CSCI 2020 BOOK of ABSTRACTS

The 2020 World Congress in Computer Science, Computer Engineering,
and Applied Computing
CSCE 2020

https://americancse.org/events/csce2020
July 27-30, 2020
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.

CSCE 2020 KEYNOTE ADDRESS
The Psychology of Artificial Intelligence: A Conversation with SANDI

Dr. James A. Crowder
Systems Fellow at Colorado Engineering, Inc., USA
(Formerly at Raytheon Technologies. Recipient of technical achievement awards from
Lockheed Martin, General Dynamics, Northrup Grumman, & Raytheon Technologies.)

The 4th International Conference on Applied Cognitive Computing (ACC 2020)
https://americancse.org/events/csce2020/conferences/acc20

An Adaptive Tribal Topology for Particle Swarm Optimization

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

Abstract: The success of global optimization rests on the ability for a compatible metaheuristic to approximate global search. Particle Swarm Optimization (PSO) is one of such heuristics, with the ideal PSO application being one which promotes swarm diversity while incorporating the global progress of the swarm in its performance. In this paper, the authors introduce an adaptive tribal topology within PSO to improve global coverage. Diversity of the swarm population was dynamically managed through an evaluation of a swarm fitness parameter, which takes into account the relative performance a swarm member and its assigned tribe has on finding better objective evaluations. The fitness function simultaneously promotes the breeding of exemplars and the elimination of swarm members stuck in local minima. The model was evaluated on a series of benchmark problems with unique and diverse search spaces, and the results demonstrate that performance relied on the distribution and scale of the local minima present.

Improving the Efficiency of Genetic-based Incremental Local Outlier Factor Algorithm for Network Intrusion Detection

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

Abstract: In the era of big data, outlier detection has become an important task for many applications, such as the network intrusion detection system. Data streams are a unique type of big data, which recently has gained a lot of attention from researchers. Nevertheless, there are challenges in applying traditional outlier detection algorithms for data streams. One of the well-known algorithms of outlier detection is Local Outlier Factor (LOF). The issue with LOF is that it needs to store the whole dataset with its distances' results in memory. In addition, it needs to start from the beginning and recalculate all processes if any change happens in the dataset, such as inserting a new data point. The Genetic-based Incremental Local Outlier Factor (GILOF)

has addressed these issues and shown significant results. In this paper, we further improved the GILOF performance in data streams by proposing a new calculation method for LOF. The experiment's results showed that our new method of calculation has led to better accuracy performance for various real-world datasets.

A Grid Partition-based Local Outlier Factor for Data Stream Processing

Raed Alsini, Omar Alghushairy, Xiaogang Ma, Terrance Soule
Department of Computer Science, University of Idaho, Moscow, Idaho, USA;
Department of Information System, King Abdulaziz University, Jeddah, Saudi Arabia;
Department of Information System and Technology, University of Jeddah, Jeddah, Saudi Arabia

Abstract: Outlier detection is getting significant attention in the research field of big data. Detecting the outlier is important in various applications such as communication, finance, fraud detection, and network intrusion detection. Data streams posed new challenges to the existing algorithms of outlier detection. Local Outlier Factor (LOF) is one of the most appropriate techniques used in the density-based method to determine the outlier. However, it faces some difficulties regarding data streams. First, LOF processes the data all at once, which is not suitable for data streams. Another issue appears when a new data point arrives. All the data points need to be recalculated again significantly. Therefore, it affects the execution time. A new algorithm is proposed in this research paper called Grid Partition-based Local Outlier Factor (GP-LOF). GP-LOF uses a grid for the LOF with a sliding window to detect outliers. The outcome of experiments with the proposed algorithm demonstrates the effectiveness in both performance accuracy and execution time in several real-world datasets compared to the state-of-the-art DILOF algorithm.

Cognitive Discovery Pipeline Applied to Informal Knowledge

Nicola Severini, Pietro Leo, Paolo Bellavista
Alma Mater Studiorum, Bologna, Italy;
IBM Active Intelligence Center, IBM Italy, Bologna, Italy

Abstract: This paper introduces an Information Extraction (IE) architecture to process unstructured data that are typically generated within non-formal contexts such as the ones that are exchanged during multiple forms of technical meetings. Information Extraction and Knowledge Management are providing practical value in all kind of industries and boosting R&D activities specically. Our project sees a clear opportunity to leverage and expand this capability to process also "informal sources" of business/technical content such as the one that is exchanged along informal channels during, for instance physical or web meetings. Given these observations, we have built a software architecture designed on the concept of expandability, a prototype has also been developed capable of showing peculiarities and the critical points.

A Hybrid Cognitive System for Radar Monitoring and Control using the Rasmussen Cognition Model

James A. Crowder, John N. Carbone
Colorado Engineering Inc., Colorado Springs, Colorado, USA;
Department of Electrical and Computer Engineering, Southern Methodist University, Dallas, Texas, USA

Abstract: To be Provided.

Assessing Cognitive Load via Pupillometry

Pavel Weber, Franca Rupprecht, Stefan Wiesen, Bernd Hamann, Achim Ebert
Department of Computer Science, University of Kaiserslautern, Germany;
insight.out GmbH - digital diagnostics, Kaiserslautern, Germany;
Department of Computer Science, University of California, Davis, California, USA

Abstract: A fierce search is called for a reliable, non-intrusive, and realtime capable method for assessing a person's experienced cognitive load. Software systems capable of adapting their complexity to the mental demand of their users would be beneficial in a variety of domains. The only disclosed algorithm that seems to reliably detect cognitive load in pupillometry signals - the Index of Pupillary Activity (IPA) - has not yet been sufficiently validated. We take a first step in validating the IPA by applying it to a working memory experiment with finely granulated levels of difficulty, and comparing the results to traditional pupillometry metrics analyzed in cognitive resarch. Our findings confirm the significant positive correlation between task difficulty and IPA the authors stated.

A Hybrid Chaotic Activation Function for Artificial Neural Networks

Siobhan Reid, Ken Ferens
Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada

Abstract: In recent years, chaos theory has been applied to neuroscience to help understand the human brain. Researchers have suggested that the brain is a dynamical system which behaves chaotically. This paper proposes the introduction of chaos into an artificial neural network (ANN) by using chaotic neurons in the network's inner layer. A chaotic neuron uses two activation functions (AF): the sigmoid function and the logistic map function. The chaotic neuron generates a unique activation value to send to each neuron in the following layer. The model was tested by solving the XOR problem and showed significant improvements over a typical ANN.

A Cognitive Unsupervised Clustering for Detecting Cyber Attacks

Kaiser Nahiyan, Samilat Kaiser, Ken Ferens
Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada

Abstract: It has always been a challenge to make meaning out of unstructured data. In the field of network intrusion detection, the availability of structured, labelled datasets is limited. The need of unsupervised learning from unlabeled datasets is of vital significance, yet there is little breakthrough achieved in the research community. Most approaches adhere to techniques that are over-exhaustive in terms of resources, and do not yield satisfactory results; hence human analysts must reexamine all the events for intrusion attempts. This study makes an effort to find an approach of making sense out of unstructured, unlabeled data, in a way that helps the human analysts to disregard a major portion of the network dataset that contains regular traffic, and isolates the finite time-windows that have been subjected to potential attacks, utilizing the concepts of cognitive science, complexity analysis and statistical higher order feature learning. In this research, use statistical higher order features from network flows to classify the network traffic into flows containing normal traffic and flows subject to attacks, using unsupervised kmeans clustering and variance fractal dimension trajectory based complexity analysis. We validate our algorithm on the UNSW dataset, and compared our results with traditional unsupervised clustering. The proposed model was able to detect errors with the accuracy of 87.27%.

Variance Fractal Dimension Feature Selection for Detection of Cyber Security Attacks

Samilat Kaiser, Ken Ferens
Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada

Abstract: In an era where machine learning algorithms are widely used in order to improve the performance of network intrusion detection system, the complexity and big volume of data available in the network are also on the rise. The cyber networks frequently encounter high dimensional, unreliable and redundant data that are often too large to process. An efficient feature selection can therefore remove the redundant and irrelevant attributes and select relevant attributes that can significantly improve the overall system performance. This research provides a variance fractal dimension feature selection method to explore the significant features of cyber security attack dataset. A complexity analysis was done to find out the cognitive discriminative features of UNSW-NB15 dataset. A performance comparison is also provided using our proposed methodology for an artificial neural network and a comparative analysis were also done that shows the proposed method helps improve the detection performance in network system. The resultant discriminative features not only consume less resource but also speed up the training and testing process while maintaining good detection rates.

Defending Aviation Cyber-Physical Systems From DDOS Attack Using NARX Model

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

Abstract: Recently, the aviation industries showed interest in transferring their aircraft models to Cyber-Physical Systems (CPS) based models. However, and as it is well known, CPS introduces security threats to the physical components of the system. Distributed Denial of Service (DDOS) attack is one of the significant security threats to the availability of the communication network, due to making the network resources unavailable. In the presence of DDOS attack, servers are overwhelmed with so many requests which increase the network traffic that causes packet loss. Therefore, this paper proposes an approach to defending the communication network system in the Aviation Cyber-Physical System (ACPS) from a DDOS attack using a Nonlinear Autoregressive Exogenous (NARX) model. NARX network is used to predict packets that were dropped due to a DDOS attack. The real-time simulation results showed that the network system in ACPS was successfully defended from the DDOS attack because the ACPS maintained the expected normal performance during the DDOS attack.

Simulated Annealing Embedded Within Personal Velocity Update of Particle Swarm Optimization

Ainslee Heim, Ken Ferens
Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada

Abstract: This paper's focus is on using Simulated Annealing (SA) to improve basic Particle Swarm Optimization (PSO) by embedding SA in the part of PSO that is responsible for a particle's inertial force, creating a new hybrid evolutionary algorithm. The paper studies the effect of this addition using two benchmark functions in a continuous solution space as well as the effect of this addition on PSO for the Travelling Salesperson Problem (TSP), which has a discrete solution. Finally, the novel hybrid algorithm is compared against basic PSO to establish a ground level performance evaluation.

The 21st International Conference on Bioinformatics & Computational Biology (BIOCOMP 2020) https://americancse.org/events/csce2020/conferences/biocomp20

Rotation-Invariant Palm ROI Extraction for Contactless Recognition

Dinh-Trung Vu, Thi-Van Nguyen, Shi-Jinn Horng
Department of Computer Science and Information Engineering,
National Taiwan University of Science and Technology, Taipei, Taiwan, ROC;
Faculty of Information Technology, Vietnam Maritime University, Haiphong, Vietnam

Abstract: The extraction of palm region of interest (ROI) is one of the most important tasks in palm recognition because the accuracy of ROI detection affects directly to the recognition performance. In contactless environment, the freely hand posture of users causes the much more difficulties to determine the accurate ROI among a number of challenges. To the best of our knowledge, the disadvantage in most conventional methods is that the fingers are required to spread and the hand faces to the camera. This restricts the flexibility of system. Besides, there is a lack of modalities that work well in the case of large hand rotation. This study proposed new robust palm ROI extraction method to handle above challenges. The experimental results on a public benchmark dataset as well as our self-collected dataset validate the high accuracy in extracting palm ROI from the proposed method.

A Novel Method for the Inverse Qsar/Qspr to Monocyclic Chemical Compounds based on Artificial Neural Networks and Integer Programming

Ren Ito, Naveed Ahmed Azam, Chenxi Wang, Aleksandar Shurbevski, Hiroshi Nagamochi, Tatsuya Akutsu
Department of Applied Mathematics & Physics, Kyoto University, Kyoto, Japan;
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Japan

Abstract: Quantitative structure activity/property relationship (QSAR/QSPR) analysis is a major approach for computer-aided drug design and has also been studied in bioinformatics. Recently, a novel method has been proposed for inverse QSAR/QSPR using both artificial neural networks (ANN) and mixed integer linear programming (MILP), where inverse QSAR/QSPR is to infer chemical structures from given chemical activities/properties. However, the framework has been applied only to the case of acyclic chemical compounds so far. In this paper, we significantly extend the framework and present a new method for inverse QSAR/QSPR of monocyclic chemical compounds. The results of computational experiments using such chemical properties as heat of atomization, heat of combustion and octanol/water partition coecient suggest that the proposed method is much more useful than the previous method.

Mathematical Modeling and Computer Simulations of Cancer Chemotherapy

Frank Nani, Mingxian Jin
Department of Mathematics and Computer Science, Fayetteville State University, North Carolina, USA

Abstract: The fundamental clinical properties of cancer chemotherapy are investigated and demonstrated by utilizing a system of clinically plausible deterministic non-linear differential equations which depicts the patho-physiology of malignant cancers. In this mathematical model, the cytokinetic properties of normal cells, cancer cells, and the pharmacokinetics of the chemotherapy drug are described respectively by biophysically measurable growth parameters, stoichiometric rate constants and Michaelis-Menten type reaction profiles. Computer simulations have been conducted to elucidate various hypothetic scenarios when the model is configured with different parametric values, including the therapeutic efficacy of the use of stealth liposomes in high dose chemotherapy.

Predicting Targets for Genome Editing with Long Short Term Memory Networks

Neha Bhagwat, Natalia Khuri
Department of Computer Science, San Jose State University, San Jose, California, USA;
Department of Computer Science, Wake Forest University, Winston-Salem, North Carolina, USA

Abstract: Naturally occurring bacterial immune system can be engineered for use in the mammalian genome editing. To assist with the design of new editing systems, we developed and evaluated three datadriven predictors of DNA targets in the mouse and human genomes. Long Short Term Memory network models outperformed classifiers trained with Support Vector Machines and Random forest algorithms. The holdout accuracy of the deep learning classifier reached 81.6% for the mouse genome and 82.5% for the human genome. We also demonstrated that classification accuracy improves when sequences surrounding a mammalian target site are incorporated into the input vector of the neural network, reaching an accuracy of 83% for both organisms.

Optimizing the Removal of Fluorescence and Shot Noise in Raman Spectra of Breast Tissue by ANFIS and Moving Averages Filter

Reinier Cabrera Cabanas, Francisco Javier Luna Rosas, Julio Cesar Martinez Romo, Ivan Castillo-Zuniga
Computer Science Department, Instituto Tecnologico de Aguascalientes, Mexico;
Systems and computation Department, Instituto Tecnologico del Llano Aguascalientes, Mexico

Abstract: Cancer is one of the main causes of death worldwide. We know that a significant percentage of cancers; including breast cancer can be cured by surgery, or chemotherapy; therefore, its detection at an early stage of the disease is essential. Raman spectroscopy is an optical technique capable of measuring vibrational modes of biomolecules, allowing their identification from the correct location of the Raman bands, one of the main challenges is the elimination of spectral noise composed of (a) fluorescence background and (b) high frequency noise. In this article we demonstrate that using ANFIS (Neuro Fuzzy Adaptive Inference System) in combination with moving averages filter on the MATLAB multicore platform we can eliminate these disturbances and optimize response time in the preprocessing of large volumes of data in order to achieve a high precision in the classification and diagnosis of breast cancer.

The Unmasking of 'Mitochondrial Adam' and Structural Variants Larger than Point Mutations as Stronger Candidates for Traits, Disease Phenotype and Sex Determination

Abhishek Narain Singh
ABioTek www.tinyurl.com/abinarain, Kuopio, Finland

Abstract: Structural Variations, SVs, in a genome can be linked to a disease or characteristic phenotype. The variations come in many types and it is a challenge, not only determining the variations accurately, but also conducting the downstream statistical and analytical procedure. Structural variations, SVs, with size 1 base-pair to 1000s of base-pairs with their precise breakpoints and single-nucleotide polymorphisms, SNPs, were determined for members of a family. The genome was assembled using optimal metrics of ABySS and SOAPdenovo assembly tools using pairedend DNA sequence. An interesting discovery was the mitochondrial DNA could have paternal leakage of inheritance or that the mutations could be high from maternal inheritance. It is also discovered that the mitochondrial DNA is less prone to SVs re-arrangements than SNPs, which propose better standards for determining ancestry and divergence between races and species over a long-time frame. Sex determination of an individual is found to be strongly confirmed using calls of nucleotide bases of SVs to the Y chromosome, more strongly determined than SNPs. We note that in general there is a larger variance (and thus the standard deviation) in the sum of SVs nucleotide compared to sum of SNPs of an individual when compared to reference sequence, and thus SVs serve as a stronger means to characterize an individual for a given trait or phenotype or to determine sex. The SVs and SNPs in HLA loci would also serve as a medical transformational method for determining the success of an organ transplant for a patient, and predisposition to diseases apriori. The sample anonymous dataset shows how the de-novo mutation can lead to non-inherited disease risk apart from those which

are known to have a disease to mutation association. It is also observed that mtDNA is highly subjected to mutation and thus the factor for a lot of associated maternally inherited diseases. "mitochondrial Adam" can be a fair reality Vas certainly the biparental mode of mtDNA puts in question the theory of "mitochondrial Eve". SVs would serve as a stronger fingerprint of an individual contributing to his traits, sex determination, and drug responses than SNPs.

MSECT: Genome Wide Massive Sequence Exhaustive Comparison Tool

Abhishek Narain Singh
ABioTek www.tinyurl.com/abinarain, Kuopio, Finland

Abstract: To be provided (current version is too long)

Feature Set Optimization by Clustering, Univariate Association, Deep and Machine Learning Omics Wide Association Study (DMWAS) for Biomarkers Discovery as Tested on GTEx Pilot Dataset for Death due to Heart-Attack

Abhishek Narain Singh
ABioTek www.tinyurl.com/abinarain, Kuopio, Finland

Abstract: Univariate and multivariate methods for association of the genomic variations with the end-orendo-phenotype have been widely used for genome wide association studies. In addition to encoding the SNPs, we advocate usage of clustering as a novel method to encode the structural variations, SVs, in genomes, such as the deletions and insertions polymorphism (DIPs), Copy Number Variations (CNVs), translocation, inversion, etc., that can be used as an independent feature variable value for downstream computation by artificial intelligence methods to predict the endo-or-end phenotype. We introduce a clustering based encoding scheme for structural variations and omics based analysis. We conducted a complete all genomic variants association with the phenotype using deep learning and other machine learning techniques, though other methods such as genetic algorithm can also be applied. Applying this encoding of SVs and one-hot encoding of SNPs on GTEx V7 pilot DNA variation dataset, we were able to get high accuracy using various methods of DMWAS, and particularly found logistic regression to work the best for death due to heart-attack (MHHRTATT) phenotype. The genomic variants acting as feature sets were then arranged in descending order of power of impact on the disease or trait phenotype, which we call optimization and that also uses top univariate association into account. Variant Id P1_M_061510_3_402_P at chromosome 3 & position 192063195 was found to be most highly associated to MHHRTATT. We present here the top ten optimized genomic variant feature set for the MHHRTATT phenotypic cause of death.

Re-Ranking of Computational Protein-Peptide Docking Solutions with Amino Acid Profiles of Rigid-Body Docking Results

Masahito Ohue
Department of Computer Science, School of Computing, Tokyo Institute of Technology, Japan

Abstract: Protein-peptide interactions, in which one partner is a globular protein and the other is a flexible linear peptide, are important for understanding cellular processes and regulatory pathways, and are therefore targets for drug discovery. In this study, we combined rigid-body protein-protein docking software (MEGADOCK) and global flexible protein-peptide docking software (CABS-dock) to establish a reranking method with amino acid contact profiles using rigid-body sampling decoys. We demonstrate that the correct complex structure cannot be predicted (< 10 A peptide RMSD) using the current version of CABSdock alone. However, our newly proposed re-ranking method based on the amino acid contact profile using rigid-body search results (designated the decoy profile) demonstrated the possibility of improvement of predictions. Adoption of our proposed method along with continuous efforts for effective computational modeling of protein-peptide interactions can provide useful information to understand complex biological processes in molecular detail and modulate protein-protein interactions in disease treatment.

Graph Network Community Based Regulatory Medicine and 'More' Using GraphBreak

Abhishek Narain Singh
ABioTek www.tinyurl.com/abinarain, Kuopio, Finland

Abstract: Graph network science is becoming increasingly popular, notably in big-data perspective where understanding individual entities for individual functional roles is complex and time consuming. It is likely when a set of genes are regulated by a set of genetic variants, the genes set is recruited for a common or related functional purpose. Grouping and extracting communities from network of associations becomes critical to understand system complexity, thus prioritizing genes for disease and functional associations. Workload is reduced when studying entities one at a time. For this, we present GraphBreak, a suite of tools for community detection application, such as for gene co-expression, protein interaction, regulation network, etc. Although developed for use case of eQTLs regulatory genomic network community study results shown with our analysis with sample eQTL data-Graphbreak can be deployed for other studies if input data has been fed in requisite format, including but not limited to gene co-expression networks, protein-protein interaction network, signaling pathway and metabolic network. GraphBreak showed critical use case value in its downstream analysis for disease association of communities detected. If all independent steps of community detection and analysis are a step-by-step sub-part of the algorithm, GraphBreak can be considered a new algorithm for community based functional characterization. Combination of various algorithmic implementation modules into a single script for this purpose illustrates GraphBreak's novelty. Compared to other similar tools, with GraphBreak we can better detect communities with overrepresentation of its member genes for statistical association with diseases, therefore target genes which can be prioritized for drug-positioning or drug-repositioning as the case be.

Cancer Diagnosis of 84 Microarrays by Rank using 100-Fold Cross-Validation

Shuichi Shinmura
Seikei University, Japan

Abstract: After we completed a new theory of discriminant analysis in 2015, we used Revised IP-Optimal Linear Discriminant Function (Revised IP-OLDF, RIP) to discriminate against the Alon microarray. The minimum number of misclassifications (Minimum NM, MNM) was zero, indicating that the data is linearly separable data (LSD). Two classes, noncancerous ("normal") patients and cancer patients, are completely separable in the 2,000 gene space. We found that LSD is a crucial signal for cancer gene diagnosis. We classify the linearly separable space and subspaces as Matryoshka. We found that LSD has two unique structures. First, LSD has a Matryoshka structure. LSD includes many smaller Matryoshkas in it up to the minimum Matryoshka (Basic Gene Set, BGS). The second structure is as follows: Program3 coded by LINGO can decompose LSD into the exclusive Small Matryoshkas (SMs) and other gene subspace (MNM>0) by the Matryoshka Feature Selection Method (Method2). Program4 coded by LINGO can decompose LSD into the many BGSs and another gene set. The Program 4 algorithm follows the same procedure as the one used to find Yamanaka four genes from 24 genes. Just as iPS cells cannot generate when one of the four genes is deleted, deleting one gene from BGS does not make its MNM zero. Because the second structure releases us from the curse of high-dimensional data, we can analyze 64 SMs and 129 BGSs with JMP statistical software and propose a cancer gene diagnosis. Although we have successfully found all the signals, we need to rank the importance of all SMs and BGS for physicians to use for diagnosis. In this paper, we validate all 193 signals via the 100-fold cross-validation (Method1) and show the rank of all signals for the cancer diagnosis that is useful for medical research.

Structural Exploration of Rift Valley Fever Virus L Protein Domain in Implicit and Explicit Solvents by Molecular Dynamics

Gideon K. Gogovi
Department of Computational and Data Sciences, George Mason University, Fairfax, Virginia, USA

Abstract: The structural conduct and preference of a protein are highly sensitive to the environment accommodating it. In this study, the solvation effect of the structure and dynamics of a domain in the C-terminal of RVFV L protein was explored by

molecular dynamics using implicit and explicit water. The force field parameters of explicit waters were taken from the TIP3P, TIP4P, SPC/E, SPCE/Fw and OPC water models, and those of the peptide taken from the AMBER ff14SB force field. The behavior of the peptide was also investigated in an implicit solvent environment using the generalized Born (GB) model by setting the dielectric constant to match that of experimental measurements of water. Several properties including the interaction energy between the peptide and solvent molecules are calculated. Structural characterization and clustering of the atomic trajectories enables a better understanding of the structural and dynamical behavior of the RVFV domain along time.

A New Literature Based Discovery (LBD) Application Using the Pubmed Database

Matthew Schofield, Gabriela Hristescu, Aurelian Radu
Computer Science Department, Rowan University, Glassboro, New Jersey;
Department of Cell, Developmental & Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, USA

Abstract: PubMed, the biomedical database of the National Institutes of Health, contains over 30 million abstracts. Comparisons between pairs of abstracts can reveal novel and highly valuable hypotheses, especially if the connected abstracts are in very different areas of biomedical knowledge. It is improbable that such hypotheses can be formulated exclusively by professional investigators, who are usually highly specialized. We developed an application that generates hypotheses by connecting pairs of facts: the first fact is contained in abstracts in the area of expertise or interest of an investigator, while the second fact can be anywhere in PubMed, inside or outside the investigator's expertise. The application is based on a natural language processing machine learning model, provided by the Python package spaCy.

MinCNE: Identifying Conserved Non-Coding Elements using Min-Wise Hashing

Sairam Behera, Jitender S. Deogun, Etsuko N. Moriyama
Department of CSE, University of Nebraska-Lincoln, Lincoln, Nebraska, USA;
School of Biological Sciences, Center for Plant Science Innovation, University of Nebraska-Lincoln, Nebraska, USA

Abstract: Conserved non-coding elements (CNEs) are non-protein-coding genomic regions that exhibit an extreme degree of conservation. CNEs are mostly clustered around the genes and found to play important roles in regulating the transcription processes. Identification of CNEs is, therefore, important for studying their functional properties. Most of the existing CNE-finding methods are pairwise-alignment based and their scalability suffers when multiple sequences need to be processed simultaneously. We propose a new efficient alignment-free CNE-finding method, MinCNE. It finds CNEs among multiple sequences where k-mers derived from each sequence are clustered using minhash signatures and locality sensitive hashing. The performance evaluation demonstrated MinCNE to be computationally efficient without compromising the accuracy. The MinCNE source codes are available at https://github.com/srbehera/MinCNE.

An Agile Pipeline for RNA-Seq Data Analysis

Scott Wolf, Dan Li, William Yang, Yifan Zhang, Mary Qu Yang
MidSouth Bioinformatics Center & Joint Bioinformatics Program of University of Arkansas at Little Rock and University of Arkansas Medical Sciences, Arkansas, USA;
Princeton University, Princeton, New Jersey, USA;
Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract: Next-generation sequencing empowers researchers to examine the transcriptome efficiently, and at the same time, it introduces significant issues for the subsequent effective analysis of sequencing data. To counteract this growing problem, we developed a robust and modular pipeline for examining whole transcriptome sequencing data quickly. Utilizing a Docker for containerization and unanimous cross-platform support, we modularized the pipeline and created a practical, automated structure for pragmatic use. Furthermore, we have developed segregated implementations of the pipelines sections to allow end users to craft custom pipelines easily. To validate our tool, we applied our tool to various data sets and verified the pipeline results as highquality informatics.

Common Motifs in KEGG Cancer Pathways

Bini Elsa Paul, Olaa Kasem, Haitao Zhao, Zhong-Hui Duan
Department of Computer Science, University of Akron, Akron, Ohio, USA

Abstract: Genes and gene products interact in an integrated and coordinated way to support functions of a living cell. In this research, we analyze these interactions in 17 different types of cancers, focusing on the interactions presented in pathway maps in Kyoto Encyclopedia of Genes and Genomes repository. We extract the gene to gene interactions from the pathway maps and integrated them to form a large integrated graph. We then utilized different techniques and filtering criteria to extract and shed lights on the gene-gene interaction patterns. We conclude the graph motifs we identified in cancer pathways provide insights for cancer biologists to connect dots and generate strong hypotheses so further biological investigations into cancer initiation, progression and treatment can be conducted effectively.

Phenome to Genome - Application of GWAS to Asthmatic Lung Biomarker Gene Variants

Adam Cankaya, Ravi Shankar
College of Computer Engineering and Science, Florida Atlantic University, Florida, USA

Abstract: We utilize the GWAS Catalog to gather a set of SNPs with the highest odds ratio values related to asthmatic phenotypes. After collecting a large set of SNPs we parse them down to a subset of ten representative SNPs. Using the odds ratio and known prevalence of asthma in the population we then calculate each SNP's individual contribution towards the lifetime genetic risk of developing asthma. For a hypothetical European patient we estimate that the SNP from our set with the smallest odds ratio value sees an increase of lifetime risk by 10%, while the SNP with the largest odds ratio value causes the risk of developing asthma to more than double. For a more complete and accurate prediction of asthma risk additional work must be done to incorporate the contribution from additional SNPs and other risk factors not considered including age, gender, lifestyle habits, and environmental influences.

The 6th International Conference on Biomedical Engineering & Sciences (BIOENG 2020)
https://americancse.org/events/csce2020/conferences/bioeng20

Stage Classification of Neuropsychological Tests Based on Decision Fusion

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

Abstract: One way to improve classification performance and reliability is the combination of the decisions of multiple classifiers, which is usually known as late fusion. Late fusion has been applied in some biomedical applications, generally, using classic fusion methods, such as mean or majority voting. This work compares the performance of several state-of-the-art fusion methods on a novel biomedical application: automated stage classification of neuropsychological tests using electroencephalographic data. Those tests were made by epileptic patients to evaluate their memory and learning cognitive function with the following stages: stimulus display, retention interval, and subject response. The following late fusion methods were considered: Dempster-Shafer combination; alpha integration; copulas; independent component analysis mixture models; and behavior knowledge space. Late fusion was able to improve the performance of the single classifiers and the most stable results were achieved by alpha integration.

An Investigation of Texture Features Based on Polyp Size for Computer Aided Diagnosis of Colonic Polyps

Yeseul Choi, Alice Wei, David Wang, David Liang, Shu Zhang, Marc J. Pomeroy
Department of Radiology, Stony Brook University, Stony Brook, New York, USA;
Stuyvesant High School, New York, USA;
Staten Island Technical High School, Staten Island, New York, USA;
Syosset High School, Syosset, New York, USA;
Ward Melville High School, East Setauket, New York, USA;
Department of Biomedical Engineering, Stony Brook University, Stony Brook, New York, USA

Abstract: Computer-aided diagnosis of polyps has played an important role in advancing the screening capability of computed tomographic colonography. Texture features, including intensity, gradient, and curvature, are the essential components in differentiation of neoplastic and nonneoplastic polyps. In this paper, we present a study to investigate the effect of separating polyps based on size on the performance of machine learning. From each polyp volume, we extracted the traditional 14 Haralick texture features plus 16 additional features with a total 30 texture features. We further employed the Random Forest classifier for polyp differentiation. Experimental results demonstrated that gradient and curvature features were ideal for differentiation the malignancy risk for medium sized polyps, whereas intensity feature was better for smaller sized polyps.

Electrocardiogram Classification using Long Short-Term Memory Networks

Shijun Tang, Jenny Tang
Department of Science and Mathematics, Alvernia University, Reading, PA, USA;
Wilson HS, Reading, PA, USA

Abstract: In this paper, a novel ECG classification algorithm based on Long Short-Term Memory (LSTM) and feature extraction was proposed to classify normal and abnormal ECG signals. We have taken RR interval, half width of QRS peak as well as their distribution as the features of ECG signals. We trained the LSTM network with feature extraction to classify the imbalanced ECG signals. This research shows how to classify electrocardiogram (ECG) data from the PhysioNet using deep learning. The experimental results demonstrate that the proposed method achieves a good classification performance. The proposed method can be applied to assist cardiologists in more accurately and objectively diagnosing ECG signals.

Cancer Gene Diagnosis of 78 Microarrays Registered on GSE from 2007 to 2017

Shuichi Shinmura
Seikei University, Japan

Abstract: From 1999 to 2004, six medical projects published papers and publicly uploaded microarrays on the Internet. We discriminated against the six data using Revised IP-OLDF (RIP) and found the six data's minimum number of misclassifications (MNMs) are zero. This fact shows linearly separable classes in gene space and the six data are linearly separable data (LSD). We classify the linearly separable space and subspaces as Matryoshka. LSD is a crucial signal and has the following two vital structures: 1) LSD has the Matryoshka structure that includes smaller Matryoshkas (SM) up to the minimum dimensional SM (Basic Gene Set, BGS) in it. 2) LSD consists of many SMs or BGSs. This fact shows we are free from the curse of the high-dimensional microarray. Thus, we can analyze all SMs and BGSs, and propose the cancer gene diagnosis. To confirm two truths, we discriminate 78 microarrays of 13 carcinomas collected from 2007 to 2017. Because we confirm that the six old and 78 new microarrays are LSD, we can open a new frontier for gene diagnosis, including microarrays and RNA-seq. Our theory is elementary for physicians. If physicians confirm whether their data are LSD, they can analyze all SMs and BGSs with statistical methods. We expect they use our methods. They establish cancer diagnoses and better treat cancer patients.

The 18th International Conference on Scientific Computing (CSC 2020)
https://americancse.org/events/csce2020/conferences/csc20

Evaluating the Effect of Compensators and Load Model on Performance of Renewable and Nonrenewable DGs

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

Abstract: Multi distributed generation and compensators are one of the practical ways to improve the performance of power network. Although each of these devices increases the efficiency of distribution system, the combine of DG and compensator can be more useful. So in this study, the performance of capacitor bank (as traditional compensator) and DSTATCOM (as modern compensator) are evaluated in the distribution system with multi-DG. The load model of the network is also considered as sensitive to voltage-frequency and various customers' daily load patterns. For getting the best result, simultaneous placement of different types of DGs and compensators is done by the combination of multi-objective whale optimization algorithm and analytical hierarchy process. Technical, economic and environmental indices are considered as objective functions of this study. The results of the simulation using the IEEE 69-bus distribution system show the proper performance of compensators in the increasing of efficiency of DG units during the different condition of operation.

POST-DS: A Methodology to Boost Data Science

Carlos J. Costa, Joao Tiago Aparicio
Universidade de Lisboa, Portugal

Abstract: As the importance of data science is increasing, the number of projects involving data science and machine learning is rising either in quantity or in complexity. It is essential to employ a methodology that may contribute to the improvement of the outputs. In this context, it is crucial to identify possible approaches. And an overview of the evolution of data mining process models and methodologies is given for context. And the analysis showed that the methodologies covered were not complete. So, we propose a new approach to tackle this problem. POST-DS (Process Organization and Scheduling electing Tools for Data Science) is a process-oriented methodology to assist the management of data science projects. This approach is not supported only in the process but also in the organization scheduling and tool selection. The methodology was employed in the context of Covid 19.

Undergraduate Research: BladeRunner

Adina Paddy, Cha Xiong, Colt Henderson, Tuu Le, Daren Wilcox*
Department of Electrical Engineering Technology, Kennesaw State University, Marietta, Georgia, USA

Abstract: In this project, students in the Electrical Engineering Technology Department at Kennesaw University worked under the guidance of advisory Professor Daren Wilcox and Mr. John Cohran, an automation engineer at Omni International Inc, to develop a prototype of the second generation of an automated liquid handling machine in current production called the Prep 96. The prototype developed was called "Bladerunner." The group researched the Prep 96 program to study its operation and worked with Omni International and the Festo Corporation to develop a prototype code, schematic, and simulation using a FESTO PLC-HMI. Students established communication as well as provide autonomous and teleoperation control between devices. What follows is a summary of the current Prep 96 industrial design, the Bladerunner prototype design, the components that are used in the Bladerunner design, a description of the software used, the prototype schematic, discovered results, and supported datasheets.

Modeling, Simulation and Verification for Structural Stability and Vibration Reduction of Gantry Robots for Shipyard Welding Automation using ANSYS Workbench and Recurdyn

Seung Min Bae, Won Jee Chung, Hui Geon Hwang, Yeon Joo Ahn
School of Mechatronics, Changwon National University, Changwon, Gyeongsangnam do, South Korea;
Robot Valley Corporation, Changwon, Gyeongsangnam do, South Korea

Abstract: With the strengthening of domestic and foreign environmental regulations in recent times, the disposal of small scrapped FRP (Fiber Reinforced Plastic) vessels is raising a problem and there is a growing interest in eco-friendly aluminum vessels. Welding is one of important things in manufacturing processes of aluminum vessels. This paper aims to research on structural stability and vibration reduction of a three axes cartesian coordinate gantry robot for improving the welding quality. Structural instability and drive unit of each axis can negatively affect vibration. Structural analysis will be performed for parts that can cause structural instability due to load. Structural analysis will verify structural stability of gantry robot using ANSYS Workbench®, based on the modeling for analysis using Solidworks®. Also, the driving parts of the x axis and y axis are rack and pinion gear models, which have the disadvantage of vibration caused by backlash. A simulation to reduce vibration caused by backlash will be conducted using Recurdyn®. By simulation with added pinion gear model, we will investigate on how the model affect the vibration of the backlash.

Dark Data: Managing Cybersecurity Challenges and Generating Benefits

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

Abstract: Data Science delivers an important role in cybersecurity by utilizing the power of data, high-performance computing and data mining to protect users against cybercrimes and threats. Due to the rapid expansion of data, there is hidden dark data in organizations that are unidentified and unmanageable. It is significant to understand the various prospects that will arise through the utilization of dark data by various organizations. They will find this data valuable to the intelligence of their business. In this paper, we are introducing the methodology which will identify dark data and forecast future cybersecurity threats.

Implementing Modern Security Solutions for Challenges Faced by Businesses in the Internet of Things (IoT)

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

Abstract: IoT connects various non-living objects through the internet and enables them to share information with their community network to automate processes for humans and make their lives easier; a trend known in the computer world as the 4th industrial revolution (Industry 4.0). The use of IoT is increasing between businesses and individuals. Businesses are currently increasing their investment in IoT for use in the coming years. Industry 4.0 will play a pivotal role in the integration of industry and IoT concepts. This paper presents the challenges of IoT and solutions. The usage of IoT is growing in marketing, given the vast benefits that IoT offers. For businesses to fully use IoT and realize its benefits, the companies will need to change their business processes. However, businesses will need to overcome three major challenges: security, privacy and network challenges.

Comparison of the IaaS Security Available from the Top Three Cloud Providers

L. Kate Tomchik
Kennesaw State University, Marietta, Georgia, USA

Abstract: A new cloud Infrastructure as a Service (IaaS) server takes less time to initiate then it takes to drive to Microcenter to buy a new laptop. For this reason, application developers are now using cloud resources for a quick route to build a new system. These servers no longer have the hardware teams ordering from the approved vendors, or the Security administrators verifying the software chosen and overseeing the installation the way they did for the physical computer purchases. The responsibility of securing the virtualization, server, storage and networking application layers is now on the cloud provider. This paper will compare the current top three cloud platforms in the area of available security tools. The need for a company to still employ Information Technology Security officers is also explained.

Numerical Modelling of a Viscous Incompressible Fluid Flow in a Channel with a Step

Saeed Dubas, Paul Bouthellier, Nihal Siriwardana
Department of Engineering and Mathematics, University of Pittsburgh at Titusville, Pittsburgh, USA;
Department of Mathematics, Prairie View A & M University, Prairie View, Texas, USA

Abstract: A stable fourth order finite difference scheme is applied for solving the problem of steady state, viscous, incompressible flow through a channel with a step. Results are presented for various step sizes and Reynolds numbers. The method converged for all parameters attempted and the results compared favorably with the literature, from a theoretical, numerical and experimental stand-point.

Analysis of all Local Pairwise Sequence Alignment Algorithms

Bharath Reddy, Richard Fields
Process Automation R&D, Schneider-Electric, Lake Forest, California, USA

Abstract: Biological sequence alignment is common today and are used in a variety of fields ranging from Bioinformatics, Computational Biology, Genome analysis, Cancer research, Stem Research and many more fields. Most of these fields use the sequence alignment to find the 'similar' regions or similarities between organisms. Since, this step is computational heavy, today, there are specialized hardware to help speed up and techniques and strategies to help speed up or improve the sensitivity (quality) of the alignment in general. The early successful algorithms in sequence alignment were focused on quality, and it produced an optimal algorithm called SmithWaterman algorithm, which we will discuss in detail later using a technique called 'Dynamic Programming'. The time complexity of this algorithms was O (mn). Later, to speedup, heuristic algorithms were developed. Heuristic algorithms gave up a little bit on the quality for speed, by calculating the near-optimal alignment rather than optimal algorithm. In this paper, we will analyze various computational approaches for local sequence alignments.

Maximum Match Subsequence Alignment Algorithm; Finely Grained (MMSAA - FG)

Bharath Reddy, Richard Fields
Process Automation R&D, Schneider-Electric, Lake Forest, California, USA

Abstract: Sequence alignment is common nowadays as it is used in many fields to determine how closely two sequences are related and at times to see how little they differ. In computational biology / Bioinformatics, there are many algorithms developed over the course of time to not only align two sequences quickly but also get good laboratory results from these alignments. The first algorithms developed were based off a technique called Dynamic Programming, which were very slow but were optimal when it comes to sensitivity. To improve speed, more algorithms today are based off heuristic approach, by sacrificing

sensitivity. In this paper, we are going to improve on a heuristic algorithm called MASAA (Multiple Anchor Staged Local Sequence Alignment Algorithm) – and MASAA -S which we published previously. This new algorithm appropriately called MMSAA – FG, stands for Maximum Match Subsequence Alignment Algorithm – Finely Grained. The algorithm is based on suffix tree data structure like our previous algorithms, but to improve sensitivity, we employ adaptive seeds, and finely grained perfect match seeds in between the already identified anchors. We tested this algorithm on a randomly generated sequences, and Rosetta dataset where the sequence length ranged up to 500 thousand.

A Note on the Sensitivity of Generic Approximate Sparse Pseudoinverse Matrix for Solving Linear Least Squares Problems

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

Abstract: During the last decades, research efforts have been focused on the derivation of effective explicit preconditioned iterative methods. In this manuscript, we review the explicit preconditioned conjugate gradient least squares method, based on generic sparse approximate pseudoinverses, in conjunction with approximate pseudoinverse sparsity patterns, based on the modified row-threshold incomplete QR factorization techniques. Additionally, modified Moore-Penrose conditions are presented and theoretical estimates for the sensitivity of the generic approximate sparse pseudoinverses are derived. Finally, numerical results concerning the generic approximate sparse pseudoinverses by solving characteristic model problems are given. The theoretical estimates were in qualitative agreement with the numerical results.

Long-Short Term Memeory Neural Network on the Trajectory Computing of Directly Dynamics Simulation

Fred Wu, Colmenares-Diaz Eduardo, Poojitha Chapala, Tejaswi Jonnalagadda, Sailaja Peruka, Pooja Sonmale
Department of Mathematics and Computer Science, West Virginia State University Institute, West Virginia, USA;
Department of Computer Science, Midwestern State University, Wichita Falls, USA

Abstract: This Directly dynamics simulation is widely used in Quantitative structure–activity relationship, virtual screening, protein structure prediction, quantum chemistry, materials design and property prediction, etc. This paper explores the idea of integrating Long-Short Term Memory (LSTM) with the trajectory computing of directly dynamics simulations to enhance the performance of the simulation and improve its usability for research and education. The idea is successfully used to predict the location, energy and Hessian of atoms in a CO2 reaction system. The results demonstrate that the artificial neural network based memory model successfully learns the desired features associated with the atomic trajectory and rapidly generates predictions that are in excellent agreement with the results from chemistry dynamics simulations. The accuracy of the prediction is better tha,n expected.

Trusted Reviews: Applying Blockchain Technology to Achieve Trusted Reviewing System

Areej Alhogail, Ghada Alhudhayf, Jood Alanzy, Jude Altalhi, Shahad Alghunaim, Shahad Alnasser
Department of Information systems, King Saud University, Riyadh, Saudi Arabia

Abstract: In today's flourishing e-business environment, reviews play a significant role in our daily products' and services' choices. Reviews can be described as trusted if the source of it was known, authentic and reliable. Trusted Reviews aims to improve the authenticity and quality of customer's reviews submitted to the system. In order to achieve this, Blockchain technology is implemented for its unique characteristics, such as the immutability of data to prevent fake reviews. In order to encourage members to write legitimate reviews, Thiqah -Trust Credit- will be used as an incentive, serving a significant role in

our reward system. Consequently, more genuine reviews will be submitted, improving the system's reputation legitimacy and enhancing the member's experience. The model has been tested using Ethereum for decentralized applications. It all starts with writing a smart contract, this contract contains the rules and conditions that will identify a review as trusted. For example, reaching a certain number of likes. Upon satisfying the contract conditions a block will be created containing all details of the review, added to the blockchain and the writer of the review gets awarded a Thiqah credit point. The implemented solution will help business owners gain a good reputation and increase customer trust.
Introduction

Large-scale Parallelization of Lattice QCD on Sunway Taihulight Supercomputer

Ailin Xu, Zhongzhi Luan, Ming Gong, Xiangyu Jiang
School of Computer Science and Engineering, Beihang University, Beijing, P. R. China;
Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, P. R. China

Abstract: Lattice quantum chromodynamics is an important method for studying strong interaction through large-scale Monte Carlo numerical simulation calculations. Common computing platforms are difficult to meet the needs of large-scale high-precision calculations of lattice quantum chromodynamics. Sunway Taihulight supercomputer, which is based on SW26010 heterogeneous many-core processor, can provide sufficient computing power, but applications still need large-scale parallelization and performance optimization according to its unique hardware architecture. Through the analysis and improvement of the previous work, we propose a new grid point data distribution method and perform efficient parallel computing. The lattice QCD application achieved the peak performance of 139.147TFlops using 1347840 cores on Sunway TaihuLight supercomputer and can maintain performance as the scale grows.

Dielectric Polymer Genome: Integrating Valence-Aware Polarizable Reactive Force Fields and Machine Learning

Kuang Liu, Antonina Nazarova, Ankit Mishra, Yingwu Chen, Haichuan Lyu, Longyao Xu, Yue Yin, Qinai Zhao, Rajiv Kalia, Aiichiro Nakano, Ken-ichi Nomura, Priya Vashishta, Pankaj Rajak
Collaboratory of Advanced Computing and Simulations, Department of Computer Science, Department of Physics & Astronomy, Department of Chemical Engineering & Materials Science, Department of Biological Sciences, University of Southern California, Los Angeles, California, USA;
Loker Hydrocarbon Research Institute and the Bridge@USC, Department of Chemistry, University of Southern California, Los Angeles, California, USA;
Mork Family Department of Chemical Engineering & Materials Science, University of Southern California, Los Angeles, California, USA;
Argonne Leadership Computing Facility, Argonne National Laboratory, Illinois, USA

Abstract: Informatics-driven computational design of advanced dielectric polymers (i.e., dielectric polymer genome) has remained a challenge. We have developed a computational framework for (i) high-throughput computational synthesis of polymer structures, (ii) evaluation of their dielectric properties using reactive molecular dynamics simulations based on a new valence-aware polarizable reactive force field (ReaxPQ-v), and (iii) learning polymer structure-property relationships using machine-learning (ML) models. The resulting large size of simulated training dataset provides an unprecedented opportunity to uncover hitherto-unknown structure property relationships purely computationally, thereby predicting new polymers with desired dielectric properties. Employing a large dataset of structurally diverse 1,276 polymers, multi-layer perceptron and random forest models achieved good accuracy for predicting the dielectric constants of these polymers, while recurrent neural network model is being developed. Such ML prediction models are indispensable for further enlarging the search space for superior dielectric polymers by orders-of-magnitude.

Orientation and Line Thickness Determination in Binary Images

Sean Matz
Claremont Graduate University, Claremont, California, USA

Abstract: This paper addresses the problems of orientation determination of lines in binary images and the determination of line thickness. The orientation problem utilizes the Radon Transform while the line thickness problem determines the thickness of lines at selected angles by considering the pattern of the pixels of those lines. The Radon transform maps a line at a given angle to a point in feature space (also known as the sinogram). The sinogram is typically generated for a wide range of angles (from 0 to 179 degrees in this case). Consequently, lines at particular angles will map to points whose sinogram value is greater than that of other points in the sinogram, thereby generating local peaks in the sinogram.

Elemental Analysis of Oil Paints

Shijun Tang, Rosemarie C. Chinni, Amber Malloy, Megan Olsson
Alvernia University, Reading, PA, USA

Abstract: Painting works have a long history and significant cultural value. Digital image processing has been introduced to analyze and identify paintings. As an important characteristic of images, image histograms are used to distinguish basic pure and mixture pigments. In this paper, we have investigated the peak location of image histograms of 21 fundamental pigments containing pure pigments and mixture pigments. Whether pure pigments or mixture pigments, the pigments' histograms have unique peak locations. Our research indicates that fundamental pigments can be effectively distinguished and separated according to their own pigments' image histograms.

Reverse Threat Modeling: A Systematic Threat Identification Method for Deployed Vehicles

Mona Gierl, Reiner Kriesten, Peter Neugebauer, Eric Sax
Karlsruhe University of Applied Sciences, Institute of Energy Efficient Mobility, International University Campus 3, Bruchsal;
Karlsruhe Institute of Technology, Institute for Information Processing Technologies, Engesserstrasse 5, Karlsruhe, Germany

Abstract: During the development phase of a vehicle, threat and risk analyses are common methods to identify potential security threats and derive applicable security countermeasures. These measures form the basis to mitigate the risk of a successful attack and are integrated and tested during the development phase of the vehicle. However, over the whole vehicle life cycle, from concept phase until decommissioning, new attack methods which have not been known at the time of design might be developed that allow an exploitation of the system. Intuitively, threat and risk assessment on a regular basis - even after deployment - is desirable. In this context, the present paper proposes a systematic threat modeling method featuring a threat identification process for deployed vehicles.

PRNG-Broker: A High-Performance Broker to Supply Parallel Streams of Pseudo-Random Numbers for Large Scale Simulations

Andre Pereira, Alberto Proenca
Algoritmi Centre, Department of Computer Science, University of Minho, Universidade do Minho, Gualtar, Portugal

Abstract: The generation of large streams of Pseudo-Random Numbers may lead to performance degradation in simulation applications. Both the PRN Generator and how it is used impact the efficiency of generating multiple PRNs. A PRNG-Broker was developed for parallel servers with/without accelerators, which transparently manages the efficient execution of PRNG implementations from CPU/GPU libraries, with an intuitive API that replaces the user PRNGs requests. PRNG-Broker allows

the development of efficient PRN intensive applications without the need of explicit parallelization and optimization of PRNG tasks. It was validated with scientific analyses of proton beam collisions from CERN, which require 30 Ki PRNs per collision. Outcomes show a performance boost over the original code: 48x speedup on a 2*12-core server and over 70x speedup when using a GPU.

The Caloric Curve of Polymers from the Adaptive Tempering Monte Carlo

Greg Helmick, Yoseph Abere, Estela Blaisten-Barojas
Center for Simulation & Modeling & Department of Computational & Data Sciences, George Mason University, Virginia, USA

Abstract: Conductive polymers are organic conjugated polymer chains with semiconducting ability that display unique mechanical properties without being thermoformable. Here we present a novel coarse-grained force field for modeling the oxidized phase of polypyrrole containing electronegative atomic dopants. The polypyrrole oligomers in this study have twelve monomers in length with a doping concentration of 25%. The polymer properties are determined using the isothermal-isobaric Adaptive Tempering Monte Carlo and the Metropolis Monte Carlo with codes optimized for GPUs. Several thermodynamic and mechanical properties are calculated along the caloric curve. When comparing with experiments, densities and bulk moduli perform very well yielding values in the range of 1.20-1.22 g/cm3 and 67-120 MPa, respectively. Comparing with our published model potential for the neutral polypyrrole phase, the oxidized phase presents about 30% increase in density, which is also in agreement with experiments. The computational implementation is easily portable for the inspection of other polymeric materials.

A New Technique of Invariant Statistical Embedding and Averaging in Terms of Pivotal Quantities for Improvement of 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: In the present paper, a new technique of invariant embedding of sample statistics in a decision criterion (performance index) and averaging this criterion via pivotal quantities (pivots) is proposed for intelligent constructing efficient (optimal, uniformly non-dominated, unbiased, improved) statistical decisions under parametric uncertainty. This technique represents a simple and computationally attractive statistical method based on the constructive use of the invariance principle in mathematical statistics. Unlike the Bayesian approach, the technique of invariant statistical embedding and averaging in terms of pivotal quantities (ISE&APQ) 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. The aim of the present paper is to show how the technique of ISE&APQ may be employed in the particular case of optimization, estimation, or improvement of statistical decisions under parametric uncertainty. To illustrate the proposed technique of ISE&APQ, application examples are given.

Shallow SqueezeNext Architecture Implementation on Bluebox2.0

Jayan Kant Duggal, Mohamed El-Sharkawy
Department of Electrical and Computer Engineering, IoT Collaboratory,
Purdue School of Engineering and Technology, IUPUI, Indianapolis, Indiana, USA

Abstract: Machine learning and its applications such as ADAS, embedded Computer Vision, image and object detection, etc made self-driving car applications possible and safer. Major factors of a hindrance for ADAS deployment are limited computational and memory resources. With the help of the DSE of CNN/DNN architectures, the Shallow SqueezeNext architecture is proposed that overcame the limitations of traditional algorithms. It achieved the least model size of 272KB with a model accuracy of 82%, a model speed of 9 seconds per epoch. Therefore, making it capable of deploying efficiently on a real-time platform Bluebox2.0 by NXP with a model size of 0.531MB with a model accuracy of 87.30% at a model speed of 11 seconds per epoch.

Greedy Navigational Cores in the Human Brain

Zalan Heszberger, Andras Majdan, Andras Biro, Andras Gulyas, Laszlo Balazs, Jozsef Biro
Department of Telecommunications & Media Informatics, Budapest University of Technology & Economics, Hungary;
MTA-BME Information Systems Research Group, Budapest, Hungary

Abstract: Greedy navigation/routing plays an important role in geometric routing of networks because of its locality and simplicity. This can operate in geometrically embedded networks in a distributed manner, distances are calculated based on coordinates of network nodes for choosing the next hop in the routing. Based only on node coordinates in any metric space, the Greedy Navigational Core (GNC) can be identified as the minimum set of links between these nodes which provides 100% greedy navigability. In this paper we perform results on structural greedy navigability as the level of presence of Greedy Navigational Cores in structural networks of the Human Brain.

