine learning algorithms and methods for cyber threat detection and prevention in cybersecurity applications.

Automotive Vehicle Security Visualization

Guillermo A. Francia, III, Darius K. Simmons
Center for Cybersecurity, University of West Florida, Pensacola, Florida, USA;
University of West Florida, Pensacola, Florida, USA

Abstract: The current hype surrounding quantum technology is not totally in accordance with reality. Quantum computing has its merits but mostly as a hybrid system of classical and quantum. The fundamental qubit has a nature that disallows pure quantum technology to be used for daily tasks. In the current state, pure quantum technology is impractical, but there are many reasons to be excited about the hybrid system. Quantum cryptography may be a technology that ensures ultimate security in the future. This paper concentrates on the principles of quantum technology and cryptography, and the way quantum crypto algorithms contribute to security. The paper also outlines the important world application implementation of quantum technology and crypto algorithms, and therefore the future direction in which quantum cryptography accelerates.

Using a Personal Device for Secure and Convenient Authentication

Michael Hutchinson, H. Karen Lu
Digital Identity & Security, Thales Austin, Texas, USA

Abstract: In modern day society, people use more and more network connected devices, from smartphones, laptops, tablets to Wi-Fi routers, thermostats, refrigerators, monitors, and so on. For security and privacy, each of these devices requires user authentication, mostly via username and password. It is difficult to use and manage complex and different passwords for each device. Using the same simple password for many devices is often the practice. This could lead to security and privacy breaches that people might not even be aware of. This paper proposes a more secure and yet convenient solution to address the problem of accessing multiple services. It leverages the fact that some devices are very personal, which people carry with them all the time, such as a smartphone, smart watch, or fitness tracker. We propose to use these personal devices to facilitate access to other services with security and convenience.

DoS/DDoS Detection Using Random Forest with Wavelet Decomposition

Ying-Feng Hsu, Morito Matsuoka
Cybermedia Center, Osaka University, Osaka, Japan

Abstract: As DoS/DDoS attacks become increasingly sophisticated and diverse, the need for attack detection models with both high detection rates and the ability to respond to new types of attacks has increased. In this study, we proposed a DoS/DDoS attack detection model which combines machine learning, wavelet decomposition, and a widely used feature extraction algorithm in an attempt to improve the detection accuracy of current machine learning models. The performance of our proposed model is compared to the models without wavelet decomposition, and the usefulness of feature selection on our proposed model is assessed. Based on our evaluation, the proposed model achieved a higher recall of 99.99% while still maintaining accuracy of 98.57%.

Exploratory Analysis of a Measurement Scale of an Information Security Management System

Rusbel Dominguez-Dominguez, Omar Flores-Laguna, Jose A. Sanchez-Valdez
Facultad de Ciencias Empresariales y Juridicas, Universidad de Montemorelos, Mexico

Abstract: The purpose of this research is to propose an instrument to measure the degree of an information security system based on ISO/IEC 27001. This research shows the analysis of multiple factors that inhibit the implementation of an Information Security Management System (ISMS). A scale of 24 items was obtained, divided into four factors: organizational policies and regulations, privacy, integrity and authenticity. This version of the instrument meets the criteria established for its validity (KMO, Bartlett's test of sphericity). An extraction was performed by the minimum residuals method, an oblique rotation was performed by the promax method, when performing the rotation 17 of the 24 items were grouped in the corresponding factor. The final reliability of the scale was calculated by the Omega coefficient, in all the dimensions the coefficients were greater than .70, therefore the reliability of the instrument is good.

The 19th International Conference on Software Engineering Research & Practice (SERP 2021)
https://www.american-cse.org/csce2021/conferences-SERP

Innovative Projects in Software Engineering for Virtual Teaching and Learning

Soma Datta, Alex Miller, Daryl Wayne Sonnier, Fernando Gonzalez
University of Houston-Clear Lake, Houston, Texas, USA

Abstract: This experience report discusses the use of several learning models during the COVID-19 pandemic. The report focuses on two courses that were taught remotely. It emphasizes two projects that were innovative and essential during this time. The students' viewpoints and the instructor's viewpoints and lessons-learned are discussed. The learning was an amalgam of coding using functional paradigm, team building, understanding software engineering concepts, and developing innovative project, and was a success with remote teaching.

