CSoNet 2023

The International Conference on Computational Data and Social Networks (CSoNet) 2023

Program Booklet

Introduction

The 12th International Conference on Computational Data and Social Networks (CSoNet) 2023 provides a premier interdisciplinary forum to bring together researchers and practitioners from all fields of big data networks, such as billion-scale network computing, data network analysis, mining, security and privacy, and deep learning. CSoNet 2023 seeks to address emerging yet important computational problems, with a focus on the fundamental background, theoretical technology development, and real-world applications associated with big data network analysis, modeling, and deep learning. The conference solicits theoretical, methodological, empirical, and experimental research reporting original and unpublished results on computational big data networks. The conference is organized at the National Economics University, Hanoi, Vietnam.

General Co-Chairs

Chuong Pham Hong, National Economics University
Sergiy Butenko, Texas A&M University

TPC Co-Chairs

Minh Hoàng Hà, National Economics University
Xingquan Zhu, Florida Atlantic University

Keynote Speaker

Roger Wattenhofer, ETH Zurich

Panel Discussion

Panelists
Roger Wattenhofer, ETH Zurich
My T. Thai, University of Florida
Xingquan Zhu, Florida Atlantic University
David Mohaisen, University of Central Florida
Moderator
Tien Mai, Singapore Management University

Local Committee

Thanh Hieu Nguyen (Chair), National Economics University
Trung Thanh To, National Economics University
Thanh Tung Dao, National Economics University
Trung Tuan Nguyen, National Economics University

TPC Members

Thien Luong, Phenikaa University
Hocine Cherifi, University of Burgundy
Akrati Saxena, Eindhoven University of Technology
Alexander Semenov, University of Florida

Tuan Anh Hoang, RMIT
Viet Lai, University of Oregon
Quoc Trung Bui, Hanoi University of Science and Technology
Xuan Thanh Le, Institute of Mathematics
Viet Sang Dinh, Hanoi University of Science and Technology
An Dang, Phenikaa University

Publicity Co-Chairs

Nguyen Phi Le, Hanoi University of Science and Technology
Min Shi, Harvard University

Web Chair

Van Canh Pham, Phenikaa University

Publication Co-Chairs

Duc Minh Vu, National Economics University
Hung Tran, Phenikaa University

Steering Committee

My T. Thai (Chair), University of Florida
Kim-Kwang Raymond Choo, University of Texas at San Antonio
Zhi-Li Zhang, University of Minnesota
Weili Wu, University of Texas, Dallas

Keynote Speaker: Professor Roger Wattenhofer

ETH Zurich, Switzerland

Short Biography

Roger Wattenhofer is a full professor at the Information Technology and Electrical Engineering Department, ETH Zurich, Switzerland. He received his doctorate in Computer Science from ETH Zurich. He also worked multiple years at Microsoft Research in Redmond, Washington, at Brown University in Providence, Rhode Island, and at Macquarie University in Sydney, Australia. Roger Wattenhofer's research interests include a variety of algorithmic and systems aspects in computer science and information technology, e.g., distributed systems, positioning systems, wireless networks, mobile systems, social networks, financial networks, deep neural networks. He publishes in different communities: distributed computing (e.g., PODC, SPAA, DISC), networking and systems (e.g., SIGCOMM, SenSys, IPSN, OSDI, MobiCom), algorithmic theory (e.g., STOC, FOCS, SODA, ICALP), and more recently also machine learning (e.g., ICML, NeurIPS, ICLR, ACL, AAAI). His work received multiple awards, e.g. the Prize for Innovation in Distributed Computing for his work in Distributed Approximation. He published the book "Blockchain Science: Distributed Ledger Technology'', which has been translated to Chinese, Korean and Vietnamese.

Learning with Graphs

Abstract

Big and complex data is often represented by a graph. Graphs play an important role in various fields: in natural sciences graphs are used to represent molecules, proteins, or genomes. In math, graphs represent algebraic groups or knots. Graphs are also widely used to model social networks or traffic, and they have dozens of applications in computer science. In this talk we will discuss how machine learning can deal with graphs. While deep learning has been successfully dealing with tables, graphs are much more challenging because the input is not fixed, and extrapolation is necessary. In the talk, we will discuss a few breakthrough stories, but also some open challenges and open problems when it comes to learning with graphs.

Panel discussion

Large Language Models: theory, applications, and challenges

Panelist 01

Roger Wattenhofer
ETH Zurich

Panelist 02

My T. Thai
University of Florida

Panelist 03

Xingquan Zhu
Florida Atlantic University

Panelist 04

David Mohaisen
University of Central Florida

Panelist 05

Tu-Bao Ho
Viasm

Moderator

Tien Mai
Singapore Management University

CSoNet-2023 PROGRAM

National Economics University, Hanoi, Vietnam, December 11-13, 2023

Conference Day 1 – December 11, 2023 (Monday)	
08:00-16:00	Registration Location: Room G, A1 Building
08:30–08:45	Opening Ceremony Location: Room G, A1 Building
	Welcome speech by the President of National Economics University
	General Chairs remarks (Prof. Xingquan Zhu)
08:45-10:00	Keynote Talk 1: **Learning with Graphs** Speaker: **Professor Roger Wattenhofer** Chair: My T. Thai Location: Room G, A1 Building
10:00–10:15	*Coffee break*
10:15-11:55	Session Mon-1A **Machine learning and prediction 1** Chair: Phi Le Nguyen Location: Room G, A1 Building
10:15-10:35	Alexandr Karpovich and Lyudmila Gadasina, Analytics for an E-commerce: Demand analysis and prediction (8010)
10:35-10:55	Quang-Linh Tran, Binh T. Nguyen, Gareth J. F. Jones and Cathal Gurrin, ViEcomRec: A Novel Dataset for Recommendation Systems in Vietnamese E-commerce For Face Cleanser Products (687)
10:55-11:15	Soohyeon Choi, Rhongho Jang, Daehun Nyang and David Mohaisen, Untargeted Code Authorship Evasion with Seq2Seq Transformation (251)
11:15-11:35	Hai Thanh Nguyen, Phat Minh Nguyen, Quang Duy Tran and Phuong Ha Dang Bui, An Approach Using Threshold-Based Noise Reduction and Fine-tuned ShuffleNetV2 for Plant Leaf Disease Detection (7995)
11:35-11:55	Ling Wang, Fangjie Song, Tie Hua Zhou, Chunxu Yang and Wanlin Zhang, Mining the Potential Temporal Features based on Wearable EEG Signals for Driving State Analysis (2073)
12:00-13:30	*Lunch* *Location: 16th floor, Building A1*
13:30-16:45	Session Mon-1B **Machine learning and prediction 2** Chair: Thanh Binh Nguyen and Thien Luong Location: Room G, A1 Building
13:30-	Kieu Anh Truong Thi, Truong-Thuy Tran, Cam-Van Thi Nguyen and Duc-Trong Le,