A Multi-Commodity Flow Formulation and Edge-Exchange Heuristic Embedded in Cross Decomposition for Solving Capacitated Minimum Spanning Tree Problem

Hansuk Sohn, Dennis Bricker
Industrial Engineering, New Mexico State University, Las Cruces, New Mexico, USA;
Industrial Engineering, University of Iowa, Iowa City, Iowa, USA

Abstract: This paper presents a new mathematical formulation, which is based on multi-commodity flow, for a classical capacitated minimum spanning tree problem. It also demonstrates the performance of Van Roy's Cross decomposition algorithm for solving the capacitated minimum spanning tree can be significantly improved by incorporating an edge-exchange heuristic algorithm at a tremendous saving in the computational effort. The results also reveal that the proposed algorithm is very competitive with Lagrangian original algorithm in terms of solution quality. The use of the new formulation and the proposed algorithm which take better advantage of the problem structure, especially that of the dual subproblem, provides a large potential for improvement.

TITLE OF SESSION:
Military and Defense Modeling and Simulation

Co-Chairs: Dr. Douglas D. Hodson, Dr. Michael R. Grimaila, Dr. Ryan D. Engle
CSE Department, US Air Force Institute of Technology (AFIT), USA;
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

Julia and Singularity for High Performance Computing

Joesph Tippit, Douglas Hodson, Michael Grimaila
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

Abstract: High performance computing (HPC) is pivotal in the advancement of modern science. Scientists, researchers, and engineers are finding an increasing need to process massive amounts of data and calculations faster and more accurately than ever before. This is especially true in our work of developing a general quantum library for researchers to use in their simulations. Much of this effort revolves around getting the maximum performance enhancements offered by GPUs as possible. We have found that the relatively newer programming language Julia has offered us a productive means of development with minimal overhead. Combined with the container engine Singularity, we can ensure maximum distributability and reproducibility.

Trojan Banker Simulation Utilizing Python

Drew Campbell, Jake Hall, Iyanu Odebode, Douglas Hodson, Michael Grimaila
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

Abstract: One of the most malicious types of malware used by hackers today are Trojans. Trojans can come in multiple different forms and are used to steal various types of information from a user. In the case of a Trojan Banker, it tricks a user by acting like something useful to the consumer, but it is secretly used to steal account information from a user and deliver that information to the hackers computer. Most of these attacks are delivered via Social Engineering, and in this paper, we will host a simulation that will identify what types of social engineering techniques are used to gain access to a user's computer, and then identify how long machines at varying levels of security take to discover and remove a Trojan from its system. This simulation should display how difficult it is for an active Trojan to go undetected by a machine as well as how much information a Trojan can steal before being caught.

CovidLock a New Form of Ransomware

Amber Modlin, Andres Gregory, Iyanu Odebode, Douglas Hodson, Michael Grimaila
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

Abstract: CovidLock is a new form of Ransomware taking advantage of the panic brought on by COVID-19. It tricks users into downloading an app claiming to track the pandemic. When downloaded, this app encrypts the files on the user's device and changes the device's password, demanding a ransom of cryptocurrency to regain access. As the problem of Ransomware grows, new forms of mitigating the problems must arise. We present a description of CovidLock, a new mitigation method for Ransomware and other Malware, and a reason why our method is better than existing methods.

The New Office Threat: A Simulation Environment of Watering Hole Cyber Attacks

Braeden Bowen, Jeremy Erqybar, Iyanu Odebode, Douglas Hodson, Michael Grimaila
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

Abstract: The focus of this paper is to develop a DEVSstyle simulation model to manipulate common variables in an Advanced Persistent Threat (APT)-style Watering Hole Attack, a style of attack that targets an organization or target group by infecting a commonly-used website or service. A simulation of the environment exposed to this specific attack was developed through Python, carrying variables of target group size, number of trusted sites, and duration of attack before discovery. Analysis of simulation averages suggest that size of the target group and the duration of the attack are the most important factors in the spread of the malware, though for each category the returns on speed of infection diminish as the size and time of the overall control groups increase.

Simulation of SYN Flood Attacks and Counter-Attack Methods Using Average Connection Times

Hai Vo, Raymond Kozlowski, Iyanu Odebode, Douglas Hodson, Michael Grimaila
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

Abstract: While SYN Flood attack leveraging TCP protocol is not new, this method is still by far the most popular attack type. In this paper, a simulation of a TCP server is introduced, as well as three different ways to gather data on connection times and number of unresolved requests to connect in the server. The results show that the most efficient way to do so would be to use an average of only successful handshake connections to apply as a server timeout.

The 19th International Conference on e-Learning, e-Business, Enterprise Information Systems, & e-Government (EEE 2020) https://americancse.org/events/csce2020/conferences/eee20

Sentiment Analysis of Product Reviews on Social Media

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

Abstract: Social media is a platform where people share their interests and opinions. Users posting their opinions about a product or a service they used is very common nowadays. This information can be important to marketers. Knowledge about how a product is perceived in a society is crucial to improve the sales and formulate marketing business strategies. The data flow in social media is of extreme high volume and this data is generated in a scale of hundreds of gigabytes each day. There are many applications which have been built to make the most use of this idle data one of the most important one being sentiment analysis. Sentiment Analysis is the process of determining classification of opinions expressed in a text as positive, negative or neutral. In this project we would like to develop an application that mimics the functions of a Sentiment Analysis system by using different third-party APIs/libraries to collect data and analyze it. A user can use this application to get the sentiment of users for a product on social media.

Enhancing Music Teachers' Cognition and Metacognition Grassroots FD Project 2019 at Music College

Chiharu Nakanishi, Asako Motojima, Chiaki Sawada
Music Department, Kunitachi College of Music, Kashiwa-cho, Tokyo, Japan

Abstract: The purpose of the present study is to outline the development of an FD (university teacher professional training) project for music teachers and to examine how it can affect the teachers. The “Grassroots FD Project 2019” was developed in which music teachers would reflect their teaching and write a paper to meet the requirements of the Ministry of Education, Culture, Sports, Science and Technology (MEXT). It was programmed with mainly six activities: observing open music lessons, answering worksheets with cognitive and metacognitive questions, discussing with colleagues, practicing and reflecting on teaching, and writing a paper. To support teacher learning, 3 worksheets were made. By participating in the project, the 13 music teachers who were novice writers could write a paper and raised their metacognition of teaching.

Emerging Interactions of ERP, Big Data and Automotive Industry

Florie Bandara, Uchitha Jayawickrama
School of Business and Economics, Loughborough University, Loughborough, United Kingdom

Abstract: Interactions of enterprise resource planning systems and big data is crucial for automotive industry in the process of quick and reliable decision making with use of large chunks of data collected by each department of the organization. Similarly, unstructured data collected by sensor systems need proper control of data to put out the best results combined to automation. This study adopts a systematic literature review conducted mainly under three phases in order to give a robust combination between the three areas i.e. ERP systems, big data and automotive industry. The three phases are determining the combination between the enterprise resource planning systems and big data and individually explaining their interaction with the automotive industry. This study has been able to identify the strict influence of large chunks of data on the automotive industry such as data management issues, trust issues and complexity in the responsiveness of enterprise resource planning systems. It is recognized that the main reasons for emergence of complexity in the responsiveness of enterprise resource planning systems is due to the

unstructured data collected by sensors of emerging concepts such as connected cars and the eventual automation of automobile functions. The study depicts the major influence of an enterprise resource planning system in order to centralize the entire organization whilst the large amount of structured and unstructured data collected.

Research on Efficient and Fuzzy Matching Algorithm in Information Dissemination System

Qinwen Zuo, Fred Wu, Fei Yan, Shaofei Lu, Colmenares-Diaz Eduardo, Junbin Liang (Heng Wu)
Information System Department, State Key Laboratory of NBC, Protection for Civilian, Beijng, P. R. China;
Department of Mathematics and Computer Science, West Virginia State University Institute, West Virginia, USA;
Department of Information Technology, Central Research Institute of History and Documents, Beijing, P. R. China;
College of Computer science and Electronic Engineering, Hunan University, Changsha, P. R. China;
Department of Computer Science, Midwestern State University, Wichita Falls, Texas, USA;
School of Computer and Electronics Information, Guangxi University, Guilin, P. R. China

Abstract: Matching algorithm is a key technology in content-based publish/subscribe system. Aiming at the lack of information sharing ability in dynamic wireless networks, an information distribution system model based on Content-based publish/subscribe technology is established. Then, according to the subjectivity and fuzziness of users' information requirements, using the logical coverage relationship between subscription constraints, an efficient fuzzy matching algorithm for battlefield information distribution and sharing is designed. Finally, the time and space efficiency of the algorithm is analyzed by experiments. Experimental results show that the algorithm is reasonable and efficient.

Tourism Service Auction Market for Saudi Arabia

Saad Nasser Almutwa, Sungchul Hong
Computer and Information Sciences Department, Towson University, Towson, Maryland, USA

Abstract: The Kingdom of Saudi Arabia is a country that contains many historical places for tourists to visit. Moreover, the Kingdom has announced it is opening the door for tourism in Saudi Arabia under the 2030 Vision. An electronic market can provide service providers and tourists a more efficient way to fulfill tourists' desired activities in Saudi Arabia at a minimum price. This paper proposed and tested a mathematical model and a heuristic algorithm using Java to match service providers consortia with tourists' wanted activity lists. The result of this proposed model and its heuristic algorithm demonstrated a feasible solution for an electronic market for Saudi Arabia's growing tourism industry.

Scaleable Undergraduate Cyber-Security Curriculum Through Auto-Graded E-Learning Labs

Aspen Olmsted
Department of Computer Science, Fisher College, Boston, MA, USA

Abstract: The US and world economies need more trained cybersecurity workers. The demand for these workers has driven prominent private universities to create large reduced cost programs for graduate students. Unfortunately, a large segment of the population does not have the pre-requisite knowledge to enter into these graduate-level programs. In this paper, we look at how to develop an undergraduate cyber-security online program through E-Learning with automatically graded labs. We develop techniques to increase the lab participation and integrity through a concept we call derivate labs. We provide preliminary evidence that these labs have, in fact, increased engagement and integrity in our online sections of courses in our undergraduate cybersecurity program.

The Use of Crowdsourcing as a Business Strategy

Hodaka Nakanishi, Yuko Syozugawa
Teikyo University, Tokyo, Japan; Tokoha University, Shizuoka, Japan

Abstract: This paper clarified the role of crowdsourcing in companies. A survey was conducted using the Internet. We analyzed the reasons and results of using crowdsourcing by procuring management resources, improving productivity, reducing costs, and responding to fluctuations. More than half of the companies use crowdsourcing to improve productivity and procure management resources while one-third of companies introduced crowdsourcing for cost reduction. Large companies are more likely to introduce crowdsourcing for productivity improvement than small and medium enterprises. Many companies use crowdsourcing to improve productivity and reduce costs for their new businesses compared to their main businesses. The results show that crowdsourcing is an innovative strategy of business management especially in large companies.

Characterizing Learner's Comments and Rating Behavior in Online Course Platforms at Scale

Mirko Marras, Gianni Fenu
Department of Mathematics and Computer Science, University of Cagliari, Italy

Abstract: Opinions expressed by learners on courses they attend play a primary role in arranging learning and teaching processes and designing data-driven educational services. The ongoing proliferation of massive open online initiatives and the increasing amount of opinions provided online are turning the exploration and exploitation of this collective intelligence into a challenging, while crucial, task. In this paper, we characterize learner's opinions conveyed by means of ratings and textual comments released after attending online courses, aiming at getting insights from multiple perspectives. Dynamics of opinion delivering are extracted from data collected along 10 years on a large-scale online learning platform. Our findings show that this domain has main distinguished peculiarities with respect to other educational domains and to other online domains wherein individuals express opinions. We expect that our findings will support the community to better understand opinion patterns in course platforms and, in turn, to devise tailored educational services.

Agile IT Service Management Frameworks and Standards - a Review

Manuel Mora, Jorge Marx-Gomez, Fen Wang, Oswaldo Diaz
Information Systems, Autonomous University of Aguascalientes, Aguascalientes, Ags., Mexico;
Informatics, University of Oldenburg, Oldenburg, Lower Saxony, Germany;
IT & Administration Management, Central Washington University, Ellensburg, Washington, USA;
Central Data Center, INEGI, Aguascalientes, Ags., Mexico

Abstract: IT Service Management (ITSM) frameworks and standards have helped to manage the planning, design, deployment, operation and improvement of IT services in business organizations in the last two decades ago. However, the current business environment has changed from stable and mid-term user demands to dynamic and short-term ones. Consequently, ITSM frameworks with an agile perspective are now emerging. This paper reviews four of the main proffered agile ITSM frameworks (ITILv4, VeriSM, FitSM and the ISO/IEC 20000-1:2018) regarding the expected agile tenets literature. It was found that ITILv4 and VeriSM adequately fit the expected agile tenets, but FitSM and the ISO/IEC 20000-1:2018 can be considered lightweight frameworks but not agile ones. Implications for the knowledge and practice, as well as conclusions are finally reported.

Software Evaluation Methods to Support B2B Procurement Decisions - An Empirical Study

F. Bodendorf, M. Lutz, J. Franke
Institute for Factory Automation & Production Systems, Friedrich-Alexander-University of Erlangen-Nuremberg, Germany;
TUM School of Management, Technical University of Munich, Munich, Bavaria, Germany

Abstract: Procurement decisions are particularly affected by the digital transformation in many industries. These decisions aim at cost savings as well as value creation of digital products coming from suppliers. The characteristics of digital components differ from conventional ones. Many companies have great difficulties in verifying the procurement costs of software products. Software evaluation becomes a more and more important part of cost engineering in companies. In theory and scientific literature, a distinction is made in software evaluation procedures between cost-, value- and license-oriented approaches. This paper presents the design and results of an empirical study to find out which of these approaches are currently being pursued and which ones are eyed for the future by companies in different industries.

Smart Low-Speed Self-Driving Transportation System

Zhenghong Wang, Bowu Zhang
School of Computer Science and Mathematics, Marist College, Poughkeepsie, New York, USA

Abstract: This paper presents a commercial application of self-driving vehicles and analyzes the strengths, weaknesses, opportunities, and threats (SWOT) related to the application. To illustrate the market potential of the application, we explore two use cases to present detailed application design, highlight prototyping activity, and identify risks and needs for risk management. In addition, we also discuss target customers and marketing strategies for the proposed application.

The Effect of Matching Learning Material to Learners' Dyslexia Type on Reading Performance

Hadeel Al-Dawsari, Robert Hendley
Department of Computer Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia;
School of Computer Science, University of Birmingham, Birmingham, United Kingdom

Abstract: Dyslexia is a universal reading difficulty. Each dyslexic suffers from individual reading problems. Some may not understand what is written, while others may omit, or transpose letters while reading. Most research has treated all dyslexics as a single class, especially in Arabic. Therefore, the aim of this re-search is to overcome these problems by training dyslexics using learning mate-rial that matches their individual needs, in order to improve their reading. To achieve this, an online training system for Arabic dyslexics has been developed and evaluated (in terms of learning gain and learner satisfaction).

Supporting Qualification Based Didactical Structural Templates for Multiple Learning Platforms

Michael Winterhagen, Benjamin Wallenborn, Dominic Heutelbeck, Matthias L. Hemmje
Multimedia and Internet Applications (MMIA), University of Hagen, Germany;
University of Hagen, Center for Media and IT (ZMI), Germany;
Forschungsinstitut fur Telekommunikation und Kooperation e.V. (FTK), Germany

Abstract: In our previous e-learning publications we introduced the concepts of Qualifications-Based Learning (QBL), Course Authoring Tools (CAT), and of Didactical Structural Templates (DST) which are a further development of the CAT. The DSTs have been combined with QBL. Therefore, the DSTs represent the pedagogical structure of learning materials. The idea is to have an abstract definition which can be implemented as a traditional course (with or without gaming content), as an applied game, or any other form or modality of learning content. In this paper we will go in deeper detail, how the DSTs can be accessed from any kind of tool or production environment.

Contingency Planning: Prioritizing Your Resources

Kathryne Burton, Necole Cuffee, Demetria Jackson, Darius Neclos, Samuel Olatunbosun, Taiwo Ajani
Department of Computer Science, Norfolk State University, Virginia, USA;
Department of Computer and Information Systems, Ferrum College, Virginia, USA

Abstract: Contingency planning has been a concept in technology for many decades, and is a more prevalent sector of information technology. Overtime, procedures have changed, but the object remains: "always have a plan." With the drastic and rapid changes in technology, companies have no choice but to have plans in case of major incidents such as natural disasters, blackouts, and certain cyber-attacks. In exploring contingency planning as a whole, this paper discusses different topics such as common successful methodologies and new discoveries.

Are Collaboration Tools Safe? An Assessment of Their Use and Risks

Cquoya Haughton, Maria Isabel Herrmann, Tammi Summers, Sonya Worrell, Samuel Olatunbosun, Taiwo Ajani
Department of Computer Science, Norfolk State University, Virginia, USA;
Department of Computer and Information Systems, Ferrum College, Virginia, USA

Abstract: The ongoing pandemic created uncertainty worldwide − stay at home orders, social distancing and business closure or reorganization to protect the health of employees and maintain business continuity. For many organizations, this meant a shift from the office to the home, and the adoption of collaboration tools for communications and sustenance of productivity. Operating safely and securely in a virtual workplace, via collaboration tools, requires distinct security considerations and configurations. The rise in the use of video and conferencing meeting applications have increased network traffic and in turn, an attraction for cybercriminals. Four collaboration tools were analyzed to identify their cyber threats, vulnerabilities and risks. We provide recommendations on how organizations and customers could mitigate risks and vulnerabilities; prevent cyber-attacks and; operate securely in a virtual workspace.

Individualized Educational System Supporting Object Oriented Programming

F. Fischman, H. Lersch, M. Winterhagen, B. Wallenborn, M. Fuchs, M. Then, M. Hemmje
FernUniversitat in Hagen, Germany; New York Institute of Technology, USA;
Wilhelm Buchner Hochschule, Germany

Abstract: The explanation of [8] will serve as the foundation for enhancing Learning Management System (LMS) [8] Moodle further with regards to generate courses automation based on students' assessments. In section 1 a description is provided on how teaching object oriented programming can be made more effective, efficient, and individualized. Furthermore, in section 2 this concept is elaborated further via the Nunamaker model [8] for addressing problem areas, research questions, challenges and research goals. In section 3 the details of the Knowledge Management Educational Portal (KM-EP) [8], Course Authoring Tool (CAT) [8], and Qualification Based Learning Model (QBLM) [8], which will provide a more detailed explanation of section 2. Section 4 will layout the conceptual work which will make use of various Unified Modeling Language (UML) [11] diagrams and normalization. The paper will conclude with section 5 and 6 that is a presentation of a protype implementation and an evaluation of the system via a cognitive walkthrough.

The 18th International Conference on Embedded Systems, Cyber-physical Systems, & Applications (ESCS 2020) https://americancse.org/events/csce2020/conferences/escs20

Wireless Blind Spot Detection and Embedded Microcontroller

Bassam Shaer, Danita L. Marcum, Curtis Becker, Gabriella Gressett, Meridith Schmieder
Electrical and Computer Engineering Department, University of West Florida, Fort Walton Beach, Florida, USA

Abstract: This paper presents the design and implementation of a blind spot detection system for installation on semi-trucks to mitigate safety risks the vehicle poses due to low visibility of the surrounding area. The system is designed for easy attachment as well as operation at highway speeds. The system will have seven transmitters and a receiver with LCD display. The seven sensors attached to the vehicle will detect traffic in the vehicle's blind-spots and transmit a signal to the receiver. The receiver LCD display will alert the driver by lighting the cab image and the obstructed sensor. All eight components of the system are powered by rechargeable batteries, which are charged via a wireless charging tray connected to the 12V power source of the vehicle. This paper will describe the design, implementation, and results of each of these objectives.

Ethical Issues of the Use Of AI in Healthcare

Suhair Amer
Department of Computer Science, Southeast Missouri State University, Cape Girardeau, Missouri, USA

Abstract: This document discusses the use of artificial intelligence in healthcare, ethical and legal issues of using artificial intelligence in healthcare, and solutions to some of the dilemmas of using AI in healthcare.

Real-Time Operating Systems: Course Development

Michael Rivnak, Leonidas Deligiannidis
Department of Computer Science and Networking, Wentworth Institute of Technology, Boston, Massachusetts, USA

Abstract: There are applications, such as industrial machine systems or airplane flight control systems, which require high accuracy that traditional operating systems are not designed to provide. Traditional operating systems are designed for fair usage among many users and many applications. Additionally, they are designed for user interaction and thus intend to provide balanced operation across all applications and users. In time-sensitive cases however, this balanced operation is not only not required, but often detrimental. Real-time applications instead focus on temporal accuracy. There is not a single definitive method of addressing this functional requirement, but rather multiple, each to be used depending on the situation. Often real-time systems run their tasks according to a time schedule in order to fit with other time-sensitive systems. Such systems come in several different forms of varying complexity. The purpose of this paper is to investigate different RTOS implementations and the associated advantages and disadvantages in terms of timing latency and accuracy; and to organize this information in an educational format for use in an operating systems course. We will incorporate this material into the curriculum to an existing undergraduate operating systems course. This course is an introduction to operating systems and thus prior knowledge of process scheduling should not be assumed. The two main implementations that will be discussed below are FreeRTOS and RT Linux, two opposites in terms of complexity and implementation among the real-time operating systems that are available for the Raspberry Pi.

BumpChat: A Secure Mobile Communication System

Brian Kammourieh, Nahid Ebrahimi Majd, Ahmad Hadaegh
Department of Computer Science & Information Systems, California State University San Marcos, California, USA

Abstract: BumpChat is a secure messaging application for Android devices. The app makes use of the Near Field Communication (NFC) hardware on devices to securely transmit cryptographic keys. Communication between devices after pairing is handled securely by a web server, through a simple PHP driven API. Data is encrypted at all stages: atrest on end devices, in-transit to the server, and at-rest on the server. A user defined password is used to protect data on the Android device and is required to be entered on every startup. There are two primary encryption algorithms used for message security (AES-256) and inbox verification with the server (RSA-2048).

On Development of Low Cost Autonomous UAVs for Generation of Topographic Maps

Michael Galloway, Elijah Sparks, Mason Galloway
School of Engineering & Applied Sciences, Western Kentucky University, Bowling Green, Kentucky, USA

Abstract: Topographic maps are used by a variety of disciplines that require detailed information about the features of the Earth's surface. These maps are traditionally created by surveyors in the field by measuring sets of points throughout a coverage area. These measurements are then converted into a topographic map representing the terrain by the use of contour lines. The focus of this research is to reduce the cost and time efforts for surveying land to generate topographic maps. We introduce an approach for developing a low-cost autonomous quadcopter UAV that uses LiDAR and CMOS image (camera) data to perform coverage area surveying to generate topographic maps. The terrain data acquired by our approach is post-processed on a dedicated server that displays the layered topographic map to the user in a web browser.

Data Collection and Generation for Radio Frequency Signal Security

Tarek A. Youssef, Guillermo A. Francia, III , Hakki Erhan Sevil
Department of Electrical and Computer Engineering, University of West Florida, Florida, USA;
Center for Cybersecurity, University of West Florida, Florida, USA;
Department of Intelligent Systems and Robotics, University of West Florida, Florida, USA

Abstract: The current proliferation of Unmanned Aerial Systems (UAS) for a wide range of applications ranging from commercial to defense purposes demands the need for their protection. The development of security tools and techniques will need realistic Radio Frequency (RF) datasets for research and testing. This paper presents an on-going research and development effort to produce RF signal datasets that can be used for the development and testing of machine learning (ML) systems. We envision that these systems will ultimately be the precursor of future autonomous and secure UAS to benefit society for many generations.

Piano Player with Embedded Microcontrollers

Bassam Shaer, Garrick Gressett, Phillip Mitchell, Joshua Meeks, William Barnes, Stone Hewitt
Electrical and Computer Engineering Department, University of West Florida, Fort Walton Beach, Florida, USA

Abstract: This paper presents the design and implementation of a piano player device which will be attached to a 61 key piano allowing a song to play from start to finish. The system translates information from a midi file or generated from a standardized music sheet utilizing an image recognition program to control the solenoids. Using solenoids controlled by two separate Arduino microcontrollers with linked serial ports, the keyboard will be actuated. The controlling program provides the user with a Graphical Interface, which provides a saved list of songs that can easily be added to and runs the image recognition program that translates and sends the midi file data to the master Arduino microcontroller.

Software Defined Global Navigation Satellite Systems and Resilient Navigation for Embedded Automation

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

Abstract: The current generation of navigation devices used on everything from robotic systems, various crewed vehicles - aircraft, watercraft, and spacecraft - integrate global navigation satellite systems (GNSS) with inertial navigation systems (INS) or measurement units (IMUs). The ubiquitous and continuous need of advanced automation and robotics for accurate and precise positioning, navigation, and timing (PNT) data require a resiliency beyond what this capability can deliver, considering the ease with which GNSS signals can be jammed and spoofed without advanced technology measures that also consume size, weight, and power (SWaP) resources. Advances in hardware such as software defined radios and single board computers enable virtualization of both GNSS and navigation systems that address the shortcomings of today's technology, allowing the emergence of resilient software defined GNSS and navigation systems for embedded automation.

Smart Automation of an Integrated Water System

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

Abstract: This research focuses on the design and development of smart automation for an integrated water system that combines two off-grid freshwater resources of Rain Harvesting System (RHS) and Atmospheric Water Generator (AWG). This system is designed to perform automatically, from the point of generation/collection to the point of use, i.e., a vertical farming unit. Smart logic control system with Arduino Uno is applied to make the water generation system adaptive to atmospheric condition and to automatically refill the supply tanks that provides nutrition solution to the plants. This smart water system offers consistent water supply, reduce energy consumption, and eliminates labor. An experimental setup was established to test if the smart logic system is working correctly with all the electrical components. Then the system is introduced on-site at the point of use.

Brief Review of Low Power GPU Techniques

Pragati Sharma, Hussain Al-Asaad
Department of Electrical and Computer Engineering, University of California, Davis, California, USA

Abstract: This paper reviews and analyzes different low-power saving techniques in GPUs. The survey attempts to analyze and understand the architectural aspect of novel GPU design suggestions that researchers have come up with to provide cost-effective solutions for increasing GPU power numbers. These power saving insights venture into the space of system-level low-power implementations, while others exploit the application dependent optimizations thereby making this study an interesting mix between architecture, application, and system-level implementations. The paper also focuses on underlining the impact of any power saving feature in terms of area overhead and timing constraint latencies. This gives the reader a complete picture of the scope and challenges in the low power GPU exploration space. The study is limited by time and hardware resources and there is an excellent opportunity to quantify reviewed techniques and results to be compared against a common baseline.

Quadratic Integer Programming Approach for Reliability Optimization of Cyber-Physical Systems under Uncertainty Theory

Amrita Chatterjee, Hassan Reza
School of Electrical Engineering and Computer Science, University of North Dakota, North Dakota, USA

Abstract: Cyber-physical systems (CPS) are an example of software and hardware components working in symphony. The greatest challenge in CPS design and verification is to design a CPS to be reliable while encountering various uncertainties from the environment and its constituent subsystems. Cost, delay and reliability of a CPS are functions of software-hardware partitioning of the CPS design. Hence, one of the key challenges in CPS design is to achieve reliability maximization while factoring in uncertainty in cost and delay. This work leverages the problem formulation developed in recent research [13], which poses CPS design as an optimization problem for reliability assurance while factoring in uncertainty in cost and delay. In this formulation cost and delay are modeled as variables with uncertainty distributions under uncertainty theory, and the reliability requirement becomes an optimization objective. Authors of [13] also show heuristic solutions of this optimization problem can produce hardware/software partitioning which has potential to offer greater reliability under uncertainty. The novel contribution of this work is the exploration of alternate heuristics to genetic algorithm used in [13] to solve the optimization problem. We conclude that treating the optimization problem as a 0-1 integer quadratic programming problem is feasible and then explore a few heuristics to solve such problems. Next, we solve this problem with an heuristic method. Preliminary results suggest that this solution method can achieve better reliability.

The 16th International Conference on Foundations of Computer Science (FCS 2020)
https://americancse.org/events/csce2020/conferences/fcs20

What Have Google's Random Quantum Circuit Simulation Experiments Demonstrated about Quantum Supremacy?

Jack K. Horner, John F. Symons
Independent Researcher, Los Alamos, New Mexico, USA;
Department of Philosophy, University of Kansas, Lawrence, Kansas, USA

Abstract: Quantum computing is of high interest because it promises to perform at least some kinds of computations much faster than classical computers. Arute et al. 2019 (informally, "the Google Quantum Team") report the results of experiments that purport to demonstrate "quantum supremacy" – the claim that the performance of some quantum computers is better than that of classical computers on some problems. Do these results close the debate over quantum supremacy? We argue that they do not. In the following, we provide an overview of the Google Quantum Team's experiments, then identify some open questions in the quest to demonstrate quantum supremacy.

Chess is Primitive Recursive

Vladimir A. Kulyukin
Department of Computer Science, Utah State University, Logan, Utah, USA

Abstract: A two-player deterministic board game is primitive recursive if there exists a primitive recursive function that returns a sequence of optimal moves for a given player and a given epoch number. The game of chess is shown to be primitive recursive.

How to Extend the Single-Processor Paradigm to the Explicitly Many-Processor Approach

Janos Vegh
Kalimanos BT, Debrecen, Hungary

Abstract: The computing paradigm invented for processing a small amount of data on a single segregated processor cannot meet the challenges set by the present-day computing demands. The paper proposes a new computing paradigm (extending the old one to use several processors explicitly) and discusses some questions of its possible implementation. Some advantages of the implemented approach, illustrated with the results of a loosely-timed simulator, are presented.

Exact Floating Point

Alan A. Jorgensen
True North Floating Point, Las Vegas, Nevada, USA

Abstract: Standard IEEE floating point, which defines the representation and calculations of real numbers using a binary representation similar to scientific notation, does not define an exact floating-point result. In contrast, here we use a patented bounded floating-point (BFP) device and method for calculating and retaining the precision of the floating-point number represented, which provides an indication of exactness, with an "exact" floating-point result defined as a result that has error within + or - ½ units in the last place (ulps). Analysis and notification of exactness is important because subtraction of "similar," but inexact, floating-point numbers can introduce an error (even catastrophic error) in the calculation. Here we also define "similar" and use bounded floating point to provide examples comparing subtraction of exact and inexact similar numbers by comparing the results from 64-bit and 128-bit standard and 80-bit bounded floating-point calculations.

Random Self-Modifiable Computation

Michael Stephen Fiske
Aemea Institute, San Francisco, California, USA

Abstract: A new computational model, called the ex-machine, executes standard instructions, meta instructions, and random instructions. Standard instructions behave similarly to machine instructions in a digital computer. Meta instructions self-modify the ex-machine's program during its execution. We construct a countable set of ex-machines; each can compute a Turing incomputable language, whenever the quantum random measurements in the random instructions behave like unbiased Bernoulli trials.

Formal Specification and Verification of Timing Behavior in Safety Critical IoT Systems

Yangli Jia, Zhenling Zhang, Xinyu Cao
School of Computer Science & Technology, Liaocheng University, P. R. China;
China National Institute of Standardization, Beijing, P. R. China

Abstract: Formal specification and verification of complex IoT systems' behavior can efficiently improve the systems' correctness and reliability. This paper presents an enhanced time behavior protocol to specify real-time components' timed interaction behaviors in IoT systems. The protocol model bound event tokens with time consumption constraint information according to requirements of practical applications, and time-related operators are added into the model language. Visualization and verification method for composited behavior is given. An application example is introduced and the experimental results show that the enhanced time behavior protocol based model can be used easily to specify, visualize and verify IoT systems' interaction behavior and timing constraint information.

Introducing Temporal Behavior to Computing Science

Janos Vegh
Kalimanos BT, Debrecen, Hungary

Abstract: The abstraction introduced by von Neumann correctly reflected the state of the art 70 years ago. Although it omitted data transmission time between components of the computer, it served as an excellent base for classic computing for decades. Modern computer components and architectures, however, require to consider their temporal behavior: data transmission time in contemporary systems may be higher than their processing time. Using the classic paradigm leaves a growing number of issues unexplained, from enormously high power consumption to days-long training of artificial neural networks to failures of some cutting-edge supercomputer projects. The paper introduces the up to now missing temporal behavior (a temporal logic) into computing, while keeps the solid computing science base. The careful analysis discovers that with considering the temporal behavior of components and architectural principles, the mystic issues have a trivial explanation. Some classic design principles must be revised, and the temporal logic enables us to design a more powerful and efficient computing.

ECM Factorization with QRT Maps

Andrew N. W. Hone
School of Mathematics, Statistics & Actuarial Science, University of Kent, Canterbury, Kent, UK

Abstract: Quispel-Roberts-Thompson (QRT) maps are a family of birational maps of the plane which provide the simplest discrete analogue of an integrable Hamiltonian system, and are associated with elliptic fibrations in terms of biquadratic curves. Each generic orbit of a QRT map corresponds to a sequence of points on an elliptic curve. In this preliminary study, we explore versions of the elliptic curve method (ECM) for integer factorization based on performing scalar multiplication of a point on an elliptic curve by iterating three different QRT maps with particular initial data. Pseudorandom number generation and other possible applications are briefly discussed.

Evaluation of Classical Data Structures in the Java Collections Framework

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

Abstract: The performance of software applications for computer networking, Web services and cloud computing, with respect to speed, scalability, fault tolerance and quality of service is critical. Designing software, involves the appropriate choice of data structures, because their performance with respect to space (memory utilization for data storage) and time (execution speed of operations) play significant roles in determining the performance of the software. This paper evaluates the performance of Java Collections implementation of classical data structures and how it scales with increase in data size. The effects of memory allocation, heap memory size and garbage collection on performance are discussed. Also, this paper identifies areas for improvement in the implementation of the classical data structures and discusses their asymptotic analysis.

The 16th International Conference on Frontiers in Education: Computer Science & Computer Engineering (FECS 2020) https://americancse.org/events/csce2020/conferences/fecs20

An Algorithm for Determining if a BST Node's Value can be Changed in Place

Daniel S. Spiegel
Kutztown University, Kutztown, USA

Abstract: A binary search tree (BST) is a strictly ordered data structure that permits quick storage and access of its data. Changing a value in a node of a BST requires ascertaining that such an update will not violate the tree's order; otherwise the update requires replacing the source node with a new node (in a different location). In this work, an algorithm for determining whether a value in a BST can be updated within an existing node will be developed. It provides experience in algorithmic development for students and may lead to improvements in more complex tree structures.

Class Time of Day: Impact on Academic Performance

Suzanne C. Wagner, Sheryl J. Garippo, Petter Lovaas
Niagara University, New York, USA

Abstract: College faculty and students often believe that student performance can be impacted by the time of day that a course is offered and may schedule their courses based on personal preference. This paper examines the relationship between the time of day for course offering and academic success, controlling for course characteristics and instructor variability. Data from a medium sized private university was studied for a course that was taught by three professors over a period of 13 years. The content of the course was consistent and the assignments and examinations followed a standard format. Results show that the time of day did not impact the academic performance of students when the instructor variable was controlled.

An Educational Guide to Creating Your Own Cryptocurrency

Paul Medeiros, Leonidas Deligiannidis
Department of Computer Science & Networking, Wentworth Institute of Technology, Boston, Massachusetts, USA

Abstract: This work presents a comprehensive guide for students and developers who are interested in developing and experimenting with a blockchain technology for cryptocurrencies. Our work allows students to rapidly deploy a new cryptocurrency and provides an accessible way for users to process transactions and mine blocks. A new blockchain can be created using the source code of Litecoin alongside with the Microsoft Azure's cloud service to deploy the requirements necessary for running the network associated with the blockchain. As Litecoin, we utilize Scrypt as the "Proof-of-Work" algorithm. Built into the codebase for Litecoin, this computationally heavy algorithm is designed to ward off large-scale brute-force attacks, preventing criminals with strong hardware capabilities from bypassing the security measures of the algorithm. This article dives into the intricacies of the steps taken to deploy a new blockchain designed to be used for cryptocurrency.

Incorporating Computer Programming into Mathematics Curricula to Enhance Learning for Low Performing, Underserved Students

Alan Shaw, Bill Crombie
Computer Science, Kennesaw State University, Georgia, USA;
The Algebra Project, Cambridge, MA, USA

Abstract: In this short research paper, we describe the outcome of our integration of computer simulations and elementary visual programming interactions with an experiential mathematics curriculum that has resulted in a system that allows students to explore conceptual and procedural knowledge within mathematics content through the interactive and exploratory nature of a particular type of simulation using mobile applications. The simulations are managed by the students using a visual programming language that we developed as a stripped-down version of MIT App Inventor. In particular, our project focused on students who were in the lower quartile of mathematics achievement, and we show how useful they found this approach in helping them better engage and explore abstract mathematics content through reflective engagement with the material. In this report, we detail the 3-dimensional approach we used to adapt an existing mathematics curricular process to a process that integrates visual programming and simulations as a type of applied mathematics. We also show how the resulting pedagogical process has the potential for helping students move beyond simply mechanically and mindlessly manipulating math symbols when solving linear equations, to applying visual reasoning and representational logic as the basis for problem solving in mathematics.

Towards Equitable Hiring Practices for Engineering Education Institutions: An Individual based Simulation Model

Marcia R. Friesen, Robert D. McLeod
Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada

Abstract: This work presents a modeling & simulation study of academic hiring policies to address equity, diversity and inclusion (EDI) in an engineering program. An agent-based modelling approach is used to investigate the comparative impacts of various EDI hiring interventions on outputs including the time associated with achieving target representation of under-represented groups, average value ('qualification score') of new hires, and number of positions that the best qualified over-represented group applicant applies for before being hired. The simulation results demonstrate that the time constants for cultural change are long even with proposals that may be considered radical. Also, the simulation results do not support a common argument that EDI initiatives will sacrifice excellence in faculty hiring.

Should the Smathphone be Considered an ICT?

Francisco Casanova-Del-Angel
Instituto Politecnico Nacional, Mexico

Abstract: This article looks at a didactic experience at undergraduate level in civil engineering; specifically, the teaching of a mathematics course through smartphones. The objective was to teach a course of an engineering degree primarily using this technology, in order to promote among the students their skills and competencies to develop and present a theme. This adaptation process in teaching practice is not easy and may not be achieved in a short period of time. The principal idea to highlight is that planning activities with new technologies cannot be carried out spontaneously and must be based on an educational model. Another important thing in this type of courses is the assessment of the teacher, priority of education administrators.

Design for Empathy and Accessibilty: A Technology Solution for Deaf Curling Athletes

Marcia R. Friesen, Ryan Dion, Robert D. McLeod
Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada

Abstract: This work presents a redeveloped introductory course in Digital Systems Design, often considered a course on embedded systems and IoT in an undergraduate computer engineering program at a Canadian university. The course was moved from a theory-driven course to a team project-based course built on a framework of sustainable design, design for accessibility, and equity principles for deaf curling athletes. Students were guided through a modified design cycle, from conceptual design to a functional prototype and physical prototype. In addition, the course included a strong emphasis on allied topics for knowledge transfer, including business development, technology marketing, intellectual property protection, and moving toward commercialization.

Tracking Changing Perceptions of Students through a Cyber Ethics Course on Artificial Intelligence

Zeenath Reza Khan, Swathi Venugopal, Farhad Oroumchian
University of Wollongong in Dubai, Dubai, United Arab Emirates, UAE;
Love That Design, Dubai, United Arab Emirates, UAE

Abstract: The advent of technological evolution, particularly in artificial intelligence has increased the responsibility of academics towards students today. It is important to understand how students view artificial intelligence, its development and use to provide ethical framework to help develop an understanding of standards, codes of conduct, right and wrong. This study records changing student choices and opinions on artificial intelligence during a capstone ethics course topic taught to tertiary students across 10 years and how introducing them to moral principles and ethical guidelines changes how they accept or question artificial intelligence. It is important to note that the study finds that students' perceptions and opinions have indeed been changing over the years, with newer generations finding the concept of artificial moral agents as viable, impressive, and exciting. The study also records how students may not grasp the full implications of artificial moral agents and that a stand-alone ethics course that focuses on learning objectives that include ethical theories and framework and runs the due course of a complete semester can help to ensure students are asking the right questions, understanding their legal, social, economic and moral responsibilities, thus emphasizing the importance of including such topics and subjects to computer and engineering degrees.

An Investigation of the Use of Whatsapp Groups as Mobile Learning System to Improve Undergraduate Performance

Rushane Jones, Sherrene Bogle
School of Computing, University of Technology, Jamaica, Kingston, Jamaica;
Humboldt State University, California, USA

Abstract: This study investigates the use of Whatsapp as a Mobile Learning System on junior and sophomore students doing multimedia-based courses. The study uses a conceptual model based on the second thesis of Anderson's Interactivity Equivalency Equation to determine the impact the use of the technology has on student learning outcomes. Results showed that the most noticeable factor of interaction was the student to teacher construct. Due to their interaction on the Whatsapp group with their peers, content and teacher, they achieved higher final results. Significant factors were the value placed on real-time chatting with peers, practical assessment on overall learning and the reading of instructor posts. 75% of the control group had GPA less than 2.70, which suggest that the lack of the additional interaction may have been a contributing factor to lower GPAs than the experimental group which had a mean GPA of 3.71.

Innovative Methods of Teaching the Basic Control Course

L. Keviczky, T. Vamos, A. Benedek, R. Bars, J. Hetthessy, Cs. Banyasz, D. Sik
Hungarian Academy of Sciences, SZTAKI, Budapest, Hungary;
Budapest University of Technology and Economics, Budapest, Hungary

Abstract: System view, understanding systems and how they are controlled is an important discipline in engineering education. Nowadays considering the ever increasing knowledge, the explosion of information, the available visual technics and software tools and the requirement for on-line distance education there is a need to revisit the content and the teaching methodology of the basic control course. Here we present our experience in renewing the basic control course. The topics of the course are given. The lectures are available electronically and can be used also in on-line teaching. The main ideas are explained on two levels: hopefully in an understandable way for everyone, and precisely, using mathematical tools. In the lectures some parts of the multilevel e-book, SYSBOOK are referred, which has been developed to present the main principles on different levels, for everyone, for the students and partly for researchers. Besides static teaching materials some interactive demonstrations can also be used in the lectures contributing to the enjoyment of the learning process. During computer laboratory exercises using MATLAB/SIMULINK software the students apply the analysis and synthesis methods discussed in the lectures. In the content of the control course a new feature is the emphasis of the YOULA parameterization method for controller design showing that other methods can be considered as its special cases. Nowadays in education a new teaching . learning paradigm is Open Content Development (OCD) which means active participation of the teachers and students creating an up-to-date teaching material. Utilizing the experiences of these pilot efforts the SYSBOOK platform has been connected to the OCD model providing the possibility for the students to develop their own control case studies. Besides it is also important to execute real-time experiments in laboratory work or using distant laboratories.

Parent Teacher Portal (PTP): A Communication Tool

Mudasser F. Wyne, Matthew Hunter, Joshua Moran, Babita Patil
School of Professional Studies, National University, San Diego, California, USA

Abstract: One of the hardest tasks facing elementary school teachers today is working with the parents of their students to make the learning experience the best it can be for the student. Most teachers and parents recognize the importance of effective parent-teacher communication. There are various ways that parents and teachers can communicate with each other, rather than relying on the scheduled parent-teacher meetings. Phone calls and visits to the classroom are also good ways to cooperate with teachers and keep informed about child's progress. Good two-way communication between parents and teachers is necessary for student's success. Having an effective parent-teacher communication helps the children do better, both socially and academically. The goal of the Parent Teacher Portal (PTP) application is to help streamline communication between elementary school teachers and the parents of their students. The PTP will allow direct communication so that any issues whether they are behavioral, learning etc. can be addressed as soon as possible. The PTP application will allow teachers to communicate with the parents of students in their classes as a whole or individually. Also, with the PTP application teachers will have the ability to create events and notify parents regarding school wide announcements.

Using Dear Data Project to Introduce Data Literacy, and Information Literacy to Undergraduates

Vetria L. Byrd
Purdue University, West Lafayette, Indiana, USA

Abstract: This paper describes an approach for introducing data and information literacy to undergraduates. This work aims to address the following research question, "What impact does the Dear Data approach have on data literacy and information literacy skills of undergraduates in an introductory data visualization course?" The initial aim of the project was to create an assignment to reinforce concepts covered in class and motivate students to think differently about data and how data is communicated. Elements of the Dear Data Project were adapted as part of a four-week assignment that paired students from two sections of the course as "Data Buddies." Dear Data Postcards were created by students with themes that aligned with data

literacy and information literacy concepts. Postcards served as a hands-on exercise for reinforcing course topics. Student self-assessment results show students found the assignment helped to broaden students' ideas of what data is or could be, helped students see data in a variety of ways that were not previously considered and helped students think critically about how to visually communicate ideas, and think of their audience when creating visualizations. This work contributes to the pedagogy of data visualization and to the knowledge base of data visualization capacity building.

Examining the Influence of Participating in a Cyber Defense Track on Students' Cybersecurity Knowledge, Awareness, and Career Choices

Michelle Peters, T. Andrew Yang, Wei Wei, Kewei Sha, Sadegh Davari
University of Houston - Clear Lake, Houston, Texas, USA

Abstract: Colleges and universities across the nation are being called upon to modernize computing programs in order to produce well-rounded professionals that are aware of and capable of mitigating the ever-increasing cyber-attacks from foreign and domestic hackers. The purpose of this mixed methods case study was to examine the influence of participating in a Cyber Defense Track (CDT) on stu-dents. cybersecurity knowledge, awareness, and career choices. Over the course of three semesters, a purposeful sample of undergraduate students enrolled in the CDT were administered pre/post surveys/as-sessments, and participated in focus groups. The findings indicated an increased level of: (a) cybersecurity knowledge, (b) awareness of cybersecurity related practices, and (c) interest in a cybersecurity re-lated career field.

A Framework for Computerization of Punjab Technical Education System for Financial Assistance to Under-represented Students

Harinder Pal Singh, Harpreet Singh
Department of Electronics and Communication Engineering, Desh Bhagat University, India;
Mandi Gobindgarh, Department of Technical Education Government of Punjab, Chandigarh, India;
Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan, USA

Abstract: Indian technical education system consists of three levels: Industrial Training Institutes (ITI), Polytechnic Colleges (Diploma) and Engineering Colleges (Degree). Also for post graduate studies Masters and PhD programs are available at University level. There are large number of under-represented students on the basis of caste, income and communities they belong. These students are not academically very sound and have average grades. Therefore government is funding their engineering education through various financial assistance schemes. Such students look for courses of their choice and for funding their education interests. Educational institutions develop solutions to handle such large amount of student data ranging from enrollments, scholarships, academic content, literature search, testing, results and finally placements of these students. There was a need for developing framework for computerization with a view to manage huge amounts of student data which is multiplying every year in the state and country as a whole. Such data require lot of integrity and scalability, memory saving schema for hardware and software platforms for the needy institutions along with lot of mobility. In this paper a new framework for computerization of Punjab Technical Education has been suggested using modern day data base technologies rather than conventional ones. Results show that the new framework is easy to use through mobile phones, tablets and laptops and also lot of memory have been saved along with demonstration of data integrity, security, and improvement of response time, scalability and also improving authorized accessibility anywhere.

Integrating the Development of Professional Skills throughout an ICT Curriculum Improves a Graduate's Competency

Nicole Herbert, David Herbert, Erik Wapstra, Kristy De Salas, Tina Acuua
University of Tasmania, Hobart, Tasmania, Australia

Abstract: The necessity to design a curriculum that improved the professional competency of graduates was a response to the growing concern that information and communication technology (ICT) graduates, while being strong in technical ICT skills,

were weak in their professional skills, particularly in relation to communication and collaboration skills. To enhance the employability of graduates an ICT curriculum was designed that integrated the development of professional skills, including communication, collaboration, creativity and critical thinking, alongside the development of ICT technical skills. This paper reports on a longitudinal study from 2012 to 2018 that provided strong evidence that integrating professional skill development across an entire ICT curriculum significantly improved a graduate's competency with professional skills without having a detrimental impact on their competency with ICT technical skills. This integrated curriculum can act as an exemplar to other ICT curriculum designers that the approach can lead to professionally competent employable ICT graduates.

Empirical Analysis of Strategies Employed within an ICT Curriculum to Increase the Quantity of Graduates

Nicole Herbert, Erik Wapstra, David Herbert, Kristy De Salas, Tina Acuua
University of Tasmania, Hobart, Tasmania, Australia

Abstract: There is an increasing demand for information and communication technology (ICT) graduates to sustain the growth of the rapidly evolving ICT industry. This demand for ICT graduates challenges higher education to be more effective with ICT curriculum design. The purpose of this study is to apply various strategies to amend student misconceptions, and improve student perceptions, motivation, engagement, and academic success within an ICT curriculum with the intent to increase the number of ICT graduates without reducing graduate competency. This empirical analysis using data collected over a significant time period has evaluated the collective changes to course commencement and attrition rates and found there was significant evidence of improvement.

Peer Assistant Role Models in a Graduate Computer Science Course

Evava Pietri, Leslie Ashburn-Nardo, Snehasis Mukhopadhyay
Department of Psychology, Indiana University Purdue University Indianapolis (IUPUI), Indiana, USA;
Department of Computer & Information Science, Indiana University Purdue University Indianapolis (IUPUI), Indiana, USA

Abstract: This paper reports the development and testing of a new intervention to recruit terminal Master's students into computer science (CS) PhD programs and to enhance diversity in academic CS. We introduced peer assistants (i.e., successful PhD students) to CSCI 549: Intelligent Systems, a popular course for Master's students with approximately 40-50% women and 90% international students. Although there is a fair amount of diversity in this course and in the CS Master's program generally (i.e., international female students), many of these students do not continue to earn a PhD, or get involved in research during the Master's program. Because increasing the diversity of CS professors is imperative for enhancing the diversity of CS majors and the CS workforce, it is critical to test new methods to recruit Master's students from underrepresented groups into academic CS. To address this need, we introduced PhD student peer assistants into Intelligent Systems to not only help the Master's students with in-class research projects, but to also act as role models to promote Master's students' interest in CS research and PhD programs. Thus, this paper suggests a new and innovative technique for an enhancing diversity in academic CS and STEM generally.

Computational Thinking and Flipped Classroom Model for Upper Division Computer Science Majors

Antonio-Angel L. Medel, Anthony C. Bianchi, Alberto C. Cruz
Department of Computer & Electrical Engineering and Computer Science,
California State University, Bakersfield, California, USA

Abstract: Graduation rates are generally low among undergraduate Computer Science (CS) majors, so we are motivated to employ flipped classroom model at CSUB to improve four-year graduation rates. There is a plethora of literature supporting improved information retention using the flipped classroom model. However, its impact on upper-division core CS courses has been understudied. An active learning environment using Just-in-Time-Teaching (JiTT), pair instruction and computational thinking has been implemented in this study. This flipped classroom model was applied to two upper-division CS courses,

Computer Architecture II and Artificial Intelligence. A year-long study measured the impact on performance through surveys, performance on group activities, and class performance such as test scores, with comparison to a control population (non-flipped classroom class) normalized for a specific instructor. An increased student performance and ultimately better graduation rates among the 40 flipped classroom model students is anticipated.

A Project-Based Approach to Teaching IoT

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

Abstract: Internet of Things (IoT) is a rapidly growing field of great interest to students and educators. However, designing and teaching a project-based IoT course can be challenging. While some textbooks and courses are already available in IoT, limited information is available on how to successfully pair the IoT curriculum with hardware and hands-on projects. More work and open discussion is needed to develop courses which successfully and effectively combine hardware and software. In the present work, we summarize how we designed and taught a graduate level IoT course at California State University, East Bay (CSUEB) by effectively pairing lectures with hands-on labs and projects. The course provided a broad introduction to IoT topics, software and hardware through the use of hands-on projects and low-cost hardware kits ($544 each). By strategically combining lectures with hands-on practice, we successfully introduced computer science students to key IoT and hardware skills such as assembling devices, debugging issues, reviewing specs and data, and optimizing performance. We also share the equipment and software packages used, topics covered, labs and projects given, and lessons learned. This course was favorably reviewed by students, led to several successful capstone projects, and even helped one student secure an internship in IoT. We anticipate that these project-based approaches to teaching and learning IoT will be beneficial to many students and educators who are interested in the IoT field.

Team-Based Online Multidisciplinary Education on Big Data + High-Performance Computing + Atmospheric Sciences

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

Abstract: Given the context of many institutions moving to online instruction due to the COVID-19 pandemic in 2020, we share our experiences of an online team-based multidisciplinary education program on big data + high performance computing (HPC) + atmospheric sciences (cybertraining.umbc.edu). This program focuses on how to apply big data and high-performance computing techniques to atmospheric sciences. The program uses both an instructional phase with lectures and team-based homework in all three areas and a multi-disciplinary research experience culminating in a technical report and oral presentation. The paper discusses how our online education program can achieve the same learning objectives as face-to-face instruction via pedagogy and communication methods including flipped classroom, online synchronous meetings and online asynchronous discussion forum.

Benchmarking the Software Engineering Undergraduate Program Curriculum at the Jordan University of Science and Technology with the IEEE Software Engineering Body of Knowledge (SWEBOK-V3.0)

Moh'd A. Radaideh
Department of Software Engineering, Faculty of Computer & IT, Jordan University of Science & Technology, Jordan

Abstract: This paper evaluates the compliance of the Software Engineering Program (SWE-Curriculum) at Jordan University of Science and Technology (JUST) with the first five of the fifteen Software Engineering Knowledge Areas (SWE-KAs #6-10) of the SWEBOK-V3.0 of the IEEE Computer Society. This research is the first to measure the coverage of the SWE-KAs in any SWE-Curriculum. It is essential to line up the said SWE-Curriculum with the IEEE view of Software Engineering (SWEBOK-

V3.0), in addition to its IET accreditation. This research was divided into three parts. The first part (P#1) focused on SWE-KAs#1-5, this second part (P#2) is focusing on SWE-KAs#6-10, and third part (P#3) will focus on SWE-KA#11-15. This paper inspected the SWE-KAs#6-10's coverage across the said SWE-Curriculum's courses. The results were identified as either Fully-Compliant, Highly-Compliant, Partially-Compliant, or Poorly-Compliant (e.g. the concerned SWE-KA is either fully, highly, partially, or poorly covered across one or more of the said SWE-Curriculum courses). This research found the compliance as Fully-Compliant in the cases of the SWE-KAs of the SWE Management, and Software Quality. While found Partially-Compliant in the cases of the SWE-KAs of the Software Configuration Management, SWE Process, and SWE Models and Methods.