An Accessible Search Software for Chinese Input Methods for Blind Persons

Hsiao Ping Lee, Yu En Lin, Tzu-Fang Sheu
Department of Medical Informatics, Chung Shan Medical University, Taiwan, R.O.C.;
Department of Computer Science and Communication Engineering, Providence University, Taiwan, R.O.C.

Abstract: The main language people use in Taiwan is Mandarin Chinese. There are more than 10,000 different characters in Chinese. People must use a Chinese input method to input Chinese characters on a computer. In the Chinese input method, each Chinese character corresponds to a specific sequence of keys, which allows the use of a standard computer keyboard to input Chinese characters. The key-sequence-search software of the Chinese input method can query the key sequence corresponding to the Chinese characters. It helps users to learn and use the Chinese input method. This project has developed a key-sequence-search software for the Chinese input method, that is compatible with screen-reading software for the blind. The software is accessible to blind persons. The assistant software for the blind developed by this project has achieved the goal of assisting the blind in learning and using Chinese input methods.

Correlation Between Reliability and Risk Priority Numbers (RPN) for Bridges

Chandrasekhar Putcha, Ahin Banerjee, Debanshee Datta, Subhrajit Datta
California State University, Fullerton, California, USA;
Indian Institute of Technology (BHU), Varanasi, India;
National Institute of Technology, Silchar, India

Abstract: This paper is an extension of the previous paper by Putcha et al. (2020). The main aim of this study is to correlate the RPN values obtained in earlier study to the well-used reliability values for bridge structures using principles of regression analysis. One of the important and a systematic method that is used for identification of failures of any system is FMEA (Failure Mode Effects Analysis). This is done through identification of the most important causes of all the failure modes of the system under consideration. This identification can be easily achieved through the description of all failure modes of the system under consideration. FMEA is based on the concepts of probabilistic analysis (Ebeling, 1997; Henley et al., 1981). A corresponding equivalent deterministic method has been developed by Dytczak and Ginda (2017). This paper describes in detail the application of this method to bridge failures. Based on actual data. This will enable the engineers to rank various failures and take necessary steps to mitigate the risk. These results will be useful to both academicians and practicing engineers.

Structural Bug Localization using Graph-based Deep Learning on UML Class Diagrams

Gelareh Meidanipour Lahijany, Manuel Ohrndorf, Johannes Zenkert, Madjid Fathi, Udo Kelter
Knowledge-Based Systems and Knowledge Management, University of Siegen, Germany;
Software Engineering Group, University of Siegen, Germany

Abstract: Bug reports are essential software artifacts that describe software bugs using natural language. Bug localization approaches mostly focus on unstructured textual information from the source codes and bug reports. This paper proposes a structural graph-based approach for bug localization using the information in a UML class diagram representing a software. We approach bug localization as a link prediction problem that predicts the bug locations as edges between a new bug report and existing classes. Our approach called IdentiBug utilizes deep learning techniques to build our bug localization model. The embeddings of a bug report and each UML class are passed through a link prediction layer to specify the existence of an edge targeting a bug location. The result is a ranked list of classes from which we generate a ranked list of diagram excerpts for supporting the developers during bug documentation and localization.

Technical Lag in OSS Integration

Thomas Weber, Rene Peinl
Institute of Information Systems, Hof University of Applied Sciences, Hof, Germany

Abstract: Software integration, especially in Open Source Software (OSS), suffers from technical debt as software evolves over time and ends in incompatibility issues like changed or removed APIs. The problem increases, as more components need to be integrated. This article analyzes common problems in OSS integration with respect to technical debt esp. externally induced technical debt and collects empirical evidence from the case study "AMiProSI" which combines different functions of various OSS systems to create a unified intranet product. As part of the analysis, this paper discusses component-based development and its related dependencies, explains technical debt (TD). By bringing down the individual problems in sub-categories of technical debt we show what kind of TD has arisen in our case study AMiProSI and gives advise on how this can possibly be prevented in other projects.