13:50	Enhancing Visual Question Answering with Generated Image Caption (2755)
13:50-14:10	Minh Tu Le, Bao Thang Ta, Nhat Minh Le, Phi Le Nguyen and Van Hai Do, A Gaussian Distribution Labeling Method for Speech Quality Assessment (2888)
14:10-14:30	Duc Nguyen Sy, Thien Huynh Ngoc, Thang Phung Dinh, Thu Bui Minh, Phuong Thai Tieu, Long Huynh Bao, Ty Nguyen Thi, An Nguyen Thai, Huu Pham Thanh and Tho Quan Thanh, VN-Legal-KG: Vietnam Legal Knowledge Graph for Legal Statute Identification on Land Law Matters (4669)
14:30-14:50	Do Quang Manh, Tran Minh Khoi, Duong Minh Hieu and Phan Duy Hung, TextFocus: Efficient Multi-Scale Detection for Arbitrary Scene Text (4976)
14:50-15:05	*Coffee break*
15:05-15:25	T. Tin Tran and Tan Loc Nguyen, CombiGCN: An effective GCN model for Recommender System (5209)
15:25-15:45	Tuan Nguyen, Hanh Nguyen, Long Nguyen, Nhung Nguyen and Kien Doan, Unveiling Sentiments in Vietnamese Education Texts: Could Large Language Model beat PhoBERT? (5947)
15:45-16:05	Nguyen Bao Anh Le, Hien Thanh Thi Nguyen, Anh Kim Su and Hai Thanh Nguyen, Fake Face Recognition on Images Generated by Various Deepfakes Tools. (1484)
16:05-16:25	An Dang, My Linh Ha and Duc Quang Vu. Multi-scale Aggregation Network for Speech Emotion Recognition (116)
16:25-16:45	The Nguyen Manh and Hoan Bui Quoc, Portfolio Construction based on Time Series Clustering Method: Evidence in Vietnamese Stock Market (186)
16:45-17:05	Huong Hoang Luong, Lê Thị Thu Lan and Hai Thanh Nguyen, An Approach for Web Content Classification with FastText. (9811)
Conference Day 2 – December 12, 2023 (Tuesday)	
08:30-09:30	Panel discussion: **Large Language Models: theory, applications, and challenges** Panelists: **Roger Wattenhofer, My T. Thai, Xingquan Zhu, David Mohaisen, Tu-Bao Ho** Moderator: **Tien Mai** Location: Room G, A1 Building
09:30-09:45	*Coffee break*
09:45-11:55	Session Tue-1A **Optimization 1** Chairs: Trung Thanh Nguyen Location: Room G, A1 Building
09:45-10:05	Raed Alharbi, Lan N. Nguyen, and My T. Thai. Continuous Length-Bounded Paths Interdiction (2010)
10:05-10:25	Guangwen Zhou, Bin Liu and Yuanyuan Qiang, Deletion-Robust Submodular Maximization under the Cardinality Constraint over the Integer Lattice (5985)
10:25-10:45	Ba Luat Le, Layla Martin, Emrah Demir and Duc Minh Vu, A* search algorithm for an optimal investment problem in Vehicle-Sharing system (3598)
10:45-11:05	Andrey Musayev and Dmitry Grigoryev, Evolutionary Parameter Optimization: A Novel Control Strategy for Chaotic Environments (9214)

11:05-11:25	Hoang Giang Pham, The Matheuristic for Building Cost-Constrained Decision Trees with Multiple Condition Attributes (1870)
11:25-11:55	Zhiyuan Dong, Yihan Wang and Yang Zhou, Maximizing regularized DR-submodular functions on a convex set (8796)
11:50-13:30	*Lunch* *Location: 16th floor, Building A1*
13:30-14:30	Session Tue-1B **Optimization 2** Chairs: Duc Minh Vu Location: Room G, A1 Building
13:30-13:50	Canh Pham, Quat Phu and Dung Ha, Fast Bicriteria Approximation Algorithm for Minimum Cost Submodular Cover Problem (4249)
13:50-14:10	Hoang Giang Pham, Toan Tran. Mixed Integer Linear Programming-based Methods for The Optimal Time-Constrained Cost-Sensitive Decision Tree (6710)
14:10-14:30	Duc Minh Vu, Mike Hewitt and Duc D. Vu, Solving Time-Dependent Traveling Salesman Problem with Time Windows under Generic Time-Dependent Travel Cost (5159)
14:30-14:50	Tan Tran, Canh Pham, Dung Pham and Uyen Nguyen. Improved Streaming Algorithm for Minimum Cost Submodular Cover Problem (3985)
14:50-15:00	*Coffee break*
15:00-17:00	Session Tue-1C **Security and blockchain** Chairs: Hung Tran Location: Room G, A1 Building
15:00-15:20	Mohammed Alqadhi and David Mohaisen, The Infrastructure Utilization of Free Contents Websites Reveal their Security Characteristics: A Correlation Analysis (4849)
15:20-15:40	Kha Nguyen Hoang, Nguyen The Anh, Quy Lu, Khoa Tran Dang, Hien Nguyen Quang, Le Khanh Tung, Nam Tran Ba, Vinh Nguyen Thanh, Khiem Huynh, Bang Le Khanh, Thuan Tran Quoc, Minh Nguyen Van, Ngan Nguyen Thi Kim, Minh Triet Nguyen and Son Ha Xuan, Towards a management system of blood and its products based on Blockchain, smart contracts, and NFT: a case study in Vietnam (3354)
15:40-16:00	Hung Tran, Hung Pham Ngoc, Van Nhan Vo, Xuan-Truong Quach, Tung Pham Huu, Long Nguyen Quoc and Giang Quynh Le Vu. Secure Conversation: A View From Physical Layer (3638)
16:00-16:20	Ayodeji Adeniran and David Mohaisen, Understanding the Utilization of Cryptocurrency in the Metaverse and Security Implications (6100)
16:20-16:40	Khiem Huynh, Huong Hoang Luong, Minh Triet Nguyen, Quy Lu, Phuc Nguyen Trong, Son Ha Xuan, Hien Nguyen Quang, Le Khanh Tung, Nam Tran Ba, Khoa Tran Dang, Vinh Nguyen Thanh, Bang Le Khanh, Kha Nguyen Hoang, Thuan Tran Quoc, Minh Nguyen Van and Ngan Nguyen Thi Kim, Blockchain and IoT for Enhanced Traceability in Waste Treatment Processes: A Microservice and Brokerless Approach (7173)
16:40-17:00	Ngan Nguyen Bui Kim, Tram Nguyen Binh Thuc, Tuan-Dung Tran, Duy Phan The and Van-Hau Pham, MIDAS: A Multi-chain Interoperable Data and Access System for Healthcare (3782)

18:00-20:00	Gala dinner - Location: 16th floor, Building A1 Speech by representatives of NEU and My T. Thai Best paper award ceremony
Conference Day 3 – December 13, 2023 (Wednesday)	
8:00-11:35	Session Wed-1A **(Social) network analysis 1** Chairs: Mai Tien Location: Room G, A1 Building
08:00-08:20	Asep Maulana and Johannes Langguth, Leveraging GNNs and Node Entropy for Anomaly Detection: Revealing Misinformation Spreader on Twitter Network. (902)
08:20-08:40	Ali Yassin, Hocine Cherifi, Hamida Seba and Olivier Togni, Air Transportation Network Backbone Extraction: A Comparative Analysis of Structural Filtering Techniques (1101)
08:40-09:00	Andrzej Jarynowski, Alexander Semenov, Łukasz Czekaj and Vitaly Belik, Multiplex network approach for modeling the spread of African swine fever in Poland (6488)
09:00	Akshat Gaurav, Brij B. Gupta, Kwok Tai Chui and Varsha Arya, Deep Learning Based Model for Stress Measurement in Online Social Networks (7608)
09:20-09:40	Tie Hua Zhou, Jinwei Wang, Ling Wang, Haoyu Hao and Tianshuo Bi, Stress Expression Identification Model for Emotion-Driven Association Calculation over Social Network Blogs (7755)
09:40-09:55	*Coffee break*
09:55-11:35	Session Wed-1B **(Social) network analysis 2** Chairs: David Mohaisen and Duc Trong Le Location: Room G, A1 Building
09:55-10:15	Muhammad Khan, Rachel Gordon, Nimra Khan, Madeline Moran, Mohammed Abuhamad, Loretta Stalans, Jeffrey Huntsinger, Jennifer Forestal and Eric Chan-Tin, Identification and Analysis of the Spread of {Mis}information on Social Media. (8597)
10:15-10:35	Liman Du, Wenguo Yang and Suixiang Gao, Modeling the change of public opinion caused by opinion posts' dissemination in social networks (8921)
10:35-10:55	Xiang Li, Gagan Agrawal, Ruoming Jin and Rajiv Ramnath, Scalable Deep Metric Learning on Attributed Graphs (9165)
10:55-11:15	Taisia Pimenova, Valeria Kolycheva, Alexander Semenov, Dmitry Grigoryev and Arsenii Pimenov, The Impact of News, Expert and Public Opinion on Painting Prices (9894)
11:15-11:35	Alexander Semenov, Alexander Nikolaev, Eduardo Pasiliao and Vladimir Boginski, Activation and Deactivation Phenomena in Online Social Media Network Cascade Propagation with Geographic Considerations (8701)
13:30-17:00	*Local tour*

List of Papers

Multi-scale Aggregation Network for Speech Emotion Recognition.