Developing a Scalable Platform and Analytics Dashboard for Manual Physical Therapy Practices Using Pressure Sensing Fabric

Tyler V. Rimaldi, Daniel R. Grossmann, Donald R. Schwartz
School of Computer Science and Mathematics, Marist College, Poughkeepsie, New York, USA

Abstract: Current manual therapy pedagogical tools do not enable instructors to objectively assess the precision of hand movements. Methods for capturing the pressure applied to specific regions of the human body are lacking. Instructors of applied manual therapy techniques will benefit from a tool that streamlines their teaching process, thereby enabling their students to be trained accurately and precisely through comparisons of their learned techniques to the instructor's mastered techniques. This project seeks to accomplish this by providing manual therapy instructors a scalable research platform that models instructor and student manual therapy data provided by a Studio 1 Labs pressure sensing fabric. The combination of the pressure sensing fabric and real-time data visualizations will enable instructors to provide immediate feedback to students to improve the quality of the therapy they provide. This paper will show the evolution of this physical therapy research platform, its development life cycle, current state, plans for future research and development, and a potential implementation of the tool in academic institutions.

Predicting the Academic Performance of Undergraduate Computer Science Students Using Data Mining

Faiza Khan, Gary M. Weiss, Daniel G. Leeds
Department of Computer and Information Sciences, Fordham University, Bronx, New York, USA

Abstract: There are myriad factors which can affect a student's academic performance as measured by Grade Point Average (GPA). Identifying characteristics of students with high GPA can help more students understand how to achieve the best grades possible. In this paper, a variety of data mining algorithms are used to predict the GPA of undergraduate students majoring in Computer Science based on survey questions. The results demonstrate that the number of hours of sleep per night, the frequency of illicit drug use, the number of hours spent studying per week, and the number of hours spent on social media platforms per week are important factors that can be used to classify student GPA. The Random Forest data mining algorithm performed the best and was able to achieve a predictive accuracy of 95% when placing students into one of four academic performance groupings.

An Educational Tool for Exploring the Pumping Lemma Property of Regular Languages

Josue N. Rivera, Haiping Xu
CIS Department, University of Massachusetts Dartmouth, Dartmouth, Massachusetts, USA

Abstract: Pumping lemma has been a very difficult topic for students to understand in a theoretical computer science course due to a lack of tool support. In this paper, we present an active learning tool called MInimum PUmping length (MIPU) educational software to explore the pumping lemma property for regular languages. For a given regular language, MIPU offers three major functionalities: determining the membership of an input string, generating a list of short strings that belong to the language, and automatically calculating the minimal pumping length of the language. The software tool has been developed to provide educational assistance to students to better understand the concepts of pumping lemma and minimum pumping length, and promote active learning through hand-on practice.

Preparing Computing Graduates for the Workplace: An Assessment of Relevance of Curricula to Industry

Ioana Chan Mow, Elisapeta Mauai, Vaisualua Okesene, Ioana Sinclair
Computing Department, Faculty of Science, National University of Samoa, Samoa

Abstract: This paper documents research which investigated how relevant the content of the Computing courses offered within undergraduate programmes of the Computing department at the National University of Samoa (NUS) were to meet the needs of industry and the workforce. The study, which was conducted in 2018 to 2019, surveyed 16 institutions and 20 graduates from the Computing programs. Findings from the survey indicated that the current course offerings within the Computing department are to a large extent, relevant to the needs of industry and the workplace. Most of the recommendations from the current survey were similar to those in the 2008 and 2013 survey and had already been incorporated into the current curriculum. The main recommendations for improving the curriculum are an improved emphasis on operating systems, cybersecurity, strategic planning, problem solving skills, and introduction of other database platforms such as Oracle as well as more web-based languages. Recommendations for improvement of the survey are i) increasing the scope to include an evaluation of the postgraduate Diploma ii) the inclusion in future surveys of questions on soft skills, iii) as well as specific consideration of the graduate profile and learning outcomes.

A Dynamic Teaching Learning Methodology Enabling Fresh Graduates Starting Career at Mid-Level

Abubokor Hanip, Mohammad Shahadat Hossain

Founder and CEO, PeopleNTech Institute of Information Technology, Inc.,USA;
Department of CSE, University of Chittagong, Chittagong, Bangladesh

Abstract: Despite a robust recent U.S. job market, new IT graduates tend to be long on theoretical knowledge yet very short on practical mastery of actual skills and knowledge needed to meet typical IT job Despite a robust recent U.S. job market, new IT graduates tend to be long on theoretical knowledge yet very short on practical mastery of actual skills and knowledge needed to meet typical IT job requirements. Thus, graduates are increasingly facing the problem of either being unable to secure full-time IT employment at all - or, if they do land a first job, it is merely a low-paying entry-level position. Typically, such newly minted graduates become frustrated and switch jobs within six to twelve months to secure a higher salary. This, in turn, causes the initial employer to lose money and time, essentially having to start all over with a new entry-level hiree. Consequently, companies are increasingly refusing to hire entry-level graduates and instead are requiring significant industry experience. Accordingly, this paper presents an innovative solution for students, universities, and technical schools alike: a unique educational model that actually provides students with suchcient practical mastery to qualify them for mid-level IT positions immediately following graduation. As the illustration below shows, and as any corporate hiring manager will readily admit, a successful IT job applicant needs to exude competence in a full range of areas in order to maximize the chances of securing a mid-level, higher-paying position. Therefore, it only makes logical sense for the educational institutions to explicitly address all of the same knowledge and skill sets as an intrinsic part of the educational experience. In fact, there is an educational institution that has been successfully applying this innovative practical-skills mastery model over the last 15 years for IT education. PeopleNTech has placed virtually all of its students in first jobs" at mid-level and seniorlevel IT positions which ordinarily require years of industry experience in order to secure.

The 16th International Conference on Grid, Cloud, & Cluster Computing (GCC 2020)
https://americancse.org/events/csce2020/conferences/gcc20

Performance Analysis of Remote Desktop Session Host with Video Playback Scenarios

Baikjun Choi, Sooyong Park
Sogang University, Seoul, Korea

Abstract: In many places, Desktop-as-a-Service under the Windows environment has been provided using Virtual Desktop Infrastructure (VDI) and Remote Desktop Session Host (RDSH). A number of studies have been conducted on analysis of sole performance of RDP or hypervisor performance, few studies have been conducted on performance analysis when a number of RDSH are running. The performance analysis of RDSH published by Microsoft is not suitable to estimate the acceptable number of users in servers running in the current use environment where videos are frequently used because RDSH employ the models that exclude video related tasks. This study aims to analyze the performance including video playback scenarios in the RDSH environment and estimate the acceptable number of servers through the performance analysis results.

Mining-RNA: WEB based System using e-Science for Transcriptomic Data Mining

Carlos Renan Moreira, Christina Pacheco, Marcos Vinicius Pereira Diogenes, Pedro Victor Morais Batista,
Pedro Fernandes Ribeiro Neto, Adriano Gomes Da Silva, Stela Mirla da Silva Felipe,
Vania Marilande Ceccatto, Raquel Martins de Freitas, Thalia Katiane Sampaio Gurgel,
Exlley Clemente dos Santos, Cythia Moreira Maia, Thiago Alefy Almeida e Sousa, Cicilia Raquel Maia Leite
UERN, Rua Almino Afonso, Centro - Mossoro/RN – Brazil;
UECE, Av. Dr. Silas Munguba, Itaperi - Fortaleza/CE – Brazil

Abstract: High throughput gene expression studies yielded a great number of large data sets and these are freely available in biological databases. Re-analyzing these studies individually or in clusters can produce new results relevant to the scientific community. The purpose of this work is to develop a WEB system based on the e-Science paradigm. The System should read massive amounts of data from the Gene Expression Omnibus database (GEO), pre-process, mine, and display it in a userfriendly interface. Thus, it is intended to mitigate the difficulty in interpreting data from transcriptomic studies made using the DNA microarray technique. Also presented will be the preliminary results obtained from the initial stages of development, as well as the proposed architecture for the system.

Secure Business Intelligence

Aspen Olmsted
Department of Computer Science, Fisher College, Boston, MA, USA

Abstract: Enterprise organizations have relied on correct data in business intelligence visualization and analytics for years. Before the adoption of the cloud, most data visualizations were executed and displayed inside enterprise applications. As application architectures have moved to the cloud, many cloud services now provide business intelligence functionality. The services are delivered in a way that is more accessible for end-users using web browsers, mobile devices, data feeds, and email attachments. Unfortunately, along with all the benefits of the cloud business intelligence services comes complexity. The complexity can lead to slow response times, errors, data leakage, and integrity issues. An information technology department or service provider must get ahead of the problems by automating the execution of reports to know when availability or integrity issues exist and dealing with those issues before they turn into end-user trouble tickets. The development of the business intelligence code must also include tools to express the privacy requirements of the data exposed in the report or document. In this paper, we present two tools we developed to help guarantee the confidentiality, integrity, and availability of business

intelligence. The first tool is our client-side correctness programming language that allows execution against many cloud documents and business intelligence services. The secBIML language enables issues to be proactively discovered before the end-users experience the problems. The other tool in our work is a server-side programming language that allows the creation of reports and business documents. The secBIrpts language enables an organization to express their privacy requirements utilizing a hierarchical security model.

Selective Compression Method for High-quality DaaS (Desktop as a Service) on Mobile Environments

Baikjun Choi, Sooyong Park
Sogang University, Seoul, Republic of Korea

Abstract: As computing is defined as a concept that various IT devices are able to be used at any time at any place, with the ongoing growth of BYOD(Bring Your Own Device) tending to use DaaS on personal smart devices and company or official work as it is, now DaaS has to be supported smoothly in mobile environment. Hence, the study suggests high quality transmission method to use smoothly on smart devices required in order to improve VDI environment to DaaS. To prove goal achievement, effectiveness results were analyzed through function and scenario tests on 'gradual terminal protocol' technology and other virtualization common use solution protocols.

The SURF System for Continuous Data and Applications Placement Across Clouds

Oded Shmueli, Itai Shaked
Computer Science Department, Technion- Israel Institute of Technlogy, Haifa, Israel

Abstract: In a hybrid cloud environment, as well as in a multi-cloud environment, an enterprise employs a number of local sites (or data centers) and cloud data center(s) that may be geographically distributed. The problem of where to place and replicate data and applications is complicated by multiple dynamically changing conditions. We describe two types of algorithms: data movement (conservative and optimistic of various types), and recovery from various system faults. They may be integrated into various system types. These systems may have their own correctness requirements. The system we provide is charged with creating an illusion that data is stationary. These algorithms are implemented on top of a ZooKeeper compact distributed database. The algorithms were extensively tested over three public clouds.

SURF - Optimized Data Distribution Technology

Oded Shmueli, Itai Shaked
Computer Science Department, Technion- Israel Institute of Technlogy, Haifa, Israel

Abstract: In a hybrid cloud environment an enterprise employs a number of local sites (or data centers) and cloud data center(s) of possibly multiple cloud service providers. SURF is a technology for controlling the distribution of data and applications in such an environment. Distribution decisions are based on (a) enterprise policy, (b) performance characteristics and (c) pricing tables. The technology may be used as the core of the multiple system types. SURF's distributed algorithms, and its data distribution decisions component, were examined via extensive simulations on a hypothetical application and is vastly superior in a changing environment to naive placement methods

The Abaco Platform: A Performance and Scalability Study on the Jetstream Cloud

Christian R. Garcia, Joe Stubbs, Julia Looney, Anagha Jamthe, Mike Packard, Kreshel Nguyen
Texas Advanced Computing Center, University of Texas at Austin, Texas, USA;
Department of Aerospace Engineering, The University of Texas at Austin, Texas, USA

Abstract: Abaco is an open source, distributed cloudcomputing platform based on the Actor Model of Concurrent Computation and Linux containers funded by the National Science Foundation and hosted at the Texas Advanced Computing Center. Abaco recently implemented an autoscaler feature that allows for automatic scaling of an actor's worker pool based on an actor's mailbox queue length. In this paper, we address several research questions related to the performance of the Abaco platform with manual and autoscaler functionality. Performance and stability is tested by systematically studying the aggregate FLOPS and hashrate throughput of Abaco in various scenarios. From testing we establish that Abaco correctly scales to 100 JetStream "m1.medium" instances and achieves over 19 TFLOPS.

Securing Mobile Cloud Computing Using Encrypted Biometric Authentication

Iehab Alrassan
College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

Abstract: Mobile Cloud Computing (MCC) is based on the integration of cloud computing and mobile devices that inherits cloud computing characteristics such as on-demand self-service, broad network access, and measured services. Also, MCC inherited the security threats of cloud computing, like loss of data, exposing the data to a third party, or unauthorized access to resources. Although, there are many researches in the area of protecting data from illegitimate access using traditional encryption techniques, this paper discusses a new methodology for preventing unauthorized access to resources by encrypting user's password along with its biometric identification (fingerprint), and storing them to the cloud. As a result, only authorized users can generate keys for encrypting their data and store them to the cloud. The proposed methodology will protect the identity of the user and keep user data from unauthorized access.

Framework for Monitoring the User's Behavior and Computing the User's Trust

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

Abstract: Traditional access control, simple methods for virus detection, and intrusion detection are unable to manage variety of malicious and network attacks. The number of users might get hacked because of limitation in basic security protection. To implement a secure, reliable, and safe cloud-computing environment, we need to consider the trust issue. A trusted cloud is guaranteed to be safe from user terminals; combined with the concept of a trusted network , it evaluates, forecasts, monitors, and manages the user's behavior to eliminate malicious datacenter attacks which are performed by unwanted cloud users and hackers; as a result, there is improved cloud security. In this paper, we propose a Framework for Monitoring the User's Behavior and Computing the User's trust (FMUBCT). this model is detecting abnormal user behavior by creating a user-behavior history pattern and compare them current user behavior. The outcome of the comparison is sent to a trust computation center to calculate a user trust value. FMUBCT is flexible and scalable as consider more evidence to monitor and evaluate user behavior. Finally, the simulation of FMUBCT shows that the model can effectively evaluate the users.

Enterprise Backend as a Service (EBaaS)

Gokay Saldamli, Aditya Doshatti, Darshil Kapadia, Devashish Nyati, Maulin Bodiwala, Levent Ertaul
Computer Engineering Department, San Jose State University, San Jose, California, USA;
Department of Computer Science, California State University East Bay, Hayward, California, USA

Abstract: In the world where we have computers and world wide web, web applications have become more and more popular. There has been a constant decrease in installed applications with people mostly relying on web applications to get their work done. With constant innovations in the field of computer, we see tons of startups everyday and what better option do they have than reaching to million people with a web application of their own. Talking about web applications we usually have 1) Frontend: what a user can see on their screen while accessing that web application and 2) Backend: what frontend communicates with to process the users' requests. Since the invention of RESTful web services, developers have relied on APIs to which frontend sends request to in order to get an appropriate response. RESTful APIs have become more of a standard in developing the backend and more often than not, they are pretty basic with only queries changing to get data from the database. This paper provides a solution to automate the development of backend and thus doesn't need any expert knowledge other than the knowledge of the underlying database and hence even a non-developer or a developer with no prior experience in developing backend can easily get access to the backend. The solution discussed here will ask user to provide database details and will create the database along with the downloadable code for backend which will be ready to use to interact with the frontend and the database.

The 6th International Conference on Health Informatics & Medical Systems (HIMS 2020)
https://americancse.org/events/csce2020/conferences/hims20

ReSmart: Brain Training Games for Enhancing Cognitive Health

Raymond Jung, Bonggyn Son, Hyseong Park, Sngon Kim, Megawati Wijaya
AKA Intelligence Corp., Seogyo-dong, Mapo-gu, Seoul, Republic of Korea

Abstract: As human beings live longer, the number of people diagnosed with dementia is growing. Many studies have proved that dementia tends to degenerate cognitive abilities. Since dementia patients endure different types of symptoms, it is important to monitor dementia patients individually. Furthermore, senior citizens are generally lack of understanding technology, which brings a low self-motivation to use technologies. To enhance the cognitive abilities of senior citizens, we propose a mobile platform called ReSmart which embeds six distinct levels of the brain training task, based on five cognitive areas to detect different types of individual symptoms. Those brain training tasks are presented in a gamelike format that aims to not lose the elder's motivation for technology use and keeping interest.

Security and Usability Considerations for a mHealth Application for Emergency Medical Services

Abdullah Murad, Benjamin Schooley, Thomas Horan
College of Computer and Information Systems, Umm A-Qura University, Mecca, Saudi Arabia;
College of Engineering and Computing, University of South Carolina, Columbia, South Carolina, USA;
School of Business, University of Redlands, Redlands, California, USA

Abstract: Mobile applications for healthcare must be both secure and usable while operating within the boundaries of a healthcare organizational network to access critical information resources, such as the EHR. However, in an increasingly inter-organizational mobile work environment, organizations are seeking provider-centered apps for continuity of care between and across organizations. Emergency Medical Services (EMS) provides one example where a single transport agency may direct patients to any number of separate health systems within a region. This research reports on the design and live field testing of an inter-organizational mobile app for EMS. Participants included 20 ambulances and 7 hospital emergency departments transmitting 1513 live patient records with multi-media data. Evaluation results were used to present a heuristic for achieving usability and security goals for practitioner-oriented inter-organizational mobile health applications, especially for time- and information-critical contexts.

Semantic Tree Driven Thyroid Ultrasound Report Generation by Voice Input

Lihao Liu, Mei Wang, Weiliang Zhao, Jian Yang, Xueyang Li
School of Computer Science and Technology, Donghua University, Shanghai, P. R. China;
Computing Department, Macquarie University, NSW, Australia

Abstract: The automatic speech recognition has achieved quite good performance in the medical domain in the past several years. However, it is still lacking of enough practical solutions with considering the characteristics of real applications. In this work, we develop an approach to automatically generate semantic-coherent ultrasound reports with voice input. The solution includes key algorithms based on a proposed semantic tree structure. The radiologists do not need to follow the fixed templates. They just need to speak their specific observations for individual patients. We have carried out a set of experiments against a real world thyroid ultrasound dataset with more than 40,000 reports from a reputable hospital in Shanghai, China. The experimental results show that our proposed solution can generate concise and accurate reports.

Internet-of-Things Management of Hospital Beds for Bed-Rest Patients

Kyle Yeh, Chelsea Yeh, Karin Li
Walnut Valley Research Institute, USA;
School of Medicine, University of California, Riverside, USA

Abstract: In this paper, we describe the application of the technologies of the Internet-of-Things (IoT) to the management of hospital beds. Specifically, it seeks to monitor the status of hospital patients assigned by their doctors to bed rest based their medical conditions on sensor data collected by embedded pressure sensors, and provide real-time alerts to the medical staff. The potential for injuries from prescribed bed-rest patients vacating the hospital bed and falling is very serious. The injuries often result in additional complications to the underlying health condition requiring the prescribed bed rest. The proposed IoT bed rest management system will alert the medical staff immediately when a bed-rest patient vacates the bed, and allow them to take immediate remedial action. The research consists of two parts. The first part created IoT-connected pressure sensors and devices used to capture the patients' presence or absence on the bed and send the data to a central server. The second part developed the medical staff alert application that runs on mobile phones and consoles located in the nurses' stations, that receive the information from the server. This system can also monitor the movement of comatose patients, who need to be moved frequently by the medical staff to avoid pressure ulcers. The server can automatically inform the medical staff when a patient is scheduled to be moved, and alert the staff when a scheduled movement is overdue for a set time.

ActiviX: Non-Invasive Solution to Mental Health

Morgan Whittemore, Shawn Toubeau, Zach Griffin, Leon Deligiannidis
Department of Computer Science and Networking, Wentworth Institute of Technology, Boston, Massachusetts, USA

Abstract: ActiviX is an at-home mental health support tool that aids people who suffer from mental health by encouraging them to perform small but meaningful tasks in hope of leading them to a happier life. ActiviX focuses on the self-care and productivity of users by keeping track of what important tasks they complete throughout their day. These tasks are defined by the user and can range from maintaining good hygiene to doing homework or eating three meals a day and staying hydrated. If the system detects that the user is skipping out on these tasks, ActiviX reminds the user to complete them. Additionally, if it detects that the user is doing well and their mood is improving, it will give bits of motivation to encourage them to keep up the good work.

Exergames for Systemic Sclerosis Rehabilitation: a Pilot Study

Federica Ferraro, Marco Trombini, Matteo Morando, Marica Doveri, Gerolamo Bianchi, Silvana Dellepiane
Department of EE, Telecom Engineering & Naval Architecture (DITEN), Universita degli Studi di Genova, Italy;
Azienda Sanitaria Locale 3, Division of Rheumatology, Department of Locomotor System, Genoa, Italy

Abstract: In this paper, a study on the use of ICT support to help therapists and Systemic Sclerosis (SSc) patients in the recovery phase is described. The ReMoVES platform is conceived in the field of assistive computing technologies and delivers engaging exergames to be performed in conjunction with traditional rehabilitation both at home and at clinical centers, thus enabling the continuity of care. The present work refers to the implementation and integration in the ReMoVES platform of standard hand rehabilitative exercises for SSc, that typically involves repetitive movements. Data related to game sessions are collected and analyzed for assessing the patients' conditions.

Classification of Craniosynostosis Images by Vigilant Feature Extraction

Saloni Agarwal, Rami R. Hallac, Ovidiu Daescu, Alex Kane
University of Texas at Dallas, Richardson, Texas, USA;
UT Southwestern Medical Center, Dallas, Texas, USA

Abstract: The development of an objective algorithm to assess craniosynostosis has the potential to facilitate early diagnosis, especially for care providers with limited craniofacial expertise. In this study, we process multiview 2D images of infants with craniosynostosis and healthy controls by computer-based classifiers to identify disease. We develop two multiview image-based classifiers, first based on traditional machine learning (ML) with feature extraction, and the other one based on CNNs. The ML model performs slightly better (accuracy 91.7%) than the CNN model (accuracy 90.6%), likely due to the availability of a small image dataset for model training and superiority of the ML features in differentiation of craniosynostosis subtypes.

Using Persuasive AI Voice Assisted Technologies to Motivate and Encourage Physical Activity

Benjamin Schooley, Dilek Akgun, Prashant Duhoon, Neset Hikmet
Health Information Technology, University of South Carolina, Columbia, South Carolina, USA;
Innovation Think Tank Lab, University of South Carolina, Columbia, South Carolina, USA;
Department of Information Technology, University of South Carolina, Columbia, South Carolina, USA;
Engineering & Computing, University of South Carolina, Columbia, South Carolina, USA

Abstract: Lack of physical activity (PA) is one major contributor to the high prevalence of modern-day chronic diseases. In this study, we design and implement an interactive PA motivation program for older adults (65+) and people with chronic conditions by adapting a commercially available artificial intelligence voice assisted (AIVA) application and activity tracker (Alexa/ Echo

Show/ Fitbit). Customized features are founded on persuasive technology theory and techniques and evidence-based PA programs. A pilot study with four older adults for five days demonstrated moderate to good usability, learnability, satisfaction, and performance; and key persuasive techniques that enhance user experience and goal achievement. The study provides a tested model for PA behavior change using AIVA for feedback, education, and motivational guidance.

Predicting Length of Stay for COPD Patients with Generalized Linear Models and Random Forests

Anna Romanova
Department of Computer Science and Quantitative Methods, Winthrop University, South Carolina

Abstract: In this study we develop a predictive model for the length of stay (LOS) for chronic obstructive pulmonary disease (COPD) patients using administrative, clinical, and operational data from a large teaching hospital in the southeastern United States. To address the issue of a large percentage of missing values for several predictor variables in the clinical data set, we carry out multiple imputation of the missing values with the bootstrapping and Expectation – Maximization (EM) algorithm from Amelia package in R and perform variable selection with the Boruta algorithm. We employ generalized linear models (GLM) and random forests to model the response variable and perform model comparison based on their generalization errors.

Smart Healthcare Monitoring Apps with a Flavor of Systems Engineering

Misagh Faezipour, Miad Faezipour
Department of Engineering Technology, Middle Tennessee State University, Tennessee, USA;
Departments of CSE and Biomedical Engineering, University of Bridgeport, Bridgeport, CT, USA

Abstract: Smart-health has the potential to create a unique platform for monitoring health. Personalized healthcare using smartphones offer seamless solutions, especially for circumstances such as the COVID-19 pandemic where physical distancing is inevitable. This paper investigates the efficiency of general smartphone-based healthcare monitoring applications (apps) from a system dynamics perspective. The ongoing research effort introduces a causal model to investigate the factors and interrelationships that impact the efficiency of smartphone-based healthcare monitoring apps. A careful study suggests that the most important factors of such systems include patient well-being, satisfaction, cost, and performance measure factors (i.e. response speed, accuracy). The proposed idea provides a novel insight of the dynamics of the model to assess the efficiency of various smartphone-based healthcare monitoring apps.

Using Artificial Intelligence for Medical Condition Prediction and Decision-Making For COVID-19 Patients

Mohammad Pourhomayoun, Mahdi Shakibi
Department of Computer Science, California State University, Los Angeles, California, USA

Abstract: Covid-19 pandemic caused by the SARS-CoV-2 has claimed numerous lives around the world. we developed a novel predictive model based on Machine Learning algorithms to predict the mortality risk of patients with COVID-19. In this study, we used documented data of 117,000 patients world-wide with laboratory-confirmed COVID-19. This study proposes a predictive model to help hospitals and medical facilities decide who has higher priority to be hospitalized, and triage patients when the system is overwhelmed by overcrowding. The results demonstrate 93% overall accuracy in predicting the mortality rate. We used a number of machine learning algorithms including Artificial Neural Networks, Support Vector Machine (SVM), and Random Forest to predict the mortality rate in patients with COVID-19. In this study, the most alarming symptoms and features were also identified.

Predicting Seizure-Like Activity Using Sensors from Smart Glasses

Sarah Hadipour, Ala Tokhmpash, Bahram Shafai, Carey Rappaport
Northeastern University, Boston, MA, USA

Abstract: In this paper we study the use smart glasses in classifying simulated epileptic seizure signals. We train a patient specific classifier using features extracted from an inertial measurement unit signals. For performance evaluation, we use the accuracy as well as the loss function values and Root-Mean-Square-Error (RMSE). Long short-term memory (LSTM) neural network is used on the data collected from the smart glasses.

Epileptic iEEG Signal Classiffication using Pre-Trained Networks

Sarah Hadipour, Ala Tokhmpash, Bahram Shafai, Carey Rappaport
Northeastern University, Boston, MA, USA

Abstract: This paper describes the use of pre-trained model classifiers in detecting epileptic seizures. A patient specific classifier was trained using features extracted from the intracranial electroencephalogram (iEEG) signals. Both accuracy and the loss function are used for performance evaluation. In order to use the pre-trained models that are commonly used for image classification the time series iEEG signal was converted into an spectrogram image associated with that signal. The model was then trained and our results are shown below.

Seizure Prediction and Heart Rate Oscillations Classiffication in Partial Epilepsy

Sarah Hadipour, Ala Tokhmpash, Bahram Shafai, Carey Rappaport
Northeastern University, Boston, MA, USA

Abstract: This paper constitutes a first step towards a wearable epileptic seizure prediction device. Since recording electrocardiogram can be accomplished fairly easily, we look into the existing correlation between epileptic pre-ictal states and heart rate variability. The intervals of extreme noise may corrupt the electrocardiogram data during the seizures, this means that we are able to use a machine learning and specifically deep learning techniques to detect the pre-ictal aka pre seizure states. The experimental results show particularly good results in terms of prediction performance. They also show the importance of a specific training for each patient. In this study we analyzed the cardiac dynamics in patients with partial epilepsy. By doing so, we discovered transient but prominent low-frequency heart rate oscillations immediately following seizures in some patients. These features have been used to for understanding cardiac and neuro-autonomic instability in epilepsy and also for classifications of such heart rates.

A Comparative Study of Machine Learning Models for Tabular Data Through Challenge of Monitoring Parkinson's Disease Progression Using Voice Recordings

Mohammadreza Iman, Amy Giuntini, Hamid Reza Arabnia, Khaled Rasheed
Department of Computer Science, Franklin College of Arts and Sciences, University of Georgia, Georgia, USA;
Institute for Artificial Intelligence, Franklin College of Arts and Sciences, University of Georgia, Georgia, USA

Abstract: People with Parkinson's disease must be regularly monitored by their physician to observe how the disease is progressing and potentially adjust treatment plans to mitigate the symptoms. Monitoring the progression of the disease through a voice recording captured by the patient at their own home can make the process faster and less stressful. Using a dataset of voice recordings of 42 people with early-stage Parkinson's disease over a time span of 6 months, we applied multiple machine learning techniques to find a correlation between the voice recording and the patient's motor UPDRS score. We approached this problem using a multitude of both regression and classification techniques. Much of this paper is dedicated to mapping the voice data to motor UPDRS scores using regression techniques in order to obtain a more precise value for unknown instances. Through this comparative study of variant machine learning methods, we realized some old machine learning methods like trees outperform cutting edge deep learning models on numerous tabular datasets.

'SasCsvToolkit' - A Versatile Parallel 'Bag-Of-Tasks' Job Submission Application on Heterogeneous and Homogeneous Platforms for Big Data Analytics such as for Biomedical Informatics

Abhishek Narain Singh
ABioTek www.tinyurl.com/abinarain, Kuopio, Finland

Abstract: The need for big data analysis requires being able to process large data which are being held fine-tuned for usage by corporates. It is only very recently that the need for big data has caught attention for low budget corporate groups and academia who typically do not have money and resources to buy expensive licenses of big data analysis platforms such as SAS. The corporates continue to work on SAS data format largely because of systemic organizational history and that the prior codes have been built on them. The data-providers continue to thus provide data in SAS formats. Acute sudden need has arisen because of this gap of data being in SAS format and the coders not having a SAS expertise or training background as the economic and inertial forces acting of having shaped these two class of people have been different. We analyze the differences and thus the need for SasCsvToolkit which helps to generate a CSV file for a SAS format data so that the data scientist can then make use of his skills in other tools that can process CSVs such as R, SPSS, or even Microsoft Excel. At the same time, it also provides conversion of CSV files to SAS format. Apart from this, a SAS database programmer always struggles in finding the right method to do a database search, exact match, substring match, except condition, filters, unique values, table joins and data mining for which the toolbox also provides template scripts to modify and use from command line. The toolkit has been implemented on SLURM scheduler platform as a `bag-of-tasks` algorithm for parallel and distributed workflow though serial version has also been incorporated. In the age of Big Data where there are way too many file formats and software and analytics environment each having their own semantics to deal with specific file types, SasCsvToolkit will find its functions very handy to a data engineer.

ICT and the Environment: Strategies to Tackle Health and Environmental Challenges in Nigeria

Tochukwu Ikwunne, Lucy Hederman
ADAPT Centre, Trinity College Dublin, Ireland

Abstract: Environmental change has become an important issue in many African countries with much recent discussion focusing on the harmful health and environmental impact resulting from human activities such as fossil fuel, oil spillages, and others, in the Global South. Many efforts are being carried out to deal with the issue of climate change and environmental degradation. One effort is the use of ICTs as an instrument for environmental protection and the sustainable use of natural resources. ICT can be leveraged to tackle harmful environmental change. This paper offers an overview of how ICTs can be used to benefit the environment in Nigeria. Guidance on how to leverage technology for the good of the environment in Nigeria is provided.

Conceptual Design and Prototyping for a Primate Health History Model

Martin Q. Zhao, Elizabeth Maldonado, Terry B. Kensler, Luci A.P. Kohn, Debbie Guatelli-Steinberg, Qian Wang
Department of Computer Science, Mercer University, Macon, Georgia, USA;
Caribbean Primate Research Center, University of Puerto Rico Medical Sciences Campus, San Juan, Puerto Rico;
Department of Biological Sciences, Southern Illinois University Edwardsville, Illinois, USA;
Department of Anthropology, The Ohio State University, Columbus, Ohio, USA;
Department of Biomedical Sciences, Texas A&M University College of Dentistry, Dallas, Texas, USA

Abstract: Primate models are important for understanding human conditions, especially in studies of ageing, pathology, adaptation, and evolution. However, how to integrate data from multiple disciplines and render them compatible with each other for datamining and in-depth study is always challenging. In a long-term project, we have started a collaborative research endeavor to examine the health history of a free-ranging rhesus macaque colony at Cayo Santiago, and build a knowledge model for anthropological and biomedical/translational studies of the effects of environment and genetics on bone development,

aging, and pathologies. This paper discusses the conceptual design as well as the prototyping of this model and related graphical user interfaces, and how these will help future scientific queries and studies.

Visualizing and Analyzing Polynomial Curve Fitting and Forecasting of Covid Trends

Pedro Furtado
DEI/CISUC, University of Coimbra, Portugal

Abstract: While predicting the evolution of Covid-19 infections on a medium and long term is mostly related to epidemiological models and considerations from disciplines of virology, immunity and epidemiology, the short term (few days ahead) evolution of Covid-19 curves can help draw the trend and understand if an outbreak is still in a dangerous exponential evolution or not. In this paper we show how we transformed the data and applied polynomial curve fitting to help on this objective, how we discovered the most appropriate polynomial and the visualizations we got from all the alternatives.

Implementation of a Medical Data Warehouse Framework to Support Decisions

Nedra Amara, Olfa Lamouchi, Said Gattoufi
Institute of Management, University of Tunis, Bardo, Tunisia;
LR-RISC-ENIT (LR-16-ES07), Tunisia

Abstract: Early detection of breast cancer is recognized as the best way to reduce mortality. Mammography is widely used to detect cancer. This paper discusses our underlying efforts to implement a mammography data warehouse to encourage medical and investigative activities. The data warehouse is a reliable combination of information from many sources. We are planning an infra-basic information systems by merging different types of breast imaging information, a decent variety of existing health system, into an advanced data warehouse center. Different types of breast imaging containing demographic data, family history, mammograms, and textual reports will be obtained from salah azaiz cancer institute picture archiving and communication system (PACS) modules, as well as the Radiological Information System (RIS). This research paper proposes a data warehouse analytical framework for exploring and analyzing data related with breast cancer in Tunisia for decision support.

Personalization of Proposed Services in a Sensors-based Remote Care Application

Mirvat Makssoud (Karine Abbas)
Lebanese University, Tripoli, Lebanon

Abstract: Nowadays, the emergence of smart technologies in healthcare domain is revolutionizing all aspects of patients. daily life. Indeed, the continual increase of patients with chronic diseases has revealed two major challenges: improving the patients' living quality, which has been motivated by their growing need to be cared in a family environment, and reducing the costs of care. Remote Patient Monitoring (RPM) at home represents a promising opportunity to face these chal-lenges. It is mainly based on using smart devices such as sensors in order to mon-itor the patient.s status anywhere and at anytime and to detect earlier any critical health situation to trigger different actions accordingly. Based on this context, we are designing a system capable of offering services in order to monitor and assist patients at home. Indeed, these services could actuate different actions according to detected situations. But it is necessary to notice that all patients do not have the same needs and preferences. So the system should be able to cover all char-acteristics that differentiate each patient as well as the devices that are used in their environment.

A Cross-blockchain Approach to Emergency Medical Information

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

Abstract: Accessing a patient's information across data sharing networks is a challenging task. For a client application to request a patient's data, it has to first refer to a centralized repository for demographic information to identify the patient. Then the search will be continued for patients' clinical and medical data that may exist in different centralized data sharing networks. This approach poses a risk on data availability especially in emergency instances because centralized data sources can be

attractive targets for cyber-attack [1] or be a single point of failure. Other problems can be data privacy and security associated with the centralized authority and governance of data [2]. In this paper, we have introduced a cross-blockchain based data search service that avoids centralized data risks. This search service consists of emergency data services that enables first responders to request and receive relevant emergency data across multiple Hyperledger Fabric (HLF) Networks. It also allows first responders' care reports to be sent back to the networks where patient's data is retrieved from. Patients' treatment data is recorded and updated on the ledger. We have implemented our approach by creating two HLF networks consisting of two hospitals and one client application, which enables first-responders to look up a patient's data and add the care report by connecting to these networks.

A Proactive Approach to Combating the Opioid Crisis using Machine Learning Techniques

Ethel Mensah, Musarath J. Rahmathullah, Pooja Kumar, Roozbeh Sadeghian, Siamak Aram
Harrisburg University of Science and Technology, Harrisburg, PA, USA

Abstract: The use of big data analytics tools and Machine Learning techniques in identifying and predicting opioid use disorder is a relatively new and emerging area. Previous studies analyzing trends in the opioid crisis have identified an increasingly expensive and worsening epidemic. Many factors contribute to opioid use, abuse, and addiction. Opioid addiction is a complex disease involving various physiological pathways, as well as environmental factors. There is some evidence to suggest that people with low education levels, high unemployment and poverty levels are at higher risk of opioid abuse. In this paper, we evaluated different conventional Machine Learning models including Support Vector Machines (SVM), Decision Tree and Logistic Regression and advanced algorithms like Gradient Boosting. The models were developed to predict opioid abuse disorder using county-level education, poverty, and unemployment data. In contrast, the results suggest that populations with higher socio-economic status are at greater risk of opioid abuse disorder compared to individuals with lowers. This can be attributed to underlying factors not previously captured increased availability of opioids and resources to acquire them. Identifying which areas and populations are at higher risk of opioid abuse disorder and underlying contributing factors will help inform judicious effective policies, programs, and solutions to tackle the worsening health crisis.

Robotic Process Automation Based Glaucoma Screening System: A Framework

Panaree Chaipayom, Somying Thainimit, Duangrat Gansawat, Hirohiko Kaneko
Department of Electrical Engineering, Kasetsart University, Bangkok, Thailand;
National Electronics and Computer Technology Center, Pathum Thani, Thailand;
Department of Information & Communications Eng., Tokyo Institute of Technology, Japan

Abstract: Robot Process Automation (RPA) is a specialized software robot wildly used for facilitating repeated tasks. The RPA can augment medical staffs in data gathering, analysis and reporting process. This paper presents a RPA framework for a mobile glaucoma screening system. With the RPA, fundus images and clinical history of patients are gathered and submitted to the machine learning based glaucoma diagnosis. If the preliminary diagnosis results in a severe condition, the application directly prompts the eye specialists for timely verification and treatments. The application facilitates the specialists in adjusting clinical parameters, making final diagnosis decision remotely, gathering and reporting information and decision, and scheduling the next hospital visit or re-imaging. The integration of the RPA framework within the eye screening system significantly reduces costs and time of operations, allows timely treatments, improves customer experiences, and promotes feasibility in large-scale population glaucoma screening.

Post-Operative Hip Fracture Rehabilitation Model

Akash Gupta, Adnan Al-Anbuky, Peter McNair
School of Engineering, Computer & Mathematical Sciences, Auckland University of Technology, New Zealand;
School of Clinical Sciences, Auckland University of Technology, Auckland, New Zealand

Abstract: Hip fracture incidence increases with age and is common among older population. It causes significant problems as there is an increased risk of mortality, restricts the movement and well-being of the injured people, loss of independence and other adverse health related concerns. Rehabilitation plays a significant role in recovery and improving the physical functionality of patients in all stages of care. This is through supporting early postoperative mobilization to secondary prevention using balance and exercise activity movements. Many studies have analysed the post-operative effect of physiotherapeutic exercise and mobility during hip fracture rehabilitation process. Nevertheless, none of them has highlighted the key stages and activities progression pathway involved during post-operative rehabilitation recovery process. Therefore, it is not clear which care and rehabilitation services accomplish the suitable outcomes for the people undergoing rehabilitation. This article presents conceptual model for the process of the post-operative hip fracture rehabilitation. The model reflects the key stages a patient undergoes straight after hospitalization. The concept could support the development of an online activity monitoring system that could track the vital changes or ongoing improvements taking place in a patient's well-being as they go through the rehabilitation program. This should also offer vital information to clinicians, related hospitals people, caregivers as well as the patient's him/her-self. Another area of interest here is that of unfolding the progressive improvement in related muscles in reference to the stages of the program and how this can evolve to offer more precise interaction with the rehabilitation process.

Introducing a Conceptual Framework for Architecting Healthcare 4.0 Systems

Aleksandar Novakovic, Adele H. Marshall, Carolyn Mcgregor
School of Mathematics and Physics, Queen's University Belfast, United Kingdom;
Faculty of Business and IT, Ontario Tech University, Oshawa, Canada

Abstract: There is an enormous number of healthcare analytics applications in existence which have been embedded into healthcare systems with varying degrees of success. One of the key challenges is their need for access to sensitive patient data in a healthcare system that has a multitude of healthcare applications. This paper introduces a new theoretical framework as an architecture in which Healthcare 4.0 applications can operate. The framework proposes using Apache Kafka as the core technology for creating data integration pipelines with the goal being to bring standardisation into the healthcare systems. The architecture offers a safe and secure environment in which multiple applications and algorithms from different organisations can seamlessly co-exist.

A Machine Learning Driven Approach to Predict the Outcome of Prostate Biopsy: Identifying Cancer, Clinically Significant Disease, and Unfavorable Pathological Features on Prostate Biopsy

John L. Pfail, Dara J. Lundon, Parita Ratnani, Vinayak Wagaskar, Peter Wiklund, Ashutosh K. Tewari
Department of Urology, Icahn School of Medicine at Mount Sinai, New York, USA

Abstract: Prostate cancer screening and diagnosis remains controversial, due to the debate regarding overdiagnosis and subsequent overtreatment of prostate cancer. Reducing unnecessary prostate biopsies is a crucial step towards reducing overdiagnosis. As we move toward more personalized medicine and individualized medical decision making, there is a fundamental need for better risk assessment tools that can aid patients and physicians in this decision-making process. The presented work here seeks to construct risk prediction models to predict the presence of prostate cancer, clinically significant cancer (Gleason ≥7), and unfavorable pathology (pT3a or pT3b and Gleason Grade Group≥3) on initial biopsy. Such multivariable risk prediction models can be used to further aid in patient counselling for those undergoing prostate biopsy.

Using Natural Language Processing to Optimize Engagement of Those with Behavioral Health Conditions That Worsen Chronic Medical Disease

Peter Bearse, Atif Farid Mohammad, Intisar Rizwan I Haque, Susan Kuypers, Rachel Fournier
Catasys Inc., Santa Monica, California, USA

Abstract: In this work, a Natural Language Processing algorithm is proposed for the analysis of outcomes of member engagement specialist.s cold calls to prospective mental and behavioral care patients. The purpose of the call is to introduce prospective members to specially designed mental healthcare therapy and get them to enroll in the program. The proposed approach is based on keywords analysis and the results are provided on the analysis of 9254 transcribed voice calls using google cloud platform (GCP) of varying duration between 30 to 50 minutes. Additionally, qualitative design was used to identify keywords and phrases for high performer and low/mid performer member engagement specialists.

DRDr: Automatic Masking of Exudates and Microaneurysms Caused By Diabetic Retinopathy Using Mask R-CNN and Transfer Learning

Farzan Shenavarmasouleh, Hamid R. Arabnia
Department of Computer Science, University of Georgia, Georgia, USA

Abstract: This paper addresses the problem of identifying two main types of lesions - Exudates and Microaneurysms - caused by Diabetic Retinopathy (DR) in the eyes of diabetic patients. We make use of Convolutional Neural Networks (CNNs) and Transfer Learning to locate and generate highquality segmentation mask for each instance of the lesion that can be found in the patients' fundus images. We create our normalized database out of e-ophtha EX and e-ophtha MA and tweak Mask R-CNN to detect small lesions. Moreover, we employ data augmentation and the pre-trained weights of ResNet101 to compensate for our small dataset. Our model achieves promising test mAP of 0.45, altogether showing that it can aid clinicians and ophthalmologist in the process of detecting and treating the infamous DR.

An Altmetric Study on Dental Informatics

Jessica Chen, Qiping Zhang
Jericho High School, Jericho, New York, USA;
Long Island University, Brookville, New York, USA

Abstract: Dental informatics is a relatively new research field, but its publication has been steadily increasing in the past decade. Altmetric is a platform that collects information from different web sources and combines it together into a citation-based metrics for researchers, publishers, and institutions. It tracks the attention that research outputs (such as scholarly articles and datasets) receive online. Altmetric Explorer is an easy-to-use web-based platform that enables users to browse and report on all attention data for a given scholarly outputs (including journal articles and dataset). Given the nature of dental informatic research is quite practical, the purpose of this study is to perform an Altmetric analysis to systematically study research trends on dental informatics. The study identified various aspects of research outputs on dental informatics (attention scores, timeline of mentions, yearly publication total, top 10 journals, top 15 affiliation of first authors, and top 5 mention categories), that would be of interest to both researchers and practitioners in the area of dental informatics. In conclusion, dental informatics needs more publicity. Although there is an increase in multiple aspects (such as social media mentions and publications) in dental informatic research, it is not enough to match other popular fields.

The 22nd International Conference on Artificial Intelligence (ICAI 2020)
https://americancse.org/events/csce2020/conferences/icai20

A Deep Learning Approach to Diagnose Skin Cancer Using Image Processing

Roli Srivasta, Musarath Jahan Rahmathullah, Siamak Aram, Nathaniel Ashby, Roozbeh Sadeghian
Harrisburg University of Science and Technology, Harrisburg, PA, USA

Abstract: Skin cancer is the most commonly diagnosed cancer in the United States with over a million cases being detected each year. Fortunately, early detection provides high odds of recovery. The traditional method of detection involves clinical screening, which is prone to false positives, followed by an invasive biopsy. While this provides for a high rate of detection, it is intrusive and costly. Artificial Intelligence for medical image analysis has proved effective in assisting in the diagnosis of many medical maladies, yet fine variations in the appearance of skin lesions has made applications to skin cancer detection difficult. We report that a Deep Convolutional Neural Network (CNN) trained over clinically labeled images (pixels) can accurately assist in the diagnosis of early stage skin cancers. Specifically, we analyze skin lesions using CNN and evaluate its performance on seven dermatologist certified clinical image types: Actinic keratoses and intraepithelial carcinoma (Bowen's disease), basal cell carcinoma, benign keratosis like lesions (solar lentigines, seborrheic keratoses, and lichen-planus like keratoses), dermatofibroma, melanoma, melanocytic nevi, and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas, and hemorrhage). The model provides significantly high levels of average accuracy, specificity, and sensitivity across these types.

White Blood Cell Classification Using Genetic Algorithm - Enhanced Deep Convolutional Neural Networks

Omer Sevinc, Serdar Guzel, Iman Askerbeli, Mehrube Mehrubeoglu
Ondokuz Mayis University, Samsun, Turkey;
Ankara University, Ankara, Turkey;
Texas A&M University-Corpus Christi, Texas, USA

Abstract: The amount of white blood cells in the blood is of great importance for disease diagnosis. White blood cells include five main classes (eosinophils, lymphocytes, monocytes, neutrophils, basophils) each of which is an important indicator for specific diseases. Deep learning models have been developed to successfully classify the different white blood cell types. The most prominent deep learning models in image classification are deep convolutional neural network (D-CNN) models. A key challenge when solving a problem using deep learning is identifying and setting the hyperparameters for the algorithm. Mostly, these hyperparameters are set manually based on experience. In this study, a new model of deep convolutional neural network is proposed for the classification of four white blood cells types. In this model, the hyperparameters are self-optimized by a genetic algorithm which provides significant improvement in the model. For the verification of the proposed model, four types of white blood cells available from the Kaggle data series were studied. The number of white blood cell images are about 12,000 and are split for training and test sets as 80% and 20%, respectively. When the proposed model was applied to the Kaggle white blood cell dataset, the four white blood cell types in the sample data set were classified with high accuracy. The genetic algorithm (GA)-enhanced D-CNN model produced above 93% classification accuracy for the test set demonstrating the success of the proposed enhancement to the D-CNN model with GA. Comparatively, D-CNN models without GA optimization, such as Inception V3 model, produced 84% accuracy, and ResNet-50 model achieved 88% accuracy.

Synthetic AI Nervous/Limbic Derived Instances (SANDI)

Shelli Friess, James A. Crowder, Michael Hirsch
School of Counseling, Walden University, Minneapolis, Minnesota, USA;
Colorado Engineering Inc., Colorado Springs, Colorado, USA;
President and CTO, ISEA TEK LLC, Maitland, Florida, USA

Abstract: Artificial feelings and emotions are beginning to play an increasingly important role as mechanisms for facilitating learning in intelligent systems. What is presented here is the theory and architecture for an artificial nervous/limbic system for artificial intelligence entities. Here we borrow from the military concept of operations management and start with a modification of the DoD Observe, Orient, Decide and Act (OODA) loop. We add a machine learning component and adapt this for processing and execution of artificial emotions within an AI cognitive system. Our concept, the Observe, Orient, Decide, Act, and Learn (OODAL) loop makes use of Locus of Control methodologies to determine, during the observe and orient phases, whether the situation constitutes external or internal controls, which will affect the possible decisions, emotions, and actions available to the artificial entity (e.g., robot). We present an adaptation of the partial differential equations that govern human systems, adapted for voltage/current regulation rather than blood/nervous system regulation in humans. Given human trial and error learning, we incorporate a Q-learning component to the system that allows the AI entity to learn from experience whether its emotions and decisions were of benefit or problematic.

Application of Modified Social Spider Algorithm (MSSA) on Unit Commitment Solution Considering Uncertainty of Wind Power in Restructured Electricity Market

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

Abstract: Nowadays, by the integration of renewable energy resources, the expenditures pertaining to energy production are diminished with respect to the vast advancements in their science and technology. Considering the integration of renewable energy resources in power system operation, particularly in the generation section, poses grave difficulties in market clearance and calculation of unit commitment problem in order to achieve the optimum objectives. The inclusion of uncertainty in a large-scale non-linear non-convex non-smooth optimization problem enhances the complexity of the problem. To solve such a sophisticated problem, some exact or heuristic methods are proposed so far. In this paper, a 10-generator test system is taken into account for the simulation of the proposed method, and the modified social spider algorithm is used to solve the unit commitment problem considering the impact of the presence of wind farms. The constraints related to the spinning reserve requirements are also incorporated to maintain system security. Finally, economic load dispatch is performed to find the optimum generation level of all committed units and obtain operational costs. The problem procedure is repeated regardless of the presence of wind units to compare the results and to assess the effectiveness of the proposed approach.

Which Scaling Rule Applies to Artificial Neural Networks

Janos Vegh
Kalimanos BT, Debrecen, Hungary

Abstract: Although an Artificial Neural Network (ANN) is a biology-mimicking system, it is built from components designed/fabricated for use in conventional computing, created by experts trained in conventional computing. Making computing systems, actively cooperating and communicating, using segregated single processors, however, has severe performance limitations. The achievable payload computing performance sensitively depends on workload type, and this effect is only poorly known. The workload type, the Artificial Intelligence (AI) systems generate, has an exceptionally low payload computational performance for ANN applications. Unfortunately, initial successes of demo systems that comprise only a few "neurons" and solve simple tasks are misleading: scaling of computingbased ANN systems is strongly nonlinear. The paper discusses some major limiting factors that affect performance. It points out that for building biology-mimicking large systems, it is inevitable to perform drastic changes in the present computing paradigm (mainly to comprehend temporal behavior of components) and architectures.

The Systems AI Thinking Process (SATP) for Artificial Intelligent Systems

James A. Crowder, Shelli Friess
Colorado Engineering Inc., Colorado Springs, Colorado, USA;
School of Counseling, Walden University, Minneapolis, Minnesota, USA

Abstract: Previous work has focused on the overall theory of Systems-Level Thinking for artificial intelligent entities in order to understand how to facilitate and manage interactions between artificial intelligent system and humans or other systems. This includes the ability to predict and produce behaviors consistent with the overall mission (duties) of the AI system, how to control the behaviors, and the types of control mechanisms required for self-regulation within an AI entity. Here we advance that work to look at the overall Systems AI Thinking Process (SATP) and the architecture design of self-regulating AI systems-level processes. The overall purpose here is to lay out the initial design and discussion of concepts to create an AI entity capable of Systems-Level thought and processing.

Judging Emotion from EEGs Using Pseudo Data

Seiji Tsuchiya, Misako Imono, Hirokazu Watabe
Department of Intelligent Info. Eng. & Science, Faculty of Science & Eng., Doshisha University, Kyoto, Japan;
Department of Information Systems, School of Informatics, Daido University, Nagoya, Aichi, Japan

Abstract: For a robot to converse naturally with a human, it must be able to accurately gauge the emotional state of the person. Techniques for estimating emotions of a person from facial expressions, intonation and speech content have been proposed. This paper presents a technique for judging the emotion of a person from EEGs by SVM using pseudo extended data. Accuracy of emotion judgment from EEG features using at random pseudo extended data was 25.0%, and using proposed extension technique was 30.0%. However, performance accuracy remains low, and continued development is required through further development of methods for both reducing noise mixed in with EEGs.