Annotation Ontology for Semantic Web Development

Hadeel Alobaidy, Ghadeer Khalil, Amani Al Heela
Department of Computer Science, University of Bahrain, Sakheer, Bahrain;
Department of Information Systems, University of Bahrain, Sakheer, Bahrain;
Department of Information Technology, Arabian Gulf University, Manama, Bahrain

Abstract: The main purpose of this paper is to examine the concept of semantic web and the role that ontology and semantic annotation plays in the development of semantic web services, which contain a lot of multimedia data. It focuses on semantic web infrastructure, illustrating how ontology and annotation work to provide learning capabilities for building multimedia content semantically. This paper applies approaches, notations and techniques offered by software engineering to improve productivity and quality of the software. It proposes a conceptual model to develop semantic web services for the infrastructure of the web information retrieval system of digital libraries. It uses ontology and annotation to build a knowledge-based system to define and link the meaning of web content to retrieve information for users' queries. The results are more relevant through keywords and ontology rule expansion that are more accurate to satisfy the requested information. The level of accuracy could be enhanced since the query is semantically analyzed with the conceptual architecture of the proposed system.

Using a Knowledge Repository and Interactive Notebooks for Deploying Data Analytics Applications across an Organisation

Fethi Rabhi, Madhushi Bandara, Lawrence Yao, Siu Lung Ng
University of New South Wales, Australia;
University of Technology Sydney, Australia;
Roar Software Pty Ltd, Australia

Abstract: There is a tendency for many organisations to address their big data analytics needs using a mix of cloud-based solutions combined with programs developed with popular open source packages such as R and Python. However, establishing and maintaining a mixed-technology infrastructure presents many challenges. This paper proposes an architecture with a knowledge repository that allows several simple analytics applications to be maintained by small teams of analysts, enabling organisations to reap the advantages of data-driven analytics insights, at a lower cost. This architecture also allows programs developed by analysts to be reused in both experimental and operational activities by leveraging computational facilities offered by interactive notebooks such as Jupyter Notebooks.

Community Driven Design in Software Engineering

Mirco Schindler, Sebastian Lawrenz
Institute for Software and Systems Engineering, Clausthal University of Technology, Clausthal, Germany

Abstract: The implementation of the circular economy as described in the literature is a slow process. Circular economy goals are usually secondary goals of companies. Furthermore, implementing these circular goals is expensive and challenging, especially for small and medium-sized businesses. This leads to a gap in the implementation of the circular economy and obviously hinders progress toward climate protection goals. However, small and medium-sized businesses could achieve their goals together by cooperating and sharing the risk of the implementation of their circular economy goals. Therefore, this paper presents a novel approach towards a circular economy community-driven ecosystem that supports the cooperation of small and mediumsized businesses for implementing the circular economy. Cooperation, in general, requires companies to share their best practices and processes, but this is contrary to their primary goals (their business models). Accordingly, technical solutions are required to allow information to be exchanged without negatively affecting the primary goals, which leads to a conflict between transparency and privacy, to name just one showstopper. We introduce the term and concept of Community-Driven Design in software engineering, we describe it with an applied example, and we use the mechanisms of the Dynamic Adaptive System Infrastructure, so-called DAiSI, to introduce our concept. Based on the DAiSI, and on the community-driven life cycle, we present five steps for achieving a common circular economy target.