An Dang (Phenikaa University), My Linh Ha (VNU University of Science) and Duc Quang Vu (Thai Nguyen University of Education)

Mon-1B 16:05-16:25

Speech emotion recognition (SER) is a challenging task due to its difficulty in finding efficient representations of emotion in speech. Most conventional speech feature extraction methods tend to be highly sensitive to factors that are emotionally irrelevant, such as the speaker, speaking styles, and background noise, rather than capturing the underlying emotional nuances. The most efficiently used feature extraction method for SER is deep convolutional neural networks (CNN), which can extract high-level features from low-level features from speech signals. However, the majority of CNN-based approaches primarily leverage single-scale features extracted from the final network layer, often falling short of adequately encapsulating the diverse spectrum of emotional characteristics inherent in speech. This paper introduces a multi-scale feature aggregation (MSA) network based on a fully convolutional neural network of the feature pyramid network (FPN) family for SER. This network aggregates multi-scale features from different layers of the feature extractor via a top-down pathway and lateral connections. This methodology empowers our proposed network to encompass a more comprehensive and nuanced understanding of emotional information embedded in spoken language. Additionally, in light of the challenges posed by limited data and data imbalances inherent in speech emotion recognition, we adopt data augmentation techniques to generate supplementary training data samples. Our experimental evaluation conducted on the interactive emotional dyadic motion capture (IEMOCAP) dataset demonstrates the efficacy of the proposed model, revealing its capacity to significantly enhance the performance of SER.

Keywords: SER, MSA, data augmentation

Portfolio Construction based on Time Series Clustering Method: Evidence in Vietnamese Stock Market.

The Nguyen Manh (National Economics University) and Hoan Bui Quoc (National Economics University)

Mon-1B 16:25-16:45

Portfolio optimization is indeed a crucial and highly pertinent topic in finance. During the process of constructing an investment portfolio, investors typically face with two important decisions on portfolio selection and portfolio allocation. This article aims to delve into the realm of portfolio optimization by applying the time series clustering method to historical data of stock returns on

the Vietnamese stock market. The VNIndex (VNINDEX) from January 2, 2013, to December 31, 2022 was collected for analysis. About methodology, we utilize hierarchical clustering, which can be applied to a single time series (UTS) or multiple time series (MTS). The resulting portfolio was constructed by selecting stocks from these clusters based on their Sharpe ratio, a measure that finely makes balance between the risk and return. Afterward, we determine the optimal portfolio weights using the Markowitz's Mean - Variance model, enabling us to create two distinct portfolios: MV_UTS and MV_MTS. Furthermore, our analysis encompasses dynamic evaluation using a rolling window, shedding light on how different methods respond to events impacting financial markets, such as crises. This study provides valuable insights into the application of clustering techniques in portfolio optimization and elucidates the performance and sensitivity of these methods under diverse market conditions.

Keywords: portfolio optimization, mean - variance model, hierarchical clustering, machine learning

Untargeted Code Authorship Evasion with Seq2Seq Transformation.

Soohyeon Choi (University of Central Florida), Rhongho Jang (Wayne State University), Daehun Nyang (Cyber Security Major of Software and Engineering Division, Ewha Womans University) and David Mohaisen (University of Central Florida)

Mon-1A 10:55-11:15

Code authorship attribution is the problem of identifying authors of programming language codes through the stylistic features in their codes, a topic that recently witnessed significant interest with outstanding performance. In this work, we present SCAE, a code authorship obfuscation technique that leverages a Seq2Seq code transformer called StructCoder. SCAE customizes StructCoder, a system designed initially for function-level code translation from one language to another (e.g., Java to C#), using transfer learning. SCAE improved the efficiency at a slight accuracy degradation compared to existing work. We also reduced the processing time by ≈ 68% while maintaining an 85% transformation success rate and up to 95.77% evasion success rate in the untargeted setting.

Keywords: CodeAuthorshipIdentification, Program Stylistic Features, Machine Learning, Code Authorship Evasion

Leveraging GNNs and Node Entropy for Anomaly Detection: Revealing Misinformation Spreader on Twitter Network.

Asep Maulana (Simula Research Laboratory) and Johannes Langguth (Simula Research Laboratory).

Wed-1A 08:00-08:20

The rapid growth of social media, misinformation propagation has become a critical challenge, especially on platforms like Facebook and Twitter. Detecting misinformation spreaders is vital to mitigate its harmful impact on users and society. This paper proposes an innovative approach to identify potential anomalous nodes of misinformation spreaders on Twitter networks by employing Graph Neural Networks (GNNs) and entropy-based method. Utilizing GNNs, we learn node embeddings that capture the intricate patterns of information diffusion and user attributes. Additionally, we analyze the entropy of node attributes on the embeddings to identify nodes exhibiting attribute distributions significantly deviating from the normal. Those anomalous nodes exhibit in the class of misinformation spreader will lead to detect potential of aggressive node in spreading further misinformation.

Through extensive experiments conducted on real-world Twitter datasets containing misinformation-related content, our novel approach showcases its efficacy in identifying potential anomalous nodes as misinformation spreaders across various categories. By harnessing the capabilities of Graph Neural Networks (GNNs) and integrating them with entropy-based techniques via node embeddings, our methodology offers a promising avenue for gaining deeper insights into the behavior of distinct misinformation spreaders and their potential influence on others.

Keywords: Graph neural network, Node embedding, Node entropy, Anomaly detection, Misinformation spreader, Twitter network analysis

Air Transportation Network Backbone Extraction: A Comparative Analysis of Structural Filtering Techniques.

Ali Yassin (University of Burgundy), Hocine Cherifi (University of Burgundy), Hamida Seba (University of Lyon) and Olivier Togni (University of Burgundy)

Wed-1A 08:20-08:40

In the age of advanced data collection tools, large-scale network analysis presents significant visualization and data processing challenges. Backbone-extracting techniques have emerged as crucial tools to tackle this challenge. They aim to reduce network size while preserving essential characteristics. One can distinguish two primary approaches: structural methods, which prioritize nodes and edges based on their topological properties, and statistical methods, which focus on their statistical relevance within the network data. This study investigates eight popular structural methods in an air transportation case study. Correlation analysis reveals that shortest path-based methods yield similar backbones, while Doubly Stochastic and H-backbone methods do not correlate with their alternatives.

Interestingly, H-backbone retains high-weight edges, and High Salience Skeleton and Doubly Stochastic backbones capture diverse weight scales. We evaluate the original network

information loss using the backbone's edge, node, and weight fraction. Doubly Stochastic and H-backbone methods keep substantially more edges compared to others. H-backbone, High Salience Skeleton, and Doubly Stochastic uncovered backbones fail to retain all nodes. Connectivity and transitivity comparisons indicate Primary Linkage Analysis, High Salience Skeleton methods disrupt the connectivity, and the Doubly Stochastic preserves the transitivity. This study sheds light on the strengths and weaknesses of these techniques, facilitating their application in real-world scenarios and inspiring future research directions in network analysis.

Keywords: Backbone Extraction, Filtering, Sparsification, Network Analysis, Complex Networks, Network Structure

Fake Face Recognition on Images Generated by Various Deepfakes Tools.