Targeted Aspect-Based Sentiment Analysis for Ugandan Telecom Reviews from Twitter

David Kabiito, Joyce Nakatumba-Nabende
Department of CS, School of Computing & Informatics Technology, Makerere University, Uganda

Abstract: In this paper we present SentiTel, a fine-grained opinion mining dataset that is human annotated for the task of targeted aspect-based sentiment analysis (TABSA). SentiTel contains Twitter reviews about three telecoms in Uganda posted in the period between February 2019 and September 2019. The reviews in the dataset have a code-mix of English and Luganda a language that is commonly spoken in Uganda. The data set in this paper consists of 5,973 human annotated reviews with the target entity which is the target telecom, aspect and sentiment towards the aspect of the target telecom. Each review contains at least one target telecom. Two models are trained for the TABSA task that is random forest which is the baseline model and BERT. The best results are obtained using BERT with an Area Under the ROC Curve (AUC) of 0.950 and 0.965 on aspect category detection and sentiment classification respectively. The results show that even though tweets are written without the intention of writing a formal review, they are rich in information and can be used for fine-grained opinion mining. Finally, the results show that fine-tuning the pre-trained BERT model on a downstream task generates better results compared to the baseline models.

Generic Object Recognition using Both Illustration Images and Real Object Images by CNN

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

Abstract: In recent years, the development of robots has been carried out for making human life more convenient and more comfortable along with the development of artificial intelligence. It is necessary for the robot to recognize the surrounding environment. However, in the surrounding environment there are objects other than real objects such as illustrations and paintings. When recognizing an image showing an illustration image with the current object recognition system which learned using real object images, the recognition rate is very low (about 65%). In this research, we aim to recognize both illustration images and real object images, and we verified whether the pseudo illustrated image which processed contour processing and the color reduction processing to the real image is effective for the recognition of the illustrated image.

When Entity Resolution Meets Deep Learning, Is Similarity Measure Necessary?

Xinming Li, John R. Talburt, Ting Li, Xiangwen Liu
Department of Information Science, University of Arkansas at Little Rock, Arkansas, USA

Abstract: In Entity Resolution (ER), more and more unstructured records impose challenge to the traditional similarity-based approaches, since existing similarity metrics are designed for structured records. Now that similarity is hard to measure for unstructured records, can we do pairwise matching without similarity measure? To answer this question, this research leverages deep learning's artificial intelligence to learn the underlying record matched pattern, rather than measuring records similarity first and then making linking decision based on the similarity measure. In the representation part, token order information is taken into account in word embedding, and not considered in Bag-of-words (Count and TF-IDF); in the model part, multi-layer perceptron (MLP), convolutional neural network (CNN) and long short-term memory (LSTM) are examined. Our experiments on both synthetic data and real-world data demonstrate that, surprisingly, the simplest representation (Count) and the simplest model (MLP) together get the best results both in effectiveness and efficiency. An F-measure as high as 1.00 in the pairwise matching task shows potential for further applying deep learning in other ER tasks like blocking.

Parallel Algorithms to Detect and Classify Defects in Surface Steel Strips

Khaled R Ahmed, Majed AlSaeed, Maryam AlJumah
School of Computing, Southern Illinois University, Illinois, USA;
King Faisal University, Hofuf, Saudi Arabia

Abstract: In steel industry, automatic defects inspection and classification is of great importance to improve the quality. This paper proposed and developed parallel algorithms using CUDA to improve the required computing time to detect defects in surface steel strips. The algorithm divides steel images into non-overlapped Region of Interest (ROI) and employs the Summed Area Table to improve the required time to extract statistical features per (block) ROI. The computation time of the proposed parallel algorithm excels the sequential one. Support Vector Machine classifier has been used to classify patches, scratches, and scale defects. The experimental results indicate significant improvements and 1.6 speed up.

Predicting Number of Personnel to Deploy for Wildfire Containment

John Carr, Matt Lewis, Qingguo Wang
College of Computing and Technology, Lipscomb University, Nashville, Tennessee, USA

Abstract: Climate change is causing longer forest fire seasons and more fires in places where they traditionally were rare. Accordingly, the U.S. Forest Service's annual budget devoted to wildfires jumped from 16% to over 50% between 1995 and 2015 and exceeded $2.4 billion in 2017. Allocating and supplying the correct amount of personnel and equipment to a fire as quickly as possible is vital in reducing acreage destroyed, costs, and saving lives. In this paper, we explore the possibility of predicting the number of personnel needed to effectively fight a wildfire, in an aim to develop a forecasting system in the future. On the IRWIN test dataset, our model obtained a coefficient of determination (R^2) 0.61 and mean absolute error 1.4 personnel. Thorough analysis and discussion are also provided in the paper.

A Path-based Personalized Recommendation using Q Learning

Hyeseong Park, Kyung-Whan Oh
Department of CSE, Sogang University, Mapo-gu, Seoul, Republic of Korea

Abstract: Recently, knowledge graph-based personalized recommendation systems have been proposed to exploit rich structured information. In particular, knowledge graph reasoning via knowledge embedding is increasingly introduced to preserve semantic information in the knowledge graph. However, due to the large dimensional representations of the millions of entities and relations, knowledge graph reasoning requires excessive computational resources. Moreover, selecting meaningful latent features in the knowledge graph between users and items is a challenging issue. In this paper, to impose efficient and accurate personalized recommendations, we propose a path-based recommendation using Q learning. Our model also offers a good explanation for recommendations. Experimental results show that the proposed method provides higher accuracy, compared to a uniform random walk based algorithm.

Deep Learning Based Constituency Parsing for Arabic Language

Amr Morad, Magdy Nagi, Sameh Alansary
Information and Communication Technologies (ICT), Bibliotheca Alexandrina, Egypt;
Faculty of Engineering, Computer & Systems Department, Alexandria University, Bibliotheca Alexandrina, Egypt;
Arabic Computational Linguistic Center, Faculty of Arts, Phonetics & Linguistics Department, Alexandria University, Bibliotheca Alexandrina, Egypt

Abstract: Constituency parse-tree is considered the backbone of several NLP tasks. Deep learning techniques are adopted because they generate parse-tree using a dataset without any predefined rules, making them extensible to any language. To capture the semantic meaning, dense words representation technique is necessary. This paper combines both dense Arabic word representations and deep learning model, to generate constituent parse-tree. The resultant tree is used in a complete workflow. It contains a web-based application to enable linguists to choose the sentence, generate its constituent, review resultant tree and edit needed parts. Moreover, the curated output sentence will be used to re-train the model for self-correction. The model is efficient and parallel, resulting in a quick training process.

The Effectiveness of Data Mining Techniques at Estimating Future Population Levels for Isolated Moose Populations

Charles E. Knadler, Jr.
Computer Science, Utah Valley University, Orem, Utah, USA

Abstract: The objective of this project is to determine if data mining techniques may be applied to small data sets to forecast population growth of isolated populations of organisms, with a useful level of accuracy. The Isle Royale moose population was

chosen to be the basis of this study, because of the quality of available data and substantial previous work studying the Isle Royale wolf/moose population dynamics.

Comparison of Similarity of Situations in Sets of Associations

Peeter Lorents
IT College, Tallinn University of Technology, Tallinn, Estonia

Abstract: Under observation is a segment of the decision-making process in which decisions justified on the similarity of situations and developments at the disposal of information of the decision-maker. Developments and situations observed and treated through descriptions that consist of statements. One form of similarity of situations or developments observed as descriptive similarity and its results through numerical evaluation. Beside descriptive similarity is under observation also structural similarity, and its relation to descriptive similarity.

Effects of Domain Randomization on Simulation-to-Reality Transfer of Reinforcement Learning Policies for Industrial Robots

Christian Scheiderer, Nik Dorndorf, Tobias Meisen
Institute of Technologies and Management of the Digital Transformation, University of Wuppertal, Germany;
RWTH Aachen University, Germany

Abstract: A requirement for a significant amount of training as well as the exploration of potentially expensive or safety-critical states limits the applicability of reinforcement learning for real-world robotics. One potential solution is given by pretraining models in simulations before transferring them to the real world. In this paper, we investigate the concept of domain randomization to train robust agents in simulation to control an industrial robot. We examine the effects of different degrees of randomization with respect to the transferability to the real world. In addition, we use attention maps to gain insights into the agents' decision-making processes. We find that attention maps enable a qualitative assessment for the data-efficiency of a pretrained agent when transferred to the real-world setup.

Lightweight Approximation of Softmax Layer for On-Device Inference

Ihor Vasyltsov, Wooseok Chang
Samsung Advanced Institute of Technology, Samsung Electronics, Republic of Korea

Abstract: In this paper we propose a method to approximate a softmax layer for computer vision applications, especially for devices with limited hardware (HW) resources, for example mobile or edge platforms. …

A Similarity-Based Decision Process for Implementation Decisions

Maryna Averkyna
Estonian Business School, Tallinn, Estonia;
The National University of Ostroh Academy, Seminarska, Ostroh, Ukraine

Abstract: The paper deals with the implementation of similarity-based approach decisions which are considered in details (specific areas and cases in real life). The role of similarity, descriptive similarity and numerical evaluation is comprehensively examined. The author pointed out that descriptive similarity is the relevant approach for equation of the statements The paper introduces the transport situation similarity results in town of Ostroh (Ukraine) and certain Estonian towns, which is of great use for Ostroh public transportation claims. The author indicated that it is important to transform textual information into formulas. Artificial intelligence application needs of two level are under appreciate importance. To justify these decisions suitable algebraic formulas of system theory as well as experts. logical steps are used.

Combination of Variational Autoencoders and Generative Adversarial Network into an Unsupervised Generative Model

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

Abstract: In this study, we explored building a generative model with the combination of Variational Autoencoders (VAE) and a Generative Adversarial Network (GAN) that offers better results when the agent interacts with the environment. Our agent model can train the unsupervised environment and increase the imaging quality. Moreover, it provides better control options and produces better accuracy results as compared to traditional systems. An experiment was performed on car racing based on the designed agent model and features that can be extracted effectively to help produce reliable information. With the combination of Variational Autoencoders and a generative adversarial network, we could provide better feature extraction to gather relevant data and solve the complexity. With the combination of VAE and GAN, elevated level visual abstraction can be made effectively.

Growing Artificial Neural Networks

John Mixter, Ali Akoglu
ECE, The University of Arizona, Tucson, Arizona, USA

Abstract: Pruning is a legitimate method for reducing the size of a neural network to fit in low SWaP hardware, but the networks must be trained and pruned offline. We propose an algorithm, Artificial Neurogenesis (ANG), that grows rather than prunes the network and enables neural networks to be trained and executed in low SWaP embedded hardware. ANG accomplishes this by using the training data to determine critical connections between layers before the actual training takes place. Our experiments use a modified LeNet-5 as a baseline neural network that achieves a test accuracy of 98.74% using a total of 61,160 weights. An ANG grown network achieves a test accuracy of 98.80% with only 21,211 weights.

Dynamic Heuristics for Surveillance Mission Scheduling with Unmanned Aerial Vehicles in Heterogeneous Environments

Dylan Machovec, James A. Crowder, Howard Jay Siegel, Sudeep Pasricha, Anthony A. Maciejewski
Department of Electrical & Computer Engineering, Colorado State University, Fort Collins, Colorado, USA;
Colorado Engineering Inc., Colorado Springs, Colorado, USA;
Department of Computer Science, Colorado State University, Fort Collins, Colorado, USA

Abstract: In this study, our focus is on the design of mission scheduling techniques capable of working in dynamic environments with unmanned aerial vehicles, to determine effective mission schedules in real-time. The effectiveness of mission schedules for unmanned aerial vehicles is measured using a surveillance value metric, which incorporates information about the amount and usefulness of information obtained from surveilling targets. We design a set of dynamic heuristic techniques, which are compared and evaluated based on their ability to maximize surveillance value in a wide range of scenarios generated by a randomized model. We consider two comparison heuristics, three value-based heuristics, and a metaheuristic that intelligently switches between the best value-based heuristics. The novel metaheuristic is shown to find effective solutions that are the best on average as all other techniques that we evaluate in all scenarios that we consider.

Traceability Analysis of Patterns using Clustering Techniques

Jose Aguilar, Camilo Salazar, Julian Monsalve-Pulido, Edwin Montoya, Henry Velasco
GIDITIC, Universidad EAFIT, Medellin, Colombia;
CEMISID, Universidad de Los Andes Merida, Venezuela;
LANTIA SAS, Medellin, Colombia

Abstract: Currently, with the high rate of generation of new information, it is important the traceability of its evolution. This paper studies techniques that allow analyzing the evolution of the knowledge, starting with analyzing the capabilities of the techniques to identify the patterns that represent the common information in datasets. From the "patterns", the evolution of their characteristics over time is analyzed. The paper considers the next techniques for the problem of tracking the traceability of the patterns: LDA (Latent Dirichlet allocation), Birch (Balanced Iterative Reducing and Clustering using Hierarchies), LAMDA (Learning Algorithm for Multivariate Data Analysis), and K-means. They are used, both for the initial task of grouping the data, as well as, to analyze the characteristics of the patterns, and the relevance of them in the patterns through their evolution (traceability). This paper uses different types of data sources of educational contents, and with these datasets, the topological models to describe the "patterns" generated from the grouping of the analyzed data, and their dynamics (evolution over time), are Studied (traceability). For the evaluation, the paper considers three metrics: CalinskiHarabasz Index,Davies-Bouldin Index and Silhouette Score.

Using Neural Networks and Genetic Algorithms for Predicting Human Movement in Crowds

Abdullah Alajlan, Alaa Edris, Robert B. Heckendorn, Terry Soule
Computer Science Department, University of Idaho, Moscow, Idaho, USA;
Technical and Vocational Training Corporation, Riyadh, KSA;
Computer Science Department, Jeddah University, Jeddah, KSA

Abstract: Safety is an important issue at large gatherings of people such as at religious gatherings or sporting events. Therefore, it's important to control crowds, and identify in advance when dangerous situations may arise. Simulations play an important role in predicting how people in crowds will react to each other and their environment. Simulations often rely on a priori models of human behavior to predict crowd behavior. We combine, Genetic Algorithms and Neural Networks to learn how people behave in crowds to avoid assumptions used in a priori models. We examine learning in two important regimes, structured crowds where individuals are moving in a specified direction, as in the Islamic Hajj or Hindu Kumbh Mela; and unstructured crowds, such as town squares and train stations. In this preliminary work we are most concerned with questions of trainability. In order to provide sufficient data and control qualitative features of our crowd data we begin use generated data based on elaborations of wildlife flocking models in NetLogo. We compared performance on structured and unstructured crowds by predicting a series of next locations. The results showed we are able to predict crowd motion, but error rates for individuals grow as time passes; however, the structured crowd gave more reliable results than the unstructured crowd.

Long Short Term Memory in Chemistry Dynamics Simulation

Heng Wu, Shaofei Lu, Colmenares-Diaz Eduardo, Junbin Liang, Jingke She, Xiaolin Tan
Department of Mathematics and Computer Science, West Virginia State University Institute, West Virginia, USA;
College of Computer Science and Electronic Engineering, Hunan University, Changsha, P. R. China;
Department of Computer Science, Midwestern State University, USA;
School of Computer and Electronics Information, Guangxi University, P. R. China

Abstract: Chemistry dynamics simulation is widely used in Quantitative structure activity relationship QSAR, virtual screening, protein structure prediction, quantum chemistry, materials design and property prediction, etc. This paper explores the idea of integrating Long-Short Term Memory (LSTM) with chemistry dynamics simulations to enhance the performance of the simulation and improve its usability for research and education. The idea is successfully used to predict the location, energy and

Hessian of atoms in a H2O reaction system. The results demonstrate that the artificial neural network based memory model successfully learns the desired features associated with the atomic trajectory and rapidly generates predictions that are in excellent agreement with the results from chemistry dynamics simulations. The accuracy of the prediction is better than expected.

Activity Recognition for Elderly using Machine Learning Algorithms

Heba Elgazzar
School of Engineering and Computer Science, Morehead State University, Morehead, Kentucky, USA

Abstract: Human activity recognition is one of the major research problems in computer science and it has a wide range of applications including healthcare applications. The goal of this research project is to design and apply machine learning methods that can be used to analyze publicly available datasets collected by wearable sensors of elderly and identify different activities such as setting and walking. Feature selection and ranking algorithms were used to select the most relevant feature and reduce the number of collected and used data from sensors. Several classification algorithms were used in this paper for activity recognition and several experiments were conducted to compare between these different algorithms using different number of features based on the ranking of features. The experimental results show a high accuracy in recognizing activities can be achieved using a fewer number of features. This can help in providing the medical services required for elderly based on the detected activity using a small number of collected features from sensors. The experimental results show that random forest algorithm gave the highest accuracy compared to the other algorithms using only three features of sensors data.

An Evaluation of Bayesian Network Models for Predicting Credit Risk on Ugandan Credit Contracts

Peter Nabende, Samuel Senfuma, Joyce Nakatumba
Makerere University, Kampala, Uganda

Abstract: Credit risk prediction is a task that continues to be studied extensively for various crediting schemes. Machine learning has become a viable option for developing credit risk prediction models for uncertain cases involving a considerable number of instances. In particular, the framework of Bayesian networks is a very suitable option for managing uncertainty. Although there have been several studies on the use of Bayesian networks in credit risk prediction, there is scarce literature about their application to cases from developing economy contexts. In this paper, we exhaustively apply and evaluate several Bayesian network models for credit risk prediction based on cases from a Ugandan financial institution. Credit risk prediction quality from some Bayesian network models is satisfactory and compares with prediction quality from other state-of-the-art machine learning methods. Evaluation results show that one Bayesian network model learned through a global optimization hill climbing method, always leads to the highest prediction quality so far on Ugandan credit contracts.

The Evolutional Treasury Hybrid Optimization System (ETHOS)

N. Loukeris, G. Chalamandaris, I. Eleftheriadis
Department of Accounting & Finance, Athens University of Economics and Business, Athens, Greece;
Department of Business Administration, University of Macedonia, Thessaloniki, Greece

Abstract: We develop an innovative method which takes into consideration advanced higher moments under a new perspective of the Portfolio Selection problem, supported by effective Computational Intelligence models. The Evolutional Treasury Hybrid Optimization System (ETHOS) mines hidden patterns from vast accounting data and financial statements restricting misguiding effects of noise or various forms of fraud, providing a new method portfolio selection optimally.

Human Motion Recognition Using Zero-Shot Learning

Farid Ghareh Mohammadi, Ahmed Imteaj, M. Hadi Amini, Hamid R. Arabnia
Department of Computer Science, University of Georgia, Athens, Georgia, USA;
School of Computing and Information Sciences, Florida International University, Miami, FL, USA

Abstract: Motion recognition: recognizing unseen and unlabeled movement patterns has become more popular and challenging in advanced machine learning. Motion recognition tackles some of the emerging challenges in computer vision problems, such as analyzing actions in a surveillance video where there is a lack of sufficient training data. Motion recognition also plays a pivotal role in human action and behavior recognition. In this paper, we propose a novel action and motion recognition method using zero-shot learning. We overcome a limitation of machine learning by recognizing unseen and unlabeled classes in the field of human action recognition. In order to evaluate the effectiveness of the proposed solution, we use a dataset available from the UCI machine learning repository. This dataset enables us to apply zero-shot learning to human motion and action recognition. Our results verify that the proposed method outperforms state-of-the-art algorithms.

Would You Turn on Bluetooth for LBA?

Heng-Li Yang, Shiang-Lin Lin, Jui-Yen Chang
Department of Management Information System, National Cheng-Chi University, Taipei, Taiwan

Abstract: In recent years, location-based advertising (LBA) can deliver advertisements to customers in targeted locations and provide product and service information of their local businesses. Thus, understanding customers' needs and considerations is essential for the popularization of LBA services. The purpose of this study is to explore the key factors that customers would concern about while they decide to turn on Bluetooth on their mobile devices to receive LBA services using fuzzy-AHP in conjunction with DEMATEL method. The analysis results in this study indicate that "personal location privacy", "get information about saving money right time and right place", and "personal preference privacy" are the top three consideration factors.

Unsupervised Classification of Cell Imaging Data Using the Quantization Error in a Self-Organizing Map

Birgitta Dresp-Langley, John M. Wandeto
Centre National de la Recherche Scientifique (CNRS), France;
Dedan Kimathi University of Technology, Nyeri, Kenya

Abstract: This study exploits previously demonstrated properties (i.e. sensitivity to spatial extent and intensity of local image contrasts) of the quantization error in the output of a Self-Organizing Map (SOM-QE). Here, the SOM-QE is applied to double-color-staining based cell viability data in 96 image simulations. The results from this study show that, as expected, SOM-QE consistently and in only a few seconds detects fine regular spatial increase in relative amounts of RED or GREEN pixel staining across the test images, reflecting small, systematic increase or decrease in the percentage of theoretical cell viability below a critical threshold. While such small changes may carry clinical significance, they are almost impossible to detect by human vision. Moreover, here we demonstrate an expected sensitivity of the SOM-QE to differences in the relative physical luminance (Y) of the colors, which translates into a RED-GREEN color selectivity. Across differences in relative luminance, the SOM-QE exhibits consistently greater sensitivity to the smallest spatial increase in RED image pixels compared with smallest increases of the same spatial magnitude in GREEN image pixels. Further selective color contrast studies on simulations of biological imaging data will allow generating increasingly larger benchmark datasets and, ultimately, unravel the full potential of fast, economic, and unprecedentedly precise predictive imaging data analysis based on SOM-QE.

Adaptive Chromosome Diagnosis Based on Scaling Hierarchical Clusters

Muhammed Akif Agca, Cihan Tastan, Kadir Ustun, Ibrahim Halil Giden
TOBB ETU / Computer Engineering, Ankara, Turkey;
Luxembourg Institute of Science and Technology - LIST, Luxembourg;
Labcell Laboratory Istanbul, Turkey; TOBB ETU / Electronics Engineering, Ankara, Turkey

Abstract: In this study, we presented how to divide chromosome data sets into scalable hierarchical clusters. Diagnosis applications of chromosomal abnormalities to identify genetic diseases are mostly implemented with semi-automatic methods, which do not provide high-throughput and fast characterizations of patient specimens. However, selecting and managing specific features of chromosome data sets require dynamic and scalable data models. Here, we adapt the variations in sets to feature units as an effective tool for our dynamic/automated and scalable approach. This method enables geneticists to implement karyotyping techniques easily into an efficient identification of hereditary disorders or characterization of novel chromosomal alterations related with genetic diseases in a trusted/scalable manner. Furthermore, we explore physical limits of available Application Specific Integrated Circuits (ASICs) and On Board Computers (OBCs) to extract novel features of chromosomes via spectral analytics for real time diagnosis and to overcome the bottlenecks of the computational complexity, as well.

Hybrid Car Trajectory by Genetic Algorithms with Non-Uniform Key Framing

Dana Vrajitoru
Department of CIS, Indiana University South Bend, Indiana, USA

Abstract: In this paper, we present a hybrid application of genetic algorithms to a problem of optimizing a car trajectory in a car race setting. The aim is to find a curve traversing the race track, or a trajectory, that minimizes the total distance, and hope that it allows a car to finish the race faster. For this, we start by mapping and segmenting the track, then setting non-uniform key frames based on significant intervals where the track is almost straight and on long curves going in a single direction. The trajectory's value is anchored in the key frame points, insuring an overall smoothness. Then on each interval, we apply the genetic algorithm to compute the trajectory, and piece it together at the end. We compare the results with a procedurally computed trajectory and with trajectories obtained by the genetic algorithm with uniform key framing.

Fine Tuning a Generative Adversarial Network's Discriminator for Student Attrition Prediction

Eric Stenton, Pablo Rivas, Eitel Lauria
Department of Computer Science, Marist College, Poughkeepsie, New York, USA

Abstract: Predicting if freshmen students will drop out of college or transfer to another is often difficult due to limited and anomalous data. This paper explores using Generative Adversarial Networks (GANs) to learn the general features of student data and uses it to produce predictions with higher accuracy and lower false positive rates than neural networks trained with traditional techniques. Here we examine the differences between a classifier.s latent space when it is trained with a GAN architecture versus traditionally for predicting if a freshman student will leave Marist College within their first year. Our experimental results suggest that GANs are an alternative to training neural models for student dropout/transfer prediction.

Application of Associations to Assess Similarity in Situations Prior to Armed Conflict

Ahto Kuuseok
Estonian Police and Border Guard Board, Estonian Business School, Estonia

Abstract: Descriptions of situations immediately preceding the outbreak of a military conflict, specifically eve of military invasion are under consideration. Situation descriptions are treated as sets of relevant statements. The descrip-tions of various situations highlight similar claims. Descriptions containing identifiable statements form associations. Developments and situations ob-served and treated through descriptions that consist of statements. One form of similarity of situations or developments observed as descriptive similarity and its results through numerical evaluation. Based on of descriptive similarity, the eve of military invasion, involving states attack to another state, are investi-gated. The calculations made, express the irrelevance of the corresponding de-scriptions from public available sources. Means provided to quantify the gen-eral similarity of descriptions of situations constituting an association. Such estimates provide a first insight into what and how similar situations may de-velop in the context of this association.

A Multigraph-based Method for Improving Music Recommendation

James Waggoner, Randi Dunkleman, Yang Gao, Qingguo Wang
College of Computing and Technology, Lipscomb University, Nashville, Tennessee, USA

Abstract: Music recommendation systems have become an important part of the user-centric online music listening experience. However, current automated systems often are not tuned for exploiting the full diversity of a song catalogue, and consequently, discovering new music requires considerable user effort. Another issue is current implementations generally require significant artist metadata, user listening history, or a combination of the two, to generate relevant recommendations. To address the problems with traditional recommendation systems, we propose to represent artist-to-artist relationships as both simple multigraphs and more complicated multidimensional networks. Using data gathered from the MusicBrainz open music encyclopedia, we demonstrate our artist-based networks are capable of producing more diverse and relevant artist recommendations.

Pathways to Artificial General Intelligence: A Brief Overview of Developments and Ethical Issues via Artificial Intelligence, Machine Learning, Deep Learning, and Data Science

Mohammadreza Iman, Hamid Reza Arabnia, Robert Maribe Branchinst
Department of CS, Franklin College of Arts and Sciences, University of Georgia, Georgia, USA;
Learning, Design, & Technology, Mary Frances Early College of Education, University of Georgia, Georgia, USA

Abstract: Today, devices and applications powered by artificial intelligence (AI) include modes of transportation, home appliances, and mobile applications; in short, they are ubiquitous. Many people will have at least heard of AI and perhaps even subdivisions of AI such as Machine Learning (ML) and Deep Learning (DL). Each of these represents an advanced tool of data science explored in this article. First, we briefly review the history of these developments in data science, tracing the history from the first mechanical computer in 1850 to the current state of DL in 2020. Each section overviews some basic definitions, tenets, and current and future developments. We discuss possible future directions of AI.including the transition to Artificial General Intelligence (AGI). Finally, we explore some of the ethical dilemmas posed by such advances and offer a call to data and social scientists to carefully consider the implications of these technologies.

A Prototype Implementation of the NNEF Interpreter

Nakhoon Baek
School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea

Abstract: The NNEF (Neural Network Exchange Format) is a de facto standard file format for the neural network description and exchange. In this work, we represent a simple implementation of the NNEF execution system, which is similar to the programming language interpreters. While the original NNEF file format focused on the data exchange format itself, our prototype implementation shows that the contents of an NNEF file can be directly executed by the underlying computing systems.

Emergent Heterogeneous Strategies from Homogeneous Capabilities in Multi-Agent Systems

R. Fernandez, E. Zaroukian, J. D. Humann, B. Perelman, M. R. Dorothy, S. S. Rodriguez, D. E. Asher
US CCDC Army Research Laboratory, Adelphi, Maryland, US;
Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

Abstract: In multi-agent systems, agents. abilities are often used to classify a system as either homogeneous or heterogeneous. In the context of multi-agent reinforcement learning (MARL) systems, the agents can also be homogeneous or heterogeneous in their strategies. In this work, we explore instances where agents with homogeneous capabilities must collaborate to achieve a common goal in a predator-prey pursuit task. We show that results from homogeneous and heterogeneous strategies associated with learning differ substantially from agents with fixed strategies that are analytically defined. Given that our agents are homogeneous in capability, here we explore the impact of homogeneous and heterogeneous strategies in a MARL paradigm.

A Low-Cost Video Analytics System with Velocity based Configuration Adaptation in Edge Computing

Woo-Joong Kim, Chan-Hyun Youn
School of EE, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

Abstract: In this paper, we propose a low-cost video analytics system analyzing multiple video streams efficiently under limited resource. The objective of our proposed system is to find the best configuration decision of frame sampling rate for multiple video streams in order to minimize the accuracy degradation in the shared limited resource, utilizing the velocity features extracted from video context in low-cost. To evaluate the proposed algorithm, we use a subset of videos from VIRAT dataset. The results show that our video analytics system outperforms the existing video analytics systems on resource-accuracy tradeoffs and reduces the high profiling cost of them.

Deep Embedded Knowledge Graph Representations for Tactic Discovery

Joshua Haley, Ross Hoehn, John Singleton, Chris Ballinger, Alejandro Carbonara
Intelligent Systems, Soar Technology Inc., Ann Arbor, Michigan, USA

Abstract: Using unsupervised Machine Learning (ML) techniques for the discovery of commonalities within a dataset is a well-established approach. However, existing clustering methods require relationships to be represented within the data, making discovery difficult a priori since unlabeled classes are discovered through exploratory clustering. To circumvent exploratory class labeling, we propose a feature-rich, connected structure (i.e., semantic graph), rather than a feature-engineered vectorization, enabling the encoding of maximal information despite class uncertainty. Expanding upon previous tactics discovery work, the authors present a systematic approach using knowledge graph representations and graph embeddings to discover tactics ab initio from data characteristics.

Using Games for Problem Solving in Complex Environments with Cultural Algorithms

Faisal Waris, Robert Reynolds
Wayne State University, Detroit, Michigan, USA

Abstract: Cultural Algorithms (CA) are population-based stochastic optimization methods that are modelled after human culture and are suited to solving problems in complex environments. The CA Belief Space stores harvested knowledge from prior generations and re-distributes it to future generations via a knowledge distribution (KD) mechanism. The population of individuals is then guided through the search space via the associated knowledge. Previously, CA implementations have used competitive KD mechanisms which have performed well for problems embedded in static environments. Relatively recently, CA research has evolved to encompass dynamic problem environments. Given increasing environmental complexity, a natural question arises about whether KD mechanisms that also incorporate cooperation can perform better in such environments than purely competitive ones? Borrowing from game theory, game-based KD mechanisms are implemented with three types of games that can employ both competition and cooperation namely, Iterated Prisoner's Dilemma, Stag-Hunt and Stackelberg. The performance of the three game mechanisms are compared using a dynamic problem generator. Weighted Majority (Wisdom of the Crowd), the default CA competitive KD mechanism is used as the benchmark. It is shown that games that support both cooperation and competition do indeed perform better but not in all cases. The results shed light on what kinds of game mechanism are suited to problem solving in complex, dynamic environments.

Event-based Keyframing: Transforming Learner Observation Data into Compact and Meaningful Form

Robert Wray, Robert Bridgman, Josh Haley, Laura Hamel, Angela Woods
Intelligent Systems Division, Soar Technology, Inc., Ann Arbor, Michigan, USA

Abstract: Learning systems that adapt to learner capabilities, needs, and preferences have been shown to improve learning outcomes. However, creating systems that can interpret learner state within the context of a dynamic learning environment is costly and often tied to the specific requirements of the learning environment. We overview a new approach for monitoring and assessing system context and learner state that is not specific to a particular domain. The process is designed to transform diverse, continuous, and multi-channel streams of heterogenous system data into a consistent, discrete, and learner-centric interpretation of the situation. Key steps in the process include discretizing the data stream into "events" and then marking some events as "keyframes" that identify important steps or changes in the learning state. This keyframing process provides a compact representation for use by learner-adaptive processes (including assessment and tailoring) and simplifies the challenges of using machine learning as a mechanism for adaptation.

Reducing the Data Cost of Machine Learning with AI: A Case Study

Joshua Haley, Robert Wray, Robert Bridgman, Austin Brehob
Intelligent Systems Division, Soar Technology Inc., Ann Arbor, Michigan, USA

Abstract: The past several years have seen a strong push toward using Deep Learning systems-Neural Networks with multiple hidden layers. Deep Learning is now used in many machine learning applications and provides leading performance on numerous benchmark tasks. However, this increase in performance requires very large data sets for training. From a practitioner prospective, the model that performs best in benchmark tasks may be too data intensive to be adapted to practical application. We describe a behavior recognition problem that was solved using a sequence-based, deep-learning system and then re-implemented using a more knowledge-driven sequence matching approach due to data constraints. We contrast the two approaches and the data required to achieve sufficient performance and flexibility.

Neural-Based Adversarial Encryption of Images in ECB Mode

P. Rivas-Perea, Prabuddha Banerjee
School of Computer Science and Mathematics, Marist College, Poughkeepsie, New York, USA

Abstract: Digital images possess rich features that are highly correlated among neighboring pixels and highly redundant bits. Thus, the study of encryption algorithms over images is a subject of study to determine encryption quality. In this paper, we study the applicability of neural networks in learning to use secret keys to protect information (e.g. images) from other neural networks. We implement a well-known adversarial neural network architecture that is designed to learn to establish its own encryption method to protect itself from an attacker that is also a neural network. Our preliminary results suggest that this type of neural architecture is capable of securing communications in the midst of a neural-based attacker using only 16-bit blocks.

Hybrid Resource Scheduling Scheme for Video Surveillance in GPU-FPGA Accelerated Edge Computing System

Gyusang Cho, Seong-Hwan Kim, Chan-Hyun Youn
School of EE, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

Abstract: Video surveillance system with object reidentification are cited as a challenge to address to enhance the safety and convenience of citizens. The system consists of a combination of complex tasks requiring a lot of computing workload. With these characteristics, efforts have continued to accelerate the system. Existing systems did not benefit from the service latency perspective to make good use of heterogeneous accelerated edge computing system. In this paper, the goal is to accelerate the system used in smart cities on limited heterogeneous edge servers, and the scheduling planning method considering them is proposed. We first identify the computational volume-based execution time model of the heterogeneous accelerators. Then, we propose a scheduling plan that distributes this task graph to resources. Finally, the planning method proposed in this paper is experimentally compared with the previous GPU-FPGA allocation scheme. We compare it to the previously proposed method, and show that queue latency can be reduced, with showing robustness to the deadline violation rate.

Brain Tumor Segmentation Using Deep Neural Networks and Survival Prediction

Xiaoxu Na, Ma Li, Mariofanna Milanova, Mary Qu Yang
MidSouth Bioinformatics Center & Joint Bioinformatics Program of University of Arkansas at Little Rock & University of Arkansas Medical Sciences, Arkansas, USA; Computer Science, University of Arkansas at Little Rock, Arkansas, USA

Abstract: Magnetic resonance image (MRI) is widely applied to the brain tumor diagnosis and treatment. Approximately 35 million MRIs are performed annually worldwide in recent years. Manual segmentation and extraction of the tumor area from MRIs are time-consuming. The emergence of deep-learning algorithms offers new opportunities to automate the medical images processing and analysis with high accuracy. In the study, we built deep-learning models for brain tumor segmentation. The MRI files were first preprocessed through the procedures including reorientation, de-noising, bias-correcting, skull stripping and co-registration. Then, the two deep-learning algorithms, DeepMedic and 3D U-Net, were used for tumor segmentation model construction. Different from the sequential DeepMedic model, 3D U-net has an encoding and decoding patch that allows shortcut connections from layers with equal resolution in the encoding path to the layers in the decoding path and can provide the high-resolution features. The dice coefficient (DICE), a most commonly used metric for validating medical volume segmentations, was adopted for performance evaluation. Our DeepMedic model achieved DICE of 0.802, whereas 3D U-Net achieved 0.876 for overall segmentation. Moreover, we built a linear regression model using shape features including size and surface area of different segmented tumor tissue sections from the results of 3D U-Net model along with clinical data of age and gross total resection status for patient survival prediction. Compared to using the clinical data alone, we found that combining shape features improved the prediction of overall survival by 7%. We further increased the overall survival prediction accuracy by an additional 9% by replacing the shape features that were not significantly correlated with survival with some selected texture. Our work provided models for automated brain tumor segmentation and patient survival prediction.

Automatic Generation of Descriptive Titles for Video Clips Using Deep Learning

Soheyla Amirian, Khaled Rasheed, Thiab R. Taha, Hamid R. Arabnia
University of Georgia, Athens, Georgia, USA

Abstract: It is nearly a decade that Deep Learning has penetrated in science to help human beings in many ways to speed up the very time-consuming tasks. For example, diagnosing diseases, financing, agriculture, search engines, robot vision, and many others. In this paper, we are proposing an architecture that utilizes image/video captioning methods and Natural Language Processing systems to generate a title or a concise abstract for a video that could be applicable for the cinema industry, search engines, supervision cameras, etc. We feed the video as frames to the captioning system. Then, the captioning system generates a caption or a document describing the objects and actions. Then, the NLP or text summarization system receives the document from the captioning system and generates a sentence that describes the entire video - Which could be assigned as a title to the video or be a concise abstract. The purpose of the proposed system is to make an automatic system that generates the title or abstract for a video to save time in time-consuming tasks.

An Incremental Learning Scheme with Adaptive Earlystopping for AMI Datastream Processing

Yungi Ha, Changha Lee, Seong-Hwan Kim, Chan-Hyun Youn
School of EE, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

Abstract: Streaming data on power usage delivered through the Advanced Metering Infrastructure (AMI) inherent the concept drift problem, in which the shape of the data changes over time. This phenomenon causes performance degradation in the processing of AMI data using deep learning models. In order to overcome this, updates of deep learning (DL) models using dataflow in real-time should be performed. Although there have been many studies tried to handle the issue so far, the problem with existing methodologies is that they have not fully considered the factors that affect the training efficiency of online learning in an environment where concept drift exists. In this paper, adaptive online learning techniques are proposed to solve the issue. This technique first determines batch size and epoch size by considering data integration latency and unstableness of the current DL model. Then it adjusts the epoch according to concept drift existence of current training input batch. By applying this technique, we were able to effectively reduce the load on learning time series power data using the DL model, while at the same time showing better performance in forecasting with newly incoming data.

Artificially Intelligent Cyber Security: Reducing Risk and Complexity

John Carbone, James Crowder
Forcepoint, LLC, USA; Baylor University, Texas, USA;
Colorado Engineering, Colorado Springs, Colorado, USA

Abstract: Historically, research shows analysis, characterization, and classification of complex heterogeneous non-linear systems and interactions have been difficult to accurately understand and effectively model. Synonymously, exponential growth of Internet of Things (IoT), Cyber Physical Systems, and the litter of current accidental and unscrupulous cyber events portray an ever-challenging security environment wrought with complexity, ambiguity, and non-linearity. Thus, providing significant incentive to industry and academia towards advanced, predictive solutions to reduce persistent global threats. Recent advances in Artificial Intelligence (AI) and Information Theoretic Methods (ITM) are benefitting disciplines struggling with learning from rapidly increasing data volume, velocity, and complexity. Research shows Axiomatic Design (AD) providing design and datum disambiguation for complex systems utilizing information content reduction. Therefore, we propose a comprehensive transdisciplinary AD, AI/ML, ITM, approach combining axiomatic design with advanced, novel, and adaptive machine-based learning techniques. We show how to significantly reduce risks and complexity by improving cyber system adaptiveness, enhancing cyber system learning, and increasing cyber system prediction and insight potential where today context is sorely lacking. We provide an approach for deeper contextual understanding of disjointed cyber events by improving knowledge density (KD) (how much we know about a given event) and knowledge fidelity (KF) (how well do we know) ultimately improving decision mitigation quality and autonomy. We improve classification and understanding of cyber data, reduce system non-linearity and cyber threat risk, thereby, increasing efficiency by reducing labor and system costs, and "peace of mind."

A Classifier of Popular Music Online Reviews: Joy Emotion Analysis

Qing-Feng Lin, Heng-Li Yang
National Cheng-Chi University, Taipei, Taiwan

Abstract: With the development of the web, a large amount of product reviews might accumulate within short time, and the problem of information overload is getting serious. Previous opinion analysis studies on online reviews have focused on finding out the positive and negative opinions on the function of commodities. But the comments, which indicate particular emotion of the reviewer, were rarely discussed. For hedonic products such as music that could invoke users' feelings, the emotional rules including such as "happiness" and "sadness" mining from the comments can further complement the traditional rules, such as "singer is good" and "cannot sing high pitch". Taking the example of joy emotion analysis, this study proposed a system structure to classify the emotional feeling of pop music online reviews. A prototype system was built. The experiment reports the satisfactory result, with F1 73.45%.

An Approach to Interactive Analysis of StarCraft: Broodwar Replay Data

Dylan Schwesinger, Tyler Stoney, Braden Luancing
Computer Science, Kutztown University, Kutztown, Pennsylvania, USA

Abstract: This paper presents an interactive approach for exploratory data analysis of StarCraft: Broodwar replay data. Our approach encodes the basic facts of StarCraft units into statements in first-order logic (FOL). The information from replay data is also encoded into FOL statements. This database of FOL facts, in conjunction with a logic programming language, enables a human user to easily query and augment the database with new facts. This approach encourages an interactive workflow to aid in the analysis of replay data. We present examples that demonstrate the utility of the approach.

Artificial Psychosocial Framework for Affective Non-Player Characters

Lawrence J. Klinkert, Corey Clark
Southern Methodist University (SMU), Guildhall, Plano, Texas, USA

Abstract: Video game designers express a need for a tool to create a human-like Non-Player Character (NPC) that makes contextually correct decisions against inputs performed at unforeseen times. This tool must be able to integrate with their current game engine, such as the Unreal Engine. To make an NPC more human-like, we use psychology theories, specifically, emotions in decision making, such as the ideas proposed by A. Damasio. This approach allows the psychology of an NPC to connect directly to the underlying intrinsic motivations that drive players. This paper presents the "Artificial Psychosocial Framework" (APF) and defines a new class of NPC known as an "Affective NPC" (ANPC). APF utilizes an ANPC's personality, perceptions, and actions to generate emotions, which in turn affects the social relationship they have with other ANPCs. Additionally, APF considers the generation of emotions as an ambivalent hierarchy and classifies the representation of a personality and social relation. APF can work in tandem with existing AI techniques, such as Fuzzy Logic, Behavior Trees, Utility Theory, and Goal-Oriented Action Planning, to provide an interface for developers to build emotion-centric game-play. APF is a library for designers to use as a plug-in to a familiar game engine, such as the Unreal Engine. On the macro-level, APF uses the cognitive appraisal theory of emotions. On the micro-level, psychological models are chained together to determine the emotional state and social relation. The personality of an ANPC is represented by the Big Five: "Openness, Consciousness, Extroversion, Agreeableness, and Neuroticism," also called the OCEAN model. The emotions of an ANPC is described by the "Ortony, Clore, and Collins" (OCC) model. The social relation between ANPCs uses the "Liking, Dominance, Solidarity, and Familiarity" (LDSF) model. This paper demonstrates APF by simulating a scenario from the "Dennis the Menace" setting, where each Dennis is an ANPC with a different personality mapped to a Myers-Briggs type indicator. The results demonstrate how a different personality profile will cause varying emotional states, responses, and actions for the two simulated ANPCs in the same simulated scenario.

Procedural Image Generation Using Markov Chain Wave Function Collapse

Pronay Peddiraju, Corey Clark
Southern Methodist University (SMU), Guildhall, Plano, Texas, USA

Abstract: Wave Function Collapse (WFC) is an iterative constraint solving algorithm that performs texture synthesis on input samples to generate procedural outputs. The two commonly used WFC implementations are the Overlapping WFC (OWFC) and Tiling WFC (TWFC) implementations. OWFC incurs a performance cost in identifying constraints whereas TWFC receives pre-determined constraints in the form of metadata. Predetermining the metadata, however, is a non-trivial process and requires substantial design time before the execution of the algorithm. The proposed Markov WFC (MkWFC) implementation aims to reduce the time involved during the metadata design stage while maintaining the performance benefits of the TWFC implementation. This is achieved by using Markov random fields to determine constraints from a set of input samples and introduces a pre-processing step to the TWFC implementation. By automating constraint identification, the MkWFC implementation reduces both the metadata generation time as well as the scope for human error. This paper compares TWFC against MkWFC on a set of 6 different procedural image generation problems using 3 sizes for inputs and outputs in a total of 720 trials. When comparing the MkWFC implementation against TWFC, there is an increase in the number of identified and used constraints from input and output respectively, which increases with image size. The performance of TWFC and MkWFC was compared by measuring their respective execution times for all trials. MkWFC was able to identify over 1.5 times more constraints for a 5x5 tiled image in 1.03 ms +/- 0.09 ms and almost 3 times more constraints in 25x25 tiled image in 28.19 ms +/- 2.58 ms. This is substantially faster than the TWFC methodology where these constraints have to be manually identified and entered into the meta-data file. The automated meta-data generation and nominal increase in execution time allows for MkWFC to scale where TWFC can not.

Machine Learning for Understanding the Relationship Between Political Participation and Political Culture

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

Abstract: How might machine learning be employed to help promote the ethos of democratic citizenship in popular society? This article uses machine learning tools to better understand the role political culture plays in determining levels of political participation among citizens spread across our democratic republic. Although many studies have utilized political culture to explore rates of political participation along class, economic, gender, and other such cleavage lines, many scholars focusing on this framework often neglect the importance of race. In what regional political culture are we more likely to find higher levels of political participation among White and Black Americans? We investigate this question. We also seek to determine whether Black Americans have their own political culture that transcends previous understandings of regional political subcultures? We show that Elazar's classification of regional political subcultures applies mostly to White Americans, and that political culture among Blacks do not adhere to geographical lines or region. Machine learning tools are deployed to sift through massive data and uncover patterns and structures embedded within it. The machine learning tools here were used to test model specification and will be used to improve its performance to better predict political participation among Americans. The information may be used by policymakers and concerned citizens in their effort to increase civic participation and combat political apathy.

Merging Deep Learning and Data Analytics for Inferring Coronavirus Human Adaptive Transmutability and Transmissibility

Jack Y. Yang, Xuesen Wu, Gang Chen, William Yang, John R. Talburt, Hong Xie, Qiang Fang, Shiren Wang, Mary Qu Yang
Graduate Colleges of Public Health and Medicine, Bengbu Medical University, Bengbu, Anhui, P. R. China;
Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, USA;
Interdisciplinary Engineering Ph.D. Program & Department of Industrial & Systems Engineering, Texas A & M University, College Station, Texas, USA; MidSouth Bioinformatics Center & Department of Information Science, Joint Bioinformatics PhD Program of University Arkansas at Little Rock & University of Arkansas for Medical Sciences, Arkansas, USA

Abstract: The World Health Organization declared a global pandemic of Covid-19 earlier this month on March 11, 2020. Upon originally animal hosted virus presented inside human bodies, over the time, human adaptive transmissibility can be induced, and once a huge number of humans get infected by the virus, its transmutability can be further enhanced to lead to a global pandemic due to nowadays ever faster flow of humans from place to place and much rapid transportation network around the world. With the advent of high-throughput of next-generation sequencing technology, virus genomes can be sequenced promptly, hence paved a way for us to develop deep learning approaches to predict human-adaptive transmutability and transmissibility. We develop a Convolutional Neural Network to impute the data and build an innovative Deep Learning model for predicting adaptive interactions with the host cells. Moreover, we construct a Heterogeneous Information Network integrating multilayer data in combination with the lineage information to infer host cell specific and lineage specific networks. The hypothesis underlying our approach is that viral mutation and the subsequent adaptive interaction networks drive the evolution of human adaptive transmutability. Hence, based on the novel representation embedded with heterogeneous information, a Simulated Annealing method is proposed to search for the global optimization to study the information fusion of communities at the interaction networks. Finally, using the fusion information, we propose a novel ranking-based clustering for master interaction detection. The results lead to better understand how viral transmutability affects the interaction networks to induce human adaptive transmissibility and better understand human adaptive transmissibility of the disease as well as advance the identification of drug targets for effective treatments.

Title of Session: Hardware Acceleration in Artificial Intelligence

Chair: Dr. Xiaokun Yang
College of Science and Engineering, University of Houston Clear Lake, Houston, Texas, USA

A LSTM and GANs based ECG Abnormal Signal Generator

Han Sun, Fan Zhang, Yunxiang Zhang
Binghamton University, State University of New York, New York, USA

Abstract: The Electrocardiogram(ECG), a recording of the electrical activity of the heart, is commonly used for cardiac analysis, but lack of abnormal ECG signal data restricts the development of high quality automatic auxiliary diagnosis. In this paper, we introduce an LSTM and GANs based ECG abnormal signal generator to alleviate the issue. By training with a small set of real abnormal signals, the proposed generator can learn and produce high quality fake abnormal signals. The fake signals are then combined with real signals to train abnormal ECG classifiers. We show that our method can significantly improve the ability of classifiers in recognizing the uncommon case with a low proportion in the database.

An IoT-Edge-Server System with BLE Mesh Network, LBPH, and Deep Metric Learning

Archit Gajjar, Shivang Dave, Xiaokun Yang
College of Science and Engineering, University of Houston Clear Lake, Houston, Texas, USA

Abstract: This paper presents a hardware architecture, IoT-Edge-Server, of a diverse embedded system including a wide variety of applications such as smart city, smart building, or smart agricultural farm. First of all, we improve computation time by integrating the idea of edge computing on Raspberry Pi and CPU which processes different algorithms. Second, the hardware processors are connected to a server that can manipulate the entire system and also possess storage capacity to save the system's important data and log files. Specifically, the hardware computes data from: 1) a non-standardized Bluetooth Low Energy (BLE) Mesh System, and 2) a Surveillance System. The BLE Mesh System has one master and three slave devices while the Surveillance System has a Passive Infrared Sensor (PIR) and a camera to detect.

An Edge Detection IP of Low-cost System-on-Chip for Autonomous Vehicles

Xiaokun Yang
College of Science and Engineering, University of Houston Clear Lake, Houston, Texas, USA

Abstract: This short paper proposes a demonstration of edge detection on field-programmable gate array (FPGA), enabling to detect the edge of 320 x 240 size of images at 1302 frames per second (fps). The future work is an integrated system-on-chip (SoC) with a low-cost bus architecture, a security engine, and an image/video processing data path including OV7670 camera and VGA-interfaced display. The end goal will be a demonstration and simulation on self-driving vehicle to detect obstacles at the network edge. The design of many intellectual properties (IPs) of the SoC have been made publicly available to serve research and teaching courses at University of Houston-Clear Lake (UHCL), as well as to bring together researchers in other universities with interests in integrated circuit design, robotics, and FPGA prototyping.

Advancing AI-aided Computational Thinking in STEAM (Science, Technology, Engineering, Arts, Math) Education (Act-STEAM)

Lei Wu, Hua Yan, Zhimin Gao, Anton Dubrovskiy, Xiaokun Yang, Andrew Yang, Bo Liu, Xiao Qin
Auburn University at Montgomery, Alabama, USA

Abstract: This paper presents a novel approach to revolutionize STEM education with the Computational Thinking (CT) and Active teaching model and its implementation for skill knowledge learning (Act-STEM). We examine the effectiveness of using computational thinking on students' understanding and competence in STEM discipline learning. The work is built upon two successful pilot projects that demonstrated a broad impact on student learning outcomes.

Realistic Drawing and Painting with AI Supported Geometrical and Computational Method (Fun-Joy)

Lei Wu, Hua Yan, Zhimin Gao, Anton Dubrovskiy, Xiaokun Yang, Andrew Yang, Bo Liu, Xiao Qin
Auburn University at Montgomery, Alabama, USA

Abstract: Studies show that at a young age, children unexceptionally exhibit intense interests in depicting and replicating objects from their external world. The most common expression is through drawing and painting. However, it is not an easy task for most of the population. Artistic expression, especially visual art, has long been prov en to be remarkably beneficial for STEM (Science, Technology, Engineering and Math) education [Tytler 2016] [Tyler-Wood 2010]. How to effectively impart such important skill knowledge to the majority of students who lack of innate artistic talent is an important and difficult task for researchers. We have developed an effective approach with software solution to help students who lack of visual art talent to effectively master the skill knowledge of realistic drawing and painting of objects in the world, including human, animal, plant, scene, architecture, machinery, design, etc. The preliminary result shows our approach is very promising.

Title of Session: Artificial Intelligence for Smart Cities

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

Training-Data Generation and Incremental Testing for Daily Peak Load Forecasting

Jihoon Moon, Sungwoo Park, Seungmin Jung, Eenjun Hwang, Seungmin Rho
School of Electrical Engineering, Korea University, Seongbuk-gu, Seoul, Republic of Korea;
Department of Software, Sejong University, Seoul, Republic of Korea

Abstract: Daily peak load forecasting (DPLF) plays a crucial role in unit commitment, security analysis, and scheduling of outages and fuel supplies in smart grid applications. Recently, various artificial intelligence-based DPLF models have been proposed for accurate electric load forecasting using sufficient datasets. However, if the available data are not sufficient for training, it is not easy to build an accurate DPLF model. Herein, we propose a novel DPLF scheme that can perform DPLF well even when the dataset for training is not sufficient. We first configured various input variables by preprocessing time and weather data, as well as the historical electric load. Then, we performed a time-series cross-validation to consider as many training datasets as possible. Simultaneously, we generated new input variables relevant to the daily peak load by using principal component analysis and factor analysis and considered them to build our DPLF model. To evaluate the performance of our approach, we employed it for day-ahead peak load forecasting and verified that our scheme can achieve better prediction performance than traditional methods.