A Data-Driven Approach to Rethinking Open Source Software Organizations as a Team of Teams

Elaine M. Raybourn, Reed Milewicz, David M. Rogers, Benjamin H. Sims, Greg Watson, Elsa Gonsiorowski, Jim Willenbring
Sandia National Laboratories, Albuquerque, New Mexico;
Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA;
Los Alamos National Laboratory, Los Alamos, New Mexico, USA;
Lawrence Livermore National Laboratory, Livermore, California, USA

Abstract: This paper introduces a data-driven approach toward rethinking the characterization of a virtual organization as an aggregate, team of teams. We developed a repository mining toolkit, Reposcanner, to assist software teams in identifying opportunities for software process improvement and connected collaborations with other teams, thus characterizing efforts as team of teams. We use the example of continuous integration (CI) testing as an indicator of software development team productivity that signals potential best practices in which team of teams collaboration exists. Our preliminary findings indicate that the degree to which continuous integration testing is present can also signal the likelihood of related best practices, and a team of teams organizational structure.

Automating Software Productivity Planning: Lightweight Tools for Upgrading Team Practices

Elaine M. Raybourn, Greg Watson, Elsa Gonsiorowski, Reed Milewicz, David M. Rogers, Benjamin H. Sims, Jim Willenbring
Sandia National Laboratories, Albuquerque, New Mexico;
Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA;
Los Alamos National Laboratory, Los Alamos, New Mexico, USA;
Lawrence Livermore National Laboratory, Livermore, California, USA

Abstract: A lightweight iterative workflow, called Productivity and Sustainability Improvement Planning (PSIP) has been developed to assist software teams with employing good software engineering practices. Three PSIP tools automate practice summarization, self-assessment, and progress tracking to enhance team engagement around software process improvement: Reposcanner, RateYourProject (RYP) and Progress Tracking Cards (PTCs) respectively. Reposcanner is a data analysis tool. RYP is an interactive self-assessment tool that provides an easy-to-use entry point to the PSIP process. Once process improvements are identified, PTCs are created by teams to measure outcomes and progression toward their goals. We present an exemplifying case study with a software team where these tools have been put to use, along with a description of the tools, and future plans for further improvements.

Evaluation of Change Impact Analysis Tools in a Real-World Case Study

Arezoo Tajmoradi, Christoph Knieke, Sebastian Lawrenz, Andreas Rausch
Institute for Software and Systems Engineering, Clausthal University of Technology, Clausthal-Zellerfeld, Germany

Abstract: There has been a wide range of research work on Change Impact Analysis techniques and their applications. Good tool support is necessary to successfully apply Change Impact Analysis for software migration. Today, there are several tools available to support these techniques. But many of these tools are just prototypes and do not scale for real-world projects. In this paper, we are interested in which tools are appropriate to be used in real-world projects in a software company. Therefore, we want to compare existing tools and apply them to a case study for software migration in this company. We are also interested in which aspects of CIA are supported and which level of maturity the tools reach today. Finally, we also want to sketch a methodology for how to integrate these tools in an existing development process.

A LibTooling-Based Refactoring Tool for CUDA Memory Management

Kian Nejadfard, Alan Hylton, Janche Sang, Haodong Wang, Jingru Zhang
Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, Ohio, USA;
Space Communications and Navigation, NASA Glenn Research Center, Cleveland, Ohio, USA

Abstract: Refactoring tools, whether automatic or semiautomatic, are an integral part of the software development life cycle. Software libraries and frameworks evolve over time. Programs that use them should also evolve over time, or risk falling behind on compatibility with modern advancements. In the case of NVIDIA CUDA's platform for general-purpose GPU programming, using the more modern unified memory architecture allows for simplifying program source code, reducing bugs associated with manual memory management between host and device memory, and benefiting from memory transfer optimization via automatic handling of memory. This paper proposes the development of a refactoring tool to handle this transition automatically, therefore freeing developers from the burden and risks involved in manually refactoring large code bases.

Cumulative Effort Estimation in an Evolving Legacy System

Lamees Alhazzaa, Anneliese Amschler Andrews, Aiman Gannous
Computer Science Department, Al-Imam Muhammad Bin Saud Islamic University, Riyadh, Saudi Arabia;
Department of Computer Science, University of Denver, Denver, Colorado, USA;
Department of Health Informatics, University of Benghazi, Benghazi, Libya

Abstract: Effort estimation is key in any software development organization. Over the past two decades there has been considerable activity in the area of effort prediction with approaches being typified as being algorithmic in nature such as COCOMO. However, many change request databases contains effort data for finished change requests but they do not contain data that would allow using these methods. We propose a statistical regression approach that uses cumulative effort data to predict future effort in an evolving legacy system. This method is then expanded to perform re-estimation of the prediction model to consider changepoints in order to get more accurate predictions of effort.