Nguyen Bao Anh Le (Can Tho University), Hien Thanh Thi Nguyen (Can Tho University), Anh Kim Su (Can Tho University) and Hai Thanh Nguyen (Can Tho University)

Mon-1B 15:45-16:05

In the era of rapid technological advancement, the emergence of Deepfake technology has transformed our interaction with digital content. Deepfakes, sophisticated synthetic media created using deep learning techniques, alter or replace visual and audio elements in images, videos, and audio recordings. While Deepfakes offer potential benefits in entertainment and training, they also raise ethical, social, and security concerns. Many people have been victims of deepfakes tools that are widely available on the internet. Such tools can cheat image recognition algorithms. Therefore, an assessment of the dangers of these tools is necessary so that researchers can focus on novel strategies for anti-deepfake. This study evaluates the ability of four famous deepfakes tools, namely Deepfakes, Face2Face, Face Swap, and Neural Textures, in cheating deep learning architectures in fake/real face recognition in images. Experimental results show that the Neural Textures tool is the most sophisticated in creating fake faces that are the most challenging for the considered fake image detection algorithms. In addition, we propose an architecture that can obtain better performance in fake/real face detection with fewer parameters.

Keywords: Deepfake detection, Convolutional Neural Network, deepfakes tools

The Matheuristic for Building Cost-Constrained Decision Trees with Multiple Condition Attributes.

Hoang Giang Pham (ORLab, Faculty of Computer Science, Phenikaa University)

Tue-1A 11:05-11:25

Cost factors are frequently essential in many real-world applications. Many previous studies in machine learning included costs, particularly when creating decision tree models. This research

also takes into account a cost-sensitive decision tree construction problem, with the premise that test costs must be spent in order to get the values of the decision attribute and that a record must be categorized without surpassing the expenditure cost threshold. Furthermore, our problem handles records with multiple condition attributes. A mathematical programming heuristic based on Variable Neighborhood Descent (VND) is introduced to compare to the existing approach in the literature. The experimental results show that our approach not only satisfactorily handles small and medium datasets with multiple condition attributes under different cost constraints but outperforms the existing method.

Keywords: Classification decision tree, Cost-sensitive learning, Multiple condition attributes, Variable Neighborhood Descent (VND)

Continuous Length-Bounded Paths Interdiction

Raed Alharbi (Saudi Electronic University), Lan N. Nguyen (University of Florida) and My T. Thai (University of Florida)

Tue-1A 09:45-10:05

Network vulnerability assessment, in which a communication between nodes is functional if their distance under a given metric is lower than a pre-defined threshold, has received significant attention recently. However, those works only focused on discrete domain while many practical applications require us to investigate in the continuous domain. Motivated by this observation, we study a Length-bounded Paths Interdiction in Continuous Domain (cLPI) problem: given a network G = (V, E), in which each edge e ∈ E is associated with a function fe(x) in continuous domain, and a set of target pairs of nodes, find a distribution x : E → R≥ with minimum ∑ e∈E x(e) that ensures any path p, connecting a target pair, satisfies ∑ e∈p fe(x(e)) ≥ T . We first propose a general framework to solve cLPI by designing two oracles, namely Threshold Blocking (TB) oracle and Critical Path Listing (CPL) oracle, which communicate back and forth to construct a feasible solution with theoretical performance guarantees. Based on this framework, we propose a bicriteria approximation algorithm to cLPI. This bicriteria guarantee allows us to control the solutions's trade-off between the running time and the performance accuracy.

Keywords: approximation algorithm, Network vulnerability assessment, assessment

Mining the Potential Temporal Features based on Wearable EEG Signals for Driving State Analysis

Ling Wang (School of Computer Science, Northeast Electric Power University), Fangjie Song (School of Computer Science, Northeast Electric Power University), Tie Hua Zhou (School of

Computer Science, Northeast Electric Power University), Chunxu Yang (School of Computer Science, Northeast Electric Power University) and Wanlin Zhang (School of Computer Science, Northeast Electric Power University)

Mon-1A 11:35-11:55

Fatigue driving is considered to be one of the main factors causing traffic accidents, so fatigue driving detection technology has an important role in road safety. Currently, EEG-based detection is one of the most intuitive and effective means for fatigue driving. We introduce a model known as EFDD (EEG-based Fatigue Driving Detection Model), in our study, by analyzing EEG signals, we extract time-domain and frequency-domain features respectively, explore the potential of different temporal EEG features for fatigue driving detection, Classification using LightGBM machine learning models, and then realize fatigue driving detection. Experiments demonstrate that our extracted features perform well in fatigue driving detection. Meanwhile, our study provides technical support for the feasibility of applying portable detection devices in the future.
Keywords: Fatigue Driving Detection, EEG Signal Processing, Feature Fusion, Temporal Analysis, Machine Learning

A Gaussian Distribution Labeling Method for Speech Quality Assessment.

Minh Tu Le (Viettel Cyberspace Center, Viettel Group), Bao Thang Ta (Viettel Cyberspace Center, Viettel Group), Nhat Minh Le (Viettel Cyberspace Center, Viettel Group), Phi Le Nguyen (Hanoi University of Science and Technology) and Van Hai Do (Viettel Cyberspace Center, Viettel Group).

Mon-1B 13:50-14:10

Speech Quality Assessment (SQA) plays a critical role in speech processing systems by predicting the quality of input speech signals. Traditionally, quality is represented as mean of scores provided by many listeners' opinions and is a floating-point value, so the straightforward approach is treating SQA as a regression task. However, speech quality score is labeled by crowdsourcing, resulting in a lack of uniformity among listeners. This discrepancy makes it challenging to accurately represent speech quality. Some recent works have proposed transforming SQA into a classification task by converting continuous scores into discrete classes. These approaches simplify the training of SQA models but not consider valuable variance information, which capture the differences among listeners and the difficulties to evaluate.

To address this limitation, in addition to directly learning score values through the regression task, we propose using a classification task in a multitask setting. We assume that quality labels for each class follow a Gaussian distribution, considering not only the mean ratings but also the variance among them.

To ensure that the improvements can be confidently attributed to our approach, we adopt a Conformer-based architecture, a state-of-the-art architectural framework for representing

speech information, as the core for our model and baseline models. Through a comprehensive series of experiments conducted on the NISQA speech quality assessment dataset, we provide conclusive evidence that our proposed method consistently enhances results across all dimensions of speech quality.

Keywords: Speech Quality Assessment, MOS, distribution labeling

Towards a management system of blood and its products based on Blockchain, smart contracts, and NFT: a case study in Vietnam.

Kha Nguyen Hoang (FPT University), Nguyen The Anh (FPT University), Quy Lu (Can Tho FPTU), Khoa Tran Dang (FPT University), Hien Nguyen Quang (FPT University), Le Khanh Tung (FPT University), Nam Tran Ba (FPT University), Vinh Nguyen Thanh (FPT University), Khiem Huynh (FPT), Bang Le Khanh (FPT University), Thuan Tran Quoc (FPT University), Minh Nguyen Van (FPT University), Ngan Nguyen Thi Kim (FPT Polytechnic), Minh Triet Nguyen (FPU University) and Son Ha Xuan (RMIT)

Tue-1C 15:20-15:40

In today's healthcare landscape, the integration of science and technology has greatly impacted various aspects of patient care, particularly in the management and preservation of blood and its derivatives. This is of paramount importance as there exists no viable natural or artificial substitute for blood. With a high demand for blood and its products, the primary source of supply comes from voluntary donors. However, in regions like the Mekong Delta in Vietnam, the process of blood donation and preservation presents significant challenges. To address these challenges, this paper proposes an innovative solution utilizing blockchain technology, smart contracts, and Non-Fungible Tokens (NFTs).

This work represents a groundbreaking approach that combines technical expertise, human collaboration, and organizational innovation to enhance the healthcare sector, particularly in blood management. Furthermore, it serves as a compelling case study that contributes to our understanding of the intricate relationship between collaboration and technology.