Attention Mechanism for Improving Facial Landmark Semantic Segmentation

Hyungjoon Kim, Hyeonwoo Kim, Seongkuk Cho, Eenjun Hwang
School of Electrical Engineering, Korea University, Seongbuk-gu, Seoul, Republic of Korea

Abstract: Various services are being embodied for smart city using IT technologies. For instance, face recognition technology can be used for effectively search for missing persons or track criminals. The key to face recognition is to detect landmarks that are the main features of faces. Many studies have been done to detect them accurately and quickly, but there is still much to be done. Deep learning methods for facial landmark detection mainly have used Convolutional Neural Networks. Recently, a new attempt to go beyond CNNs is being tried by considering relationships among all pixels in the images to improve missing long-range context problem in CNNs. In this paper, we propose a scheme for improving the performance of facial landmark detection based on *attention* and show its performance through various experiments

A Person Re-identification Scheme Using Multiple Input Images and Cross-Input Neighborhood Differences

Hyeonwoo Kim, Hyungjoon Kim, Bumyeon Ko, Eenjun Hwang
School of Electrical Engineering, Korea University, Seongbuk-gu, Seoul, Republic of Korea

Abstract: Intelligent CCTV-based surveillance is becoming an essential element in smart cities. Despite the recent explosion of CCTV installed for security pur-poses is rapidly increasing, its monitoring still depends on people. Person re-identification is a technique to find an image in disjoint camera views that con-tains the previously detected pedestrian. Conventional methods for person re-identification used the similarity based on hand-crafted features and their perfor-mance heavily relies on lighting or camera angle. In recent years, deep learning-based methods have shown good performance in person re-identification. How-ever, deep learning-based models using two input images have a limitation that they cannot detect similarities and differences between images simultaneously. In this paper, we propose a model that calculates similarities and differences be-tween images simultaneously by extracting features from input of three images and reconstructing the extracted feature map.

Variational AutoEncoder-based Anomaly Detection Scheme for Load Forecasting

Sungwoo Park, Seungmin Jung, Eenjun Hwang, Seungmin Rho
School of Electrical Engineering, Korea University, Seongbuk-gu, Seoul, Republic of Korea

Abstract: Smart grids can optimize their energy management by analyzing data collected from all processes of power utilization in smart cities. Typical smart grids consist of diverse systems such as energy management system and renewable energy system. In order to use such systems efficiently, accurate load forecasting should be carried out. However, if there are many anomalies in the data used to construct the predictive model, the accuracy of the prediction will inevitably decrease. Many statistical methods proposed for anomaly detection have had difficulty in reflecting seasonality. Hence, in this paper, we propose VAE (Variational AutoEncoder)-based scheme for accurate anomaly detection. We construct diverse artificial neural network-based load forecasting models using different combinations of anomaly detection and data interpolation, and then compare their performance. Experimental results show that using VAE-based anomaly detection with a random forest-based data interpolation shows the best performance.

Prediction of Clinical Disease with AI based Multiclass Classification using Naive-Bayes and Random Forest Classifier

V. Jackins, S. Vimal, M. Kaliappan, Mi Young Lee
Department of Software, Sejong University, Seoul, Republic of Korea

Abstract: Healthcare practices include collecting all kinds of patient data which would help the doctor correctly diagnose the health condition of the patient. This data could be simple symptoms observed by the subject, initial diagnosis by a physician or a detailed test result from a lab. Thus, far this data is only utilized for analysis by a doctor who then ascertains the disease using his/her personal medical expertise. The Artificial Intelligence has been used with Naive Bayes classification and Random forest classification algorithm to classify disease datasets of Heart disease, to check whether the patient is affected by that disease or not. A performance analysis of the disease data for both algorithms is calculated and compared. The results of the simulations show the effectiveness of the classification techniques on a data set, as well as the nature and complexity of the data set used.

A Hybrid Deep Learning Approach for Detecting and Classifying Breast Cancer using Mammogram Images

K. Lakshminarayanan, Y. Harold Robinson, S. Vimal, Dongwann Kang
Department of CSE, Seoul National University of Science & Technology, Seoul, Republic of Korea

Abstract: Next to skin cancer the most common cancer disease found in women is the breast cancer and which is very rarely can found in men. This type of cancer will be developed on the upper part of the breast on the lobules which causes lobular carcinoma or in the milk ducts which causes ductal carcinoma which will be usually in form of tumor cells which is visible in mammogram x-ray image or feel like lumps. The cancer becomes deadly when the cell starts growing and it spreads and invade the skin tissues which are around the breast. About 15% of the cancer deaths are because of breast cancer as per the survey of WHO (World Health Organization). At the same time all the lumps cannot be as deadly as the breast cancer. So identifying the disease and classify its type in the early stage will reduce the impact. In this research we propose a hybrid deep learning based approach for detecting and classify the breast cancer cells in the early stage. The algorithm used in this proposed work is a hybrid approach which is a combination of Convolutional Neural Networks (CNN) and Random Forest (RF) Algorithm which are the most popular deep learning algorithms for detection of breast cancer. The mammogram dataset for this research was from various online dataset repositories. Experimental results show that the proposed method gains an overall accuracy of 98.6% with Sensitivity and specificity of 96.5% and 98% respectively.While compared with the other state of the art methods the superior performance of the proposed hybrid technique is visible. The overall system was implemented in Matlab 2018b software with deep learning toolbox.

Trend Analysis using Agglomerative Hierarchical Clustering Approach for Time Series Big Data

P. Subbulakshmi, S. Vimal, M. Kaliappan, Y. Harold Robinson, Mucheol Kim
School of CSE, Chung-Ang University, Dongjak-gu, Seoul, Republic of Korea

Abstract: To be provided later.

Food Type Recognition and Estimation of Calories Using Neural Network

R. Dinesh Kumar, E. Golden Julie, Y. Harold Robinson, Sanghyun Seo
School of Computer Art, College of Art & Technology, Chung-Ang University, Kyunggi-do, Republic of Korea

Abstract: .In the fast moving world obesity become the major health issue to the human beings.BMI defines the obesity when it is greater than 30 kg/m^2.Obesity leads to many disease like high cholesterol, liver failure, knee problems, diabetes and sometimes cancer. When the patient eats healthy food the obesity can be controlled. The obesity problem can be addressed when there is a system that monitor the food consumed by the patient automatically and gives the suggestion periodically to the patient in treatment of obesity. Many of the people find difficulty in monitoring their food intake periodically, due to less knowledge in nutrition and self control. In this paper identification of food type is made and estimation of calorie is done using MLP and proposed the results. Single food item types are considered previously, but here mixed food item types are considered.ROI (Region of Interest) is used to identify the mixed food item type. The next step includes feature extraction process. The extracted feature image is fed into MLP classification to classify the food image. The volume of the food is used to calculate the calories present in the food. The implementation is processed in MATLAB with 1000 fruit images containing 6 food classes with good accuracy. The automatic dietary control is made available for the diabetic patients.

Progression Detection of Glaucoma using K-Means and GLCM Algorithm

S. Vimal, Y. Harold Robinson, M. Kaliappan, K. Vijayalakshmi, Sanghyun Seo
School of Computer Art, College of Art & Technology, Chung-Ang University, Kyunggi-do, Republic of Korea

Abstract: Diabetes Mellitus (DM) is one of the main medical issues far and wide causing national financial weight and low personal satisfaction. People with diabetes have an extended possibility of glaucoma. The hurt brought about by glaucoma is irreversible. This can occur if abnormal vein growth, which can occur because of diabetic retinopathy, is the significant consequence of diabetic illness is diabetic retinopathy (DR), it will hinder the characteristic misuse of the eye which impacts the retina of diabetic individuals and the basic period of diabetic retinopathy can prompt changeless vision misfortune. The early discovery and observing of diabetic retinopathy is critical to forestall it or for compelling treatment, yet the issue related to early identification of Diabetic Retinopathy (DR) is minor changes on retinal fundus picture, it incorporates hemorrhages, exudates, red sore, cotton fleece spots, drusen and so forth. To early location or screening of changes on the retinal picture is exceptionally testing and tedious for ophthalmologists, as the size and shading changes are at first coordinated with neighborhood veins I retinal picture. So the Glaucoma is one of the most unsafe visual maladies, keeps on influencing and weight a huge area of our populace. Accordingly, it is basic to distinguish glaucoma at a beginning time ahead. The proposed frameworks have focused on the parameter cup to plate proportion (CDR) for identification of glaucoma, that might be the best methodology for building proficient, vigorous and precise computerized framework for glaucoma diagnosis, and this strategy advocates the utilization of half and half methodology of manual element making with profound learning. It holds the guarantee of improving the precision of glaucoma conclusion through the robotizedsystems.

Demand Response: Multi-agent System-based DR Implementation

Faisal Saeed, Anand Paul, Seungmin Rho, Muhammad Jamal Ahmed
Department Software, Sejong University, Gwangjin-gu, Seoul, Republic of Korea

Abstract: A successful implementation of DR (Demand Response) always depends on proper policy and their empower technologies. This paper proposed an intelligent multiagent system to idealize the residential demand response in distributed network. In our model, the primary stakeholders (Smart Homes and Retailers) are demonstrated as a multifunctional intelligent agents. Home agents (HAs) are able to predict and schedule the energy load. Both HAs and RAs are modelled to predict the real time pricing. We used LSTM model (artificial neural networks) to predict the electricity load and energy price. Simulation results presents present significant reduction in electricity payments.

T-SNE Based K-NN: A New Approach for MNIST

Muhammad Jamal Ahmed, Faisal Saeed, Anand Paul, Seungmin Rho
Department Software, Sejong University, Gwangjin-gu, Seoul, Republic of Korea

Abstract: K nearest neighbors (K-NN) is an ective lazy and instance-based learner. It is natural to scale the K-NN to big data. In this paper, we propose to conduct spectral clustering through t-Distributed Stochastic Neighboring Embedding (t-SNE) and then apply k nearest neighbor to improve its performance and accuracy on big data.

Short to Mid-Term Prediction for Electricity Consumption using Statistical Model and Neural Networks

Malik Junaid Jami Gul, Malik Urfa gul, Yangsun Lee, Seungmin Rho, Anand Paul
Department Software, Sejong University, Gwangjin-gu, Seoul, Republic of Korea

Abstract: Electricity is one of the key role players to build an economy. Electricity consumption and generation can affect overall policy of the country. This opens an area for some intelligent system that can provide future insights. Intelligent management for electric power consumption requires future electricity power consumption prediction with less error. These predictions provide insights for making further decisions to smooth line the policy and helps to grow economy of the country. Future prediction can be categorized into three categories namely 1) Long-Term 2) Short-Term and 3) Mid-Term predictions. For our study, we consider Mid-Term electricity consumption prediction. Dataset provided by Korea Electric power supply to get insights for metropolitan city like Seoul. Dataset is in time-series so we require to analyze dataset with Statistical and machine learning models that can support time-series dataset. This study provides experimental results from the proposed models. Our proposed models for provided dataset are ARIMA and LSTM, which look promising as RMSE for training is 0.14 and 0.20 RMSE for testing.

BI-LSTM-LSTM Based Time Series Electricity Consumption Forecast for South Korea

Malik Junaid Jami Gul, M. Hafid Firmansyah, Seungmin Rho, Anand Paul
Department Software, Sejong University, Gwangjin-gu, Seoul, Republic of Korea

Abstract: Electricity is playing an important factor to drive economy of the nation. Every country trying to find fuel resources alternative to gasoline. Electricity is the promising resource because of low carbon footprints as compared to other fuel resources. Right now, biggest electricity consumer are households and industries. Forecasting the need of the respective sectors, governments can decide the future direction. This can result in better planning. As the second phase of our project, we have test LST with Bi-LSTM to the check overall performance of the neural network model. Dataset provided by Korea Electric power supply to get insights for metropolitan city like Seoul. Dataset is in time-series so we require to analyze dataset with Time distributed machine learning models that can support time-series dataset. This study provides experimental results from the proposed models. Our model shows RMSE score of 0.15 on training and 0.19 for testing with tuning hyperparameters of the model to optimum level.

Title of Session: XX Technical Session on 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

Inside Blockchain and Bitcoin

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

Abstract: Blockchain and Bitcoin have both become a very important topic in academia as well as in the economic world. This is due to the main characteristics of this system: transparency, anonymity, security and low transactional costs. This paper studies the blockchain network, with special attention to Bitcoin. Its uses, advantages and challenges are all presented, with special focus in instability and volatility, trust and security, cyber attacks and hacking as well as illegal activities. Within the conclusion, the main future research trends are also analyzed.

Priority Management in a Cybernetic Organization: A Simulation Study

Puche J., Costas J., Ponte B., Pino R., De la Fuente D.
Department of Applied Economics, Faculty of Economics and Business, University of Burgos,
Plaza Infanta Dona Elena s/n, Burgos, Spain

Abstract: This extended abstract shows an overview of the development of a model-driven decision support system (DSS) to ensure efficient management of workload priorities in organizations. The DSS has been built through discrete-event simulation (DES) using guidelines dictated by Beer's Viable System Model (VSM). The prioritization policy considers four key factors: customer importance, nature of task, value of task, and window of opportunity. They can be appropriately combined by the use of the DSS according to the set of indicators defined by the organization.

A Model for the Strategic Management of Innovation and R&D based on Real Options Valuation: Options to Abandon and Expand Clinical Trials per Disease

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

Abstract: In this short research paper, we conceptually propose a valuation model that is suitable for assessing R&D projects and managing uncertainty in pharmaceutical laboratories. Taking this into consideration, this model would allow these organizations to take better strategic decisions that will affect the pipeline of clinical trials (per phase) and the portfolio of innovative drugs. More specifically, our valuation methodology would help decision makers to identify if and when to promote and abandon clinical trials and new drug developments. To this end, we adopt a real options valuation approach, which is combined with fuzzy techniques and simulation. Our proposal incorporates some additional features in relation to the previous literature that aim to make our model more adaptable to deal with real-world uncertainties in the development of new drugs.

Smart Marketing on Audio-visual content platforms - Intellectual Property Implications

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

Abstract: The exponential growth of multimedia content on the Internet has led to the generation of search engines and user-specific recommender systems. However, not all practices carried out through these systems are legal; a condition that will depend on the object of the search or recommendation and the country where such practice is carried out, presenting a special prosecution as far as certain aspects of intellectual property are concerned. In this paper we analyze the legality of search engines and recommender systems in audiovisual content platforms from the point of view of copyright and trademarks mainly, distinguishing the factors to be taken into account when incorporating into their system the titles of works that do not have in their catalog.

Title of Workshop: Intelligent Linguistic Technologies (ILINTEC'20)

Co-Chairs: 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 Mathematics in Research and 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: To be provided later.

Discourse Profiling based on Analytical Textology

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: To be provided later.

Contrastive Study of Spatial Constructions on NPloc in the Russian Language and NP in the Chinese Language in the Cognitive Aspect

Irene M. Kobozeva, Lie Dan
Moscow State University, Moscow, Russia;
Jilin Huaqiao Foreign Languages Institute, Changchun, P. R. China

Abstract: Establishment of locations is a very important task in the Named Entity Recognition technology. The given paper describes a contrastive study of the prepositional constructions with a spatial meaning in the Russian and Chinese languages. Basing on the principles of cognitive semantics, a spatial use of the Russian preposition «na» (meaning «on» in English) is investigated in combination with NP in the Prepositional case in the Russian language compared with the frame construction "在NP上" (zài NP shàng) in Chinese. For semantic representation of these constructions, the pictures of image schemata are used, i.e. the mental structures that reflect the spatial relationships of "Trajector" and "Landmark". The analysis of the markers compared in Russian and Chinese shows that although the underlying schemata they reflect are largely the same, there is no full translational equivalence between these markers. This is due not only to the fact that for one of the image schemata expressed by the Chinese construction in the Russian language the preposition «nad» («over») is used, but also by a number of other factors: competition of alternative schemes for the same referent spatial configuration and preference by the Russian language of the variant encoded by another preposition (lampa pod potolkom 'a lamp under the ceiling' > lampa pod potolkom 'a lamp on the ceiling'; dyrka v podoshve 'a hole in the sole' > dyrka na podoshve «a hole of the sole»), by a stronger influence of the topological type of a Landmak object and the nature of the Trajector movement on the choice of the preposition in the Russian language.

Methods and Algorithms for Generating Sustainable Cognitive Systems Based on Thematic Category Hierarchies for the Development of Heterogeneous Information Resources in Technological and Social Spheres

Michael M. Charnine, Elena B. Kozerenko
Institute of Informatics Problems of the Federal Research Center, Computer Science and Control of the Russian Academy of Sciences, Moscow, Russia

Abstract: To be provided later.

Mental Model of Educational Environments

Natalia R. Sabanina, Valery S. Meskov
Moscow Pedagogical State University (MPSU), Moscow, Russia

Abstract: To be provided later.

The 16th International Conference on Data Science (ICDATA 2020)
https://icdatascience.org/
https://americancse.org/events/csce2020

Revealing the Relation between Students' Reading Notes and Examination Scores with NLP Features

Zhenyu Pan, Yang Gao, Tingjian Ge
University of Massachusetts Lowell, Massachusetts, USA

Abstract: Predicting students' exam scores has been a popular topic both in educational psychology and data mining areas for many years. Currently, many researchers devote efforts to predict exam score precisely with student behavior data and exercise content data. In this paper, we present the Topic-based Latent Variable Model (TB-LVM) to predict the midterm and final scores with students' textbook reading notes. We compare the Topic-based Latent Variable Model with the Two-step LDA model. For TB-LVM, the standard deviations of the maximum likelihood estimation and method of moments for the midterm exam are 7.79 and 7.63, respectively. The two standard deviations for the final exam are 8.68 and 7.72, respectively. The results are much better than the results of the Two-step LDA model, which is 14.38 for the midterm exam and 16.55 for the final exam. Finally, we also compare with the knowledge graph embedding method to predict exam scores.

Meta-Learning for Industrial System Monitoring via Multi-objective Optimization

Parastoo Kamranfar, Jeff Bynum, Amarda Shehu
George Mason University, USA

Abstract: The complexity of data analysis systems utilizing machine learning for industrial processes necessitates going beyond model selection and learning over the entire system space. It posits learning over various algorithms considered for featurization and feature-based learning, the interplay of these algorithms, and the model space for each algorithm. This problem, often referred to as meta-learning, has not been addressed in industrial monitoring. In this paper, motivated a real-world problem of quantifying actuations of an industrial robot, we address meta-learning. Our contribution generalizes beyond the specific application; we propose a Pareto-based, multi-objective optimization approach that can be easily generalized to system diagnostics. A detailed evaluation compares solutions of this approach and other existing approaches and shows them superior in distinguishing movements of a robot from recorded acoustic signals.

Data Analysis for Supporting Cleaning Schedule of Photovoltaic Power Plants

Chung-Chian Hsu
National Yunlin University of Science and Technology, Taiwan

Abstract: To reduce the extent of dependence on nuclear power and thermal power, the government in Taiwan has aggressively provoking the use of green energy such as solar power and wind power in recent years. Solar energy has in fact become an indispensable part in human daily life in Taiwan. One critical issue on photovoltaic (or PV) power plant operation is to determine when to clean dirty solar panels caused by dust or other pollutants. Overly frequent cleaning can lead to excessive cleaning fee while insufficient cleaning leads to reduced production. In a tropical island-type climate in Taiwan, it rains frequently in some seasons, which results in cleaning effect of the dirty solar panels, referred to as natural cleaning in contrast to manual cleaning by maintenance personnel. In this paper, we investigate the panels cleaning issues in Taiwan under uncontrolled, operational configuration. We propose methods to estimate solar power loss due to dust on panels and further estimate the cumulative revenue loss. When the loss exceeds the cleaning fee, manual cleaning shall be scheduled. The preliminary result demonstrated that the proposed approach is promising.

A Comparison of Important Features for Predicting Polish and Chinese Corporate Bankruptcies

Yifan Ren, Gary Weiss
Fordham University, New York, USA

Abstract: This study generates data mining models to predict corporate bankruptcy in Poland and China, and then examines these models to determine the financial characteristics that are of the greatest predictive value. These financial features are then compared for the two countries. The study finds that there while there are some common financial indicators for bankruptcy between the two diverse financial markets, there are also key differences. In particular, asset-related features play a much larger role in predicting bankruptcy in China, while operations-related features play a larger role in predicting bankruptcy in Poland.

Improving Physician Decision-Making and Patient Outcomes Using Analytics: A Case Study with The World's Leading Knee Replacement Surgeon

Anish Pahwa, Shikhar Jamuar, Varun Kumar Singh, Matthew Lanham
Purdue University, USA

Abstract: Every year in the United States, more than 300,000 knee replacements are performed. According to Time magazine, this number is expected to increase by 525 percent by the year 2030. Although knee surgeries are highly effective, patients are still prone to post-surgery complications. We address this problem in collaboration with one of the world.s leading knee replacement technology company. We show how analysis of insurance codes, patient demographics, and health data can help better support physicians in the diagnosis phase, by assessing patients. risk of developing complications. We identified the factors that led to successful knee surgeries by utilizing classification algorithms. We developed a composite KPI to track surgery failure rates by combing 3 important factors. Namely, a number of post-op visits, direct complications from ICD codes, and whether a revision surgery has been carried out. Our study found factors such as BMI, smoking, blood pressure, and age were statistically significant parameters for determining a surgery outcome. The surgeon performing the surgery was also a significant factor in ascertaining the outcome. This could be due to the different techniques used by different surgeons. Our model can save millions of dollars per year by detecting two-thirds of actual complications that would occur. We believe healthcare providers and consulting firms who are developing analytics-driven solutions in the healthcare space will find our study novel and inspiring.

Integrated Plant Growth and Disease Monitoring with IoT and Deep Learning Technology

Jonathan Fowler, Soheyla Amirian
Department of Computer Science, Colorado Technical University, Colorado Springs, Colorado, USA;
Department of Computer Science, University of Georgia, Athens, Georgia, USA

Abstract: At present, the time, labor, and inaccuracies in plant and seedling care make feasibility a major concern in large-scale agricultural operations. Developments in Internet of Things (IoT) technology and image classification by deep learning have made it possible to monitor various aspects of plant conditions, but an integrated solution that combines IoT sensor data, high-resolution imagery, and manual intervention data in a synchronized time-series database environment has not yet been brought to market. In this paper, we propose such an integrated solution. The overall system architecture is outlined, as well as the individual components including sensors, drone imagery, image processing, database framework, and alerting mechanism. These components are brought together and synchronized in a time-series database. By synchronizing all the variables, this solution presents a comprehensive view and better means for intervention. Finally, opportunities for research and specific component improvements are identified.

Tornado Storm Data Synthesization using Deep Convolutional Generative Adversarial Network (DCGAN)

Carlos Barajas, Matthias Gobbert, Jianwu Wang
University of Maryland, Baltimore County, Maryland, USA

Abstract: Predicting violent storms and dangerous weather conditions with current models can take a long time due to the immense complexity associated with weather simulation. Machine learning has the potential to classify tornadic weather patterns much more rapidly, thus allowing for more timely alerts to the public. A challenge in applying machine learning in tornado prediction is the imbalance between tornadic data and non-tornadic data. To have more balanced data, we created in this work a new data synthesization system to augment tornado storm data by implementing a deep convolutional generative adversarial network (DCGAN) and qualitatively compare its output to natural data.

The Effectiveness of Pre-Trained Code Embeddings

Ben Trevett, Donald Reay, Nick Taylor
Heriot-Watt University, Edinburgh, Scotland, United Kingdom

Abstract: Few machine learning applications applied to the domain of programming languages make use of transfer learning. It has been shown that in other domains, such as natural language processing, that transfer learning improves performance on various tasks. We investigate the use of transfer learning for programming languages, focusing on two tasks: method name prediction and code retrieval. We find that transfer learning provides improved performance, as it does to natural languages. We also find that these models can be pre-trained on programming languages that are different from the downstream task language and even pre-training models on English language data is sufficient to provide similar performance as pre-training on programming languages.

Nested Named Sets in Database Retrieval

H. Paul Zellweger
ArborWay Labs, USA

Abstract: To be provided later.

Discovery of Urban Mobility Patterns

Ivan Dario Penaranda Arenas, Hugo Alatrista-Salas, Miguel Nunez-del-Prado Cortez
Pontificia Universidad Catolica del Peru

Abstract: The detection of mobility patterns is crucial for the development of urban planning policies and the design of business strategies. Some of the proposed approaches to carry out this task use surveys, registration data in social networks or mobile phone data. Although it is possible to infer through the latter the place of residence of the clients and estimating the probability of visiting a store, it cannot be guaranteed, based on this information, that a purchase was actually made. This paper develops a proposal for the discovery of urban mobility patterns by adapting the trade areas approach using bank transaction data. The main advantages of our approach are the estimation of the probability of purchasing in a shop showing the importance of taking into account the business category and including individuals of all social classes. Likewise, different metrics were used to determine commercial attractiveness according to the category to which the business belongs. Some of the real-world benefits of this work include, but are not limited to, serving as a tool to facilitate business decision making such as the location of a new retail store or the design of marketing campaigns.

Deep Learning Approach to Extract Geometric Features of Bacterial Cells in Biofilms

Md Hafizur Rahman, Jamison Duckworth, Shankarachary Ragi, Parvathi Chundi, Govinda Chilkoor, Venkata R. Gadhamshetty
South Dakota School of Mines & Technology, South Dakota, USA

Abstract: We develop a deep learning approach to estimate the geometric characteristics of bacterial biofilms grown on metal surfaces. Specifically, we focus on sulfate-reducing bacteria (SRB) which are widely implicated in microbiologically influenced corrosion (MIC) of metals; costing billions of dollars annually. Understanding the growth characteristics of biofilms is important for designing and developing protective coatings that effectively combat MIC. Our goal here is to automate the extraction of the shape and size characteristics of SRB bacterial cells from the scanning electron microscope (SEM) images of a biofilm generated at various growth stages. Typically, these geometric features are measured using laborious and manual methods. To automate this process, we use a combination of image processing and deep learning approaches to determine the geometric properties. This is done via image segmentation of SEM images using deep convolutional neural networks. To address the challenges associated with detection of individual cells that form clusters, we apply a modified watershed algorithm to separate the cells from the cluster. Finally, we estimate the number of cells as well as their length and width distributions.

Virtual Machine Performance Prediction Based on Transfer Learning of Bayesian Network

Wang Bobo
Yunnan University, P. R. China

Abstract: The performance of virtual machines in the cloud is fraught with uncertainty due to the complexity of the environment, which poses a challenge for accurate prediction of virtual machine performance. In addition, a well-performing virtual machine performance prediction model cannot be multiplexed in either the temporal or spatial dimensions. In this paper, we build a virtual machine performance prediction model based on Bayesian Network to solve the problem of accurate prediction of virtual machine performance. Furthermore, to achieve multiplexing of the performance prediction model in both temporal and spatial dimensions, we proposes a Bayesian Network transfer learning approach. Experiments show that our transfer learning approach, contrast with reconstruction, the amount of data in the training set was reduced by 90%, and the training time was reduced by 75%, while the macro average precision maintaining 79%.

Novel Community Detection and Ranking Approaches for Social Network Analysis

Matin Pirouz, Pujitha Reddy
California State University, Fresno, California, USA

Abstract: Enterprises are collecting, procuring, storing, curating and processing increasing quantities of Big Data. This facilitates the detection of new insights capable of driving more efficient and effective operations and provide management with the ability to steer the business proactively. Currently, Big Data Analysis blends a traditional statistical data analysis approach with computational methods. Since the dataset types are growing exponentially in popularity, various Community Detection and Community Ranking algorithms are developed for target marketing. Such analyses utilize the concepts of shortest path, closeness centrality, and clustering coefficient. In this study, we developed a community detection algorithm based on centrality and node closeness. We performed Exploratory Analysis, i.e. the graphical representation of data, to depict an interconnected collection of entities among people, groups, or products. We also performed Network Analysis (Community Detection and Ranking Algorithms) to analyze the relationships among the entities. The proposed algorithms were applied to multiple datasets and the hidden patterns of each were identified. Among the benchmark datasets, the algorithms were implemented on the American College Football and Karate Club datasets. We were able to predict the next matches, the most popular member of the club, and their relevant connections with a high accuracy as compared to the ground truth. In addition to the high accuracy, it encompasses all the features and predicts the importance of the community leader, which is a key differentiating factor for the propose algorithms. Modularity was used as the metric to compare the effectiveness of the proposed methods with state-of-the-art frameworks. The proposed Community Detection and Community Ranking algorithms outperformed the existing frameworks on scale-free networks. We were also able to identify the hidden patterns of friendships on social media, frequent itemsets purchased together, which can be used to develop recommendation systems to e-commerce portals.

An analysis of Flight Delays at Taoyuan Airport

Shwu-Min Horng, S. K. (Kyle) Hwang, Chia-Ling Chao
National Chengchi University, Taiwan

Abstract: This study aims to find trends and probabilities of factors resulting in on-time performance. This study uses two models to address the factors affecting flight delays from two different managerial viewpoints. Using flight data at Taoyuan Airport from 2014-2016, a linear regression is used to analyze delays in a detailed way, and allows airlines to draw comparisons to their peers. Secondly a data mining model uses association rules to find probabilities of flight delays which can be used from an airport's perspective to improve on-time performance. The models applied in this study show that operation factors such as flight origin and turnaround times, are related to and can affect delays. Regardless of which method is employed, results show that low cost carrier business models has successfully undercut their full-service carrier peers, even in a primary airport setting, to produce better on-time performace.

Extending Micro-Mobility Deployments: A Concept and Local Case Study

Zhila Dehdari Ebrahimi, Raj Bridgelall, Mohsen Momenitabar
North Dakota State University, North Dakota, USA

Abstract: Micromobility is a recent phenomenon that refers to the use of small human- or electric-powered vehicles such as scooters and bikes to travel short distances, and sometimes to connect with other modes of transportation such as bus, train, or car. Deployments in major cities of the world have been both successful and challenging. This paper reviews the evolution of micromobility services from shared bicycles, dockless systems, and shared electric scooters. The authors evaluated benefits, deficiencies, and factors in adoption to inform more rigorous and extensive geospatial analysis that will examine intersections with land-use, public transit, socio-economic demographics, road networks, and traffic. This work conducted exploratory spatial data analysis and correlation of publicly available datasets on land use, trip production, traffic, and travel behavior. Data from Washington D.C. served as a case study of best practices for scaling deployments to meet the social, economic, and mobility needs of the city.

Modelling and Analysis of Network Information Data for Product Purchasing Decisions

Md Asaduzzaman, Uchitha Jayawickrama, Samanthika Gallage
Staffordshire University, United Kingdom

Abstract: Technology has enabled consumers to gain product information from different product reviews and other digital media. Large manufacturers and retailers can make use of this network information to forecast accurately, to manage the demand and thereby to improve profit margin, efficiency, etc. This paper proposes a novel framework to model and analyses consumers' purchase decision for product choices based on information obtained from two different information networks. This model has also taken into account variables such as socio-economic, and demographic characteristics. We develop a utility-based discrete choice model (DCM) to quantify the effect of consumers' attitudinal factors from two different networks, namely, social network and product information network. The network information modelling and analysis are discussed in detail taking into account the model complexity, heterogeneity and asymmetry due to the dimension, layer and scale of information in each type of network. The likelihood function, parameter estimation and inference procedures of the full model are also derived for the model. Finally, extensive numeric investigations were carried out to establish the model framework.

Analyzing the Impact of Foursquare and Streetlight Data with Human Demographics on Future Crime Prediction

Fateha Khanam Bappee, Lucas May Petry, Amilcar Soares, Stan Matwin
Dalhousie University, Canada;
Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada

Abstract: Finding the factors contributing to criminal activities and their consequences is essential to improve quantitative crime research. To respond to this concern, we examine an extensive set of features from different perspectives and explanations. Our study aims to build data-driven models for predicting future crime occurrences. In this paper, we propose the use of streetlight infrastructure and Foursquare data along with demographic characteristics for improving future crime incident prediction. We evaluate the classification performance based on various feature combinations as well as with the baseline model. Our proposed model was tested on each smallest geographic region in Halifax, Canada. Our findings demonstrate the effectiveness of integrating diverse sources of data to gain satisfactory classification performance.

Hierarchical POI Attention Model for Successive POI Recommendation

Lishan Li
Tsinghua University, P. R. China

Abstract: The rapid growth of location-based social networks developed a large number of point-of-interests (POIs). POI recommendation task aims to predict users' successive POIs, which has attracted more and more research interests recently. POI recommendation is achieved based on POI context, which contains a variety of information, including check-in sequence pattern, POIs' textual contents and temporal characteristics. Existing efforts only model part of them, and loses valuable information of other aspects. In this paper, we propose a hierarchical POI attention model (HPAM), which jointly takes advantage of both POIs' text contents, temporal characteristics and sequential patterns. Specifically, HPAM proposes a lower-level POI representation layer to explore textual content with word attention mechanism, and a higher-level contextual sequence layer to depict the temporal characteristics with a temporal-level attention mechanism. Experimental results on a public dataset show that HPAM consistently outperforms the state-of-the-art methods. Experiment results on HPAM variants evaluate the effectiveness of the proposed multiple attention mechanisms.

Leveraging Insights from "Buy-Online Pickup-in-Store" Data to Improve On-Shelf Availability

Sushree Patra, Matthew Lanham, Shantam Mogali, Pranav Saboo, Sachin Arakeri, Zaid Ahmed
Purdue University, USA

Abstract: This research provides insights on how to leverage "Buy-Online Pickup-in-Store" data to understand customer preferences, demand patterns and when products go out-of-stock (OOS) to improve replenishment decisions for grocery chains. The motivation for this study is to reduce lost sales opportunities by improving on-shelf availability (OSA), subsequently improving the overall revenue and profitability of the retail store. In collaboration with a national grocery chain having over 240 stores in the USA, our team developed and assessed different predictive models to improve on-shelf availability rate. The solution uses various product categories based on the grocer's business segments, and then specific predictive models are implemented to predict stockouts for each category. While some research has been performed in this area, our work is novel in how OOS data from brick-and-click is utilized to advise the grocery stores on timely replenishment of stock to reduce overall lost sales. This research aims to evaluate and compare multiple classification algorithms for predicting OOS at a store-product level. Subsequently, the study performed an in-depth analysis to ascertain which business segments rendered better prediction accuracy.

Dynamic Pricing for Sports Event Ticket

Wei-Cheng Chen, Matthew Lanham, Ziyun Huang, Wenyin Huang
Purdue University, USA

Abstract: This research studies the market demand for sports tickets of a major NFL team and develops a dynamic pricing model for the price of the tickets based on the understanding of the market demand. The authors utilized R together with packages like h2o and ggplot2 to develop predictive models that could reflect future demand of tickets and then developed an optimization strategy based on this model for the use of dynamic pricing. A Tableau dashboard was also created using simulation data from one of the previous games to demonstrate the potential revenue increase of this model.

Performance Analysis of Deep Neural Maps

Boren Zheng, Lutz Hamel
University of Rhode Island, USA

Abstract: Deep neural maps are unsupervised learning and visualization methods that combine autoencoders with self-organizing maps. An autoencoder is a deep artificial neural network that is widely used for dimension reduction and feature extraction in machine learning tasks. The self-organizing map is a neural network for unsupervised learning often used for clustering and the representation of high-dimensional data on a 2D grid. Deep neural maps have shown improvements in performance compared to standalone self-organizing maps when considering clustering tasks. The key idea is that a deep neural map outperforms a standalone self-organizing map in two dimensions: (1) Better convergence behavior by removing noisy/superfluous dimensions from the input data, (2) faster training due to the fact that the cluster detection part of the DNM deals with a lower dimensional latent space. Traditionally, only the basic autoencoder has been considered for use in deep neural maps. However, many different kinds of autoencoders exist such as the convolutional and the denoising autoencoder and here we examine the effects of various autoencoders on the performance of the resulting deep neural maps. We investigate five types of autoencoders as part of our deep neural maps using three different data sets. Overall we show that deep neural maps perform better than standalone self-organizing maps both in terms of improved convergence behavior and faster training. Additionally we show that deep neural maps using the basic autoencoder outperform deep neural maps based on other autoencoders on non-image data. To our surprise we found that deep neural maps based on contractive autoencoders outperformed deep neural maps based on convolutional autoencoders on image data.

Using Matrix Factorization and Evolution Strategy to Develop a Latent Factor Recommendation System for an Offline Retailer

Shwu-Min Horng, Chia-Ling Chao, Yuan-Yao Chang
National Chengchi University, Taiwan

Abstract: Recommendation systems have been developed for online services with a simple product mix. This study extended its application to an offline retailer with a more complex product mix. Purchasing records of two thousand members within one year from an offline retailer in Taiwan were used as the dataset for the study. Datasets of the first nine months were used for training and models were tested by the records of the last three months. This study developed a recommendation system by integrating a matrix factorization to uncover latent factors from both customers and items, and an evolutionary program to optimize the parameter settings of duration adjustment functions that were applied to assign weights so that past purchasing records closer to the testing period would receive higher weights. The objective of the system is to predict the likelihood of customers' purchasing behavior toward items they never purchased during the training period. By measuring the average percentage-ranking of items for two thousand members, the recommendation system developed in this study outperformed two other approaches, popularity and item-based nearest neighborhood systems. In addition, academic and practical contributions were also discussed.

Time Series Modelling Strategies for Road Traffic Accident and Injury Data: A Case Study

Ghanim Al-Hasani, Md Asaduzzaman, Abdel-Hamid Soliman
Staffordshire University, United Kingdom

Abstract: The paper aims to provide insights of choosing suitable time series models and analysing road trac accidents and injuries taking road tracc accident (RTA) and injuries (RTI) data in Oman as a case study as the country faces one of the highest numbers of road accidents per year. Data from January 2000 to June 2019 from several secondary sources were gathered. Time series decomposition, stationarity and seasonality checking were performed to identify the appropriate models for RTA and RTI. SARIMA (3, 1, 1)(2, 0, 0)(12) and SARIMA (0, 1, 1)(1, 0, 2)(12) models were found to be the best for the road tracc accident and injury data, respectively, comparing many different models. AIC, BIC and other error values were used to choose the best model. Model diagnostics were also performed to confirm the statistical assumptions and two-year forecasting was performed. The analyses in this paper would help many Government Departments, academic researchers and decision-makers to generate policies to reduce accidents and injuries.

GFDLECG: PAC Classification for ECG Signals Using Gradient Features and Deep Learning

Hashim Abu-gellban, Long Nguyen, Fang Jin
Texas Tech University, Texas, USA

Abstract: ECG signal classification is a popular topic in healthcare for arrhythmia detection. Recently, ECG signal analysis using supervised learning has been investigated with the goal to help physicians to automatically identify the Premature Atrial Complex (PAC) heartbeats. PAC may be a sign of underlying heart conditions, which may change to supraventricular tachycardia that increases the possibility of sudden death. In this paper, we propose a data-driven approach, GFDLECG, which is based on ECG behavior to detect abnormal beats. We extract further features from ECG using the gradient feature generation algorithm. We also build the classification model by utilizing the Gated Recurrent Unit (GRU) and the residual fully convolutional networks with GRU to learn long-short term patterns of ECG behaviors.

Optimizing Network Intrusion Detection using Machine Learning

Sara Nayak, Anushka Patil, Reethika Renganathan, Lakshmisudha Kondaka
SIES Graduate School of Technology, India

Abstract: Machine Learning (ML) techniques are essential in the detection of network attacks and enhancing network security. A device or software that recognizes any unusual pattern in the network and alerts the user about the same is an Intrusion Detection System (IDS). In this paper, we have described the use of ML classification algorithms on the UNSW-NB15 dataset, leading to the generation of a network intrusion detection model which classifies incoming traffic into malicious or non-malicious traffic and issues an alert to the user. We have implemented the following ML Algorithms - Support Vector Machine, Artificial Neural Network and One-Class Support Vector Machine with an average accuracy of 89.25%, 91.54% and 93.05% respectively. Two graphical user interfaces (online and offline versions) have been developed for the system. Thus, the paper proposes an optimized intrusion detection system that improves upon the existing intrusion detection systems which detect malicious packets in the network.

Estimating the Effective Topics of Articles and journals Abstract Using LDA And K-Means Clustering Algorithm

Shadikur Rahman, Umme Ayman Koana, Aras M. Ismael, Karmand Hussein Abdalla
Daffodil International University, Dhaka, Bangladesh

Abstract: Analyzing journals and articles abstract text or documents using topic modeling and text clustering becomes modern solutions for the increasing number of text documents. Topic modeling and text clustering are both intensely involved tasks that can benefit one another. Text clustering and topic modeling algorithms are used to maintain massive amounts of text documents. In this study, we have used LDA, K-Means cluster and also lexical database WordNet for keyphrases extraction in our text documents. K-Means cluster and LDA algorithms achieve the most reliable performance for Keyphrases extraction in our text documents. This study will help the researcher to make searching string based on journals and articles by avoiding misunderstandings.

Towards a Reference Model for Artificial Intelligence Supporting Big Data Analysis

Thoralf Reis, Marco X. Bornschlegl, Matthias L. Hemmje
University of Hagen, Germany

Abstract: This publication will introduce the reference model AI2VIS4BigData for the application domains Big Data analysis, AI and visualization. Without a reference model, developing a software system as well as other scientific and industrial activities in this topic field lack a common specification and a common basis for discussion and thus pose a high risk of inefficiency, reinventing the wheel and solving problems that have already been solved elsewhere. To prevent these disadvantages, this publication systematically derives the reference model AI2VIS4BigData with special focus on use cases where Big Data analysis, Artificial Intelligence (AI) and visualization mutually support each other: AI-powered algorithms empower data scientists to analyze Big Data and thereby exploit its full potential. Big Data enables AI specialists to comfortably design, validate and deploy AI models. In addition, AI's algorithms and methods offer the opportunity to make Big Data exploration more efficient for both, involved users and computing and storage resources. Visualization of data, algorithms and processing steps improves comprehension and lowers entry barriers for all user stereotypes involved in these use cases.

Implicit Dedupe Learning Method on Contextual Data Quality Problems

Alladoumbaye Ngueilbaye, Hongzhi Wang, Daouda Ahmat Mahamat, Roland Madadjim
School of Computer Science and Technology, Harbin Institute of Technology, Harbin, P. R. China;
Departement d'Informatique, Universite de N'Djamena, N'Djamena, Chad;
School of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

Abstract: Variety of applications such as information extraction, data mining, e-learning, or web applications use heterogeneous and distributed data. As a result, the usage of data is challenged by deduplication issues. To harmonize this issue, the present study proposed a novel Dedupe Learning Method (DLM) and other algorithms to detect and correct contextual data quality anomalies. The method was created and implemented on structured data. Our methods have been successful in identifying and correcting more data anomalies than current taxonomy techniques. Consequently, these proposed methods would be important in detecting and correcting errors in broad contextual data (Big Data).

Deep Metric Similarity Clustering

Shuanglu Dai, Pengyu Su, Hong Man
Stevens Institute of Technology, USA

Abstract: Effective data similarity measures are essential in data clustering. This paper proposes a novel deep metric clustering method with simultaneous non-linear similarity learning and clustering. Unlike pre-defined similarity measures, this deep metric enables more effective data clustering on high-dimensional data with various non-linear similarities. In the proposed method, a similarity function is firstly approximated by a deep metric network. The graph Laplacian matrix is introduced to make data cluster assignments. A stochastic optimization is then proposed to efficiently construct the optimal deep metric network, and calculate data similarity and cluster assignment on large-scale data set. For N data samples, the proposed optimization effectively reduces the computation of N^2 pairs of data to M^2 (M<<N) pairs at each step of the approximation. A co-training method is further introduced to optimize the deep metric network on a portion of semi-supervised data for clustering with targeted purposes. Finally, this paper shows theoretical connections between the proposed method and spectral clustering in subspace learning. This method is able to achieve about 20% higher accuracies than the best existing multi-view and subspace clustering methods on Caltech and MSRCV1 object recognition data sets. Further results on benchmark and real-world visual data show competitive performance of the proposed method over deep subspace clustering network, and many related and state-of-the-art methods.

A Holistic Approach for Determining Effective Promotional Product Groupings

Rahul Raj, Matthew Lanham, Vinitha Ravindran, Siddharth Harisankar, Mehul Zawar, Xuanming Hu
Purdue University, USA

Abstract: With companies across industries continually striving to get ahead of the competition, product pricing could be the deciding factor in driving or destroying the margins of a company. Promotional grouping of products is an effective pricing strategy used across multiple industries such as retail, healthcare, and many more. Promotional product groupings or bundling can be seen everywhere, from buffets served at restaurants to the suite of products sold together by MS Office. The fact that the component products are readily available means that bundling is one of the most flexible elements of product strategy. However, some caveats come with bundling, most of which stem from inadequate planning. Bundling could lead to the cannibalization of products that are not present in bundles. Furthermore, it could lead to customers not choosing to buy the desired product because she would have to buy the other product bundled with it. The study encapsulates the selection and creation of labels for promotional product groupings for individual SKUs of a consumer product goods company. The groupings are based on historical data of the company's incremental sales and competitors' sales data collected in the same time frame. Currently, product grouping analysis is done manually, which could be compromised by human error and an individual's unique decision framework that could be biased. A more pertinent issue faced is that the company would fail to recognize the life of a successful promotion. Failure to do so could lead to stagnant promotional groupings that would not only fail to gain traction with customers but also siphon off the already existing sales, eventually leading to the company being overtaken by its competitors and lose market share.In order to develop recommendations for an ideal product grouping strategy, the study initially delves into the existing promotional groupings of the company and compares it with those of its competitors. Detailed competitive analysis provides an idea of the company's success with its past bundling strategy. The study uses machine learning models to identify the drivers of a successful promotion and finally uses optimization to suggest an ideal bundling strategy that would maximize revenue.

A Personalized Recommender System Using Real-Time Search Data Integrated with Historical Data

Hemanya Tyagi, Matthew Lanham, Robin Jindal, Mohinder Pal Goyal, Dibyamshu Shrestha
Purdue University, USA

Abstract: With companies focusing intensively on customer experience, personalization and platform usability have become crucial for a company's success. Hence, providing appropriate recommendations to users is a challenging problem in various industries. We work towards enhancing the recommendation system of a timeshare exchange platform by leveraging real-time search data. Previously, the recommendation model utilized historical data to recommend resorts to users and was deployed online once a day. The limitation of this model was that it did not consider the real-time searches of the user, hence losing context. This directly impacted the click-through rate of the recommendations, and the users had to navigate the website excessively to find a satisfactory resort. We build a model such that it utilizes not only the historical transactional and master data but also the real-time search data to provide multiple relevant resort recommendations within five seconds.

Automated Prediction of Voter's Party Affiliation using AI

Sabiha Mahmud
Mercyhurst University, USA

Abstract: The goal of this research is to develop the foundation of a cross-platform app, Litics360, that helps political election campaigns utilize data-driven methods, and high-performance prediction models, to align candidates with voters who share similar socio-political ambitions. To attain this goal, the first step is to understand a voter's currently aligned political party affiliation, based primarily on historical records of their turnout at previous elections, and basic demographic information. This research paper aims to find a solution to this first step, by comparing varied performance measures to find a reliable prediction model from learning algorithms, including decision tree, random forest and gradient boosting machine XGBoost binary classifiers. Significant correlations between independent variables and the target prediction class, i.e., voter's registered party affiliation, contribute towards the development of an automated predictive ML model. The Ohio Secretary of State public voter database was used to collect voter demographics and election turnout data, then prepared using preprocessing methods, and finally used to identify the best performing ML model. Hyperparameter grid search with XGBoost proved to be the superior binary logistic classifier, reproducing a nearly perfect skillful model. Tracking the alignment between voters and PEC candidates the proposed future of Litics360; i.e., to develop an application that promotes a healthy and transparent platform for voters to communicate their socio-political grievances to PECs, enabling efficient appropriation of a PEC's funds and resources to engineer successful marketing campaigns.

Hyperparameter Optimization Algorithms for Gaussian Process Regression of Brain Tissue Compressive Stress

Folly Patterson, Raj Prabhu, Osama Abuomar
Mississippi State University, USA

Abstract: Traumatic brain injury (TBI) is modeled using in vitro mechanical testing on excised brain tissue samples. While such testing is essential for understanding the mechanics of TBI, the results can vary by orders of magnitude due to the varying testing condition protocols. Gaussian process regression (GPR) provides good predictive accuracy of the compressive stress state. Here, the efficacy of different search algorithms in optimizing GPR hyperparameters was evaluated. Bayesian optimization, grid search, and random search were compared. Grid search reached the minimum objective function in fewer iterations and the final regression model was comparable to that of Bayesian optimization in terms of RMSE and log likelihood in the prediction of compressive stress.

Competitive Pokemon Usage Tier Classification

Devin Navas, Dylan Donohue
Fordham University, USA

Abstract: This paper investigates competitive Pokémon usage tier classification given a Pokémon's stats and typing. Pokémon were classified into the usage tiers defined by the competitive battling website Pokémon Showdown based on their individual base stats, the sum of all their base stats (BST), and their number of type weaknesses and type resists. Classifications were done using Weka's J48 decision tree, Lazy IBk 1-nearest neighbor, and Logistic Regression algorithm. The algorithms were evaluated by the metrics of accuracy and precision. The results of this study give insight into what factors most impact a Pokémon's use on competitive teams, and could give future insights on how Pokémon may perform as other Pokémon fall in and out of popularity, and as more Pokémon are added in future games.

Mining Modern Music: The Classification of Popular Songs

Caitlin Genna
Fordham University, New York, USA

Abstract: The rising popularity of streaming services has made in-depth musical data more accessible than ever before and has created new opportunities for data mining. This project utilizes data from 19,000 songs made available by Spotify. Several data mining algorithms (including J48 Decision Trees, Random Forest, Simple K Means, NaiveBayes, ZeroR, and JRIP) were used to assess the data as a classification task with the target class being popularity. The data was pre-processed and the popularity class was split into two different schemes, both of which were used to train the aforementioned algorithms with the goal of attaining the highest possible classification accuracy. Once reliable models were produced, the best performing algorithms were used in conjunction with association algorithms and Information Gain evaluation in order to assess the importance of features such as key, acousticness, tempo, instrumentalness, etc. in the prediction of the popularity class. Through this lens certain groups of attributes emerged as indicators of what makes a song popular or unpopular, and relationships between the attributes themselves were revealed as well. Overall it was concluded that popular music does in fact have patterns and a formulaic nature, making the 'art' of creating music seem more like a science. However, within those patterns enough variation can be seen to account for different genres and musical moods that still persist in this era of pop music, and support the idea that as a modern musical community we still maintain some diversity.

How is Twitter Talking about COVID-19?

Jesus Leopoldo Llano Garcia, Hector G. Ceballos, Francisco J. Cantu
Tec de monterrey, Mexico

Abstract: The novel coronavirus COVID-19 (SARS-CoV-2) virus spread rapidly, both as a pandemic and as a viral topic of conversation. Social networks, especially after the boom of smartphones, completely revolutionised the speed and channels where information spreads. A clear example of this is how fast information spreads across Twitter, a platform famous for creating trends and spreading the news. This work focuses on the analysis of the overall opinion of the COVID-19 pandemic on Twitter. We attempted to study the polarity and emotional impression of the people applying a series of Natural Language Processing techniques to a total of 270,000 tweets identified as related to Covid-19.

Improving Model Accuracy with Probability Scoring Machine Learning Models

Juily Vasandani, Saumya Bharti, Deepankar Singh, Shreeansh Priyadarshi
Purdue University, West Lafayette, Indiana, USA

Abstract: Binary classification problems are exceedingly common across corporations, regardless of their industry, with examples including predicting attrition or classifying patients as high-risk vs low-risk. The motivation for this research is to determine techniques that improve prediction accuracy for operationalized models. Collaborating with a national partner, we conducted feature experiments to isolate industry-agnostic factors with the most significant impact on conversion rate. We also

use probability scoring to highlight incremental changes in accuracy while we applied several improvement techniques to determine which would significantly increases a model's predictive power. We compare five algorithms: XGBoost, LGBoost, CatBoost, and MLP, and an Ensemble. Our results highlight the superior accuracy of the ensemble, with a final log loss value of 0.5784. We also note that the highest levels of improvement in log loss occurs at the beginning of the process, after downsampling and using engineered custom metrics as inputs to the models.

Deep Ensemble Learning for Early-Stage Churn Management in Subscription-Based Business

Sijia Zhang, Peng Jiang, Azadeh Moghtaderi, Alexander Liss
Ancestry.com Operations Inc., USA

Abstract: Churn prediction provides the opportunity to improve customer retention via early intervention. Previous research on churn prediction focused on two types of methods, classification and survival analysis. The comparison and combination of algorithms in these two types have not been fully explored. In this paper, we explore two stacking models to combine predictive capabilities of XGBoost, RNN-DNN classifiers and survival analysis. We first apply a standard stacking model, where predictions from base learners are fed into a meta-classifier. Furthermore, we propose a novel ensemble model, Deep Stacking, that integrates neural networks with other models. We evaluate the stacking models for early stage churn prediction at Ancestry, the global leader in family history and consumer genomics, with metrics dictated by business needs.

Obstacle Detection via Air-Disturbance in Autonomous Quadcopters

Jason Hughes, Damian Lyons
Robotics and Computer Vision Lab, Fordham University, Bronx, New York, USA

Abstract: Autonomous drones can detect and avoid walls as the technology stands today but they can only do this with camera, ultrasonic or laser sensors. This paper highlights how data mining classification techniques can be used to predict which side of a drone an object is located from the air-disturbance created by the drone being near such an object. Data was collected from the drone's IMU while it flew near a wall to its immediate left, right and front. The IMU includes gyroscope, accelerometer, roll, pitch and yaw data. Position and barometer data was also collected. The data was then fed to NearestNeighbor, Gradient Boosting and Random Forest classifiers.

The Evaluation of Rating Systems in Online Free-for-All Games

Arman Dehpanah, Muheeb Faizan Ghori, Jonathan Gemmell, Bamshad Mobasher
DePaul University, Chicago, USA

Abstract: Online competitive games have become increasingly popular. To ensure an exciting and competitive environment, these games routinely attempt to match players with similar skill levels. Matching players is often accomplished through a rating system. There has been an increasing amount of research on developing such rating systems. However, less attention has been given to the evaluation metrics of these systems. In this paper, we present an exhaustive analysis of six metrics for evaluating rating systems in online competitive games. We compare traditional metrics such as accuracy. We then introduce other metrics adapted from the field of information retrieval. We evaluate these metrics against several well-known rating systems on a large real-world dataset of over 100,000 free-for-all matches. Our results show stark differences in their utility. Some metrics do not consider deviations between two ranks. Others are inordinately impacted by new players. Many do not capture the importance of distinguishing between errors in higher ranks and lower ranks. Among all metrics studied, we recommend Normalized Discounted Cumulative Gain (NDCG) because not only does it resolve the issues faced by other metrics, but it also offers flexibility to adjust the evaluations based on the goals of the system.