Unified Framework for Smart Rigs for Multi-Role UAV Avionics Ground Evaluation

Iqsaam Mukhtar, Sudha Rani S.V., Y. Dilip, B. Ramadoss Balakrishnan
Aeronautical Development Establishment, Defence Research Development Organisation, Bangalore, India;
National Institute of Technology, Tiruchirappalli, Tamil Nadu, India

Abstract: Unmanned Aerial Vehicles (UAVs) execute multiple military roles, involving complex communication among their Integrated Avionics systems with GBs of data movement with precise hard timelines between them. The Integrated Evaluation of these systems is a challenging task. This paper brings out a Unified Framework for Smart Rigs for UAV Avionics Ground Evaluation namely Unified Intelligent Multi-purpose Integrated Test System overtaking the test design philosophy legacy of one customized test equipment for one system. This is a comprehensive common test platform capable of multi-purpose testing of both single, integrated systems and multi-level testing stages right from Depot to Squadron, implements Artificial Intelligence, Data Mining, Test-Automation, user-friendly GUI features, thus Integrated Systems Performance Evaluation with improved speed, accuracy and minimum human involvement.

An Agile Approach to Automate Real Estate CRM (Pre-Sales) using Scrum and Essence

Parthasarathi Ray, Pinakpani Pal
Applied Statistics Unit, Indian Statistical Institute, Kolkata, India;
Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata, India

Abstract: Customer relationship management (CRM) is both a leading enterprise application software and a pivotal business strategy in many industries. This paper addresses an industry use case, automation of the pre-sales functionalities of a CRM system in Real Estate business, by using Scrum, the most widely used agile framework. There are challenges with Scrum implementations, primarily with adoption of Scrum and putting it into practice. That's where a formal Object Management Group (OMG) standard called Essence, a software engineering kernel resulting from the SEMAT (Software Engineering Method and Theory) community's efforts, comes in. We describe how Essence can be leveraged as the foundation for defining software engineering with practices like Scrum being adopted on top of it in the given context.

AI Enabled Design Agent for Software Engineering

R. Muralidharan, Vincent J. Kovarik, Jr.
TATA Advanced Systems, Mumbai, India;
VINMA Systems LLC, West Melbourne, Florida, USA

Abstract: The evolution and adoption of processes and tools supporting model-based system engineering (MBSE), software design and development has enhanced the productivity of the systems and software engineering domain. However, although the capabilities of tools have improved significantly, many of the advances have been in support of the engineering process with little support for easing the engineering cognitive workload. This paper presents a tool-based approach that leverages artificial intelligence (AI) technology as an embedded AI Design Agent ADAge coupled to model-based development tools. The primary focus of this research is the application of AI as an intelligent assistant supporting the system and software engineer by off-loading some of the analysis tasks. Examples of initial research progress are presented.

Code Visualization for Replacing Code Complexity

Janghwan Kim, Je Hyeong Cho, R. Young Chul Kim
Software Engineering Laboratory, Department of Software and Communication Engineering, Hongik University, Jochiwon-eup, Sejong-si, Republic of Korea

Abstract: An 'UNTACT' stream blows around the world due to the coronavirus. In this situation, it will be hugely expanding the use of smart electronics and software with low power and good performance. As these electronics are controlled through software, it may be most important factor to consume less power and use less memory in the same environment. In this paper, we visualize and measure the complexity of internal code based on defined metrics through our code visualization. Then, we propose a research method on performance improvement through reducing power consumption of the complexed code unit against the basic core code and building block structures such as nested if, multi loop, module, and calling statements. Therefore, we identify good code patterns with low power and good performance. Through the proposed method, we expect to improve the consumption power and performance efficiency through the simplification of complex code patterns.