Keywords: Blood donation, Blockchain, blood products supply chain, Smart contract, NFT, Ethereum, Fantom, Polygon, Binance Smart Chain

A search algorithm for an optimal investment problem in Vehicle-Sharing system.*

Ba Luat Le (Phenikaa Univeristy), Layla Martin (Department of Industrial Engineering & Eindhoven AI Systems Institute, Eindhoven University of Technology), Emrah Demir (Cardiff Business School, Cardiff University) and Duc Minh Vu (Phenikaa University)

Tue-1A 10:25-10:45

We study an optimal investment problem that arises in the context of the vehicle-sharing system. Given a set of locations to build stations, we need to determine i) the order of stations to be built and the volume of vehicles to buy in order to get the target state where all stations are built and ii) the number of vehicles to acquire and their allocation in order to maximize the profit returned by operating the system when some or all stations are open. In this problem,
the profit (per period of time) when operating open stations is modeled as a linear optimization problem over a set of open stations. Then, with operating capital, the system owner can open new stations. This characteristic makes the transition time needed for opening a new station a set-dependent function, and the optimal investment problem can be seen as a variant of the Traveling Salesman Problem with set-dependent cost. We propose an A* search algorithm to address this particular Traveling Salesman Problem variant. The experiment results show the advantages of the proposed algorithm over the well-known classic Dijkstra algorithm and open new directions for both exact and approximate A* in future work

Keywords: Autonomous Mobility On-Demand, vehicle-sharing system, traveling salesman problem, A* algorithm

Secure Conversation: A View From Physical Layer

Hung Tran (Faculty of Computer Science, Phenikaa University), Hung Pham Ngoc (Faculty of Computer Science, Phenikaa University), Van Nhan Vo (Duy Tan University), Xuan-Truong Quach (Thai Nguyen University of Information and Communication Technology Thai Nguyen, Vietnam), Tung Pham Huu (Faculty of Information Technology Hanoi University of Civil Engineering Hanoi, Vietnam), Long Nguyen Quoc (Duy Tan University) and Giang Quynh Le Vu (Faculty of Information Technology National Academy of Education Management Hanoi, Vietnam)

Tue-1C 15:40-16:00

In this paper, we investigate the performance of a two-way secure conversation among users under the threat of an eavesdropper. To be more specific, we assume that the source and destination engage in a conversation to exchange confidential information. An eavesdropper exists, seeking to exploit the conversation illicitly. The messages of sender are split into multiple packets and then delivered to the receiver via wireless channels. To ensure communication security, we introduce a new concept called the "secure conversation constraint" to analyze the trade-off between security and system performance. Most importantly, the proposed approached can be extended to the study of other wireless networks such as cognitive radio, non-orthogonal multiple access communication, and cooperative relay networks.

Keywords: Secure conversation, Security constraint, Physical layer security, Packet transmission time

MIDAS: A Multi-chain Interoperable Data and Access System for Healthcare

Ngan Nguyen Bui Kim (VNUHCM - University of Information Technology (UIT)), Tram Nguyen Binh Thuc (VNUHCM - University of Information Technology (UIT)), Tuan-Dung Tran (VNUHCM - University of Information Technology (UIT)), Duy Phan The (VNUHCM - University of Information Technology (UIT)) and Van-Hau Pham (VNUHCM - University of Information Technology (UIT)).

Tue-1C 16:40-17:00

The emergence of Blockchain technology has inaugurated a transformative era by its multifaceted advantages and wide-ranging applications across diverse industries. Nevertheless, while holding great promise, Blockchain encounters a significant challenge in achieving interoperability within the complex landscape of multi-blockchain ecosystems. The imperative necessity for seamless data and digital asset exchange is evident, yet it is accompanied by the escalating complexities of managing network entity identities dispersed across disparate systems resulting in a stark dearth of cohesive connectivity and agile interaction. Henceforth, in response to these challenges, we present 'MIDAS' as a captivating research endeavor aimed at fostering seamless interoperability among a multitude of blockchain networks. MIDAS acts as a key intermediary, skillfully facilitating cross-chain interactions, elevating user experience, and ensuring secure data access for authorized users. In order to augmenting this architecture's prowess while also focusing on user-centric data control, we introduce a pioneering decentralized identity management system known as "Blockchain Interoperability Decentralized Identifier" (BIDI). BIDI offers a range of powerful features, including strict entity management, seamless connections, and customized data access controls for authorized users, with a focus on user-centricity and data ownership. In other words, our framework is a significant step towards improving the effectiveness and versatility of blockchain technology, breaking down barriers for harmonious coexistence among multiple blockchains.

Keywords: Blockchain interoperability, Data access control, Decentralized identifiers, Sidechain, Healthcare

Improved Streaming Algorithm for Minimum Cost Submodular Cover Problem.

Tan Tran (University of Engineering and Technology, VNU), Canh Pham (Phenikaa University), Dung Pham (Phenikaa University) and Uyen Nguyen (Institute of Engineering and Technology, Vinh University).

Tue-1B 14:30-14:50

This paper introduces an efficient streaming algorithm for a well-known problem named Minimum cost Submodular Cover ($\MSC$). Our algorithm makes $O(\log n)$ passes over the ground set and takes $O(n \log n)$ query complexity and returns a $(1/\epsilon, 1-\epsilon)$-bicriteria approximation solution with the ground set of size n and the precise parameter $\epsilon>0$. Therefore, it is better than the existing streaming algorithms in terms of solution guarantees and query complexity. Besides the theoretical performance, the experiment results on two applications, Revenue Maximization, and Maximum Cover, further show the superiority of our algorithm with the state-of-the-art ones.

Keywords: Submodular Cover, Approximation Algorithm, Streaming Algorithm

Fast Bicriteria Approximation Algorithm for Minimum Cost Submodular Cover Problem

Canh Pham (Phenikaa University), Quat Phu (VNU University of Engineering and Technology, Hanoi, Vietnam) and Dung Ha (University of Engineering and Technology)

Tue-1B 13:30-13:50

This paper studies the Minimum Cost Submodular Cover ($\MCSC$) problem over the ground set of size n, which aims at finding a subset with the minimal cost required so that the utility submodular function exceeds a given threshold. The issue has recently attracted a lot of attention due to its applications in various domains of artificial intelligence and combination optimization. However, the best approximation algorithms for the problem require an expensive query complexity of $O(n^2)$ that may become infeasible for some applications with large data.

In this work, we propose a bicriteria approximation algorithm that archives the performance guarantees of the state-of-the-art ones but reduces the required number of queries to $O(n \log n)$. Besides the theoretical, the experiment results on two applications, Twitter feed summarization, and influence threshold in social networks, further show the superiority of our algorithm with the state-of-the-art algorithm intern of both solution quality and query complexity.

Keywords: Submodular Cover, Combinatorial Optimization, Approximation Algorithm

VN-Legal-KG: Vietnam Legal Knowledge Graph for Legal Statute Identification on Land Law Matters

Duc Nguyen Sy, Thien Huynh Ngoc, Thang Phung Dinh, Thu Bui Minh, Phuong Thai Tieu, Long Huynh Bao, Ty Nguyen Thi, An Nguyen Thai, Huu Pham Thanh and Tho Quan Thanh

Mon-1B 14:10-14:30

Legal Statute Identification (LSI) is a critical task within the realm of law, involving the identification of relevant statutory laws based on the natural language descriptions found in legal documents. Traditionally, this challenge has been approached as a single-class text classification problem. However, due to the inherent complexity of legal information, characterized by intricate connections and associations between various legal entities and concepts, we propose that a graph based representation offers a more suitable and informative solution. In response to this need, our paper introduces VN-Legal-KG, an innovative Legal Statute Identification Knowledge Graph tailored to meet the specific requirements of Vietnamese users seeking clarity on Land Law matters. Leveraging cutting-edge graph neural network techniques, we also present a link prediction mechanism integrated into VN-Legal-KG, which addresses the LSI task as a multiple-label classification problem, better aligning with real-world legal practices. Through experimentation with real-world data, our approach demonstrates favorable performance when compared to previous models reported in the literature.