Ensemble Learning for Early Identification of Students at Risk from Online Learning Platforms

Li Yu, Tongan Cai
The Pennsylvania State University, University Park, PA, USA

Abstract: Online learning platforms has made knowledge easily and readily accessible for people, yet the ratio of students withdrawing or failing a course is relatively high comparing to in-class learning as students do not get enough attention from the instructors. We propose an ensemble learning framework for the early identification of students who are at risk of dropping or failing a course. The framework fuses student demographics, assessment results and daily activities as the total learning statistics and considers the slicing of data with regard to timestamp. A stacking ensemble classifier is then built upon eight base machine learning classification algorithms. Results show that the proposed model outperforms the base classifiers. The framework enables the early identification of possible failures at the half of a course with 85% accuracy; with full data incorporated an accuracy of 94.5% is achieved. The framework shows great promise for instructors and online platforms to design interventions before it is too late to help students to pass their courses.

Phoenix: A Scalable Streaming Hypergraph Analysis Framework

Kuldeep Kurte, Neena Imam, S. M. Shamimul Hasan, Ramakrishnan Kannan
Oak Ridge National Laboratory, USA

Abstract: We present Phoenix, a scalable hypergraph analytics framework for data analytics and knowledge discovery that was implemented on the leadership class computing platforms at Oak Ridge National Laboratory (ORNL). Our software framework comprises a distributed implementation of a streaming server architecture which acts as a gateway for various hypergraph generators/external sources to connect. Phoenix has the capability to utilize diverse hypergraph generators, including HyGen, a very large-scale hypergraph generator developed by ORNL. Phoenix incorporates specific algorithms for efficient data representation by exploiting hidden structures of the hypergraphs. Our experimental results demonstrate Phoenix's scalable and stable performance on massively parallel computing platforms. Phoenix's superior performance is due to the merging of high performance computing with data analytic.

Real-Time Spatiotemporal Air Pollution Prediction with Deep Convolutional LSTM through Satellite Image Analysis

Pratyush Muthukumar, Mohammad Pourhomayoun, Emmanuel Cocom, Jeanne Holm, Dawn Comer, Anthony Lyons, Irene Burga, Christa Hasenkopf
California State University, Los Angeles, California, USA

Abstract: Air pollution is responsible for the early deaths of 7 million people every year in the world. The first and the most important step in mitigating the air pollution risks is to understand it, discover the patterns and sources, and predict it in advance. Real-time air pollution prediction requires a highly complex model that can solve this spatiotemporal problem in multiple dimensions. Using a combination of spatial predictive models (deep Convolutional Neural Networks) and temporal predictive models (deep Long Short-Term Memory), we utilized the Convolutional LSTM structure that learns correlations between various points of location and time. We created a sequential encoder-decoder network that allows for accurate air pollution prediction 10 days in advance using data of 10 days in the past in the county of Los Angeles on a Nitrogen Dioxide metric. Through a 5D tensor reformatting of air quality satellite image data, we provide a prediction for Nitrogen Dioxide in various areas of Los Angeles over various time periods.

Comprehensive Performance Comparison between Flink and Spark Streaming for Real-Time Health Score Service in Manufacturing

Seungchul Lee, Daeyoung Kim, Donghwan Kim
Research, BISTel Inc., Republic of Korea

Abstract: We investigated two powerful streaming processing engines, Apache Flink and Apache Spark Streaming for real-time health score service in the perspective of computational performance. A health score is an important index because it represents machine's lifetime that is significant measurement for detecting machine's failures. Many services have attempted to adopt streaming processing engines in order to compute a health score in real-time but there is limited literature on studying streaming engine's performance for computing health scores in the aspects of the computational resources. In this paper, we extensively studied two main open-source streaming projects, Apache Flink and Apache Spark Streaming when it comes to obtaining health scores. To obtain the health scores, we equipped the deep-learning models with the streaming engine and evaluated the performance of the model calculation time for datasets with milliseconds interval. Especially, we tested our service with a dataset consisting of 10,000 assets to verify two stream engines could process large-scale level of the data sets, demonstrating the process of the massive sensor data records in real time. We anticipate that our study will be serve as a good reference for selecting streaming engines in computing health score.

An Improved Oversampling Method based on Neighborhood Kernel Density Estimation for Imbalanced Emotion Dataset

Gague Kim, Seungeun Jung, Jjyoun Lim, Kyoung Ju Noh, Hyuntae Jeong
Electronics and Telecommunications Research Institute (ETRI), Republic of Korea

Abstract: Classification problem of imbalanced dataset is one of the main research topics. Imbalanced dataset where majority class outnumbers minority class is more difficult to handle than balanced dataset. The ADASYN approach has tried to solve this problem by generating more minority class samples for a few samples around the border between two classes. However, it is difficult to expect good classification with ADASYN when the imbalanced dataset contains noise samples instead of real minority class samples around the border. In this study, to overcome this problem, a new oversampling approach deals with the probability that a minority class sample belongs to dangerous set, not noise samples by using kernel density estimation. The proposed method generates appropriate synthetic samples to train well the learning model for minority class samples. Experiments are performed on ECG dataset collected for emotion classification. Finally, the experimental results show that our method improves the overall classification accuracy as well as recall rate for minority class.

Detecting Asian Values in Asian News via Machine Learning Text Classification

Li-Jing Arthur Chang
Jackson State Univeristy, Jackson, Mississippi, USA

Abstract: The study is aimed at developing supervised machine learning models to automatically detect Asian values of harmony and support in Asian English-language newspaper articles. Harmony has two classes ("harmony" vs. "conflict"), with "harmony" defined as a void of conflict. Support has two classes ("supportive" vs. "critical"), with "supportive" defined as supporting economic, political and communal unity and strength at home. Nine algorithms, with their parameters tuned, were compared for performance. Results showed logistic regression is the top performer for both the "harmony" and "support" models, with 93.0% and 91.2% accuracy rates, respectively. Logistic regression models were then deployed through web pages to make classifications of unseen, unlabeled data. The testing of the deployed models demonstrated the utilities of the models developed.

The 21st International Conference on Internet Computing & IoT (ICOMP 2020)
https://americancse.org/events/csce2020/conferences/icomp20

Information Diffusion Models in Micro-blogging Networks Based on Hidden Markov Theory and Conditional Random Fields

Chunhui Deng, Siyu Tang, Huifang Deng
School of Computer Engineering, Guangzhou College of South China University of Technology, Guangzhou, P. R. China;
School of Computer Science and Engineering, South China University of Technology, Guangzhou, P. R. China

Abstract: There are few studies on information diffusion models in micro-blogging (Weibo) networks that take into account the contents, the users as well as network structure. And existing multi-information (a sequence of posted information pieces sorted by time) diffusion models are seldom considering both competition and collaboration among information. Therefore, we propose an information diffusion model based on the hidden Markov theory (IDMBHMT) and a multi-information diffusion model based on conditional random fields (MIDMBCRF). Further, we use the graph partition technique to optimize performance of these two models and applied them to users' reposting prediction. Finally, we analyze the factors that affect performance of our models via experiments. The results show that the optimized models outperform the reference models in comparison.

Comparative Study of Hybrid Machine Learning Algorithms for Network Intrusion Detection

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

Abstract: This paper introduces promising network intrusion detection techniques, mainly based on an ensemble of machine learning and feature dimensionality reduction algorithms. We applied hybrid techniques using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) as dimension reduction techniques, followed by LDA or Random Forest as classifiers. We then applied LDA, Random Forest, Naïve Bayes and Artificial Neural Network (ANN) as independent classifiers. A thorough investigation and comparative study of the models on the Kitsune family of datasets; representing different attacks; is then provided in terms of accuracy and confusion matrix elements. The experimental results and analysis indicate favorable results in network intrusion detection systems to identify security threats, especially when there is limited training data for imbalanced classes.

ISLSTM: An Intelligent Scheduling Algorithm for Internet of Things

Heng Wu, Shaofei Lu, Qinwen Zuo, Jon Musselwhite, Raj Vijeshbhai Patel, Sweya Reddy Dava
Department of Mathematics and Computer and Science, West Viginia State University Insititute, West Viginia, USA;
College of Electrical and Information Engineering, Hunan University, Changsha, P. R. China;
State Key Laboratory of NBC, Protection for Civilian, Beijing Institute of Chemical Defense, Beijing, P. R. China

Abstract: Ascheduler is a key component of Internet of Things and edge computing. Some scholars and papers proposed many schedulers and scheduler algorithms.In this paper, we propose a smart scheduling algorithm based on Long-Short Term Memory (LSTM) of Deep Learning in this paper. After experimental testing of the positions of the moving object and server performance, the algorithm we proposed in here can efficiently predict the positions of the moving object in the near future and the server performance at the new location.This greatly solves the problem of communication and computing bottleneck in the Internet of Things and edge computing.

The Implementation of Application for Comparison and Output of Fine Dust and Public Database using Fine Dust Sensor

YunJung Lim, WoongSik Kim, Yongsuk Kim
University of Konyang, Republic of Korea

Abstract: To be provided later.

Dynamic Clustering Method for Massive IoT Systems

Yunseok Chang
Department of Computer Engineering, Daejin University, Pocheon, Republic of Korea

Abstract: This study suggests a dynamic clustering method that can regroup, change, add or remove groups of specific IoT devices according to the user's demands without re-installing or reconfiguring the system in an environment where multiple IoT devices are collectively used. In the dynamic clustering method, IoT devices can be freely and logically connected to other cluster controllers, since IoT devices are connected to a cluster controller through a logical network. Since a cluster that performs a specific group control can be dynamically configured without changing a physical network connection, IoT devices included in the cluster can also be dynamically organized or logically relocated as needed. Through this dynamic clustering method, the user can change the control groups of IoT devices as necessary and perform group control. The dynamic clustering method requires a dynamic cluster controller that supports logical connections with IoT devices. The dynamic cluster controller physically connects several IoT devices. However, the actual function control is performed in a logically connected group unit through a PLM table shared by all cluster controllers. The PLM table is implemented so that the dynamic cluster controller can be composed of any logical group regardless of the physical connection. Through this method, the user can dynamically configure the logical segment suitable for the required situation. The dynamic clustering method proposed in this study can implement efficient and organic, dynamic group control as it can control the IoT devices of the cluster in logical group units. Thus the method can be developed as a vital technology to be refined and utilized as an essential advanced technology to construct an IoT environment that adapts the rapidly changing modern IT environment in real-time.

Per-user Access Control Framework for Link Connectivity and Network Bandwidth

Shogo Kamata, Chunghan Lee, Susumu Date
Graduate School of Information Science and Technology, Osaka University, Japan;
Toyota Motor Corporation, Japan;
Cybermedia Center, Osaka University, Japan

Abstract: Recently, a wide variety of Internet of Things (IoT) devices and their corresponding IoT applications have emerged. Such IoT applications have adopted some access control mechanisms on the IoT devices, but they do not consider access control of network connectivity to the IoT devices. In addition, since network resources such as links and bandwidth are used by a variety of users, access control must consider the security requirements and the attributes of users. In this study, we propose a per-user access control framework to network bandwidth and links as network resources. The proposed framework aims to control network connectivity to resources including IoT devices using network programmability achieved by Software Defined Networking (SDN). To realize fine-grained access control, we have built an access control framework based on a Role Based Access Control (RBAC) concept. In the evaluation, we investigated the feasibility and practicality of the proposed framework under different experimental scenarios. The simulation results show the proposed framework provides fine-grained dynamic access control based on user roles. Moreover, we found that the impact of authentication and authorization is small under our scenarios.

UnQuantize: Overcoming Signal Quantization Effects in IoT Time-Series Databases

Matthew Torin Gerdes, Kenny Gross, Guang Chao Wang
Oracle Physical Sciences Research Center, Oracle Corporation, San Diego, California, USA

Abstract: Low-resolution quantized time-series signals present a challenge to big-data Machine Learning (ML) prognostics in IoT industrial and transportation applications. The challenge for detecting anomalies in monitored sensor signals is compounded by the fact that many industries today use 8-bit sample-and-hold analog-to-digital (A/D) converters for almost all physical transducers throughout the system. This results in the signal values being severely quantized, which adversely affects the predictive power of prognostic algorithms and can elevate empirical false-alarm and missed-alarm probabilities. Quantized signals are dense and indecipherable to the human eye and ML algorithms are challenged to detect the onset of degradation in monitored assets due to the loss of information in the digitization process. This paper presents an autonomous ML framework that detects and classifies quantized signals before instantiating two separate techniques (depending on the levels of quantization) to efficiently unquantize digitized signals, returning high-resolution signals possessing the same accuracy as signals sampled with higher bit A/D chips. This new "UnQuantize" framework works in line with streaming sensor signals, upstream from the core ML anomaly detection algorithm, yielding substantially higher anomaly-detection sensitivity, with much lower false-alarm and missed-alarm probabilities (FAPs/MAPs).

A Network Traffic Reduction Method for Smart Dust IoT by Sensor Clustering

Joonsuu Park, Keehyun Park
Department of Computer Engineering, Keimyung University, Daegu, Republic of Korea

Abstract: In Smart Dust environments with millions of the Smart Dust nodes are scattered over an area, particularly in areas that are difficult to access by human, wherein controlling the Smart Dust nodes and gathering the sensed data are challenging tasks. In this paper, we propose a network traffic reduction method for a dust IoT system by sensor clustering as an attempt to solve the problem. We propose a method to lessen network traffic loads further and hence to alleviate the bottleneck problem by the clustering of Smart Nodes. The method proposed in this paper reduces the size of transmission data from a Relay Dust Device to the Smart Dust IoT Server by sending a single representative sensed data value, instead of sending the data from all the SDDs. The proposed method is based on the assumption that the Smart Dust Devices that are in very close proximity to each other may send very similar sensed data values. Experiments show that the transmission data size of the proposed work is reduced by as much as 24% to 26% of that of our earlier work.

The 19th International Conference on Wireless Networks (ICWN 2020) https://americancse.org/events/csce2020/conferences/icwn20

A Tool for the Analysis of MANET Routing Protocols based on Abstract State Machines

Alessandro Bianchi, Emanuele Covino, Giovanni Pani, Sebastiano Pizzutilo
Dipartimento di Informatica, Universita di Bari, Bari, Italy

Abstract: We introduce MOTION, a Java tool for modeling and simulating mobile ad-hoc networks (MANET); it allows to define the abstract state machine model of the network, to translate it into the ASMETA framework, and to simulate it, finally.

Autonomous Vehicle Security Model

Noha Hazzazi, Kevin Daimi, Hanady Issa
Department of Electrical Engineering and Computer Science, Howard University, Washington, USA;
Department of Electrical, Computer Engineering, and Computer Science, University of Detroit Mercy, Michigan, USA;
Electronics and Communications Department, Arab Academy for Science, Technology and Maritime Transport, Egypt

Abstract: Autonomous vehicle (AV) is capable of sensing its surroundings and progressing safely with no human input. The massive production and improvements of such vehicles have the aptitude to revolutionize transportation mobility and safety. Safety could suffer due to technical problems associated with design, manufacturing, or operation. Such problems can be alleviated with improved design and advanced manufacturing. However, if these issues are caused by a security attack, solving them maybe be time consuming and very costly as casualties are involved. Examples of some wrong decisions arising from security attacks include, moving to the wrong lane, not stopping at intersection, and not recognizing pedestrians. Therefore, autonomous vehicle's security needs to be enforced during the design and manufacturing processes and treated as a continuous effort. This paper suggests a security model based on cryptography. The perception, planning, and control are protected together with AV's communication with other vehicles and the infrastructure.

A Self-adaptivity Indoor Ranging Algorithm Based on Channel State Information with Weight Grey Prediction Model

Jingjing Wang, Joon Goo Park
Electronics Engineering, Kyungpook National University, Daegu, Republic of Korea

Abstract: With the development of Wi-Fi technology, the IEEE 802.11n series communication protocol and the subsequent wireless LAN protocols use multiple-input multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) and other technologies. Channel characteristics between Wi-Fi transceivers can be estimated at the physical layer and stored in the form of channel state information (CSI). In this paper, we propose a CSI based indoor ranging method using a grey prediction method that generates CSI predictions. This paper also provides experimental comparisons of our proposed data generation method with existing indoor ranging methods. Experimental results show that the proposed data generation approach achieves significant ranging accuracy improvement over using an ective CSI ranging method, while it incurs much less computational complexity. Meanwhile, the proposed method also can obtain more accurate ranging results in corner areas of the indoor environment where WIFI signals cannot be obtained.

WiFi Direct Issues and Challenges

Rabiah Alnashwan, Hala Mokhtar
Department of IT, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

Abstract: WiFi Direct is a promising technology for ad hoc connectivity that was standardized by the WiFi Alliance in October 2010. Its purpose is to allow device-to-device (D2D) direct communication that does not require a central access point (AP). It outperforms other existing ad hoc technologies and shows a good performance in terms of connecting a small number of devices in a group. It does not, however, provide a solution for connecting different groups to each other which limits the network size and range. It also suffers from its full dependence on the group owner (GO) that acts as a soft AP to the group. If the GO disconnects, the whole group is destroyed. The objective of this study was to identify the main limitations and challenges of WiFi Direct technology and analyse the existing solutions for these challenges. An in-depth analysis of the inter-group communication issue and its existing solutions was conducted. Group formation optimization techniques that could be applied were also investigated. Open research issues for the enhancement of WiFi Direct technology were also addressed.

Regular Plans with Differentiated Services using Cuckoo Algorithm

John Tsiligaridis
Mathematics and Computer Science, Heritage University, Toppenish, Washington, USA

Abstract: The management of a server's differentiated services is one of the top interesting roles in mobile technology. A server with cluster resources serves clients in groups of multiple classes according to their service level agreement. The server will have the ability to create a Regular Broadcasting Plan (RBP) after some data transformation for either a single or multiple channels. Based on theorems the Cuckoo Search (CS) algorithm discovers RBPs with delay differentiated services (CSDS). Given the delay ratio of just two services, the Extended version of CSDS (ECSDS), can discover the RBP with the smallest waiting time for low priority mobile services. These algorithms will enhance servers' ability for self-monitoring, and self-organizing adapting their services to predefined ratio parameters.

RFID Assisted Vehicle Navigation Based on VANETs

Yang Lu, Miao Wang
Department of Electrical and Computer Engineering, Miami University, Oxford, Ohio, USA

Abstract: With the trend of global urbanization, complex urban structure leads to the positioning inaccuracy issue in GPS or even failure for conventional navigation. Thanks to the developments of Vehicular ad hoc networks (VANETs) and RFID systems, the real time vehicular and positioning information can be collected to deal with the position inaccuracy problem, and the navigation system can be improved by local information exchanging with vehicles and infrastructure. In this paper, we proposed an RFID-enhanced VANETs system to provide accurate path planning, which is determined at a server in a real-time manner based on the shared vehicular information and positioning information. Specifically, the RFID system is utilized for reading accurate positioning information, and vehicular information can be collected and delivered to the server via VANETs. Simulations demonstrate that our system can perform a more accurate positioning and efficient path planning than the traditional GPS, ...

Using Multimodal Biometrics to Secure Vehicles

Kevin Daimi, Noha Hazzazi, Mustafa Saed
Department of ECE & CS, University of Detroit Mercy, Detroit, Michigan, USA;
Department of EECS, Howard University, Washington, USA;
HATCI Electronic Systems Development, Hyundai-Kia America Technical Center, Superior Township, USA

Abstract: Vehicle security is becoming increasingly critical and complex. Biometrics can be used for various identifications and authentications in vehicles. Multimodal biometrics offers complementary information to enhance the identification performance with regards to accuracy and reliability to overcome the shortcomings of using a single biometric. In this paper, multimodal biometrics are adopted for both vehicle entry and vehicle start. For vehicle entry, fingerprint and face biometrics are relied on. The multimodal biometrics; iris, face, and voice are used for vehicle start. The biometrics samples and templates are protected with cryptography. Both symmetric and asymmetric approaches are applied. Cryptographic protocols to protect various units of the security architecture in the proposed vehicle biometrics system are introduced. Drivers will have the option to override the use of biometrics should they prefer that. Furthermore, passwords will override the multimodal biometrics approach whenever the biometrics sensors malfunction.

A New Real-time Geolocation Tracking Tool Enhanced with Signal Filtering

Erkan Meral, Mehmet Serdar Guzel, Mehrube Mehrubeoglu, Omer Sevinc
Software Systems, Gazi University, Ankara, Turkey;
Computer Engineering, Ankara University, Ankara Turkey;
Computer Programming, Ondokuz Mayis University, Samsun, Turkey;
Department of Engineering, Texas A&M University-Corpus Christi, Corpus Christi, Texas, USA

Abstract: Global Positioning System (GPS) is a satellite network which transmits encoded data and makes it possible to pinpoint the exact location on Earth by measuring the distance between satellites and the receiver. As the GPS satellites constantly emit radio signals, appropriately designed receivers are able to receive these signals. In this study, a tool is proposed in which an electronic circuit that integrates SIM908 shield and Arduino card work as a GPS receiver enhanced by signal filtering. The positional data obtained from GPS satellites yield error due to the noise in signals. Here, Kalman and averaging filters are applied and compared to reduce the overall positional error to improve results. Several experiments have been carried out to verify the performance of the filters with the GPS data. The results of the proposed enhanced system are compared with the initial configuration of the system. The results obtained are quite encouraging, demonstrating improvement in both clear and cloudy weather. Although the averaging filter shows higher performance at the beginning using up to four data points, Kalman filter shows higher error improvement rate when more temporal data points are included in the filtering operations.

Extended RTS/CTS Control based on Transmission Request Distribution in Wireless Ad-Hoc Networks

Momoka Hara, Hiroaki Higaki
Department of Robotics and Mechatronics, Tokyo Denki University, Tokyo, Japan

Abstract: In a wireless ad-hoc network where wireless nodes exchange data messages without help of stationary base stations, collisions of control and data messages are reduced and/or avoided by CSMA/CA and RTS/CTS control of wireless LAN protocols. Random backoff timers for avoidance of collisions among RTS control messages provides equally opportunities to transmit data messages to neighbor wireless nodes since the value of the backoff timer monotonically decreases. In usual 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 transmissions and receipts requests for data messages, it is not always possible to receive a response CTS control message even though a wireless node 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 a novel RTS/CTS control to equally

provide opportunities to transmit data messages whose receiver node is hard to transmit a CTS control message on response to an RTS control message. Here, a transmission of a CTS control message precedes a transmission of an RTS control message in cases that transmissions of a CTS control message fail repeatedly.

Interference of Overhearing by Eavesdropper Nodes for Secure Wireless Ad-Hoc Networks

Hinano Amano, Hiroaki Higaki
Department of Robotics and Mechatronics, Tokyo Denki University, Tokyo, Japan

Abstract: In ad-hoc networks, data messages are transmitted from a source wireless node to a destination one along a wireless multihop transmission route consisting of a sequence of intermediate wireless nodes. Each intermediate wireless node forwards data messages to its next-hop wireless node. Here, a wireless signal carrying the data message is broadcasted by using an omni directional antenna and it is not difficult for an eavesdropper wireless node to overhear the wireless signal to get the data message. Some researches show that it is useful to transmit a noise wireless signal which collides to the data message wireless signal in order for interfering the overhearing. However, some special devices such as directional antennas and/or high computation power for complicated signal processing are required. For wireless multihop networks with huge number of wireless nodes, small and cheap wireless nodes without such special devices are mandatory for construction of the network. This paper proposes a novel method for interfering the overhearing by the eavesdropper wireless nodes by a routing protocol and a data message transmission protocol with cooperative noise signal transmissions by 1-hop and 2-hop neighbor wireless nodes of each intermediate wireless node. The results of simulation experiments show that the proposed intentional collision method provides enough coverage of noise wireless signals especially by help of part of 2-hop neighbor wireless nodes.

The 19th International Conference on Information & Knowledge Engineering (IKE 2020)
https://americancse.org/events/csce2020/conferences/ike20

Viability of Water Making from Air in Jazan, Saudi Arabia

Fathe Jeribi, Sungchul Hong
College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia;
Computer and Information Sciences Department, Towson University, Towson, Maryland, USA

Abstract: Water is very important resource in daily life activities of people. It is one of the most significant resources that individuals need every day. Water can be used for individuals, industry, or farms. Unfortunately, many regions of the world do not have sufficient water to use. Some areas like Jazan, Saudi Arabia do not have enough water and the necessary water is generated by desalination from sea water. This water can be supplied to cities and farms. However, this desalination plant should be located near the seashore to provide water to the inland and there should be transportation costs for that water. In this paper, water from air, a new water generation method, is proposed. The Jazan area has a lack of rainfall but it has relatively high humidity and constant wind. Using this wind and sun light, the electricity will be generated, and water will be generated from the electricity. The goal of this paper is to check the viability of generation of water from air.

Comparative Analysis of Sampling Methods for Data Quality Assessment

Sameer Karali, Hong Liu, Jongyeop Kim Indiana University Kokomo, Indiana, USA;
Southern Arkansas University, Arkansas, USA

Abstract: Data quality assessment is an integral part of maintaining the quality of a system. The purpose of implementing an assessment and accountability application is to make it easier to read and analyze a given data set. Data quality assessment builds trust in data, regardless of whether it was originally of poor quality or good quality. The assessment spares expenses and fixes issues before problems become more substantial and get out of hand. Some databases are prone to errors, and most contain a large number. These errors can mitigate fast analysis, which can cause financial issues for countries, companies, and hospitals. Thus, data quality assessment is seen as the best-suited technology for addressing data quality. It can handle the fast analysis of large scale data while reducing the risk of data errors. Clean data can help speed up analysis and detect valuable information, and data that is free from errors and issues become trusted data and helps speed up analysis. In this paper, we determined the sample size first. A control data set was then created to validate the proposed method. After this, four methods were used to obtain samples of different sizes. Then, based on the three quality dimensions, we compared the sampling results using statistical methods.

Concept into Architecture - A Pragmatic Modeling Method for the Acquisition and Representation of Information

Sebastian Jahnen, Wolfgang Bein, Stefan Pickl
Fakultat fur Informatik, Universitat der Bundeswehr Munchen, Neubiberg, Germany;
Department of Computer Science, University of Nevada, Las Vegas, Nevada, USA

Abstract: Process models are very important in a wide range of application areas, for example software development or operational management. The first step in modeling a process is gaining information and knowledge engineering. This information documented as process models are a very important part of Enterprise Architectures. All forces in the NATO use the NATO Architecture Framework (NAF). Process models in the NAF are represented as part of the operational view, the sub view NOV-5. Process models are often the starting point for modeling or creating an EA. According to the principles of proper modeling, not only the correct use of the syntax is necessary, but also the relevance of a model. This is inseparable from the underlying information from which the model is created. The present article deals with the creation of a modeling method that allows subject matter experts (SME) in the area to be modeled to get started with EA. The aim of presentation is to use the method to obtain the necessary information for the creation of the process model and represent the information in a human and machine-readable form. Also, goal of this contribution is to transform the information syntactically correctly as a NOV-5 process model, using a software developed as part of this contribution. The transformed NOV-5 model is similar to the originally representation of the gained information, so this enables the SME to check the created NAF-compliant model for correctness of content and to use the model without knowledge of the NAF.

Content-Based Image Retrieval Using Deep Learning

Tristan Jordan, Heba Elgazzar
School of Engineering and Computer Science, Morehead State University, Morehead, Kentucky, USA

Abstract: The problems of content-based image retrieval (CBIR) and analysis is explored in this paper with a focus on the design and implementation of machine learning and image processing techniques that can be used to build a scalable application to assist with indexing large image datasets. The CBIR application will be able to search large image datasets to retrieve digital images that are like pre-defined specifications such as a given digital image, or a given image type. The search is based on the actual contents of images and not the metadata of these images. Feature extraction techniques are used in this research project to analyze images and extract important features of images. The extracted features reflect the important characteristics of images that are related to contents (such as colors, shapes, edges, and textures) that can identify the image type. Supervised Machine learning techniques are used in this project to analyze these extracted features and to retrieve similar images in the form of a convolutional neural network. This network classifies images and using the statistics that made these classifications, similarities can be drawn between the query image and entities within the database. The developed CBIR algorithms were able to analyze and classify images based on their contents.

Nutrition Intake and Emotions Logging System

Tony Anusic, Suhair Amer
Department of Computer Science, Southeast Missouri State University, Cape Girardeau, Missouri, USA

Abstract: Unbalanced diet is a major risk for chronic diseases such as cardiovascular, metabolic, kidney, cancer and neurodegenerative diseases. Current methods to capture dietary intakes may include food-frequency questionnaires, 7-day food records and 24-hour recall. These methods are expensive to conduct, cumbersome for participants, and prone to reporting errors. This document discusses the development of a Personal Nutrition Intake and Emotions logging system. The following will be discussed: Establishing requirements and designing a simple interactive system, implementing a simple interactive system, performing data analysis and evaluation of the simple interactive system.

Geographical Labeling of Web Objects through Maximum Marginal Classification

Anjan Kumar K N, Satish Kumar T, J. Reshma
Department of CSE, RNS Institute of Technology, Bangalore, Karnataka, India;
Department of CSE, BMS Institute of Technology and Management, Bangalore, Karnataka, India

Abstract: Web search engines have become extremely popular in providing requested information to the user. The result set effectiveness of Web search engines has been continuously improving over the years. However, the documents of the result set may also contain irrelevant information having no importance to the user. So, the user has to spend some effort in searching for relevant information in these result set documents. To overcome this searching overhead, Web object search engines have been proposed. Such systems are built by extracting object information from various Web documents and integrating them into object repository. The user is provided with the facility to submit object search queries, and the required object information is retrieved. Unlike, Web search engines, providing results to geography specific queries is still in nascent stage for Web object search engines. Recently, Gaussian Mixture Model based technique for geographical labeling of Web objects was proposed in the literature. However, there is significant scope to improve the labeling accuracy results obtained in this technique. In this work, maximum marginal classifier-based technique for Web object geographical labeling is proposed. The advantages of this proposed technique are empirically exhibited on real world data set. This proposed technique, outperforms the contemporary technique by at-least 40% in labeling accuracy, and twice better in execution efficiency.

Improving Knowledge Engineering Through Inter-Organisational Architecture, Culture, Agility and Change in e-Learning Project Teams

Jonathan Bishop, Kamal Bechkoum
Centre for Research into Online Communities, e-Learning and Socialnomics, UK;
University of Gloucestershire, UK

Abstract: Project management is systematically conducted in order to achieve a deliverable outcome. In the case of information technology projects, it is the case that project failure is very common – this is also the case with those IT projects that involve the development of e-learning systems, whether they involve minimal or intensive use of ICTs. The aim of this study therefore is to propose an approach to project management that involves the creation of a toolkit so that people without a background in project management, or e-learning, can more successfully run projects within their organisation. The toolkit enhances through approach to knowledge engineering through tailoring project management methodologies to an organisation and its partner organisations.

A Human Resources Competence Actualization Approach for Expert Networks

Mikhail Petrov
SPIIRAS, Saint Petersburg, Russia

Abstract: Expert networks are popular and useful tool for many organizations. They organize the storage and use of information about employees and their skills. Correspondence of the data stored in expert networks to the real experts competencies is a crucial for project management. Irrelevant information may lead to unexpected the project performance results. Analysis of such results can help to keep the information up to date. The approach presented in this paper uses information about the projects results and its participants competencies to change experts competencies. The reference model and the algorithm used in the approach are described in this paper.

Automatic Brand Name Translation Based on Hexagonal Pyramid Model

Yangli Jia, Zhenling Zhang, Xinyu Cao, Haitao Wang
School of Computer Science, Liaocheng University, Liaocheng, P. R. China;
China National Institute of Standardization, Beijing, P. R. China

Abstract: Brand name translation is of great importance for international corporations when their goods entering exotic markets. In this article, we investigate the strategies and methods of brand name translation from west languages to Chinese, and propose a hexagonal pyramid brand name translation model, which provides a comprehensive summary of brand name translation methods and makes the classification of some translated brand names from vague to clear. Based on the model and similarity calculating, an efficient automatic translation method has been proposed to provide help of finding adequate translated words in Chinese. And an experiment has been done by the way of a dedicated program with results of a cluster of recommended Chinese brand words with a good potential to be used.

Smart Health Tracking Emotions System

Geetika Koneru, Suhair Amer
Department of Computer Science, Southeast Missouri State University, Cape Girardeau, Missouri, USA

Abstract: Many diseases such as mental health disorders, can be linked to variable mood. Applications for smart health are important allowing patients with abnormal health conditions to be monitored and provided rapid help. Advances in IOT plays a major role in the field of health care by empowering people to connect their health and wealth in a smart way. A simple Smart health tracking emotions system is developed and that records scheduled activities and student emotions over a period. It allows them later to have access to this data that can be analyzed. The emotions are measured based on the scale of emotions ranged from 0 to 10.

Human-Computer Interaction Interface for Driver Suspicious Action Analysis in Vehicle Cabin

Igor Lashkov, Alexey Kashevnik
SPIIRAS, Saint Petersburg, Russia

Abstract: The paper presents an approach for monitoring in-vehicle driver behavior focused on finding vulnerabilities in interaction interfaces between human and systems built up with artificial intelligence in transport environment. We propose a reference model aimed at analyzing driver suspicious actions, comprising explicit intentional driving behavior and implicit unintentional one. The former kind of actions refers to human attitude placed at a conscious level and are easy to self-report. While, the latter are human attitudes, that are at the unconscious level, are involuntarily formed and generally unknown. We develop the prototype of software application focused on suspicious actions detection in driver behavior is responsible for analyzing video stream recorded from the camera located in vehicle cabin.

Development and Evaluation of a Machine Learning Based Value Investing Methodology

Jun Yi Derek He, Joseph Ewbank
Academy of Science and Technology, The Woodlands College Park High School, Texas, USA

Abstract: The majority of approaches to utilize computers for fundamental analysis in stock investing is plagued with scalability and profitability issues. The present work tested four machine learning algorithms to overcome them. Random Forest and a soft voting ensemble obtained strong risk-reward ratios over 12 test years. An innovative process involving picking exclusively the top 20 stocks based on the algorithms' softmax confidence enhanced returns by approximately 30%. Methods such as comparing in sample and out of sample precision performance as well as distributions of test and training data suggested that the algorithm can be scaled to out of sample data. Finally, the returns of the algorithms were found to generally outperform most mutual and hedge funds.

A Resampling Based Semi-Supervised Learning Analysis for Identifying School Needs of Backpack Programs

T. M. Bashir, S. Kim, L. Liu, L. B. Davis
Department of Mathematics & Statistics, North Carolina Agricultural & Technical State University, Greensboro, NC, USA;
Department of Industrial & Systems Eng., North Carolina Agricultural & Technical State University, Greensboro, NC, USA

Abstract: School-based backpack programs, which supply children with food to take home on weekends and holiday breaks when school cafeterias are unavailable, have shown positive effects evidenced by families and schools' reports. However, some studies show some negative impacts when the program is not well designed and executed. To aid with the appropriate design of the program, this study focuses on statistical modeling of the limited data to identify schools with true needs of the backpack programs. The data in this study was collected from a backpack program organization and various public data websites in Guilford County, North Carolina. Utilizing various classification models such as logistic regression, Naive Bayes, decision tree, random forest, and support vector machine, a resampling-based semi-supervised learning (RSSL) method is developed and employed. After extensive numerical simulations, the proposed RSSL was able to create several ranking systems for the food backpack needs for schools in the county not currently being serviced by the backpack program. Random forest and support vector machine in the proposed RSSL outperformed other selected classifiers in reporting probabilities useful for decision making. The resampling-based semi-supervised learning method developed in this study can be easily implemented to analyze the same backpack programs in the other regions. The RSSL can be applied to other similar problems for the potential identification of misclassified labels with some slight or no modification and limited available data.

Using Entropy Measures for Evaluating the Quality of Entity Resolution

Awaad Alsarkhi, John R. Talburt
Department of Information Science, The University of Arkansas at Little Rock, Arkansas, USA

Abstract: This research describes some of the results from an unsupervised ER process using cluster entropy as a way to self-regulate linking. The experiments were performed using synthetic person references of varying quality. The process was able to obtain a linking accuracy of 93% for samples with moderate to high data quality. While results for low-quality references were much lower, there are many possible avenues of research that could further improve the results from this process. The purpose of this research to allow ER processes to self-regulate linking based on cluster entropy. The results are very promising for entity references of relatively high quality, using this process for low-quality data needs further improvement. The best overall result obtained from the sample was just over 50% linking accuracy.

Fake News Detection through Topic Modeling and Optimized Deep Learning with Multi-Domain Knowledge

Vian Sabeeh, Mohammed Zohdy, Rasha Al Bashaireh
Computer Science and Engineering Department, Oakland University, Rochester, Michigan, USA;
Electrical and Computer Engineering Department, Oakland University, Rochester, Michigan, USA

Abstract: Increased internet access has exacerbated the severity of fake news on social media leading to employing advanced deep learning methods using large-scale data. Most of these methods rely on supervised models, demanding a large volume of training data to avoid overfitting. This paper presents Fake news Identification using Bidirectional Encoder Representations from the Transformers model (BERT) with optimal Neurons and Domain knowledge (FIND), a two-step automatic fake news detection model. To accurately detect it, the FIND approach applies a deep transformer model such as the BERT with the large-scale unlabeled text corpus to facilitate the classification model, and the Latent Dirichlet allocation (LDA) topic detection model to examine the influence of the article's headline and the body individually and collaboratively. The proposed FIND approach outperforms the existing exBAKE approach in terms of 10.78% of the greater F-score.

GeoDataLinks: A Suggestion for a Replacement for the ESRI Shapefile

Vitit Kantabutra
Department of Electrical Engineering, Idaho State University, Pocatello, Idaho, USA

Abstract: The ESRI Shapefile system of geographical information storage was essential to the development of GIS near the end of the previous millennium. However, Shapefile is now frustrating to many professionals who use GIS because the simple array-based data structures used in Shapefile lack the organizational and expressive powers of modern object-oriented programming technology. Alternatives to Shapefile have been proposed. However, even the best of those extant proposed alternatives are based on older technologies such as the Relational database system. In this paper a new system of geographical information data storage is proposed that is superior to Shapefile as well as the best extant Shapefile alternatives in various important ways such as superior data organization and the ability to securely and naturally express more complex relationships amongst geographic entities. The new geographical data storage system, called GeoDataLinks, that is presented here is based on the author's Intentionally-Linked Entities database system (ILE).

Accuracy Evaluation: Apply Existing Different Prediction Methods for COVID-19

Sameer Karali, Hong Liu
School of Sciences and Informatics, Indiana University Bloomington, Bloomington, Indiana, USA;
School of Sciences, Indiana University Kokomo, Kokomo, Indiana, USA

Abstract: A coronavirus is a type of contagious virus that can infect the nose, upper throat, or sinuses. It can spread just like a viral infection. The China branch of the World Health Organization (WHO) received the first evidence of an unidentified virus behind several cases of pneumonia identified as COVID 19 in Wuhan. As an outbreak, primarily confined to China is now actually worldwide pandemic. John Hopkins University COVID-19 website [13], which collects information from domestic and also global health agencies, there have now been over 2 Million reported cases and 153,822 deaths. More than 200 countries and territories have reported the epidemic, with the US, Italy, and Spain suffering the most acute cases outside of China. In this period, the scientist trying best to bring vaccines to the market, which is a relatively slow process, and any new vaccination will have to go through multiple phases of efficacy and safety testing. Moreover, once we know a vaccine is ready, it needs to be manufactured to a sufficiently large scale to be used worldwide. Each vaccination is possibly about 18 months far. In this work, we proposed a framework for classifying patients and evaluating accuracy. We conducted data transformation, descriptive analysis, and inference analysis to understand the data characteristics of COVID-19. In addition, we evaluated three prediction methods, namely KNN algorithm, K-mean algorithm, and Decision Tree algorithm. The best algorithm we found is that KNN algorithm with high accuracy.

A Dynamic Data and Information Processing Model for Unmanned Aircraft Systems

Mikaela D. Dimaapi, Ryan D. L. Engle, Brent T. Langhals, Michael R. Grimaila, Douglas D. Hodson
US Air Force Institute of Technology (AFIT), USA

Abstract: Dynamic Data and Information Processing (DDIP) involves symbiotic control feedback through sensor reconfiguration to integrate real-time data into a predictive method for system behavior. Fundamentally, DDIP presents opportunities to advance understanding and analysis of activities, operations, and transformations that contribute to system performance, thereby aiding in decision-making and event prediction. Previously examined DDIP application domains include weather monitoring and forecasting, supply chain system analysis, power system and energy analysis, and structural health monitoring. Currently, there is limited existing work that implements DDIP in support of Unmanned Aircraft Systems (UAS). Developing a DDIP model for a UAS application could further demonstrate DDIP capabilities and provide information relevant to improving maintenance operations and resource management.

Improving Performance of Machine Learning on Prediction of Breast Cancer over a Small Sample Dataset

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

Abstract: The application of machine learning (ML) algorithms aim to develop prognostic tools that could be trained on data that is routinely collected. In a typical scenario, the ML algorithm based prognostic tool is utilized to search through large volumes of data to look for complex relationships in the training data. However, not much attention has been devoted to scenarios where small sample datasets are a widespread occurrence in research areas involving human participants such as clinical trials, genetics, and neuroimaging. In this research, we have studied the impact of the size of the sample dataset on the model performance of different ML algorithms. We compare the model fitting and model prediction performance on the original small dataset and the augmented dataset. Our research has discovered that the model fitted on a small dataset exhibits severe overfitting during the testing stage, which reduces when the model is trained on the augmented dataset. However, to different ML algorithms, the improvement of the model performance due to trained by the augmented dataset may vary.

Clearview, an Improved Temporal GIS Viewer and Its Use in Discovering Spatiotemporal Patterns

Vitit Kantabutra
Department of Electrical Engineering, Idaho State University, Pocatello, Idaho, USA

Abstract: A GIS must have correct, efficient temporal functionality to be an ective visualization tool for spatiotemporal processes. Nevertheless, the major commercial GIS software providers have yet to produce GIS software packages that consistently function correctly with temporal data. This paper explains some of the serious errors in these existing software packages. The paper then shows how an efficient and correct temporal GIS viewer called Clearview can be designed and implemented. We also look at how Clearview can uncover new spatiotemporal patterns and knowledge in an HGIS (Historical GIS) database, Harvard and Fudan Universities' CHGIS (China Historical GIS).

Utilizing Economic Activity and Data Science to Predict and Mediate Global Conflict

Kaylee-Anna Jayaweera, Jens Mache, Caitlin Garcia, Quinn Vinlove, Alain Kagi, Kris Gado
Lewis and Clarke College, USA

Abstract: The year 2020 has left many individuals finding that their lives are continually being changed based on the state of global circumstances. Some believe that these changes have given many citizens the opportunity to understand the interconnected nature of global actions and domestic consequences. Our preliminary hypothesis and research centers around the belief that an informed global population produces a safer, and better prepared, global society. It is our understanding that when individuals are able to reasonably prepare or expect conflict, early mediation and resource management can not only save tremendous funds, but also save numerous lives. We believe that creating a source of accessible predictive models is not only possible, but can be done without tremendous resource demand by tracking key pointers within the global economy and historic conflict triggers.

Data-driven Environmental Management: A Digital Prototype Dashboard to Analyze and Monitor the Precipitation on Susquehanna River Basin

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

Abstract: Streamlined access to data makes forecasting, monitoring, and timely action much easier for any organization. Whether in business, education, and even environmental protection, quick access to data defines the difference between achieving and not achieving a centralized goal. For the Susquehanna River Basin Commission (SRBC), data is key to their core mission, which is to enhance public welfare through comprehensive planning, water supply allocation, and management of the water resources of the Susquehanna River Basin. River basin management involves multiple stakeholders, and scientific management requires significant forecasting capabilities. This research built the requisite digital prototype dashboard to monitor the precipitation as one of the environmental features. In this work, we applied several Machine Learning techniques, visualization, and data mining to identify relationships of related environmental parameters as well as indicators to facilitate better decision making. It will help to develop decision-support tools and methods for governments and businesses that help them make better, more informed, more accurate, and more pragmatic decisions.

Title of Session:
The 2020 International Workshop on Electronics and Information Technology (IWEIT-2020)

Chair: Prof. Cheng-Ying Yang; University of Taipei, Taiwan

Do Sarcastic News and Online Comments Make Readers Happier?

Jih-Hsin Tang, Chih-Fen Wei, Ming-Chun Chen, Chih-Shi Chang
Department of Information Management, National Taipei University of Business, Taiwan

Abstract: Newsvendors cultivate online communities to encourage online users' comments worldwide. However, the incivility of online commenting is an important issue for both researchers and practitioners. This study focuses on the impact of news with and without online comments on readers' emotions. An online experiment was designed with news sarcasm (sarcastic vs. neutral) and comments (civil, uncivil, and none) to examine participants' emotions. Two pretests were administered to determine the target news and the incivility of online comments. Five hundred and twenty-nine subjects took part in the formal online experiment, and the results demonstrated both sarcasm in news and incivility in comments made readers significantly unhappy. The interaction effect between sarcasm in news and incivility in comments was also significant, implying that news might form a "frame" in readers' minds and influence how they judge comments and emotions. Implications and discussions are also included.

Image Resizing in DCT Domain

Hsi-Chin Hsin, Cheng-Ying Yang, Chien-Kun Su
Department of Computer Science and Information Engineering, National United University, Taiwan

Abstract: There is a high demand for effective resizing of images in order to preserve the region of interest as much as possible on various screens with different dimensions and aspect ratios. In addition, image compression is important in many applications. This paper presents a compressed domain image-resizing algorithm by converting the discrete cosine transform (DCT) blocks of the original image into the DCT blocks of the resized one. Experimental results show that the proposed method outperforms other existing algorithms.

The 24th International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV 2020) https://americancse.org/events/csce2020/conferences/ipcv20

Apple Leaf Disease Classification Using Superpixel and CNN

Manbae Kim
Department of Computer & Communications Engineering, Kangwon National University, Chunchon, Republic of Korea

Abstract: Apple leaf disease is one of the main factors to constrain the apple production and quality. It takes a long time to detect the diseases by using the traditional diagnostic approach, thus farmers often miss the best time to prevent and treat the diseases. The classification of apple leaf disease is an essential research topic. Compared with the conventional approaches depending upon either region segmentation or end-to-end-learning of full-image by a neural network, we propose superpixel-based disease classification. CNN (Convolutional Neural Network) is used as a classifier. Based on PlantVillage images, we show that the proposed method is able to match, achieving a 92.43% accuracy as well as F1-score of 0.93 compared with 98.28% and 0.98 of full-image.

A Deep Learning Framework for Blended Distortion Segmentation in Stitched Images

Hayat Ullah, Irfan Tahir, Kyungjin Han, Jong Weon Lee
Department of Software Convergence, Sejong University, Seoul, Republic of Korea

Abstract: The visual quality of stitched images is play an important role to provide the high-quality immersive viewing experience of Virtual Reality (VR) contents. There are several image stitching algorithms that generates panoramas from multiples image taken with different cameras and angle of views. The performance of these stitching algorithms can be measure by estimating the quality of generated panoramas. This paper presents a segmentation based Stitched Image Quality Assessment (SIQA) approach that capture the blended distortion in stitched images and segment the distorted region using binary mask. The segmented regions provide the location and total area of the distorted region. The results obtained from the experimental evaluation validate the reliability of our method for capturing the blended distortions in stitched images.

Clustering Method for Isolate Dynamic Points in Image Sequences

Paula Niels Spinoza, Andriamasinoro Rahajaniaina, Jean-Pierre Jessel
Department of Mathematics, Computer Science and Applications, University of Toamasina, Madagascar;
IRIT, REVA, Paul Sabatier University Toulouse, France

Abstract: In this paper, we propose an optimization of the a-contrario clustering method using the probabilistic Guillaume Khenchaff Measure (M_{GK}) quality technique. A-contrario is used for tracking salient objects in the scene in real time. This method analyzes the data contained in a motion vector, which contains the scattered optical flow accumulated points of interest. The aim of our study is to improve the first results obtained from the Number of False Alarm (NFA) criterion by using M_{GK} to bring together the group of points endowed with a coherent movement of the binary tree. The idea is to isolate dynamic points so that we can use static points in the future.

Intraocular Pressure Detection Using CNN from Frontal Eye Images

Afrooz Rahmati, Mohammad Aloudat, Abdelshakour Abuzneid, Miad Faezipour
Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT, USA;
Applied Science University, Amman, Jordan;
Department of Biomedical Engineering, University of Bridgeport, Bridgeport, CT, USA

Abstract: As high Intraocular Pressure (IOP) is the main cause of glaucoma which can result in irreversible vision loss, early detection is extremely important for prevention. This paper is a research work in-progress, built upon our prior work, to distinguish healthy eye images from high IOP cases using a deep learning approach solely from frontal eye images. We propose a novel computer vision-based technique using a convolutional neural network (CNN) to extract common features of high IOP and glaucoma cases automatically from frontal eye images. The dataset used in this work contains 473 normal and high IOP frontal eye images. The proposed idea has the potential to minimize the patient's presence at healthcare facilities and prevent glaucoma causes by automating early detection.

Utilizing Quality Measures in Evaluating Image Encryption Methods

Abdelfatah A. Tamimi, Ayman M. Abdalla, Mohammad M. Abdallah
Faculty of Science and Information Technology, Al-Zaytoonah University of Jordan, Amman, Jordan

Abstract: The metrics or measures used in evaluating image quality usually serve applications that aim at enhancing image appearance or at preventing the image quality from degrading after image processing operations. On the other hand, the main goal of cryptography applications is to produce unrecognizable encrypted images that can resist various kinds of attacks. Furthermore, cryptography must consider extra measures such as keyspace and key sensitivity. This paper discusses the most useful quality metrics used in image cryptography and explains the type of values that indicate good encryption according to each metric. These metrics include statistical analysis measures, sensitivity measures, and other metrics.

Object Detection and Pose Estimation from RGB and Depth Data for Real-time, Adaptive Robotic Grasping

Shuvo Kumar Paul, Muhammed Tawfiq Chowdhury, Mircea Nicolescu, Monica Nicolescu, David Feil-Seifer
Department of Computer Science and Engineering, University of Nevada, Reno, Nevada, USA

Abstract: In recent times, object detection and pose estimation have gained significant attention in the context of robotic vision applications. Both the identification of objects of interest as well as the estimation of their pose remain important capabilities in order for robots to provide effective assistance for numerous robotic applications ranging from household tasks to industrial manipulation. This problem is particularly challenging because of the heterogeneity of objects having different and potentially complex shapes, and the difficulties arising due to background clutter and partial occlusions between objects. As the main contribution of this work, we propose a system that performs real-time object detection and pose estimation, for the purpose of dynamic robot grasping. The robot has been pre-trained to perform a small set of canonical grasps from a few fixed poses for each object. When presented with an unknown object in an arbitrary pose, the proposed approach allows the robot to detect the object identity and its actual pose, and then adapt a canonical grasp in order to be used with the new pose. For training, the system defines a canonical grasp by capturing the relative pose of an object with respect to the gripper attached to the robot's wrist. During testing, once a new pose is detected, a canonical grasp for the object is identified and then dynamically adapted by adjusting the robot arm.s joint angles, so that the gripper can grasp the object in its new pose. We conducted experiments using a humanoid PR2 robot and showed that the proposed framework can detect well-textured objects, and provide accurate pose estimation in the presence of tolerable amounts of out-of plane rotation. The performance is also illustrated by the robot successfully grasping objects from a wide range of arbitrary poses.

Theoretical Applications of Magnetic Fields at Tremendously Low Frequency in Remote Sensing and Electronic Activity Classification

Christopher Duncan, Olga Gkountouna, Ron Mahabir
Computational Science and Informatics, George Mason University, Virginia, USA;
Computer Engineering, George Mason University, Virginia, USA;
Earth Systems and Geoinformation Science, George Mason University, Virginia, USA

Abstract: This study was conducted to demonstrate the potential for magnetic fields to serve as a method of remotely sensing electronic activity and to evaluate the potential for classifying the electronic activity. To demonstrate this, a radio frequency generator, antenna and oscilloscope were placed inside a Faraday cage with varying frequencies transmitted in the range of 1 to 1000 hertz. A standard radio frequency antenna and magnetic loop antennas were placed outside the Faraday cage and the results were compared to each other as well as natural ambient signals. The results showed positive detection of magnetic field activity outside of the Faraday cage in the transmitted frequencies, with no detection with the radio frequency antenna. They also demonstrated the inability of a Faraday cage to attenuate magnetic fields of objects inside the cage. Errors that produced anomalies in the first attempt served to further validate the collection of the data by generating positive detection on both antennas. Ultimately the use of magnetic field antennas to detect electronic activity demonstrated potential use in a radio frequency adverse environment.

Deep Image Watermarking with Recover Module

Naixi Liu, Jingcai Liu, Xingxing Jia, Daoshun Wang
Department of Computer Science and Technology, Tsinghua University, Beijing, P. R. China

Abstract: Image watermarking based on deep learning has been proposed in the last few years. Typical framework for image watermarking consists of embedding network, extracting network and attack stimulating module. Adversarial discriminator is sometimes used to make watermarked images much more similar to cover images. To improve the robustness of CNN-based work against attacks of different type and strength, we proposed a novel model, introducing recover module into the framework to compensate some damaged watermark information during attacks and improve extracting accuracy. For non-differentiable JPEG compression, we propose a new approximation approach based on previous methods. Experimental results show that the proposed model performs better than the state-of-the-art in bit accuracy of message extraction while the image quality does not deteriorate.

Computer-aided Industrial Inspection of Vehicle Mirrors Using Computer Vision Technologies

Hong-Dar Lin, Hsu-Hung Cheng
Department of Industrial Engineering and Management, Chaoyang University of Technology, Taiwan

Abstract: Vehicle mirrors can reflect objects behind the cars and play an important role in driving security. In manufacturing process of vehicle mirrors, certain tasks operated unusually will cause producing surface and profile defects on vehicle mirrors. Those appearance defects sometimes will severely have an impact on standard of the mirror reflection and grow the driving hazard. At traditional examination of vehicle mirrors in manufacturing process, almost all works are performed by human examiners. This study works toward investigating the automatic appearance defect detection of vehicle mirrors. We propose a defect enhancement technique based on Fourier high-pass filtering and the convex hull arithmetic to inspect appearance defects on vehicle mirrors. This approach only utilizes their own information of testing images to judge whether there are any irregular appearance changes without the need of standard patterns for matching. Experimental results show that performance of the Fourier based approach in the defect detection is effective and efficient.

Evolution of Convolutional Neural Networks for Lymphoma Classification

Christopher David Walsh, Nicholas Kenelm Taylor
Department of Computer Science, Heriot-Watt University, Edinburgh, Scotland

Abstract: There are more than 60 subtypes of Lymphoma. This diversity usually requires a specialised pathologist for diagnosis. We aimed to investigate the effectiveness of Artificial Neural Networks (ANNs) and Deep Learning at Lymphoma classification. We also sought to determine whether Evolutionary Algorithms (EAs) could optimise accuracy. Tensorflow and Keras were used for network construction, and we developed a novel framework to evolve their weights. The best network was a Convolutional Neural Network (CNN); its 10-fold cross-validation test accuracy after training and weight evolution was 95.64%. The best single run test accuracy was 98.41%. This suggests that ANNs can classify Lymphoma biopsies at a test accuracy higher than the average human pathologist. The EA consistently improved accuracy, demonstrating that they are a useful technique to optimise ANNs for Lymphoma classification.

Axial Symmetry Detection Using AF8 Code

Jimenez-Ibarra Cesar Omar, Vazquez-Martin del Campo Miguel, Sanchez-Cruz Hermilo
Universidad Autonoma de Aguascalientes, Departamento de Ciencias de la Computacion, Mexico

Abstract: Symmetry is a very common geometric feature in natural objects and, highly marked, in artificial objects; particularly, mirror symmetry is a relevant topic in fields such as computer vision and pattern recognition. Previous work on symmetry has shown that there are useful solutions to the problem of mirror symmetry detection; however, there are still important challenges to successfully model symmetry objects and properly detect multiple axes. And there is also the challenge of assigning a level of symmetry to quasi-symmetrical objects. In this paper we propose an algorithm that detects level and axes of symmetry for quasi-mirror and mirror symmetry in 2D contours represented by AF8 code.

Polyhedral Approximation for 3D Objects by Dominant Point Detection

Miguel Vazquez-Martin del Campo, Cesar Omar Jimenez-Ibarra, Hermilo Sanchez-Cruz, Mario Alberto Rodriguez-Diaz
Universidad Autonoma de Aguascalientes, Departamento de Ciencias de la Computacion, Mexico; ITA, Mexico

Abstract: A new method for polyhedral approximation is presented in this paper. The representation of 3D objects is a complicated task, this is the reason why the object is organized as a slices set. The proposed method takes the slices, obtains the chain code from the contour, and uses the existing context-free grammar method to find the dominant points from contour of each slice, obtaining a point cloud from the selected slices. These dominant points are strategically joined to create the approximate polyhedron of the object. Finally, we adapt an existing error criterion to evaluate the approximated polyhedron with the original object.

Superpixel-based Stereoscopic Video Saliency Detection Using Support Vector Regression Learning

Ting-Yu Chou, Jin-Jang Leou
Department of Computer Science & Information Engineering, National Chung Cheng University, Chiayi, Taiwan

Abstract: In this study, a superpixel-based stereoscopic video saliency detection approach is proposed. Based on the input stereoscopic video sequences containing left-view and right-view video sequences, a sequence of right-to-left disparity maps are obtained. First, the simple linear iterative clustering (SLIC) algorithm [15] is used to perform superpixel segmentation on all video frames. Second, the spatial, temporal, depth, object, and spatiotemporal features are extracted from video frames to generate the corresponding feature maps. Third, all feature maps are concatenated and support vector regression (SVR) learning using LIBLINEAR tools is employed to generate the initial saliency maps of video frames. Finally, the initial saliency maps are refined by using the center bias map, the significant increased map, visual sensitivity, and Gaussian filtering. Based on the experimental results obtained in this study, the performance of the proposed approach is better than those of three comparison approaches.

Superpixel-based Multi-focus Image Fusion

Kuan-Ni Lai, Jin-Jang Leou
Department of Computer Science & Information Engineering, National Chung Cheng University, Chiayi, Taiwan

Abstract: Due to the finite depth-of-field of optical lenses, it is difficult to make all objects in an image sharp and clear. Only objects within the depth-of-field are in focus and sharp, while the others are defocus and blurred. In this study, a superpixel-based multi-focus image fusion approach is proposed. First, each multi-focus source image is performed superpixel segmentation, and the saliency, depth, and difference image information are computed. Third, each superpixel is classified into one of three types (focus, defocus, and undefined), and each undefined superpixel is determined as focus or defocus by sum-modified-Laplacian (SML). The initial focus maps are estimated and refined by matting Laplacian-based image interpolation. Third, the boundaries between focus and defocus regions are employed to generate the weighting maps, followed by fused image generation. Based on the experimental results obtained in this study, the performance of the proposed approach is better than those of five comparison approaches.

Sensor Scheduling for Airborne Multi-Target Tracking with Limited Sensor Resources

Simon Koch, Peter Stutz
Institute for Flight Systems, University of the Bundeswehr Munich, Germany

Abstract: Aerial surveillance systems have become an important aspect in a range of civilian and military applications. For instance, reliable location data of individuals or objects is essential for use cases like traffic surveillance, search and rescue, relief efforts after natural disasters or patrol and border control missions. Unmanned aerial vehicles (UAVs) provide great value in such situations, but conventionally require at least one human operator whose insights and intuition are necessary for performing the task. With this contribution, we conceptualize a functional design that leverages performance models to track multiple targets and generate situational awareness autonomously. Further, we devise a testing harness for different scheduling schemes and provide some preliminary experimental data on basic scheduling policies found in literature.

Multi-Sensor Fusion based Action Recognition in Ego-Centric Videos with Large Camera Motion

Radhakrishna Dasari, Karthik Dantu, Chang Wen Chen
Department of Computer Science & Engineering, University at Buffalo, Buffalo, New York, USA

Abstract: Real-time action recognition on smartphones and wearable cameras is a challenging task. The effect of camera motion on action recognition algorithms is not trivial. To our advantage, smartphones and wearable devices are often augmented with depth, audio, geo-location, and inertial data, which can be incorporated into action recognition framework. Our study aims to quantify the effect of ego-centric motion on standard action recognition algorithms. As a part of the study, we collected multi-sensor video dataset with seven actions from the same subjects captured in three different camera motion scenarios - minimal, pure rotational and both rotational and translational. We plan to create a multi-sensor fusion based action recognition framework to improve the recognition accuracy on smartphones and wearable cameras, which are equipped with inertial sensors. We present preliminary experiment results on our multi-sensor video dataset.

Deep Convolutional Likelihood Particle Filter for Visual Tracking

Reza Jalil Mozhdehi, Henry Medeiros
Marquette University, Milwaukee, Wisconsin, USA

Abstract: We propose a novel particle filter for convolutional-correlation visual trackers. Our method uses correlation response maps to estimate likelihood distributions and employs these likelihoods as proposal densities to sample particles. Likelihood distributions are more reliable than proposal densities based on target transition distributions because correlation response maps provide additional information regarding the target's location. Additionally, our particle filter searches for multiple modes in the likelihood distribution, which improves performance in target occlusion scenarios while decreasing computational costs by more efficiently sampling particles. In other challenging scenarios such as those involving motion blur, where only one mode is present but a larger search area may be necessary, our particle filter allows for the variance of the likelihood distribution to increase. We tested our algorithm on the Visual Tracker Benchmark v1.1 (OTB100) and our experimental results demonstrate that our framework outperforms state-of-the-art methods.

Similar Multi-Modal Image Detection in Multi-Source Dermatoscopic Images of Cancerous Pigmented Skin Lesions

Sarah Hadipour, Siamak Aram, Roozbeh Sadeghian
Northeastern University, Boston, MA, USA;
Harrisburg University of Science and Technology, Harrisburg, PA, USA

Abstract: The pursuit in similar image detection is constantly increasing in the computer vision field. This is even more prominent in the medical imaging field. Medical image sets such as cancerous pigmented skin lesions have a long tendency of including images with multiple modalities. The diagnostic algorithms such as classifiers that categorise the skin lesions rely on unique set of images for best performance. In this paper we developed a distance-based approach to image similarity detection. We applied six methods of image distances which utilized features from image histogram. The sensitivity and ectiveness of each method as well as the determination of threshold is discussed.

Deep Learning for Plant Disease Detection

Matisse Ghesquiere, Mkhuseli Ngxande
Stellenbosch University, Western Cape, South Africa

Abstract: In today's world, plant diseases are a major threat to agriculture crops and their production rate. These are difficult to spot in early stages and it's not feasible to inspect every leaf manually. We tested different convolutional neural networks on their ability to classify plant diseases. The best model reaches an accuracy of 99.70 %, made with a deep training method. We also developed a hybrid training method, reaching a 98,70 % accuracy with faster training times, reducing the gap between accuracy and training time. This was made possible due to the freezing of layers at a predefined step. In general, detecting plant diseases using deep learning models is an excellent approach and much more practical than detection with the human eye.

Application of Image Processing Tools for Scene-based Marine Debris Detection and Characterization

Mehrube Mehrubeoglu, Farha Pulukool, Dekwaan Wynn, Lifford McLauchlan, Hua Zhang
Department of Engineering, Texas A&M University-Corpus Christi, Texas, USA;
Department of Computer Sciences, Texas A&M University-Corpus Christi, Texas, USA;
Department of EE & CS, Texas A&M University-Kingsville, Texas, USA

Abstract: Pollution encompasses any substance that has a negative impact on the environment or the organisms that live within an affected area. Marine environmental pollution includes debris, which can be natural or man-made. In this paper, four distinct debris scenes are analyzed using image processing tools to detect man-made marine debris with different image and object properties. The scenes vary in their challenges from multiple floating surface and subsurface debris to single pieces of debris. A successful image processing chain is described that includes image preprocessing such as conversion to HSV color model, filtering, thresholding, and post processing operations such as blob analysis and statistical computations to detect and characterize the man-made debris in each scene. We demonstrate detection of multiple debris and computation of its percent cover of the scene, its size distribution, as well as identification of single debris pieces. The applied methods demonstrate successful debris detection and characterization in each scene, which can be extended to other debris images for automated detection.

DeepMSRF: A novel Deep Multimodal Speaker Recognition Framework with Feature Selection

Ehsan Asali, Farzan Shenavarmasouleh, Farid Ghareh Mohammadi, Prasanth Sengadu Suresh, Hamid R. Arabnia
Department of Computer Science, University of Georgia, Georgia, USA

Abstract: For recognizing speakers in video streams, significant research studies have been made to obtain a rich machine learning model by extracting high-level speaker.s features such as facial expression, emotion, and gender. However, generating such a model is not feasible by using only single modality feature extractors that exploit either audio signals or image frames, extracted from video streams. In this paper, we address this problem from a different perspective and propose an unprecedented multimodality data fusion framework called DeepMSRF, Deep Multimodal Speaker Recognition with Feature selection. We execute DeepMSRF by feeding features of the two modalities, namely speakers. audios and face images. DeepMSRF uses a two-stream VGGNET to train on both modalities to reach a comprehensive model capable of accurately recognizing the speaker's identity. We apply DeepMSRF on a subset of VoxCeleb2 dataset with its metadata merged with VGGFace2 dataset. The goal of DeepMSRF is to identify the gender of the speaker first, and further to recognize his or her name for any given video stream. The experimental results illustrate that DeepMSRF outperforms single modality speaker recognition methods with at least 3 percent accuracy.

The 17th International Conference on Modeling, Simulation & Visualization Methods (MSV 2020) https://americancse.org/events/csce2020/conferences/msv20

On the Parameterizations of Closed-Loop Control Systems

Cs. Banyasz, L. Keviczky, R. Bars
Institute for CS & Control, Laboratory of Systems & Control Engineering, Budapest, Kende, Hungary;
Budapest University of Tech. & Economics, Department of Automation & Applied Informatics, Budapest, Magyar, Hungary

Abstract: The optimization of simple two-degree-of-freedom control systems is quite easy with the new parameterizations, as YOULA and KEVICZKYBANYASZ. The comparison of their model based versions is important at the practical applications.

Modern Control Methods of Time-Delay Control Systems

R. Bars, Cs. Banyasz, L. Keviczky
Budapest University of Tech. & Economics, Department of Automation & Applied Informatics, Budapest, Magyar, Hungary;
Institute for CS & Control, Laboratory of Systems & Control Engineering, Budapest, Kende, Hungary

Abstract: It is shown that the SMITH predictor is a subclass of the YOULA parameterization based generic two-degree of freedom controllers. Comparing the algorithms the application of the new approach is suggested.

Ground Vehicle Suspension Optimization Using Surrogate Modeling

Jeremy Mange
US Army - CCDC - Ground Vehicle Systems Center, Analytical Methods Team, Warren, Michigan, USA

Abstract: Using surrogate models in place of more computationally expensive simulations is a common practice in several contexts. In this paper, we present an optimization task of finding ideal spring coefficient values for ground vehicle suspensions with respect to a particular metric of driver safety and comfort, develop a set of surrogate models based on sampling full system simulations to calculate this metric, and present and compare the results of using these surrogate models to perform the optimization. We show that the medium-fidelity model, as defined for this study, is of sufficient fidelity for the optimization and that additional fidelity offers little benefit, but that the underlying objective function is noisy enough to limit the usefulness of the surrogate model approach for the optimization.

Step Response Method for Transmission Line Parameter Estimation

Louiza Sellami
Electrical and Computer Engineering Department, US Naval Academy, Annapolis, Maryland, USA

Abstract: In this paper a method is proposed to determine parameters for circuits of known structure through the use of step responses. The method is developed in the frequency as well as time domains based on nonlinear minimum least squares extraction to determine the unknown parameters. Pspice circuit simulations and MATLAB simulations are performed on circuits constructed from passive elements of transmission lines, these being resistors, inductors and capacitors, to validate the method. Additionally, comparison in terms of accuracy is made with the direct method which uses time specific output data points.

A Virtual Serious Game for Nursing Education

Youxin Luo, Abel A. Reyes, Parashar Dhakal, Manisa Baker, Julia Rogers, Xiaoli Yang
Electrical and Computer Engineering Department, Purdue University Northwest, Hammond, Indiana, USA;
Electrical & Computer Science Engineering, The University of Toledo, Toledo, Ohio, USA;
College of Nursing, Purdue University Northwest, Hammond, Indiana, USA

Abstract: Procedure skill competency is a crucial element in nursing education. However, there are several barriers to skill competency and proficiency, which include limited simulation faculty, lab time, and resources. Serious gaming has proven to be an effective approach to enhance knowledge attainment, promote problem-based active learning, and encourage critical thinking in many industries, including nursing. Therefore, we propose a virtual serious game featured with a realistic environment, considerable interaction and animation, well-designed userfriendly graphic interface, and an innovative evaluation system for nursing education. This game is designed as complementary learning material to be utilized along with traditional pedagogical methods. This game provides a supplemental learning methodology for undergraduate nursing students, enabling them to practice the skill in a safe and innovative environment. Incorporating a serious game into undergraduate nursing education can circumvent the problems of scheduling, individualized instruction, physical requirements of facilities, equipment, and supplies, and geography constraints.

An Interactive Software to Learn Pathophysiology with 3D Virtual Models

Abel A. Reyes, Youxin Luo, Parashar Dhakal, Julia Rogers, Manisa Baker, Xiaoli Yang
Electrical and Computer Engineering Department, Purdue University Northwest, Hammond, Indiana, USA;
Electrical and Computer Engineering Department, University of Toledo, Toledo, Ohio, USA;
College of Nursing, Purdue University Northwest, Hammond, Indiana, USA

Abstract: Educators have accessibility to a repository of resources to guide pedagogical instruction within graduate nursing curriculum. Multiple modalities are utilized within graduate education including face-to-face, online, and alternative platforms, such as web page, video, or mobile applications. Supplemental resources include e-learning applications, mobile learning, and game-based learning using internet accessible devices, smart displays and even virtual reality headsets. The use of interactive and innovative methods have shown positive results in student engagement and cognitive learning. However, the implementation of 3D visualization strategies has been limited within healthcare education, specifically graduate advanced practice nursing. It is vital in graduate nursing education to provide students with advanced knowledge of disease processes and critical reasoning skills. While some programs include the ability to display text and image, as well as present animated contents, they lack sufficient interactive features to enhance learning. Therefore, an efficient and effective modality of information delivery is required to achieve this goal. In this paper, we describe the development of an interactive 3D visualization software tool that provides an innovative approach for education within graduate nursing. The visualization software provides a framework to enhance the teaching and learning process in graduate nursing pathophysiology utilizing 3D virtual models.

Vehicle Test Rig Modeling and Simulation

Sara Boyle
Ground Vehicle Systems Center (GVSC), Combat Capabilities Development Command (CCDC), Warren, Michigan, USA

Abstract: Vehicle test rigs can be used to evaluate vehicle subsystems for mobility characteristics. These vehicle test rigs can be simulated using multi-body physics tools, such as Chrono. This paper details the method used for developing a user input file format and wrapper methods for running the Chrono::Vehicle tire, track, and suspension test rigs, and how to build and run these various test rigs.

Dealing Bridge Hands: A Study in Random Data Generation

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

Abstract: In this study we examine the problem of dealing bridge hands and producing output in elegant usable form using the Data Generation Language (dgl). Although this is a simple problem, the solution lead to the inclusion of several new features for the dgl language. These features can be useful for many other problems as well. Several techniques for dealing bridge hands are discussed, along with the new dgl features that were used in each technique. Examples are given of actual hands dealt using the dgl grammars.

An Empirical Study of the Effect of Reducing Matching Frequency in High Level Architecture Data Distribution Management

Mikel D. Petty
Department of Computer Science, University of Alabama in Huntsville, Alabama, USA

Abstract: The High Level Architecture (HLA) is an interoperability protocol standard and implementation architecture for distributed simulation. Using HLA, concurrently executing simulation models collaboratively simulate a scenario by exchanging messages over a network. In HLA simulations, large volumes of messages are possible, potentially limiting scalability. The HLA Data Distribution Management services are designed to reduce message volume. When using those services, the HLA software determines how to route messages by repeatedly solving a computationally expensive computational geometry problem known as "matching": given a set of axis-parallel hyper-rectangles in a multi-dimensional coordinate space, find all intersecting pairs of rectangles. Much effort has been devoted to performing matching as *efficiently* as possible. This study investigates a different approach, namely, performing matching as *seldom* as possible, without compromising the simulation results. A constructive entity-level combat model, similar to production military semi-automated forces systems, was developed and verified. It was then used to experimentally assess the effect of performing matching at various frequencies. The primary metric of effect was the mean time at which battlefield entities first sighted each other. Experimental results regarding the delay of first sighting times suggest that matching frequency, and thus its computational expense, can be substantially reduced without negative effects.

Enhanced Freehand Interaction by Combining Vision and EMG-based Systems in Mixed Reality Environments

Carol Naranjo-Valero, Sriram Srinivasa, Achim Ebert, Bernd Hamann
Technical University of Kaiserslautern, Kaiserslautern, Germany;
University of California, Davis, California, USA

Abstract: This paper studies the capabilities, limitations and potential of combining a vision-based system with EMG sensors for freehand interaction in mixed reality environments. We present the design and implementation of our system using the Hololens and Myo armband, conduct a preliminary user study with 15 participants to study the usability of our model and discuss the advantages, potentials and limitations of this approach, and discuss our findings and its implications for the design of user interfaces with a similar hardware setup. We show that the flexibility of interaction in our proposal has positive effects on the user performance for the completion of a complex user task, although measured user performance for the individual gestures was worse on average than the performance obtained for the gestures supported by the standard Holotoolkit. One can conclude that the presented interaction paradigm has great potential for future use in mixed reality, but it still has some limitations regarding robustness and ergonomics that must be addressed for a better user acceptance and broader public adoption.

A Simulation-Optimization Technique for Service Level Analysis in Conjunction with Reorder Point Estimation and Lead-Time Consideration: A Case Study in Sea Port

Mohammad Arani, Saeed Abdolmaleki, Maryam Maleki, Mohsen Momenitabar, Xian Liu
Department of Industrial Management, University of Shahid Beheshti, Tehran, Iran;
Department of Systems Engineering, University of Arkansas at Little Rock, Arkansas, USA;
Department of Transportations and Logistics, North Dakota State University, Fargo, North Dakota, USA

Abstract: This study offers a step-by-step practical procedure from the analysis of the current status of the spare parts inventory system to advanced service level analysis by virtue of simulation-optimization technique for a real-world case study associated with a seaport. The remarkable variety and immense diversity on one hand, and extreme complexities not only in consumption patterns but in the supply of spare parts in an international port with technically advance port operator machinery, on the other hand, have convinced the managers to deal with this issue in a structural framework. The huge available data require cleaning and classification to properly process them and derive reorder point (ROP) estimation, reorder quantity (ROQ) estimation, and associated service level analysis. Finally, from 247000 items used in 9 years long, 1416 inventory items are elected as a result of ABC analysis integrating with the Analytic Hierarchy Process (AHP), which led to the main items that need to be kept under strict inventory control. The ROPs and the pertinent quantities are simulated by Arena software for all the main items, each of which took approximately 30 minutes run-time on a personal computer to determine near-optimal estimations.

Research on Repair Strategy of Heterogeneous Combat Network

Yanyan Chen, Yonggang Li, Shangwei Luo, Zhizhong Zhang
School of Communication and Information Enginnering,
Chongqing University of Posts and Telecommunications, Chongqing, P. R. China

Abstract: The purpose of restoration strategy is to repair the combat system in time after the node is attacked so as to reduce the impact of functional failure. The combat system is modeled into a heterogeneous combat network and its combat capability is measured by functional reliability based on the functional chain structure. This paper proposes an edge increasing strategy with the goal of maximizing functional reliability, and combines optional edges and connection costs as constraints. The artificial colony algorithm was used to solve the repair model. Simulation experiments were carried out under four node attack strategies. The results verify the effectiveness of the repair model and prove that the proposed method is superior to other repair algorithms.

Sustainability, Big Data, and Local Community: a Simulation Case Study of a Growing Higher Education Institution

Anatoly Kurkovsky
University System of Georgia, Georgia Gwinnett College, Greater Atlanta, Georgia, USA

Abstract: Higher education institutions are core elements in the Sustainable Development paradigm because they prepare the future decision-makers of the local, national, or international societies. At the same time, many Colleges and Universities just declare their sustainability support and rare use a quantitative approach to analyze/ manage this area. It is happening because of sustainable development paradigm complexity, reusability problems of already created tools/models, and some methodological difficulties to use already available big data. This paper introduces an approach where we use simulation as a united methodological platform to combine sustainability, big data, and local community needs for particular numerical analysis. Within the simulation case study, we analyze the transportation system as a part of the sustainability of a young fast-growing higher education institution, USA.

Workflow for Investigating Thermodynamic, Structural and Energy Properties of Condensed Polymer Systems

James Andrews, Estela Blaisten-Barojas
Center for Simulation & Modeling & Department of Computational & Data Sciences,
George Mason University, Fairfax, Virginia, USA

Abstract: Soft matter materials and polymers are widely used in the controlled delivery of drugs. Simulation and modeling provide insight at the atomic scale enabling a level of control unavailable to experiments. We present a workflow protocol for modeling, simulating, and analyzing structural and thermodynamic response properties of poly-lactic-coglycolic acid (PLGA), a well-studied and FDA approved material. We concatenate a battery of molecular dynamics, computational chemistry, highly parallel scripting, and analysis tools for generating properties of bulk polymers in the condensed phase. We provide the workflow leading to the glass transition temperature, enthalpy, density, isobaric heat capacity, thermal expansion coefficient, isothermal compressibility, bulk modulus, sonic velocity, cohesive energy, and solubility parameters. Calculated properties agree very well with experiments, when available. This methodology is currently being extended to a variety of polymer types and environments.

Modelling and Simulation of MEMS Gyroscope with Coventor MEMS+ and Matlab/SIMULINK Software

Jacek Nazdrowicz, Adam Stawinski, Andrzej Napieralski
Department of Microelectronics and Computer Science, Lodz University of Technology, Poland

Abstract: In this paper authors presents heterogeneous environment for modeling and simulations created with use Coventor MEMS+ and Matlab/SIMULINK software. The big advantage of this solution is possibility to merge with Cadence software what gives in effect big solution for modeling, simulation and design MEMS structure with ROIC (Read Out Integrated Circuit) for further fabrication. This environment was created for needs of multidyscipinary project (medicine, electronics and computer sciences areas) realized by three scientific institutions and two companies.

The Influence of Decorations and Word Appearances on the Relative Size Judgments of Words in Tag Clouds

Khaldoon Dhou, Robert Kosara, Mirsad Hadzikadic, Mark Faust
Texas A&M University Central Texas, Texas, USA;
Tableau Software, USA;
University of North Carolina Charlotte, North Carolina, USA

Abstract: A tag cloud is a representation of the word content of a source document where the importance of the words is represented by visual characteristics such as color and size. Tag clouds can be used for several purposes such as providing a high-level understanding of a document. Although previous research has indicated that the relative size of tags is a strong factor in communicating the importance of the words in the underlying text, there are still many unanswered questions. Examples of these questions are: how do viewers perceive the relative size of the words in a tag cloud? how is the judgment of the relative size is influenced by other characteristics of the words in a tag cloud? and do viewers make their judgments based on the area, the height, or the length of the words? In this paper, we investigate the viewers. estimation of the tag word relative sizes given while varying the size, types of letters, and the surrounding text box decorations around the target word pairs. The results indicate a range of relative sizes where relative size judgments may be approximately correct, but also a large region of relative sizes where the relative size judgments are increasingly underestimated as the true size ratio increases. This underestimation bias was only modestly influenced by appearance characteristics. The results have implications for tag cloud design, and for reliance on relative size judgment of words as a visualization technique.

Automation of an Off-Grid Vertical Farming System to Optimize Power Consumption

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

Abstract: The world's resources are finite. As the population increases, these resources must be better utilized to maintain or improve living standards. One area ripe for improvement is agricultural farming. Farms use the most available freshwater and vast swaths of land. Additionally, fertilizer runoff pollutes nearby rivers and reservoirs. One solution is vertical farms. Vertical farms can be stacked on top of each other, reducing land, water, and fertilizer usage significantly. However, this comes at the tradeoff of increased energy usage. Renewable energy sources can supply the power, reducing the compromise, but they are unreliable, and their power production varies hourly and by season. An automated system can adapt to the changing power availability without harming plants because plants acclimate to a wide variety of growing conditions in nature. This research focused on automating a vertical farm system to adjust to weather variations without experiencing blackouts while maximizing power usage and improving growing conditions for the plants in the systems.

Modeling Digital Business Strategy During Crisis

Sakir Yucel
NetApp, Wexford, Pittsburgh, USA

Abstract: Covid-19 is a major health, economic and social crisis in the modern age. Even before the Covid-19 pandemic, digitization had changed the consumer behavior and habits, regulations, supply side factors, demand side factors, and costs of information structure and coordination. We have been experiencing shifts from physical interactions to digital interactions, and transitioning from physical, predictable and slow world into digital, virtual, fast and agile world. Covid-19 will likely accelerate the digitization but will also force corporations to refine and possibly redefine their digital business strategies. Corporations have to address many questions about how to deal with the crisis in digital age and refine their digital business strategy while going through some transformations already. The main question is how corporations could navigate through this crisis when traditional economy and even digital economy assumptions and approaches do not necessarily apply. In this paper, we study how corporations could characterize the digital business strategy during crisis and devise a framework for how they could model and evaluate various strategic options. Due to so many complex dynamics involved and many uncertainties, we argue a layered intelligence framework that supports qualitative analysis should be used to model the digital business strategy during crisis.

The 26th International Conference on Parallel & Distributed Processing (PDPTA 2020)
https://americancse.org/events/csce2020/conferences/pdpta20

Refactor Business Process Models with Redundancy Elimination

Fei Dai, Huihui Xue, Zhenping Qiang, Lianyong Qi, Mohammad R. Khosravi, Zhihong Liang
School of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming, P. R. China;
School of Information Science and Engineering, Qufu Normal University Jinning, P. R. China;
Department of Computer Engineering, Persian Gulf University, Bushehr, Iran;
Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran

Abstract: Since business processes are important assets, enterprises must be able to deal with their quality issues. Since understandability is one important quality criterion, a question arises here is how to improve the understandability of these models? In this paper, we propose a novel approach to refactor business process models represented as Petri nets with redundancy elimination for improving their understandability. More specifically, we first propose a process model smell for identifying redundant elements in a business process model using the unfolding technique, where the metric of this smell is an implicit place (IP). To avoid the state explosion problem caused by concurrency, we present a novel algorithm for computing an IP from the Complete Fintie Prefix Unfolding (CFPU) rather than the reachability graph (RG) of a net system. Then, we propose three refactoring operations to eliminate an IP from the business process model without changing their external behavior. After refactoring, the size of the model is decreased such that the model is easier to be understood, that is, the understandability of the model can be improved. Experiments show our approach can eliminate IPs from business process models efficiently and preserve the behavior of these models.

Toward a Numerically-Robust and Efficient Implicit Integration Scheme for Parallel Power Grid Dynamic Simulation Development in GridPACK

Shuangshuang Jin, Shrirang G. Abhyankar, Bruce J. Palmer, Renke Huang, Wiliam A. Perkins, Yousu Chen
School of Computing, Clemson University, North Charleston, South Carolina, USA;
Optimization and Control Group, Pacific Northwest National Labtoratory, Richland, Washington, USA;
High Performance Computing, Pacific Northwest National Labtoratory, Richland, Washington, USA;
Electricity Infrastructure, Pacific Northwest National Laboratory, Richland, Washington, USA;
Hydrology Group, Pacific Northwest National Labtoratory, Richland, Washington, USA

Abstract: GridPACK is a highly modular parallel computing package for developing power grid simulations that run on high performance computing platforms. As one of the key modules in GridPACK, dynamic simulation assesses the transient behavior of power systems and plays an important role in determining the performance of dynamic security assessment applications which rely heavily on the computational speed and scalability of dynamic simulation. This paper presents an ongoing effort on redesigning the existing "fixed-step" Modified-Euler explicit numerical integration scheme based dynamic simulation module in GridPACK to incorporate numerically robust and efficient "variable-step" implicit integration schemes. Promising computational performance over the explicit integration method in addition to the improved usability is presented in the paper as the outcome of this preliminary study.

Directive-based Hybrid Parallel Power System Dynamic Simulation on Multi-Core CPU and Many-Core GPU Architecture

Cong Wang, Shuangshuang Jin, Yousu Chen
School of Computing, Clemson University, North Charleston, South Carolina, USA;
Electricity Infrastructure, Pacific Northwest National Laboratory, Richland, Washington, USA

Abstract: High-Performance Computing (HPC) based simulation tools for large-scale power grids are important to the improvement of future energy sector resiliency and reliability. However, the application development complexity, hardware adoption, and maintenance cost with large HPC facilities have hindered the wide utilization and quick commercialization of HPC applications. This paper presents a hybrid implementation of power system dynamic simulation . a time-critical function for transient stability analysis using directive-based parallel programming models to showcase the advantage of leveraging multi-core CPU and many-core GPU computing with superior floating-point acceleration performance and cost-effective architecture to lower this barrier. Real-time modeling and simulation with least modifications on the legacy sequential program are achieved with significant speedup performances on two test cases.

An NPGA-II based Multi-objective Edge Server Placement Strategy for IoV

Xuan Yan, Zhanyang Xu, Mohammad R. Khosravi, Lianyong Qi, Xiaolong Xu
School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, P. R. China;
Engineering Research Center of Digital Forensics, Ministry of Education, P. R. China;
Department of Computer Engineering, Persian Gulf University, Bushehr, Iran;
Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran;
School of Information Science and Engineering, Qufu Normal University, P. R. China

Abstract: With the emergence of crowded traffic conditions, edge computing appears to deal with resource provision in the Internet of Vehicles (IoV). The tasks are offloaded from the vehicles to the nearby roadside units (RSUs) and transferred from the RSUs to the edge servers (ESs) for computing. Since the total number of the ESs is a constant, placing an ES on the remote area would improve the coverage rate while the workload variance of the ESs and waiting time of the tasks deteriorate. An ideal ES location is supposed to achieve a balance among these three aspects. Therefore, An NPGA-II based Multi-objective Edge Server Placement Strategy, named NMEPS, is proposed in this paper to obtain the proper schemes for the ES placement. Technically, the coverage rate, the workload variance of the ESs and the waiting time of the tasks are formulated as several fitness functions. Then, niched pareto genetic algorithm II (NPGA-II) and roulette algorithm are applied to seek out the optimal solutions for ES placement. Furthermore, an evaluation function is designed to assess the performance of the solutions obtained. Finally, experimental evaluations are conducted to prove the validity of this method by using big data from Nanjing, China.

A Shortest-path Routing Algorithm in Bicubes

Masaaki Okada, Keiichi Kaneko
Graduate School of Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan

Abstract: Recently, an explosive increase of demand on space- and time-consuming computation makes the research activities of massively parallel systems enthusiastic. Because in a massively parallel system a huge number of processors cooperate to process tasks by communicating among others, it forms an interconnection network, which is a network that interconnects the processors. By replacing a processor and a link with a vertex and an edge, respectively, many problems regarding communication and/or routing in interconnection networks are reducible to the problems in the graph theory. There have been many topologies proposed for interconnection networks of the massively parallel systems. The hypercube is the most popular topology and many variants were proposed so far. The bicube is a such topology, which can connect the same number of vertices with the same number degree as the hypercube while its diameter is almost half of that of the hypercube keeping the vertex-symmetric property. Therefore, we focus on the bicube and propose a shortest-path routing algorithm. We give a proof of correctness of the algorithm and demonstrate its execution.

MPI Communication Performance in a Heterogeneous Environment with Raspberry Pi

Oscar C. Valderrama Riveros, Fernando G. Tinetti
III-LIDI, Facultad de Informatica, UNLP, Argentina;
Universidad Cooperativa de Colombia, Ibague, Colombia;
CIC, Prov. de Buenos Aires, Argentina

Abstract: The Raspberry Pi SBC (Single Board Computer) has been used for distributed memory parallel computing mainly as a low cost teaching environment and as a low energy consumption/green computing platform. In this paper, we take a heterogeneous approach, where the Raspberry Pi is used along with standard computers. In the heterogeneous environment, computing as well as communication performance have to be taken into account in order to get the best results. We focus the work in this paper on the communication performance because it provides one of the best guidelines for parallel computing successful granularity. We have carried out our experiments with a standard MPI (Message Passing Interface) implementation as well as using the currently most powerful Raspberry Pi models in order to analyze the communication performance. We have experimented with classical Send-Receive MPI operations as well as the so called one-sided MPI communication operations. Also, we document several details specifically related to the heterogeneous configuration environment we found necessary for inter-operation of MPI.

A FPGA-based Heterogeneous Implementation of NTRUEncrypt

Hexuan Yu, Chaoyu Zhang, Hai Jiang
Department of Computer Science, Arkansas State University, Jonesboro, Arkansas, USA

Abstract: Nowadays, the lattice-based cryptography is believed to thwarting future quantum computers. The NTRU (Nth Degree Truncated Polynomial Ring Unit) encryption algorithm, abbreviated as NTRUEncrypt, is belonging to the family of lattice-based public-key cryptosystems. Comparing to other asymmetric cryptosystems such as RSA and Elliptic-Curve Cryptography (ECC), the encryption and decryption operations of NTRU significantly rely on basic polynomial multiplication, which makes it faster compared to those alternatives. This paper proposes the first heterogeneous implementation of NTRUEcrypt on FPGA (Altera Stratix V) and CPU using OpenCL, which has shown that this kind of lattice-based cryptography lends itself excellently for parallelization and achieves high throughput as well as energy-efficiency.

High Performance and Energy-Efficient FPGA-GPU-CPU Heterogeneous System Implementation

Chaoyu Zhang, Hexuan Yu, Yuchen Zhou, Hai Jiang
Department of Computer Science, Arkansas State University, Jonesboro, Arkansas, USA

Abstract: Since Moore's law is slowing down, CPU Optimizations and Multi-core architectures are exposing more and more limitations in energy efficiency and high performance. No single architecture can be best for every workload due to incredible diversity. Inspired by that GPUs have been widely deployed as accelerators in the past few years to speed up different types of tasks, FPGA-GPU (Field-Programmable Gate Array and Graphic Processing Unit) heterogeneous computing can optimize traditional system architecture. In this paper, we port six benchmarks kernels to the FPGA-GPU-CPU heterogeneous system. Selecting the most suitable hardware architecture for every task. Implement performance-oriented and energy-efficient-oriented kernel launch on this system. Due to the recent improvements in high-level synthesis and Intel FPGA SDK for OpenCL, it is convenient for FPGA to cooperate with GPU within a heterogeneous computing system.

Automatic Mapping of a Physical Model into a Conceptual Model for a NoSQL Database

Fatma Abdelhedi, Amal Ait Brahim, Rabah Tighilt Ferhat, Gilles Zurfluh
CBI2 - TRIMANE, Paris, France;
Toulouse Institute of Computer Science Research (IRIT), Toulouse Capitole University, Toulouse, France

Abstract: NoSQL systems have proven their efficiency to handle Big Data. Most of these systems are schema-less which means that the database doesn't have a fixed data structure. This property offers an undeniable flexibility allowing the user to add new data without making any changes on the data model. However, the lack of an explicit data model makes it difficult to express queries on the database. Therefore, users (developers and decision-makers) still need the database data model to know how data are stored and related, and then to write their queries. In previous works, we have proposed a process to extract the physical model of a document-oriented NoSQL database. In this paper, we aim to extend this work to achieve a reverse engineering of NoSQL databases in order to provide an element of semantic knowledge close to human understanding. The reverse engineering process is ensured by a set of transformation algorithms. We provide experiments of our approach using a case study example taken from the health care field. We also propose a validation of our solution in a real context; the results of this validation show that the generated conceptual model provide a good assistance to the user to express their queries on the database while saving a lot of time.

Composition of Parent-Child Cyberattack Models

Katia P. Maxwell, Mikel D. Petty, Tymaine S. Whitaker, Walter A. Cantrell, C. Daniel Colvett
Department of Mathematical, Computing and Natural Sciences, Athens State University, Athens, Alabama, USA;
Department of Computer Science, University of Alabama in Huntsville, Huntsville, Alabama, USA;
College of Computing and Technology, Lipscomb University, Nashville, Tennessee, USA;
Department of Systems Engineering, University of Alabama in Huntsville, Huntsville, Alabama, USA

Abstract: In today's world every system developer and administrator should be familiar with cyberattacks and possible threats to their organizations systems. Petri nets have been used to model and simulate cyberattacks allowing for additional knowledge on the planning stages of defending a system. Petri nets have been used since the 1960's and there exists several extensions and variations of how they are designed, in particular Petri Nets with Players, Strategies and Cost has been recently proposed to model individual cyberattacks on target systems. A formalism on composing these models has also been introduced as long as the attacks are performed either in a sequential order or parallel order. However, cyberattacks are also documented as having a parent-child relationship. The model composition described in this study provides a formalism that addresses cyberattacks that have this type of relationship and the process in which they should be composed through the use of inheritance concepts from object oriented programming. An example is provided by composing a Sniffing attack (parent) with a Sniffing Application Code attack (child).

Contrived and Remediated GPU Thread Divergence using a Flattening Technique

Lucas Vespa, Genevieve Peters
University of Illinois, Springfield, Illinois, USA

Abstract: General Purpose GPU applications have become mainstream. However, to this day, some code with major thread divergence can ruin GPU performance. In this work, we demonstrate examples of such code. We also propose a solution in the form of a flattening technique, which, although creates poor CPU performance, can be used to revive a GPU ruined by extreme thread divergence. We show the effect of data input on divergence and performance, and compare this to the flattened approach called algorithm flattening (AF). AF trades off best-case performance for deterministic performance, and works well in the average case, where extreme divergence exists.

Preliminary Performance and Programmability Comparison of the Thick Control Flow Architecture and Current Multicore CPUs

Martti Forsell, Sara Nikula, Jussi Roivainen
VTT, Oulu, Finland

Abstract: Multicore CPUs integrate a number of processor cores on a single chip to support parallel execution of computational tasks. These CPUs improve the performance over single core processors for independent parallel tasks nearly linearly as long as the memory bandwidth is sufficient. Speedup is, however, difficult to find when dense intercommunication between the cores is required. This forces programmers to use more complex and error-prone programming techniques instead of straight-forward parallel processing patterns. To solve these problems, we have introduced the Thick Control Flow (TCF) Processor Architecture TPA. TCF is an abstraction of parallel computation that combines self-similar threads into computational entities. While there are already a number of performance studies for TPA, it is not known how well TPA performs against commercial multicores. In this paper, we compare the performance and programmability of TPA and Intel Skylake multicore CPUs with kernel programs. Code examples and qualitative observations on the included programming approaches are given.

Survey on Recent Active Learning Methods for Deep Learning

Azar Alizadeh, Pooya Tavallali, Mohammad R. Khosravi, Mukesh Singhal
Computer Science Department, University of California, Merced, California, USA;
Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran

Abstract: The motivation of active learning is that by providing limited labeled training samples a machine learning algorithm can provide higher accuracy. The provided training samples are selected from a large or streaming dataset. The selection procedure often incorporates some measure of informativeness of samples. This measure is also defined based on the machine learning model itself. The data used in active learning is usually unlabeled; hence, the selected samples have to be labeled by an oracle (e.g., a human or a machine annotator). This is in case that labeling data is time-consuming, or expensive in some way. In this paper, active learning is first introduced and a general introduction is given and several query strategy frameworks are reviewed. Several recent papers on the topic of active learning and deep learning are studied, analyzed and categorized based on their query strategy and applications. Specially, the overview of active learning and recent Deep Learning techniques are explored.

Cloud-Edge Centric Service Provisioning in Smart City Using Internet Of Things

Manoj Kumar Patra, Bibhudatta Sahoo, Ashok Kumar Turuk, Sampa Sahoo
Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

Abstract: Three highly discussed and researched computing technology in recent times are cloud computing, edge computing, and the Internet of Things (IoT). In cloud computing, service seekers request computing resources from cloud servers connected over the Internet. In edge computing, the edge devices placed between the cloud server and service seekers resulting in faster access to the computing resources and hence reducing the computational time and cost. Internet of Things is a technology where several devices are connected and communicate with each other over the Internet. In this paper, we try to integrate these three technologies and propose a cloud-edge centric Internet of Things architecture for service provisioning in a smart city. The integration of these technologies improves the overall performance of the smart city system by efficiently utilizing the computing resources resulting in reduced makespan, response time, implementation cost, increased throughput, and better security of data.

Two-Phase Commitment Problem in Distributed Systems with Fail-Stop Model

Sung-Hoon Park, Yong-cheol Seo, Su-Chang Yoo
Department of Computer Engineering, Chungbuk National University, Chungbuk, Republic of Korea

Abstract: This paper defines the Non-Blocking Atomic Commitment problem in a message-passing asynchronous system and determines a failure detector to solve the problem. This failure detector, which we call the modal failure detectorstar, and which we denote by M*, is strictly weaker than the perfect failure detector P but strictly stronger than the eventually perfect failure detector P. The paper shows that at any environment, the problem is solvable with M*.

Challenges for Swarm of Drone-Based Intelligence

Muhammed Akif Agca, Peiman Alipour Sarvari, Sebastien Faye, Djamel Khadraoui
Luxembourg Institute of Science and Technology - LIST, Esch-Sur-Alzette, Luxembourg

Abstract: Swarms of UAVs/Drones are efficient resources for swarm intelligence, especially for monitoring/detect/react mechanisms. However, the increasing number of nodes in the system inflates the complexity of swarm behaviour, due to computation, communication and control limitations for monitoring and security purposes. In order to maintain the high performance of such a system, mission/safety/operation-critical applications must be verified via the elaboration of critical check-points. To make it resilient, this requires real-time updates in different system layers reflected in this paper and therefore, scalability (from the networking viewpoint) and memory speed limitations (from the processing viewpoint), as well as security controls are challenging. In the context of swarms of UAVs, this can be accomplished via big data technologies and ledger base chained structures, which is one part of the contribution of this paper. In order to assure resilience against manipulation threats, the other parts of the contribution concern end-to-end trust mechanism (integrated view of the three pillars: networking, processing/optimization as well as security) and swarm controller methods guaranteeing safety, which aims at enabling the trusted scalability of the swarm systems.

Parallel Computation of Grobner Bases on a Graphics Processing Unit

Mark Hojnacki, Andrew Leeseberg, Jack O'Shaughnessy, Michael Dauchy, Alan Hylton, Leah Gold, Janche Sang
Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, Ohio, USA;
Department of Mathematics, Cleveland State University, Cleveland, Ohio, USA;
Space Communications and Navigation, NASA Glenn Research Center, Cleveland, Ohio

Abstract: Solving polynomial systems of equations of both many degrees and variables is not a simple computation. A method of solving these systems is to transform them into a Gröbner basis. Gröbner bases have desirable mathematical properties that make it possible to solve systems of polynomial equations. The computations necessary to compute Gröbner bases are many and can sometimes take days if not longer. Existing implementations are unable to handle the high degree polynomial systems necessary for specialized applications. Graphics Processing Units specialize in fast parallel computations by means of using many (several hundred to a couple thousand) computing cores. Utilizing these cores in parallel allows difficult problems to be solved quickly, versus using a Central Processing Unit, when optimized properly. The goal of this project is to implement a Gröbner basis algorithm that is optimized for GPUs which in turn will allow for faster computations of Gröbner bases.

Improving Analysis in SPMD Applications for Performance Prediction

Felipe Tirado, Alvaro Wong, Dolores Rexachs, Emilio Luque
Computer Architecture and Operating System Department, Universidad Autonoma de Barcelona, Barcelona, Spain;
Departamento de Computacion e Industrias, Universidad Catolica del Maule, Talca, Chile

Abstract: The analysis of parallel scientific applications allows us to know the details of their behavior. One way of obtaining information is through performance tools. One such tool is PAS2P, which is based on parallel application repeatability, focusing on performance analysis and prediction using the application signature. The analysis is performed using the same execution resources of the parallel application to create an independent machine model and identify common patterns. The analysis stage of the PAS2P tool is costly in terms of runtime, due to the high number of communications it performs, degrading performance by increasing the number of execution processes. To solve this problem, we propose designing a module that reduces the data dependency between processes, reducing the number of communications, and taking advantage of the characteristics of the SPMD applications. For this, we propose an analyzer module that is independent of data between processes. Our proposal allowed us to decrease the analysis time when the application scales.

Single Core vs. Parallel Software Algorithms on a Multi-core RISC Processor

Austin White, Micheal Galloway
Western Kentucky University, Bowling Green, Kentucky, USA

Abstract: Algorithms with order-independent instructions have the opportunity to be executed on multiple cores at the same time in a process called parallelism. For this reason, multicore processors are the standard in contemporary computer architecture. To understand the benefits and drawbacks of multicore processors, we analyzed the performance of three algorithms that are important workloads in the general use of computers - sorting, password hashing, and graphics rendering - when computed as single core and multi-core workloads. We found that in the first and last examples for small workloads, the benefits of parallelism did not outweigh the performance drawbacks of coordination, but in the second example, they did.

Prototype of MANET Network with Ring Topology for Mobile Devices

Ramses Fuentes-Perez, Erika Hernandez-Rubio, Diego D. Flores-Nogeira, Amilcar Meneses-Viveros
Instituto Politecnico Nacional-ESCOM, Mexico;
Instituto Politecnico Nacional, SEPI-ESCOM. Mexico;
Departamento de Computacion, Cinvestav-IPN, Mexico

Abstract: This work presents the design of a MANET network for heterogeneous mobile devices, based on the ring topology. The net is implemented as a distributed system for handling errors such as crash and net recovery. There are different mobile applications, such as collaborative or emergency applications, that require support of a Mobile ad hoc network when the internet infrastructure is not available.

Tree-Based Fixed Data Transmission for Healthcare Sensor Networks

Susumu Shibusawa, Toshiya Watanabe
Ibaraki University, Hitachi, Ibaraki, Japan;
National Institute of Technology, Gunma College, Gunma, Japan

Abstract: The ability to obtain health-related information at any time through the use of sensor network technology for healthcare instead of taking measures after becoming ill can greatly improve an individual's life. The processing of healthcare sensor data requires effective studies on inter-node transmission in the sensor network and collective data processing. This paper introduces asynchronously operating data transmission with a fixed number of transmission data (fixed data transmission) on trees and evaluates the execution times of fixed data transmission and level data transmission. Tree-based fixed data transmission can continue transmission operations with an average number of data at either a shallow-level edge or deep-level edge. Level data transmission, on the other hand, begins data transmission from edges near leaves and transmits a large number

of integrated data at a level near the root. The execution time of fixed data transmission on a complete binary tree with a maximum number of transmission data of 2 is equivalent to or smaller than the execution time of level data transmission, and as the number of nodes increases, fixed data transmission approaches a value 1.5 times faster than level data transmission.

Title of Workshop: Mathematical Modeling and Problem Solving (MPS)

Co-Chairs: Dr. Masakazu Sekijima + Prof. Kazuki Joe***
** Tokyo Institute of Technology, Japan*
***chair of Life Computing and Communication Department, Nara Women's University, Japan*

New state-of-the-art Results on ESA's Messenger Space Mission Benchmark

Martin Schlueter, Mohamed Wahiby, Masaharu Munetomo
Hokkaido University, Japan

Abstract: This contribution presents new state-of-the-art results for ESA's Messenger space mission benchmark, which is arguably one of the most difficult benchmarks available. The European Space Agency (ESA) created a continuous mid-scale black-box optimization benchmark which resembles an accurate model of NASA's 2004 launched Messenger interplanetary space probe trajectory. By applying an evolutionary optimization algorithm (MXHPC/MIDACO) that relies on massive parallelization, it is demonstrated that it is possible to robustly solve this benchmark to a near global optimal solution within one hour on a computer cluster with 1000 CPU cores. This is a significant improvement over the previously in 2017 published state-of-the-art results where it was demonstrated for the first time, that the Messenger benchmark could be solved in a fully automatic way and where it took about 12 hours to achieve a near optimal solution. Presented here are the results that fortify the effectiveness of massively parallelized evolutionary computing for complex real-world problems which have been previously considered intractable.

Crawling Low Appearance Frequency Characters Images for Early-Modern Japanese Printed Character Recognition

Nanami Fujisaki, Yu Ishikawa, Masami Takata, Kazuki Joe
Nara Women's University, Japan

Abstract: To be provided later.

Application of the Orthogonal QD Algorithm with Shift to Singular Value Decomposition for Large Sparse Matrices

Hiroki Tanaka, Taiki Kimura, Tetsuaki Matsunawa, Shoji Mimotogi, Masami Takata, Kinji Kimura, Yoshimasa Nakamura
Nara Women's University, Japan

Abstract: In semiconductor manufacturing process, lithography simulation modeling is known as an ill-posed problem. A normal solution of the problem is generally insignificant due to measurement constraints. In order to alleviate the difficulty, we introduced a regularization method using a preconditioning technique which consists of scaling and uniformization based on prior information. By regularizing the solution to prior knowledge, an accurate model can be achieved because the solution using truncated singular value decomposition from a few larger singular values becomes a reasonable solution based on the physically appropriate prior knowledge. The augmented implicitly restarted Lanczos bidiagonalization (AIRLB) algorithm is suitable for the purpose of the truncated singular value decomposition from a few larger singular values. Thus, the AIRLB algorithm is important for obtaining the solution in lithography simulation modeling. In this paper, we propose techniques for improving the AIRLB algorithm for the truncated singular value decomposition of large matrices. Specifically, we implement the improvement of the AIRLB algorithm by Ishida et al. Furthermore, instead of using the QR algorithm, we use the orthogonal-qd-with-shift algorithm for the singular value decomposition of the inner small matrix. Several numerical experiments demonstrate that, compared with AIRLB using the original QR algorithm, the proposed improvements provide highly accurate truncated singular value decomposition. For precise discussion, both large-scale sparse matrices and large-scale dense matrices are included in the experiments.

On an Implementation of the One-Sided Jacobi Method with High Accuracy

Masami Takata, Sho Araki, Kinji Kimura, Yoshimasa Nakamura
Nara Women's University, Japan

Abstract: The one-sided Jacobi method for performing singular value decomposition can compute all singular values and singular vectors with high accuracy. Additionally, the computation cost is insignificant for comparatively small matrices. However, in the case of the conventional implementation in Linear Algebra PACKage, the subroutine may not be able to compute a singular vector with sufficient orthogonality. To avoid this problem, we propose a novel implementation of the one-sided Jacobi method. In the proposed implementation, a Givens rotation with high accuracy and fused multiply-accumulate are adopted.

Improvement of Island GA using Multiple Fitness Functions

Shigeka Nakajima, Masami Takata
Nara Women's University, Japan

Abstract: In this paper, we propose an island genetic algorithm (GA) that promotes a unique evolution. In a conventional island GA, all objective functions are combined into a single fitness function. Hence, offspring generations are generated using the same fitness function. In the natural world, each should evolve in a manner that suits the environment, and owing to the various environments on Earth, various organisms have been diversified. Therefore, we propose an improved island GA with different fitness functions to create a distinctive evolution.