Keywords: Legal Stature Identification, Multi-Label Classification, Deep Learning, Knowledge Graph, Graph Neural Networks, Heterogeneous Graph, Graph Attention Network

The Infrastructure Utilization of Free Contents Websites Reveal their Security Characteristics: A Correlation Analysis

Mohammed Alqadhi (University of Central Florida) and David Mohaisen (University of Central Florida)

Tue-1C 15:00-15:20

Free Content Websites (FCWs) are a significant element of the Web, and realizing their use is essential. This study analyzes FCWs worldwide by studying how they correlate with different network sizes, cloud service providers, and countries, depending on the type of content they offer. Additionally, we compare these findings with those of premium content websites (PCWs). Our analysis concluded that FCWs correlate mainly with networks of medium size, which are associated with a higher concentration of malicious websites. Moreover, we found a strong correlation between PCWs, cloud, and country hosting patterns. At the same time, some correlations were also observed concerning FCWs but with distinct patterns contrasting each other for both types. Our investigation contributes to comprehending the FCW ecosystem through correlation analysis, and the indicative results point toward controlling the potential risks caused by these sites through adequate segregation and filtering due to their concentration in the network, cloud, or country.

Keywords: Websites Security, Correlation Analysis, Network analysis, Country-level analysis, Free Content, Premium Content

TextFocus: Efficient Multi-Scale Detection for Arbitrary Scene Text

Do Quang Manh (FPT University), Tran Minh Khoi (FPT University), Duong Minh Hieu (FPT University) and Phan Duy Hung (FPT University)

Mon-1B 14:30-14:50

With the advancement of deep learning, there are more and more datasets related to scene text detection that are high-resolution and have small texts in them and that gradually become common tasks. When the normal solution is to scale the image smaller, the texts inside the images will become more abstract and unrecognizable, as well as make the performance worse. In this research, we present TextFocus, an algorithm exploiting multi-scale training strategy but in a more efficient way. Instead of analyzing each pixel in an image pyramid, TextFocus will attempt to identify context regions surrounding ground-truth instances, or "chips," and then process for finding all text regions in the image sample. All text information from every chip from the model

will then be combined with careful post processing methodology to obtain the final results for text detection. As a result of TextFocus' ability to resample very large image samples (4000x4000 pixels) into low resolution chips (640x640 pixels), our model can train twice as quickly and handle batches as large as 50 on a single GPU when scaled normally. When the larger the training size, the better the result is basic tactic, our method demonstrates that training on high resolution scale might not be ideal. Our implementation using ResNet-18 backbone with segment-like head achieves 0.828 F1 score on the SCUT-CTW1500 [1] dataset, 0.611 F1 score on the Large CTW [2] dataset with acceptable FPS for realtime purpose.

Keywords: Computer Vision, Scene Text Detection, Multi-scale Detection, High Resolution Image

Solving Time-Dependent Traveling Salesman Problem with Time Windows under Generic Time-Dependent Travel Cost

Duc Minh Vu (Phenikaa Univeristy), Mike Hewitt (Loyola University Chicago) and Duc D. Vu (University of Michigan - Flint)

Tue-1B 14:10-14:30

In this paper, we present formulations and an exact method to solve the Time Dependent Traveling Salesman Problem with Time Window (TD-TSPTW) under a generic travel cost function where waiting is allowed. A particular case in which the travel cost is a non-decreasing function has been addressed recently. With that assumption, because of both First-In-First-Out property of the travel time function and the non-decreasing property of the travel cost function, we can ignore the possibility of waiting. However, for generic travel cost functions, waiting after visiting some locations can be part of optimal solutions. To handle the general case, we introduce new lower-bound formulations that allow us to ensure the existence of optimal solutions. We adapt the existing algorithm for TD-TSPTW with non-decreasing travel costs to solve the TD-TSPTW with generic travel costs. In the experiment, we evaluate the strength of the proposed lower bound formulations and algorithm by applying them to solve the TD-TSPTW with the total travel time objective. The results indicate that the proposed algorithm is competitive with and even outperforms the state-of-art solver in various benchmark instances

Keywords: time-dependent travel time, time-dependent travel cost, traveling salesman problem, dynamic discretization discovery

CombiGCN: An effective GCN model for Recommender System

T. Tin Tran (Ton Duc Thang University) and Tan Loc Nguyen (Ton Duc Thang University)

Mon-1B 15:05-15:25

Graph Neural Networks (GNNs) have opened up a potential line of research for collaborative filtering (CF), the key power of GNNs based on injecting collaborative signal into user and item embeddings which will contain information about user-item interactions after that. However, there are still some unsatisfactory points for a CF model that GNNs could have done better. The way in which the collaborative signals are extracted through an implicit feedback matrix that is

essentially built on top of the message-passing architecture of GNNs and it only helps to update the embedding based on the value of items (or users) embeddings neighboring. By identifying the similarity weight of users through their interaction history, a key concept of CF, we endeavor to build a user-user weighted connection graph based on their similarity weight of them.

In this study, we propose a recommendation framework CombiGCN, item embeddings only linearly propagated on the user-item interaction graph, while user embeddings are propagated simultaneously on both user-user weighted connection graph and user-item interaction graph graphs with Light Graph Convolution (LGC) and combined in an simpler method by using the weighted sum of the embeddings for each layer. Experiments comparing CombiGCN with several state-of-the-art models on three real-world datasets show that CombiGCN improves the recommendation results by at least about 3% on the Ciao dataset and up to about 15% on the Epinions dataset, when considering recall@20 metric.

Keywords: Recommender System, Collaborative Filtering, Collaborative signal, Graph Convolution Network, Embedding Propagation

Unveiling Sentiments in Vietnamese Education Texts: Could Large Language Model beat PhoBERT?

Tuan Nguyen (National Economics University), Hanh Nguyen (National Economics University), Long Nguyen (National Economics University), Nhung Nguyen (National Economics University) and Kien Doan (National Economics University)

Mon-1B 15:25-15:45

Text classification in general, and sentiment analysis specifically is a major branch of Natural Language Processing with various uses and challenges. Classifying sentiments from Vietnamese text presents numerous challenges as Vietnamese is inherently and linguistically dissimilar to English. Therefore, there urges measures to recognize subtleties in Vietnamese texts and accurately categorize them into suitable labels. In this work, we leverage the power of the pre-trained model PhoBERT to classify a collected dataset of education-related texts into three separate classes, namely positive, negative, and neutral. It is expected that the model will have positive social implications by bringing people updated information on the current education system and ensuring their comprehension of educational policies, thus accordingly adapting itself towards more progressive growth. The model is first trained on two datasets VLSP 2019 and ViHSD. We then use the fine-tuned model to train on our data gathered from Vietnam online news sites and social media platforms. The input is either a sentence or a paragraph and the expected output is among the three aforementioned labels. Through testing and fine-tuning different models including Large Language Models (LLMs) with one-shot and few-shot approaches, our

model consistently outperformed them in sentiment classification tasks with an accuracy of up to 94%

Keywords: Classification, Sentiment analysis, Vietnamese text, Pho-BERT, Large Language Model

Deletion-Robust Submodular Maximization under the Cardinality Constraint over the Integer Lattice

Guangwen Zhou (Ocean University of China), Bin Liu (Ocean University of China) and Yuanyuan Qiang (Ocean University of China)

Tue-1A 10:05-10:25

Submodular optimization is a classical problem of combinatorial optimization. The objective functions of many combinatorial optimization problems are submodular functions and they also have significant applications in real life. Since some practical problems such as budget allocations that are hard to be modeled over set functions, submodular functions over the integer lattice have been widely and intensively studied subject to various classical constraints for decades. In this paper we study the robustness of maximizing a monotone diminishing return submodular function over the integer lattice under the cardinality constraint. We propose a robustness model over the integer lattice and design algorithms under this specific model by utilizing stochastic strategy combined with binary search approach. The algorithms we designed in centralized settings can achieve a $(1/2-\delta)$-approximation and maintain robustness against deleting any d elements adversarially. While in streaming settings the algorithms we designed can still achieve the same approximation and be robust against deleting any d elements adversarially as well.