High-Performance Cloud Computing for Exhaustive Protein-Protein Docking

Masahito Ohue, Kento Aoyama, Yutaka Akiyama
Tokyo Institute of Technology, Japan

Abstract: Public cloud computing environments, such as Amazon AWS, Microsoft Azure, and the Google Cloud Platform, have achieved remarkable improvements in computational performance in recent years, and are also expected to be able to perform massively parallel computing. As the cloud enables users to use thousands of CPU cores and GPU accelerators casually, and various software types can be used very easily by cloud images, the cloud is beginning to be used in the field of bioinformatics. In this study, we ported the original protein-protein interaction prediction (protein-protein docking) software, MEGADOCK, into Microsoft Azure as an example of an HPC cloud environment. A cloud parallel computing environment with up to 1,600 CPU cores and 960 GPUs was constructed using four CPU instance types and two GPU instance types, and the parallel computing performance was evaluated. Our MEGADOCK on Azure system showed a strong scaling value of 0.93 for the CPU instance when H16 instance with 100 instances were used compared to 50, and a strong scaling value of 0.89 for the GPU instance when NC24 instance with 20 were used compared to 5. Moreover, the results of the usage fee and total computation time supported that using a GPU instance reduced the computation time of MEGADOCK and the cloud usage fee required for the computation. The developed environment deployed on the cloud is highly portable, making it suitable for applications in which an on-demand and large-scale HPC environment is desirable.

HoloMol: Protein and Ligand Visualization System for Drug Discovery with Augmented Reality

Atsushi Koyama, Shingo Kawata, Wataru Sakamoto, Nobuaki Yasuo, Masakazu Sekijima
Tokyo Institute of Technology, Japan

Abstract: To develop effective drugs against various diseases, it is vital to understand the three-dimensional (3D) structures of proteins and drug candidates that serve as drug targets. In the field of drug discovery, molecular structure display systems that are displayed on computer displays are used. In these systems, the 3D structures of the proteins and drug candidates are projected and visualized in two dimensions. In this study, we construct a molecular structure visualization system that visualizes the 3D structures of proteins and drug candidates that are essential for drug discovery with augmented reality (AR) using HoloLens.

Leave-one-element-out Cross Validation for Band Gap Prediction of Halide Double Perovskites

Hiroki Igarashi, Nobuaki Yasuo, Masakazu Sekijima
Tokyo Institute of Technology, Japan

Abstract: Perovskite solar cells have attracted much attention as a new type of solar cell that can be smaller and thinner than conventional silicon solar cells. However, the development of lead-free perovskite solar cells is required because currently most of them contain lead, which is harmful to the human body and the environment. In addition, the field of materials informatics, which combines materials development with information technology and computational science, has become active in recent years. Research on materials development that incorporates machine learning methods has become common in order to develop better materials quicker. In this paper, we aim to predict the band gap, one of the properties of unknown lead-free perovskite materials, by using machine learning methods. We focused on an element and constructed a prediction model to evaluate the case where the element is not included in the training data.

Interpretation of ResNet by Visualization of Preferred Stimulus in Receptive Fields

Genta Kobayashi, Hayaru Shouno
The University of Electro-Communications, Japan

Abstract: One of the methods used in image recognition is the Deep Convolutional Neural Network (DCNN). DCNN is a model in which the expressive power of features is greatly improved by deepening the hidden layer of CNN. The architecture of CNNs is determined based on a model of the visual cortex of mammals. There is a model called Residual Network (ResNet) that has a skip connection. ResNet is an advanced model in terms of the learning method, but it has not been interpreted from a biological viewpoint. In this research, we investigate the receptive fields of a ResNet on the classification task in ImageNet. We find that ResNet has orientation selective neurons and double opponent color neurons. In addition, we suggest that some inactive neurons in the first layer of ResNet affect the classification task.

Bayesian Sparse Covariance Structure Analysis for Correlated Count Data

Sho Ichigozaki, Takahiro Kawashima, Hayaru Shouno
The University of Electro-Communications, Japan

Abstract: In this paper, we propose a Bayesian Graphical Lasso for correlated countable data and apply it to spatial crime data. In the proposed model, we assume a Gaussian Graphical Model for the latent variables which dominate the potential risks of crimes. To evaluate the proposed model, we determine optimal hyperparameters which represent samples better. We apply the proposed model for estimation of the sparse inverse covariance of the latent variable and evaluate the partial correlation coefficients. Finally, we illustrate the results on crime spots data and consider the estimated latent variables and the partial correlation coefficients of the sparse inverse covariance.

Gaze Analysis of Modification Policy in Debugging an Embedded System

Takeru Baba, Erina Makihara, Hirotaka Yoneda, Kiyoshi Kiyokawa, Keiko Ono
Nara Institute of Science and Technology, Japan

Abstract: In embedded system development, the debugging is difficult for novice because developers must consider the state of the hardware and the software. Therefore, this study analyzed the gaze transition in the debugging of the embedded system by experts and novices. The gaze data reveals the difficult points of novices and the hidden technique of experts in the debugging process. The analysis segmented the time-series data of gaze-object using GP-HSMM, which is an unsupervised learning and time-series data division method with great accuracy. The results showed that experts tend to debug in three phases, namely, circuit debugging in the early stage, source code debugging in the middle stage, and confirmation of both the circuit and source code in the final stage. Based on the temporal trend of gazing at an object, we proposed that the teaching contents of modification policy for the novices in order to increase debugging efficiency.

The 19th International Conference on Security and Management (SAM 2020)
http://sam.udmercy.edu/sam20/

A Comprehensive Survey on Fingerprint Liveness Detection Algorithms by Database and Scanner Model

Riley Kiefer, Ashok Patel
Department of Computer Science, Florida Polytechnic University, Lakeland, Florida, USA

Abstract: This comprehensive survey highlights the state-of-the-art solutions to fingerprint liveness detection across a variety of datasets and scanner models. This paper includes most algorithms published between 2014 and 2019, which are ranked according to the Average Classification Error (ACE, the average of the statistical Type I and II errors), Error Rate (ER, the ratio of misclassified fingerprints to total fingerprints), or Accuracy Rate (AR, the ratio of correctly classified fingerprints to total fingerprints), for each scanner model in each dataset. Most algorithms surveyed in this paper were tested on the various LivDet datasets, but other popular datasets such as ATVS and FVC2000 are included in this survey as well. This paper reviews the LivDet competition series and its progress over time, the various published algorithm performances on all available LivDet datasets (2009 – 2017), the performance of traditional machine learning algorithms and their variants, and the performance on miscellaneous datasets. This paper aims to facilitate the research and development of novel liveness classification algorithms through a clear comparison of algorithm performance.

Security of DBMSs

Suhair Amer
Department of Computer Science, Southeast Missouri State University, Cape Girardeau, Missouri, USA

Abstract: Database management systems utilize various security measures but still have some known weaknesses. A security risk of ADABAS occurs from direct calls to the database from an unauthorized third-generational language program. Adaptive Server Enterprise is weak against denial of service attacks. Advantage Database Server security issue results from Dynamic-link library injections. Access's security weakness is the result of use of macros. InterBase's built in backdoor left databases vulnerable for full access to the system. Other security risks will be discussed for Datacom, File Maker, Integrated Database Management System, Informix, Ingres, InterSystems Cache, SQL Server.

Phishing Prevention Using Defense in Depth

Joel Williams, Job King, Byron Smith, Seyedamin Pouriyeh, Hossain Shahriar, Lei Li
Information Technology Department, Kennesaw State University, Marietta, Georgia, USA

Abstract: Phishing attacks are very damaging, costing even moderately sized companies an average of over $1 million per year. It is also a fast growing and quickly evolving threat with nearly 1.5 million phishing websites created each month while each phishing site having an average duration of 54 hours. The combination of the potential damage done by phishing attacks and the complex task of "staying ahead of the phishers" requires a layered approach composed of multiple complimentary components. In this paper, we investigate phishing prevention from three different perspectives including web security gateway, email security gateway and in depth security awareness training.

Automotive Vehicle Security Metrics

Guillermo A. Francia, III
Center for Cybersecurity, University of West Florida. Pensacola, Florida, USA

Abstract: The emergence of connected and autonomous vehicles is at a rapid and unprecedented pace. A recent state-sponsored initiative calls for building a transportation information network that utilizes intelligent sensors on sections of state highways. Indeed, vehicle to infrastructure communication systems are slowly evolving. Peripheral sensors that are used to assist the human operator in lane changing, obstacle avoidance, and parking is becoming a norm in modern automotive vehicles. Although this newly found convenience is a boon to society, both socially and economically, it presents security challenges that are endemic to connected technologies. These challenges underscore the need to look closely at the state of automotive vehicle security. In conjunction with this effort, it is imperative that we investigate the metrics with which security can be measured. As a major component of continuous improvement, quantitative and qualitative measures must be devised to be able to make a full appreciation of the process. This chapter describes automotive vehicle security metrics and derives sample attack calculations to illustrate their applicability.

An Analysis of Applying STIR/Shaken to Prevent Robocalls

James Yu
DePaul University, Chicago, Illinois, USA

Abstract: The robocall epidemic has caused millions of phone scam victims resulting in billions of financial loss in US. To address this issue, Federal Communication Commission (FCC) mandates all Internet Telephony Service Providers (ITSP) to implement Secure Telephony Identity Revisited (STIR) with the Signature-based Handling of Asserted information using toKENs (SHAKEN) to authenticate Voice of Internet Protocol (VoIP) calls on their networks. This paper provides an analysis of the effectiveness of STIR/SHAKEN in protecting users from being victims of robocalls which are mostly scam calls with fake caller ID. Our analysis shows three majors issues that could have impact on the effectiveness of STIR/SHAKEN. These issues are (a) poorly protected enterprise IP-PBX, (b) unscrupulous service providers, and (c) lack of support of Q.1912.5 which is the interworking standard between IP and Public Switch Telephone Network (PSTN).

A Practice of Detecting Insider Threats within a Network

David Velez, Harry Staley, Navin Mathew, Daniel De Leon II, Jeong Yang
Department of Computing and Cyber Security, Texas A&M University - San Antonio, Texas, USA

Abstract: Internal threats within a work environment are becoming more common, while external threats are no longer the only means of concern for a network. The rise of social media, websites, and the availability of Open Source Intelligence (OSINT) is being utilized for blackmail, extracting personal data, and scoping out potential targets. Likewise, data leaks and hacks are daily news, legitimizing security and privacy concerns with identity theft, tarnishing corporate/individual reputation, and theft of intellectual property. Behind firewalls, routers, switches, and multiple layers of security, employees are seldom aware of what goes on within their network. Because the security of the network is placed solely within the hands of IT Professionals, the impression is that security is not also the responsibility of every person on the network. This study investigates the source of privacy and security issues to help identify when, where, and how attacks are initiated within a network. As all exploits begin with reconnaissance, the investigation is conducted by detecting active reconnaissance, over the wire, in real-time using existing technical controls and fingerprinting that can detect threats before an attack. Furthermore, the study explores common, real-world vulnerabilities for exploitation in use cases such as port and vulnerability scanners. The study also identifies these attack vectors through filters and internal network traffic detectors. This research is intended to classify potential threats inside the network from collected reconnaissance scans to thwart impending attacks and illuminate how everyone in a work environment plays a role in protecting against the insider threat.

Phishing Detection using Deep Learning

Beatrice M. Cerda, Shengli Yuan, Lei Chen
University of Houston - Downtown, Houston, Texas, USA;
Georgia Southern University Statesboro, Georgia, USA

Abstract: The rapid advancements in technology come with complex security challenges. One such challenge is phishing attack. Often a fake website is deployed to trick users into believing the website is legitimate, and is safe to give away sensitive information such as their passwords. Anti-Phishing frameworks have been developed in various forms. The most recent implementation involves datasets used to train machines in detecting phishing sites. This paper focuses on implementing a Deep Feedforward Artificial Neural Network using supervised learning to detect phishing URLs. Several models were created that used a single feature to train. We compared how effective each feature was in detecting phishing URLs. Groups of features were also used to train models. Most models using only one feature yielded low accuracies, while models using more features showed better accuracies.

Chor-Rivest Knapsack Cryptosystem in a Post-Quantum World

Raul Duran Diaz, Luis Hernandez Encinas, Agustin Martin Munoz, Araceli Queiruga Dios
Departamento de Automatica, Universidad de Alcala, Alcala de Henares, Spain;
Consejo Superior de Investigaciones Cientificas (CSIC), Madrid, Spain;
Departamento de Matematica Aplicada, Universidad de Salamanca, Salamanca, Spain

Abstract: To be provided later.

Towards Home Area Network Hygiene: Device Classification and Intrusion Detection for Encrypted Communications

Blake A. Holman, Joy Hauser, George T. Amariucai
Kansas State University, Manhattan, Kansas, USA

Abstract: With the abundance of Internet of Things (IoT) devices on the market, proper home area network (HAN) hygiene is not only desirable for easy management and maintenance, but also a requirement at the foundation of any security measures. To ensure HAN hygiene, a method is proposed for automatic device detection and classification. Given the popularity of dynamic IP address allocation, and the increasing popularity of end-to-end encrypted communications, this method relies solely on communication metadata that can be extracted from network tracffic. But rather than extract explicit statistical features of traffic over sliding or hopping windows, this method instead uses entire sequences of packets, where each packet is represented by a tuple describing its length and the duration of the associated subsequent inter-packet pause. The proposed classifier is implemented as a recurrent neural network, and achieves encouraging accuracy, demonstrating that even the simplest form of communication metadata (and thus the least privacy-invasive) is a valuable resource for keeping track of the devices on our networks.

Suitability of Voice Recognition within the IoT Environment

Salahaldeen Duraibi, Fahad Alqahtani, Frederick T. Sheldon, Wasim Alhamdani
Jazan Uinversity, Jazan, Saudi Arabia, KSA; Prince Sattam Bin Abdulaziz University, Saudi Arabia, KSA;
University of Idaho Moscow, Idaho, USA; University of the Cumberlands Williamsburg, USA

Abstract: In this paper, the usability of voice biometric authentication is investigated. In doing so, we first studied the current status of Internet of Things [IoT] authentication technologies. We subsequently determined that existing authentication systems are suffer from a number of deficiencies. Some systems need higher computing and energy resources to perform their work, which exacerbates the resource-constrained nature of IoT technology. Others suffer from security deficiencies. Our implementation of the voice biometric shows that it can be used remotely to authenticate the user on their IoT devices.

Requirements for IoT Forensic Models – A Review

Nawaf Almolhis, Abdullah Mujawib Alashjaee, Micheal Haney
Computer Science Department, Jazan University, Jazan, Saudi Arabia, KSA;
Computer Science Department, Northern Borders University, Arar, Saudi Arabia, KSA;
Computer Science Department, University of Idaho Moscow, Idaho, USA

Abstract: Based on the elasticity, virtualization, volatility, multi-tenancy and multi-jurisdiction characteristics of IoT environments, implementation of conventional digital forensic process models to security incidents is challenging. To address the challenges, new IoT-specific forensic process models are required. In this paper, issues and challenges in digital forensics brought by IoT technologies are presented. Based on the challenges authors derive and present requirements that should be met by forensic process models designed for investigations in the IoT ecosystem. IoT forensic process models proposed in the literature are evaluated based on these requirements. Through this evaluation, gaps that are left by the evaluated IoT forensic process models are brought to light.

Mobile Malware Forensics Review: Issues and Challenges

Abdullah Mujawib Alashjaee, Nawaf Almolhis, Micheal Haney
Computer Science Department, Northern Borders University, Arar, Saudi Arabia, KSA;
Computer Science Department, Jazan University, Jazan, Saudi Arabia, KSA;
Computer Science Department, University of Idaho Moscow, Idaho, USA

Abstract: In this paper, recent scholastic research conducted by various groups in the field of mobile malware detection, analysis, and associated forensics techniques are reviewed. Information on mobile malware evolution, investigative procedures, methodologies on detection, current tools and related implications are presented. The purpose of this work is to provide insights into the mobile malware industry and a better understanding of the current tools and techniques in mobile malware forensics. Special consideration has been given to mobile malware detection and analysis architecture to gain a clear picture of how specialized tools and techniques are used in investigating incidents involving mobile malware. Practical implications of implementing conventional malware forensics techniques in mobile malware forensics are also presented.

Secure Authentication Protocol for Drones in LTE Networks

Dayoung Kang, Gyuhong Lee, Jin-Young Choi
The Graduate School of Information Security, Korea University, Korea;
The Department of Cyber-warfare, Korea Army Academy at Yeongcheon, Korea;
The Graduate School of Information Security, Korea University, Korea

Abstract: As LTE networks account for more than 50% of GSM subscribers in 2020, drone pilots have used LTE for communication with drones. The pilots could operate the drones in large spaces, and the possibility of using drones has spontaneously increased in various industries with LTE. If a malicious adversary exploits the leakage of the IMSI of drones, which existed as a security threat in the LTE networks, it may be vulnerable to the location privacy and security of the drone. In this paper, we define the architecture of drones using LTE networks, and show that the plaintext IMSI leaks out of LTE networks. Later, while using the LTE network, we explain that the leakage of IMSI poses a threat to drone location privacy and security. Next, we propose a protocol for the drone authentication in LTE that would improve the security of drones, especially related to location privacy. Finally, we formally specify the proposed authentication protocol and verify that the proposed protocol satisfies the security properties by using the protocol verification tool Scyther.

The Organizational Cybersecurity Success Factors: An Exhaustive Literature Review

Nancy Poehlmann, Kevin Matthe Caramancion, Mehdi Barati, Irem Tatar, Terry Merz
Department of Information Science, University at Albany, State University of New York, Albany, New York, USA

Abstract: To identify gaps in cybersecurity research, the authors reviewed the secondary literatures of organizational cybersecurity posture, focusing on five factors: technology design, management procedures, organizational structure, the legal environment, and human competencies. Their findings show security technology can be improved with advanced technology, usability, and design process. Organizations should employ a unified security platform for all security tools and consider insider threats in the design of cybersecurity-related organizational procedures. Legal literature highlights two major concerns for the future: small and mid-sized businesses will be major cyberattack victims and human error is the largest factor in cyberattack success. Given the user as the weakest link of cybersecurity, human-centric technology designers should create human-centric designs in addition to training personnel against cyber incidents.

Enhancing Data Security in the User Layer of Mobile Cloud Computing Environment: A Novel Approach

Noah Oghenfego Ogwara, Krassie Petrova, Mee Loong Yang, Stephen MacDonell
Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand

Abstract: This paper reviews existing Intrusion Detection Systems (IDS) that target the Mobile Cloud Computing (MCC), Cloud Computing (CC), and Mobile Device (MD) environment. The review identifies the drawbacks in existing solutions and proposes a novel approach towards enhancing the security of the User Layer (UL) in the MCC environment. The approach named MINDPRES (Mobile-Cloud Intrusion Detection and Prevention System) combines a host-based IDS and network-based IDS using Machine Learning (ML) techniques. It applies dynamic analysis of both device resources and network traffic in order to detect malicious activities at the UL in the MCC environment. Preliminary investigations show that our approach will enhance the security of the UL in the MCC environment. Our future work will include the development and the evaluation of the proposed model across the various mobile platforms in the MCC environment.

A Posteriori Access Control with an Administrative Policy

Farah Dernaika, Nora Cuppens-Boulahia, Frederic Cuppens, Olivier Raynaud
IMT Atlantique, Rennes, France; Be-ys Research, Geneve, Switzerland;
Polytechnique Montreal, Montreal, Canada

Abstract: The a posteriori access control is a particular type of access control in which security auditors tend to verify policy compliance by analyzing logs. However, in this type of access control, the investigations are done after the occurrence of a log event. Therefore, the rules defined in the security policy might not be the same as the ones that were at the time of the access. Furthermore, administrators should have the right privileges to perform a modification on the security policy, and these privileges might be altered over time as well. This paper treats the a posteriori access control in case of an evolving policy. It proposes an approach, based on the Event Calculus, that consists in getting the security rules that were in place at the time of the access, as well as monitoring administrators. actions since themselves can be held accountable.

A Hybrid AI and Simulation-Based Optimization DSS for Post-Disaster Logistics

Gonzalo Barbeito, Dieter Budde, Maximilian Moll, Stefan Pickl, Benni Thiebes
Fakulat fur Informatik, Universitat der Bundeswehr Muchen, Germany;
German Committee for Disaster Reduction (DKKV), Bonn, Germany

Abstract: This paper describes a new simulation-based optimization framework for developing and assessing relief distribution strategies following a disaster where citizen's relocation is necessary. This work has been conceptualized and is currently being developed with and for experts in the field of disaster and security management, in order to tackle the real issues arising during such crises.

A Hybrid Recommender System for Cybersecurity Based on a Rating Approach

Carlos Ayala, Edison Loza-Aguirre, Kevin Jimenez, Roberto O. Andrade
Escuela Politecnica Nacional, Quito, Ecuador

Abstract: The main function of a security analyst is to protect and make the best decisions for preserving the integrity of computer systems within an organization. To provides a quick response, the analyst usually depends on his good judgement, which should lead him to execute manual processes in a limited time. By dealing with too many anomalies, responses are only provided to those threats with the highest level of criticality. This research aims to propose a tool for helping analysts to filter out anomalies and latent risks. To meet this objective, a recommendation system based on collaborative filtering and knowledge was developed, generating ratings of the worst cases with the best available recommendations based on expert judgement. During tests, the system allowed an improvement in the response time from analysts to solve problems. It also eliminated subjectivity and reduced the number of manual processes.

Vulnerability of Virtual Private Networks to Web Fingerprinting Attack

Khaleque Kamal, Sultan Almuhammadi
King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

Abstract: Virtual private networks (VPN) are used to maintain secrecy of internet usage. They provide end-to-end encrypted traffic to hide the content and destination details from potential eavesdroppers. Recent studies show that 72% of VPN users apply it to access blocked content or hide identity from government. Concerned government departments and other organizations need to analyze the encrypted traffic of VPN to observe whether people are using blocked content or not. Typical traffic analysis fails in this case as traffic is encrypted and the destination IP address is hidden. However, traffic metadata and some packet attributes, like packet size and time, could be considered as fingerprint of any specific web service. In this paper, we analyze five commonly used VPN services, namely: Psiphone, Softether, HotspotShield, OpenVPN, and AviraPhantom. Our goal is to identify which VPN service is vulnerable to Cai et al. fingerprinting attack, and what types of web services are most appropriate to detect using this attack. The results show that Open VPN is more vulnerable to this attack compared to the other VPNs in this study. The efficiency of the web traffic classification through VPN is also estimated for four different web services with Cai et al. Useful recommendations are provided as a result of this study.

Secure-Stor: A Novel Hybrid Secure Edge Server Architecture and CDN to Enhance the Security and Response Time for Edge Devices

Mais Nijim, Raghavar Reddy, Muhammad Aurangzeb
Electrical Engineering and Computer Science Department, Texas A&M University-Kingsville, Texas, USA

Abstract: The Internet of Things (IoT) is grabbing the attention of scientists all over the globe. It is explosive growth along with its applications that require real-time computing power has emerged the edge computing. As a result, Edge Computing changed the way data is processed and handled back and forth for millions of devices around the world, such as autonomous vehicles and electric cars. Although, the confinement of the cloud computing technology, such as a Content Delivery Network (CDN), contributed significantly to the development of Edge Computing. Nowadays, this technology can meet the demands of the ever-growing mobile devices and IoT. In this paper, we propose an innovative framework that consists of a hybrid storage architecture-containing could provide computing (CDN), edge computing, and centralized storage. The centralized storage consists of Solid-State Drives (SSDs), Hard Disk Drives (HDDs), which can provide an ideal data storage solution for a wide variety of real-time data processing applications. Transforming the data back and forth between the SSDs and the HDDs plays a critical role in achieving high performance while meeting the deadline of the edge device request. Additionally, we introduced a new dynamic solid-state disk partitioning to optimize security for the proposed framework, among hard disk drives where we assign a partition from the solid-state disks to hard disk drives based on the workloads of the hard disk drives.

Intrusion Detection through Gradient in Digraphs

S. S. Varre, Mais Nijim, Muhammad Aurangzeb
Electrical Engineering and Computer Science Department, Texas A&M University-Kingsville, Texas, USA

Abstract: Unmanned autonomous vehicles (UAVs) driven by information obtained through sensor networks believe to have a pivotal part in every area of life in near future. In order to fulfil their objectives, the UAV will move through the data obtained by the associated network of sensors and of UAVs. The data includes heading to reach a destination, the heading to hit a goal, the volume of the goal, and the min path to reach a destination. The accomplishment of such a task will be highly dependent upon the precision and legitimacy of any such data obtained on the ground. The significance of the data invites the adversary to interrupt the data collection network. The most shocking way to interrupt a data collection network is to insert malicious node(s) in the networks to contaminate the data and render it worthless. This sort of attacks upon networks are serious and quiet. Moreover, such attacks are straightforward to unveil as they need nominal resources from the adversary's part. Gradient, level surfaces, and scalar fields are known concepts in thermodynamics and physics. This paper extends these concepts to the networks. By using the extended concepts of gradients, level surfaces, and scalar point functions this paper provides a novel linear time algorithm to find the min path to a malicious node within a network.

Statistical Analysis of Prime Number Generators Putting Encryption at Risk

Aykan Inan
University of Applied Sciences Ravensburg-Weingarten, Baden-Wurttemberg, Germany

Abstract: To be provided later.

Enhancing the Cybersecurity Education Curricula Through Quantum Computation

Hisham Albataineh, Mais Nijim
Department of Physics and Geoscience, Texas A&M University-Kingsville, Texas, USA;
Department of Electrical Engineering and Computer Science, Texas A&M University-Kingsville, Texas, USA

Abstract: Governmental and corporate networks are increasingly at risk, malicious acts are escalating abruptly, and organized crime and terrorist groups are expanding and improving their cyber capabilities. Individuals and countries have initiated attacks against others, including the public and private sectors. Hiring qualified cybersecurity experts is currently considered to be the highest priority in the United States. Quantum computation is a revolutionary technology for Cybersecurity education. The traditional digital computers run on encoded data according to the standard binary system 0 or 1. With the emerging of quantum computation, quantum computers run qubits or quantum bits instead of binary data 0 1nd 1. Quantum computation is vital for Cybersecurity education. The overarching goal of this paper is to present a cybersecurity educational model with a focus on quantum computation. The given education model is designed according to the National Institute of Standard and Technology (NIST), and the industry's expectations regard the knowledge, training, and skills that Cybersecurity experts should possess.

Memorable Password Generation with AES in ECB Mode

Timothy Hoang, Pablo Ravis
School of Computer Science and Mathematics, Marist College, Poughkeepsie, New York, USA

Abstract: This paper presents ALP: an adjustable Lexicon-generated password mechanism. ALP is a password encryption web application. In ALP, a password is generated randomly from an imported dictionary of words. For example, with an entire dictionary, it will pick out random strings that add up to a 128-bit hex key and return the key to the user. The password will be "easy to remember" since it will be made of a chosen pool of words. It will then take in the user's message to be encrypted and output the encrypted message according to the key. The password generator, if need be, can be switched to a completely random mode which will make it output a string of random numbers and letters if the user does not choose to use the easy to remember password generator function. The application will also be able to decrypt the AES-128 encrypted message and transform it back into normal text. This experiment suggests that the proposed model is a block cipher that can successfully generate a memorable random password.

Supervised Learning for Detecting Stealthy False Data Injection Attacks in the Smart Grid

Mohammad Ashrafuzzaman, Saikat Das, Yacine Chakhchoukh, Salahaldeen Duraibi, Sajjan Shiva, Frederick T. Sheldon
Department of Computer Science, University of Idaho, Idaho, USA;
Department of Computer Science, University of Memphis, Tennessee, USA;
Department of Electrical and Computer Engineering, University of Idaho, Idaho, USA

Abstract: The largest and the most complex cyber-physical systems, the smart grids, are under constant threat of multi-faceted cyber-attacks. The state estimation (SE) is at the heart of a series of critical control processes in the power transmission system. The false data injection (FDI) attacks against the SE can severely disrupt the power systems operationally and economically. With knowledge of the system topology, a cyber-attacker can formulate and execute stealthy FDI attacks that are very difficult to detect. Statistical, physics-based, and more recently, data-driven machine learning-based approaches have been undertaken to detect the FDI attacks. In this paper, we employ five supervised machine learning models to detect stealthy FDI attacks. We also use ensembles, where multiple classifiers are used and decisions by individual classifiers are further classified, to find out if ensembles give any better results. We also use feature selection method to reduce the number of features to investigate if it improves detection rate and speed up the testing process. We run experiments using simulated data from the standard IEEE 14-bus system. The simulation results show that the ensemble classifiers do not perform any better than the individual classifiers. However, feature reduction speeds up the training by many fold without compromising the model performance.

Vulnerability Analysis of 2500 Docker Hub Images

Katrine Wist, Malene Helsem, Danilo Gligoroski
Norwegian University of Science and Technology (NTNU), Norway

Abstract: The use of container technology has skyrocketed during the last few years, with Docker as the leading container platform. Docker's online repository for publicly available container images, called Docker Hub, hosts over 3.5 million images at the time of writing, making it the world's largest community of container images. We perform an extensive vulnerability analysis of 2500 Docker images. It is of particular interest to perform this type of analysis because the vulnerability landscape is a rapidly changing category, the vulnerability scanners are constantly developed and updated, new vulnerabilities are discovered, and the volume of images on Docker Hub is increasing every day. Our main findings reveal that (1) the number of newly introduced vulnerabilities on Docker Hub is rapidly increasing; (2) certified images are the most vulnerable; (3) official images are the least vulnerable; (4) there is no correlation between the number of vulnerabilities and image features (i.e., number of pulls, number of stars, and days since the last update); (5) the most severe vulnerabilities originate from two of the most popular scripting languages, JavaScript and Python; and (6) Python 2.x packages and jackson-databind packages contain the highest number of severe vulnerabilities. We perceive our study as the most extensive vulnerability analysis published in the open literature in the last couple of years.

Leveraging Security Management with Low-Level System Monitoring and Visualization

Karlen Avogian, Basel Sababa, Ioanna Dionysiou, Harald Gjermundrod
Computer Science Department, University of Nicosia, Nicosia, Cyprus

Abstract: Preventing security breaches in today's heterogeneous and diverse environment is nontrivial, despite the abundance of security products and technologies in the market, as the attack surface is simply too broad. This paper presents SMAD, an open-source security monitoring tool that monitors kernel and system resources data and aims to detect abnormal activity on a Linux server. As it is not uncommon for users to maintain personal home servers, SMAD empowers these novice administrators with a tool to track their Linux server's health in an intuitive and user-friendly manner. The user-centric SMAD environment allows the specifications of monitors, alerts, and anomaly detection rules to be done in a free-of-errors manner.

An Effective Tool for Assessing the Composite Vulnerability of Multi-Factor Authentication Technologies

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

Abstract: With multi-factor authentication technologies continue to advance and adoption rates for those technologies increase, there exists a need to characterize the composite vulnerability score for complete authentication solutions. To meet this need, we propose an extension to the Common Vulnerability Scoring System (CVSS) v3 calculator to provide an aggregate score for any metric category, enabling organizations and researchers to succinctly determine the composite vulnerability impact of authentication factor multiplicity. This paper has presented a novel mathematical approach and demonstrated the approach through a real-world application which is a comparative study on the composite vulnerability of two different multi-factor authentication technologies.

CyberCheck.me: A Review of a Small to Medium Enterprise Cyber Security Awareness Program

Craig Valli, Ian Martinus, Jayne Stanley, Michelle Kirby
Security Research Institute, Edith Cowan University, Australia;
WA AustCyber Innovation Hub, Australia

Abstract: To be provided later.

Static Analysis for Software Reliability and Security

Hongjun Choi, Dayoung Kang, Jin-Young Choi
School of Cybersecurity, Korea University, Seoul, Republic of Korea

Abstract: Because people make software, they cannot avoid software errors. So, developers often use static analysis in the implementation phase to diagnose and correct these errors. In this paper, we divide the input value requirements of a code of binary search into two perspectives: reliability and security, and deal with the static analysis for the implemented program. The result of static analysis can be either true positive or false positive, depending on which aspect the developer sees. Regardless of what point of view, distinguishing between true positive and false positive is carried out. It takes a lot of time and effort to determine whether it's a true positive or a false positive. After performing static analysis, it is more efficient for programmers to modify all alarms to the correct code without dividing their views.

The Impact of 21st Century Skills and Computing Cognition Cyber Skills on Graduates' Work Readiness in Cyber Security

Anna Griffin, Nicola Johnson, Craig Valli
Cyber Security CRC, Australia;
Security Research Institute, Edith Cowan University, Australia

Abstract: This study aims to explore the impact that 21st century skills and cyber security skills have on graduates' perceived work readiness. By surveying 3rd year undergraduate cyber security student's, it will explore the connection between the technical skills and 'soft skills', and how they individually and jointly, impact on how students' perceive their work readiness. The survey was developed using previously established survey tools for 21st century skills, concentrating on teamwork, critical thinking and problem solving. The cyber security skills were adapted from the NIST NICE cyber security framework and organized using Bloom's adapted taxonomy. Using SEM statistics to develop a model that can be used to identify the significant impact the skill development has on students' perceived work readiness.

Analysis of Conpot and its BACnet Features for Cyber-Deception

Warren Z. Cabral, Craig Valli, Leslie F. Sikos, Samuel G. Wakeling
Cyber Security Co-Operative Research Centre, Australia;
Security Research Institute, Edith Cowan University, WA, Australia

Abstract: There is an increasing awareness of the cybersecurity issues around SCADA/ICS systems that are the cyber-physical connectors of our industrialized world to the Internet. Alerton's BACtalk VAV-DD controller is an example of a multipurpose SCADA device that provides autonomous pressure control for any dual-duct VAV container using the BACnet protocol. The majority of devices functioning on the BACnet protocol are legacy deployments, which are difficult to secure. Therefore, SCADA honeypots such as Conpot are significant tools not only for regulating threats affecting SCADA devices such as the VAV-DD controller, but also for the early detection of probable malicious tampering within a SCADA environment. This paper analyzes the templates of the Conpot honeypot with special emphasis on the default template.xml file and the bacnet.xml protocol file, and their potential to be used deceptively.

Lightweight Network Steganography for Distributed Electronic Warfare System Communications

Tim Lei, Jeremy Straub, Benjamin Bernard
Department of Computer Science, San Francisco State University, San Francisco, California, USA;
Institute for Cyber Security Education and Research, North Dakota State University, Fargo, North Dakota, USA;
Department of Computer Science, North Dakota State University, Fargo, North Dakota, USA

Abstract: This paper presents the application of a modified implementation of the StegBlocks TCP method as part of the Distributed Electronic Warfare System. The existing system is not equipped with a secure information communications mechanism for transmission between assimilated hosts and from hosts back to the server. The method implemented utilizes network steganography to provide covert data transmission through the network using network packets. The proposed implementation is compared to another implementation of the same method on the aspects of the implementations' usability, versatility, and applicability. Discussion on how the proposed implementation is more suitable than the alternate implementation is presented. Future proposed improvements to the implementation are also discussed.

The 18th International Conference on Software Engineering Research and Practice (SERP 2020)
https://americancse.org/events/csce2020/conferences/serp20

A Preliminary Study of Transactive Memory System and Shared Temporal Cognition in the Collaborative Software Process Tailoring

Pei-Chi Chen, Jung-Chieh Lee, Chung-Yang Chen
Trend Micro Inc., Texas, USA; Beijing Normal University Zhuhai, Zhuhai City, P. R. China;
Department of Information Management, National Central University, Taoyuan city, Taiwan

Abstract: Software project teams often need to customize standard software processes to fit projects' particularities. This process customization is called software process tailoring (SPT). SPT is a critical issue in contemporary software development such as agile; its importance is widely recognized in both practice and the academia, yet most of existing SPT literature focus on technical aspect at developing better tailoring guidelines and approaches. However, from the human and managerial perspective, SPT is a collaborative yet highly conflicting process involving task and temporal conflicts, and how teams' operational mechanisms, under these situations, can increase SPT performance remain unknown. To address the aforementioned gap, this study bases on transactive memory systems (TMS) and shared temporal cognitions (STC) to conduct an introductory review on the theories and concepts in the SPT context, and explore how a team's TMS and STC may be applied in the conflictual SPT process to facilitate the tailoring performance.

Overall Scheduling Requirements for Scheduling Synthesis in Automotive Cooperative Development

Arthur Strasser, Christoph Knieke, Andreas Rausch
Technische Universitat Clausthal, Institute for Software and Systems Engineering, Arnold-Sommerfeld-Strasse, Clausthal-Zellerfeld, Germany

Abstract: In cooperative development, vehicle manufacturers follow the AUTOSAR standard to synthesize the overall scheduling of a componentbased real-time software system. However, the vehicle manufacturer is not able to define overall scheduling requirements as the decomposition of subsystems into software components is determined in later development steps by subsystem developers. The determination of an overall scheduling can only take place when integrating the subsystem models into the system model. Due to the missing scheduling requirements costand time-consuming quality assurance measures are required to find a correct overall scheduling. To overcome this challenge, this paper presents the PortChain as a specification for overall scheduling requirements extending the AUTOSAR standard and an approach for overall scheduling synthesis based on that specification.

How to Test Interoperability of Different Implementations of a Complex Military Standard

Andre Schobel, Philipp Klotz, Christian Zaschke, Barbara Essendorfer
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Interoperability and Assistance Systems (IAS), Karlsruhe, Germany

Abstract: To access and use the functionality of a software system, different interfaces must be provided. When establishing a system-to-system communication, a clear definition and specification of the interfaces must exist. In the NATO (North Atlantic Treaty Organization) environment different Standardization Agreements (STANAGs) are defined. When exchanging reconnaissance and surveillance data between NATO partners, the STANAG 4559 is used. To ensure interoperability, it is necessary to be able to test the corresponding implementations regarding the adherence of the guidelines defined in the standard. Within this paper, an idea of a test center is presented which can be used as an independent instance to test and validate a system's conformance to the STANAG 4559 standard.

Water Market for Jazan, Saudi Arabia

Fathe Jeribi, Sungchul Hong, Ali Tahir
College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia;
Computer and Information Sciences Department, Towson University, Towson, Maryland, USA

Abstract: Water is a limited resource and it is a necessary of the life necessaries. Lives of individuals could be hard without water. Consequently, the diversity of ways of getting water is required. Water can be utilized for humans, farms, or industry. The problem is that some regions around the world do not have enough amount of water to utilize. The water in the Jazan region is generated using desalination plants, which is costly. Water can be generated using desalination plant, water well, or Water from Air (WFA) method. Rainfall in the Jazan region is limited; however, it is hot area and it has fairly high humidity as well as constant wind. People or industry can utilize solar energy and wind energy to generate electricity. After that, they can use them for operate water making devices. The goal of this paper is to demonstrate the possibility of water market by computer simulations. In addition, it aims to help individuals to take benefit of water making in dry area with some humidity.

Securing a Dependability Improvement Mechanism for Cyber Physical Systems

Gilbert Regan, Fergal McCaffery, Pangkaj Chandra Paul, Ioannis Sorokos, Jan Reich, Eric Armengaud, Marc Zeller
Lero @DKIT, Dundalk, Ireland; Fraunhofer IESE, Kaiserslautern, Germany;
AVL List GmbH, Austria; AVL List GmbH, Turkey; Siemens, Munich, Germany

Abstract: The open and cooperative nature of Cyber-Physical Systems (CPS) poses a significant new challenge in assuring dependability. A European funded project named DEIS addresses this important and unsolved challenge by developing technologies that facilitate the efficient synthesis of components and systems based on their dependability information. The key innovation that is the aim of DEIS is the corresponding concept of a Digital Dependability Identity (DDI). A DDI contains all the information that uniquely describes the dependability characteristics of a CPS or CPS component. In this paper we present an overview of the DDI, and provide the protocol for ensuring the security of the DDI while it is in transit and rest. Additionally, we provide confidentiality, integrity and availability validation of the protocol.

Melody Based Pitch Correction Model for a Voice-Driven Musical Instrument

John Carelli
Computer Science & Information Technology, Kutztown University of Pennsylvania, Kutztown, Pennsylvania, USA

Abstract: An investigation is presented into musical note recognition and pitch correction using only recently sung notes captured in real time. The goal is to improve, by correcting inaccuracies in singing execution, the performance of a voice driven musical instrument that translates sung pitch into notes played by a separate virtual instrument. This is accomplished without a-priori knowledge of musical key and, in order to enable real-time response, with a minimal number of recent sung notes. The technique involves inferring a tonal center, or set of keys, then performing analysis and pitch correction based on expected note usage in that tonal center. It provides an interesting and potentially useful study in human/computer interaction with application in live musical performance.

Extracting Supplementary Requirements for Energy Flexibility Market Place

Tommi Aihkisalo, Kristiina Valtanen, Klaus Kansala
VTT Technical Research Center of Finland, Kaitovayla, Oulu, Finland

Abstract: Finding even a partially covering set of requirements for a system may be challenging some times. The modern quick prototyping and minimum viable product approaches may not favor a full scale requirement analysis and respective design work. For this study we initially created a minimum working example of an electricity flexibility market place targeting mainly household level custom-ers. It was successful in a technical sense as many of the features were based on the generic patterns of online trading especially on the stock markets. However, after the very brief analysis and implementation phase it was apparent that do-main expertise was necessary in some of the energy related and other general details. The expert panel was created from the research project participants rep-resenting expertise on the electrical energy, related ICT solution and grid gov-ernance field. The panel was somewhat familiar with the already existing solution and was able to produce feedback and their expert views that mostly was recorded on the free formed worksheets provided for the occasion. This research is based on the analysis of those recorded worksheets and presents an extraction of the supplementary requirements for the electricity flexibility trading market place.

Technical Personality as Related to Intrinsic Personality Traits

Marwan Shaban, Craig Tidwell, Janell Robinson, Adam J. Rocke
Seminole State College, Sanford, Florida, USA

Abstract: Technical personality is introduced and defined as a profile of the broad technical preferences of an individual. Intrinsic personality is the customary personality profile as measured by a standard personality test such as the Big 5 personality test. The relationship between intrinsic and technical personalities is evaluated in this research. The proposed technical personality traits are defined and statistical data about each technical trait is provided, and evaluated in order to identify any correlation between intrinsic and technical personality traits. The result of this research provides a useful tool for self-discovery and can also confirm or dispel certain stereotypes in fields related to information technology.

Mixed-integer Linear Programming Models for the Simultaneous Unloading and Loading Processes in a Maritime Port

Ali Skaf, Sid Lamrous, Zakaria Hammoudan, Marie-Ange Manier
Univ. Bourgogne Franche-Comte, FEMTO-ST Institute/CNRS, (UTBM), Belfort, France

Abstract: This paper discusses the jointly quay crane and yard truck scheduling problems (QCYTSP) with unloading and loading containers from/to vessel(s) in the same time. Yard trucks transport the containers to/from yard locations with all containers that are homogeneous. We propose a mixed integer linear programming model to solve the scheduling problem. We consider in this study, the quay crane interference, containers precedence and safety margin. The main objective is to minimize the total completion time of the vessels.

Analysis of Bug Types of Textbook Code with Open Source Software

Young Lee, Jeong Yang
Department of Computing and Cyber Security, Texas A&M University - San Antonio, Texas, USA

Abstract: This study aims to analyze the code examples of two widely adopted collegiate level programming textbooks using static analysis tools and compared them with the bugs found in real-world Open Source projects. Overall, 42.6% of the bugs found in the Texts relate to Internationalization while the Open Source Software (OSS) has 39.8% of their source code associated with Dodgy Code. Bad Practice issues consistently present in both Texts (18.0%) and OSS (26.4%) groups. DM_DEFAULT_ENCODING violation type ranked first in Texts. SE_NO_SERIALVERSIONID ranked the first in OSS. DLS_DEAD_LOCAL_STORE in a Dodge Code category ranked 4th in both Texts and OSS. Textbooks are missing certain code examples that are related to the high ranked bug types in OSS.

A Dynamic Scaling Methodology for Improving Performance of Data Intensive Systems

Nashmiah Alhamdawi, Yi Liu
Department of EE & CS, South Dakota State University, Brookings, South Dakota, USA;
Department of Computer and Information Science, University of Massachusetts Dartmouth, Massachusetts, USA

Abstract: The continuous growth of data volume in various fields such as, healthcare, sciences, economics, and business has caused an overwhelming flow of data in the last decade. The overwhelming flow of data has raised challenges in processing, analyzing, and storing data, which lead many systems to face an issue in performance. Poor performance of systems, such as slow processing speed, creates negative impact such as delays, unprocessed data, and increasing response time. This paper presents a novel dynamic scaling methodology to improve the performance of data-intensive systems. The dynamic scaling methodology is developed to scale up the system based on the several aspects from the data-intensive perspective. Moreover, these aspects are used by the helper project algorithm which is designed to divide a task into small pieces to be processed by the system. These small pieces run on several virtual machines to work in parallel to enhance the system's runtime performance. In addition, the dynamic scaling methodology does not require many modifications on the applied, which makes it easy to use.

Implications of Blockchain Technology in the Health Domain

Merve Vildan Baysal, Ozden Ozcan-Top, Aysu Betin Can
The Scientific and Technological Research Council of Turkey, Ankara, Turkey;
Information Systems, Middle East Technical University, Ankara, Turkey

Abstract: Blockchains are tamper evident and tamper resistant digital ledgers. Due to their distributed and shared nature, and the cryptographic functions; blockchains are resilient to alterations. They provide a trustworthy environment for the data stored in the ledgers. However, each new technology brings its own challenges along with the opportunities. Our minds are still busy with the question of "how could blockchain technology potentially benefit us?" In this paper, we approach to this question from the health domain perspective. Based on a systematic literature review, we discuss to what extent blockchain could provide solutions for the challenges inherited in the health domain and if blockchain technology introduces new challenges for development of health applications. The review included 27 publications which share experiences of practitioners from 2016 to 2020.

Benchmarking the Software Engineering Undergraduate Program Curriculum at Jordan University of Science and Technology with the IEEE Software Engineering Body of Knowledge (Software Engineering Knowledge Areas #1-5)

Moh'd A. Radaideh
Department of Software Engineering, Faculty of Computer & IT, Jordan University of Science & Technology, Jordan

Abstract: This paper evaluates the compliance of the Software Engineering Program (SWE-Curriculum) at Jordan University of Science and Technology (JUST) with the first five of the fifteen Software Engineering Knowledge Areas (SWE-KAs) of the SWEBOK-V3.0 of the IEEE Computer Society. This research is the first to measure the coverage of the SWE-KAs in any SWE-Curriculum. Although the SWE-Curriculum is accredited by the Institute of Engineering and Technology (IET), it is essential to line up the said Curriculum with the IEEE view of Software Engineering (SWEBOK-V3.0). Research Questions: (1) What is the gap in the coverage of the SWE-KAs#1-5 topics across the said SWE-Curriculum? (2) What can be done to eliminate that gap? This research was divided into three parts. This paper focused on SWE-KAs#1-5, the second and third (P#2 and P#3) shall focus on SWE-KA#6-10 and SWE-KA#11-15, respectively. The coverage of SWE-KAs#1-5 was inspected across the SWE-Curriculum courses. The results are identified as either Fully-Compliant (e.g. the SWE-KA is fully covered across one or more of the SWE-Curriculum courses); Highly-Compliant (e.g. the SWE-KA is highly covered); Partially-Compliant (e.g. the SWE-KA is partially covered); or Poorly-Compliant (e.g. the SWE-KA is poorly covered). The compliance was found as Fully-Compliant in the cases of the Software Requirements, Software Design, and Software Testing SWE-KAs. While it was found as Partial-Compliant in the cases of the Software Construction, and Software Maintenance SWE-KAs.

A Framework for Developing Custom Live Streaming Multimedia Apps

Abdul-Rahman Mawlood-Yunis
Physics and Computer Science Department, Wilfrid Laurier University, Waterloo, Canada

Abstract: The rise in the number of mobile-connected devices and the emergence of new streaming models, such as on-demand, on-the-go and interactive streaming, has an impact on the viewing practice of consumers. Mobile devices will play a major role in new viewing habit changes. In this paper, we present a generic framework for developing custom live streaming multimedia apps for mobile devices, i.e., we identify the principle characteristics of mobile live streaming apps, the components, components' interactions, and the design decision needed to create such apps. To demonstrate how the generic framework can be used, we created an app for live streaming audio for Android devices using URLs. The app acts as an instance of the framework and validates it. The paper has two main contributions: 1) The generic framework which can be adapted when developing live streaming multimedia apps or similar apps. 2) Developers and end users can reuse the instance app to their own specific needs using live streaming URLs.

Change Request Prediction in an Evolving Legacy System: A Comparison

Lamees Alhazzaa, Anneliese Amschler Andrews
Computer Science Department, Al-Imam Muhammad Bin Saud Islamic University, Riyadh, Saudi Arabia;
Department of Computer Science, University of Denver, Denver, Colorado, USA

Abstract: Software Reliability Growth Models (SRGM) have been used to predict future defects in a software release. Modern software engineering databases contain Change Requests (CR), which include both defects and other maintenance requests. Our goal is to use defect prediction methods to help predict CRs in an evolving legacy system. CRs include both corrective and perfective requests. Limited research has been done in defect prediction using curve-fitting methods an evolving software systems, with one or more change-points. This work demonstrates the use of curve-fitting defect prediction methods to predict CRs. We compare future CR predictions for three different approaches. Our data show that the Time Transformation (TT) approach provides more accurate CR predictions than the other existing curve-fitting approaches.

Using Clients to Support Extract Class Refactoring

Musaad Alzahrani
Albaha University, Albaha, Saudi Arabia

Abstract: Extract Class refactoring refers to the process of separating the different responsibilities of a class into different classes. Existing approaches of the Extract Class refactoring are based on factors internal to the class, i.e., structural and semantic relationships between methods. However, using the internal factors to identify and separate the responsibilities of the class are inadequate in many cases. Therefore, the paper proposes a novel approach that exploits the clients of the class to support the Extract Class refactoring. The proposed approach is useful and complementary to the existing approaches because it involves factors external to the class to be refactored, i.e., the clients.

Analyzing Technical Debt of a CRM Application by Categorizing Ambiguous Issue Statements

Yasemin Doganci, Ozden Ozcan-Top, Altan Kocyigit
Information Systems Department, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey

Abstract: Poor decisions and suboptimal actions taken in software development result in technical debt. In service business, technical debt may become more evident and destructive. Customer Relationship Management (CRM) platforms is one example of such businesses where several customizations are performed to adapt the software to customers' processes and needs. The purpose of this study is to investigate technical debt in customizations made in different Salesforce CRM organizations based on ambiguous issue statements. We categorized 300 anonymous confessions of Salesforce consultants, administrators, and developers by using three different technical debt categorization approaches. This study would improve awareness among CRM teams for potential technical debts and may serve as a starting point to determine appropriate strategies to deal with technical debt.

Applying DevOps for Distributed Agile Development: A Case Study

Asif Qumer Gill, Devesh Maheshwari
University of Technology Sydney, Australia

Abstract: Agile software engineering principles and practices have been widely adopted by the software-intensive organizations. There is an increasing interest among organizations in adopting DevOps for improving their distributed agile software environments. However, the challenge is how best to adopt and integrate DevOps in their software development environments – especially in distributed agile environment. This paper presents one such successful case study of DevOps adoption by the distributed agile teams for the development and deployment of a real-time high-performance gaming platform. (1) Small teams, (2) trust, (3) active communication and collaboration culture, (4) shared product vision and roadmap, (5) continuous feedback and learning culture, (6) appreciation and excellent senior management support are some of the key success

factors of DevOps. The experiences and learnings discussed in this paper can be used by other organizations to effectively plan and adopt DevOps for their environment.

Further Examination of Youtube's Rabbit-Hole Algorithm

Matthew Moldawsky
Marist College, Poughkeepsie, New York, USA

Abstract: In the past couple of years, YouTube has been criticized for its algorithm radicalizing users. The purpose of this research is to further refute the claims that YouTube's algorithm radicalizes people. An analysis of previous research by Mark Ledwich will be used as a base. Anecdotes gathered by the Mozilla Foundation will also be analyzed. An overview of the problems with the algorithm, as stated by a developer, will be introduced. All of this is to illustrate that the current algorithm is a "rabbit-hole" and is not healthy for the user. Rather, the algorithm does a good job at prioritizes a user's preferences for content which could lead to unintended problems. The unintended problems are what need to be fixed with the YouTube algorithm. Potential solutions are provided and could be used as suggestions for YouTube to investigate to improve the algorithm.

Modeling Unmanned Aircraft System Maintenance Using Agile Model-Based Systems Engineering

Justin R. Miller, Ryan D. L. Engle, Brent T. Langhals, Michael R. Grimaila, Douglas D. Hodson
Air Force Institute of Technology, USA

Abstract: Model-Based System Engineering (MBSE) and agile software development are gaining popularity in the Department of Defense (DoD) because of their promise of more flexible and higher quality products. However, these newer methods have yet to be implemented across the DoD which leads to questions of effectiveness in different environments. This research will seek to determine how effective the use of MBSE and agile software development is in fielded, distributed systems to close a capability gap. A small team will lead the development of both the MBSE products and the agile software prototype. The use of iterative development with continuous customer feedback is expected to yield a more effective product with higher user satisfaction than traditional methods.

A Study of Third-Party Software Compliance and the Associated Cybersecurity Risks

Rashel Dibi, Brandon Gilchrist, Kristen Hodge, Annicia Woods, Samuel Olatunbosun
Department of Computer Science, Norfolk State University, Virginia, USA

Abstract: Companies have to assess the security risk when purchasing Third-Party Software (TPS). There have always been needs for TPS use within companies however, issues like server space and maintenance up-keep have hindered companies from purchasing TPS. The evolution of the cloud has made it easier for companies to make the necessary TPS purchases that they need. Companies use the TPS to modernize certain processes and make others more efficient however; there is a cost. Purchases of TPS that reside in the cloud brought new security risks and compliance issues for companies. This research paper analyzed the associated cybersecurity risks of purchasing TPS and the importance of TPS compliance with companies' policies.