Keywords: Submodular maximization, Cardinality constraint, Robustness

Understanding the Utilization of Cryptocurrency in the Metaverse and Security Implications

Ayodeji Adeniran (University of Central Florida) and David Mohaisen (University of Central Florida)

Tue-1C 16:00-16:20

We present our results on analyzing and understanding the behavior and security of various metaverse platforms incorporating cryptocurrencies. We obtained the top metaverse coins with a capitalization of at least 25 million US dollars and the top metaverse domains for the coins, and augmented our data with name registration information (via whois), including the hosting DNS IP addresses, registrant location, registrar URL, DNS service provider, expiry date and check each metaverse website for information on fiat currency for cryptocur-rency. The result from virustotal.com includes the communication files, passive DNS, referrer files, and malicious detections for each metaverse domain. Among other insights, we discovered various incidents of malicious detection associated with metaverse websites. Our analysis highlights indicators of (in)security, in the correlation sense, with the files and other attributes that are potentially responsible for the malicious activities.

Keywords: Metaverse, security, cryptocurrencies, data analysis

Multiplex network approach for modeling the spread of African swine fever in Poland

Andrzej Jarynowski (Freie Universität Berlin), Alexander Semenov (University of Florida), Łukasz Czekaj (Aidmed) and Vitaly Belik (Freie Universität Berlin)

Wed-1A 08:40-09:00

African Swine Fever (ASF) is a viral infection which causes acute disease in Sus scrofa – domestic pigs and wild boar. Although the virus does not cause disease in humans, the impact it has on the economy, especially via trade and farming disturbance, is substantial. We analyze 3487 ASF notifications of wild boars and pigs in Poland (infection events registered to OiE) from February 2014 to April 2019 comprising event time, longitude, latitude and administrative unit: county/poviat. We propose a spatial modeling approach incorporating phenomenological analysis of multiplex transmission networks: 1) domestic pig abundance, 2) disease vectors (wild boar) abundance, 3) human mobility related to disease propagation. We use a pseudo gravity model to simulate the future epidemic projection and calculated the most probable infection paths for all counties (poviats) as well as estimated the most likely disease arrival times with or without countermeasures such as border fencing and animals corridors blocking on the A1 motorway. According to our model, the ASF spread in Poland had been continuing and investigated jump to wschowski poviat 320km from the c losest previously affected area manifests its complex behavior. The proposed complex network approach promise to be useful to practitioners, farmers and veterinarians, helping them to choose the optimal mitigation strategies.

Keywords: African Swine Fever, multiplex networks, spreading processes, epidemiological modelling

Mixed Integer Linear Programming-based Methods for The Optimal Time-Constrained Cost-Sensitive Decision Tree

Hoang Giang Pham, Toan Tran (ORLab, Faculty of Computer Science, Phenikaa University)

Tue-1B 13:50-14:10

Cost management and completion timelines are common classification problem applications. For instance, a blood test requires human and equipment resources, costs money, and must be done quickly. To avoid major patient risks, medical diagnosis should minimize positive-to-negative misclassification. There has been extensive research on how to build a decision tree from training data that minimizes misclassification and test costs, but not on how to design an optimal decision tree with time constraints. Mixed Integer Programming creates the Optimal Time-Constrained Cost-Sensitive Decision Tree in this study. We adapt several existing approaches in the literature with time constraints, misclassification cost, and feature selection cost, and introduce new

mathematical models designed for these problems. Our experiments compare all methods and examine how time and cost constraints affect optimal solution identification and demonstration.

Keywords: Optimal Decision Tree, Time-Constrained, Cost-Sensitive, Mixed Integer Programming

The Role of Social Media Knowledge Sharing, Leadership, and Exploitation-Exploration in High-Performance Work System

Idria Maita, Saide Saide, Nita Antasari and Kurnia Prayogi

Sat-2B 15:00 – 15:25

Research on knowledge sharing, participation leadership, and exploitation-exploration has shown that it has an influence on the high-Performance work system (HPWS) or the quality of employee performance. In employee ambidexterity, there are two innovations, namely exploitation and exploration, but not much research has been done that directly focuses on these two parts of employee ambidexterity. Data from target employees was collected using an offline questionnaire- based survey with 56 valid responses from the sample of employees taken. 56 representative data from employees working in the top two universities in the Riau region. This study aims to determine how much and important the influence of supporting variables on employee ambidexterity. By using structural equation modeling, the benefit of this research is that it will find out that knowledge sharing, participatory leadership, and exploitation-exploration are important to improve the high-performance work system (HPWS) or the quality of employee performance, so that companies can improve or implement new things that can improve employee performance. By using structural equation modeling, the research findings show that participative leadership has a positive and significant effect on high-performance work system (HPWS) or the quality of employee performance. Furthermore, exploration also has a positive and significant effect on the high-performance work system (HPWS) or the quality of employee performance. So, from the results of this study, there are strategies that can be used to improve the competitive quality of employee performance, namely with leadership participation, sharing knowledge between one another, and innovating exploration.

Keywords: Ambidexterity, social media, exploitation, exploration, participative leadership, high-performance work system, knowledge sharing

Blockchain and IoT for Enhanced Traceability in Waste Treatment Processes: A Microservice and Brokerless Approach

Khiem Huynh, Huong Hoang Luong, Minh Triet Nguyen, Quy Lu, Phuc Nguyen Trong, Son Ha Xuan, Hien Nguyen Quang, Le Khanh Tung, Nam Tran Ba, Khoa Tran Dang, Vinh Nguyen Thanh,

Bang Le Khanh, Kha Nguyen Hoang, Thuan Tran Quoc, Minh Nguyen Van and Ngan Nguyen Thi Kim

Tue-1C 16:20-16:40

The management of medical waste, especially in developing countries, presents substantial challenges that have been further intensified by the Covid-19 pandemic. Traditional waste treatment methods often bypass the crucial step of segregating waste at the source, which results in environmental pollution and health hazards. To tackle these is- sues, we propose a model that leverages Hyperledger Fabric to establish a transparent, secure, and unalterable record of waste treatment processes, thereby ensuring data integrity, transparency, and traceability. The in- corporation of Internet of Things (IoT) devices facilitates real-time data collection and control, thereby further enhancing the traceability of the medical waste treatment process. Our model adopts a microservice architecture, which provides several advantages over traditional approaches, including improved scalability, simplified maintenance and updates, and the flexibility to use different technologies for different services. This re- search makes a five-fold contribution to the field, namely i) presenting a robust and scalable solution based on the blockchain platform; ii) implementing a proof-of-concept via Hyperledger Fabric; iii) providing an integration of IoT devices for real-time data collection and control; iv) proposing the microservice architecture for medical waste treatment; and v) evaluating the effectiveness of the proposed model via the Round Trip Time and Hyperledger Caliper (i.e., data creation & query) test cases.

Keywords: Waste treatment, Blockchain, IoTs, Microservice, Hyperledger Fabric, Hyperledger Caliper

Activation/Deactivation Phenomena in Online Social Media Network Cascade Propagation with Geographic Considerations

Alexander Semenov (University of Florida), Alexander Nikolaev (University at Buffalo), Eduardo Pasiliao (Air Force Research Laboratory) and Vladimir Boginski (University of Central Florida)

Wed-1A 11:15-11:35

Activation and influence propagation cascades are pervasive in the modern planetary-scale online social media networks, where users influence each other's opinions, beliefs, etc. The most commonly used influence propagation models are based on activation thresholds, including linear threshold models, where a node in a network would "activate" if a certain minimum fraction of its neighbors are active. As an extension of such models, we consider the model of activation/influence propagation in networks based on the concept of "bandpass" thresholds: a node will "activate" if at least a certain minimum fraction of its neighbors are active and no more than a certain maximum fraction of neighbors are active. Such models are more flexible than

standard "single" (linear) threshold models as they allow to incorporate more complex dynamics of diffusion processes where nodes can activate and deactivate, which have natural interpretations in the context of social networks. The respective hypothesis was originally mentioned in the seminal work on threshold models by Granovetter (1978); however, to our knowledge there have not been any rigorous mathematical/computational studies of double threshold models prior to our current and recent work. In addition, the considered activation/deactivation processes can be affected by geographic proximity factors (i.e., social media users being physically closely located). We present the results of real-world cascade traces of single- and bandpass-threshold cascades in city-wide and country-wide online social media networks based on anonymized user friendships data collected from VK.com in December 2016. We investigate the penetration of the considered cascading processes under various initial conditions on threshold values and seed node choice strategies, both in terms of the percentage of activated nodes in a network, and in terms of the geographic reach of a cascade.

Keywords: Activation cascades, Social media networks, Bandpass threshold models, Geographic considerations

Deep Learning Based Model for Stress Measurement in Online Social Networks

Akshat Gaurav (Ronin Institute, Montclair, New Jersey 07043, U.S.), Brij B. Gupta (Department of Computer Science and Information Engineering, Asia University, Taiwan), Kwok Tai Chui (School of Science and Technology, Hong Kong Metropolitan University, Hong Kong) and Varsha Arya (Asia University, Taiwan).

Wed-1A 09:00–09:20

Online Social Networks (OSNs) have become ubiquitous platforms for individuals to express their thoughts and emotions, making them valuable sources for studying mental health. This paper presents a novel Deep Learning-based approach for stress measurement in OSNs. We leverage a comprehensive dataset collected from Kaggle, specifically curated for stress analysis in social media. The proposed model demonstrates remarkable accuracy in identifying stress levels, paving the way for proactive mental health interventions and more targeted support systems in the digital age. This research contributes to the growing body of knowledge addressing mental health challenges in the online world, emphasizing the potential of AI and deep learning techniques in this critical domain.

Keywords: Stress Detection, Online Social Networks, Deep Learning, Reddit, Mental Health

Stress Expression Identification Model for Emotion-Driven Association Calculation over Social Network Blogs

Tie Hua Zhou, Jinwei Wang, Ling Wang, Haoyu Hao and Tianshuo Bi

Wed-1A 09:20-09:40

Emotion and stress expression are deeply intertwined. Analyzing blog data from social network platforms reveals that specific emotions often appear simultaneously with expressions indicating stress, and emotional expressions in different stress scenarios also display unique patterns. This paper introduces the Stress Expression Identification (SEI) model based on Emotion Sequence Pattern Analysis (ESPA), which can identify stress and stress level accurately by by exploring the corresponding rules between emotion and stress. We constructed a stress emotion dictionary, mined the pattern rules of emotion sequence under different stress scenarios, and established a stress emotion rule set, aiming at identifying and understanding stress more accurately. The experimental results show that this model achieved a higher accuracy rate in stress identification and can identify three levels of stress.
Keywords: Stress expression, Emotion sequence, Association rules, NLP, Data mining

An Approach Using Threshold-Based Noise Reduction and Fine-tuned ShuffleNetV2 for Plant Leaf Disease Detection

Hai Thanh Nguyen (Can Tho University), Phat Minh Nguyen (Can Tho University), Quang Duy Tran (Can Tho University) and Phuong Ha Dang Bui (Can Tho University)

Mon-1A 11:15-11:35

Diagnosing plant leaf diseases is essential for agricultural development. Leaves are an important part of the plant and are often where signs of disease appear. With the support of image processing algorithms, researchers have widely used them to support disease detection tasks on plant leaves. Transfer learning approaches have revealed encouraging results in many domains but require fine-tuning hyperparameter values. Additionally, a combination with noise reduction can lead to positive potential effects in improving performance. This study presents an approach leveraging a noise reduction technique based on Soft-Thresholding with Lasso regression and then performing the disease classification with a fine-tuned ShuffleNetV2. The experimental results on 14,400 images of 24 leaf types of 10 various plant species show that the Threshold-based noise reduction combined with a fine-tuned ShuffleNetV2 can obtain better performance in disease classification on plant leaves than the original model and several considered transfer learning methods.
Keywords: plant leaf diseases, threshold, noise reduction, ShuffleNetV2

Analytics for an E-commerce: Demand analysis and prediction

Alexandr Karpovich (Saint-Petersburg State University) and Lyudmila Gadasina (Saint-Petersburg State University)

Mon-1A 10:15-10:35

The problem of balancing supply and demand in different market segments is one of the fundamental ones in the online retailer. The research considers the problem of predicting the demand for products of the kitchen knives category from the kitchen accessories segment in the WildBerries online store. To create a sales prediction model, information about sales for 13 months and product characteristics are collected. The LightGBM method showed the best values

of qual-ity metrics. To improve the quality of the model, an exploratory analysis was carried out. Based on its results, additional features for the model are designed. This made it possible to improve all quality metrics. The estimation of SHAP values allows making management decisions to regulate demand and achieve business objectives. The analysis of the demand elasticity carried out in the study makes it possible to trace the effect of price changes on the demand for products.
Keywords: E-commerce, online retailer, predictive model, boosting, LightGBM, elasticity of demand

Identification Model for Emotion-Driven Association Calculation over Social Network Blogs

Muhammad Khan, Rachel Gordon, Nimra Khan, Madeline Moran, Mohammed Abuhamad, Loretta Stalans, Jeffrey Huntsinger, Jennifer Forestal and Eric Chan-Tin

Wed-1A 09:55-10:15

With unfolding crises such as the COVID-19 pandemic, it is essential that factual information is dispersed at a rapid pace. One of the major setbacks to mitigating the effects of such crises is misinformation. Advancing technologies such as transformer-based architectures that can pick up underlying patterns and correlational information that constitutes information provide tools that can be used to identify what is misinformation/information. To identify and analyze the spread of misinformation, this work performs a quantitative analysis that uses X (previously Twitter) as the data source and a BERT-based model to identify misinformation. The information of the posts, users, and followers was collected based on hashtags and then processed and manually labeled. Furthermore, we tracked the spread of misinformation related to COVID-19 during the year 2021 and determined how communities that spread information and/or misinformation on social networks interact from an analytical perspective. Our findings suggest that users tend to post more misinformation than information, possibly intentionally spreading misinformation. Our model showed good performance in classifying tweets as information/misinformation, resulting in an accuracy of 86%.
Keywords: BERT, Spread of Misinformation, Social Networks

Activation and Deactivation Phenomena in Online Social Media Network Cascade Propagation with Geographic Considerations

Alexander Semenov (University of Florida), Alexander Nikolaev (University at Buffalo), Eduardo Pasiliao (Air Force Research Laboratory) and Vladimir Boginski (University of Central Florida)

Wed-1A 11:15-11:35

Activation and influence propagation cascades are pervasive in the modern planetary-scale online social media networks, where users influence each other's opinions, beliefs, etc. The most commonly used influence propagation models are based on activation thresholds, including linear threshold models, where a node in a network would "activate" if a certain minimum fraction of its neighbors are active. As an extension of such models, we consider the model of activation/influence propagation in networks based on the concept of "bandpass" thresholds: a node will "activate" if at least a certain minimum fraction of its neighbors are active and no more

than a certain maximum fraction of neighbors are active. Such models are more flexible than standard "single" (linear) threshold models as they allow to incorporate more complex dynamics of diffusion processes where nodes can activate and deactivate, which have natural interpretations in the context of social networks. The respective hypothesis was originally mentioned in the seminal work on threshold models by Granovetter (1978); however, to our knowledge there have not been any rigorous mathematical/computational studies of double threshold models prior to our current and recent work. In addition, the considered activation/deactivation processes can be affected by geographic proximity factors (i.e., social media users being physically closely located). We present the results of real-world cascade traces of single- and bandpass-threshold cascades in city-wide and country-wide online social media networks based on anonymized user friendships data collected from VK.com in December 2016. We investigate the penetration of the considered cascading processes under various initial conditions on threshold values and seed node choice strategies, both in terms of the percentage of activated nodes in a network, and in terms of the geographic reach of a cascade.

Keywords: Activation cascades, Social media networks, Bandpass threshold models, Geographic considerations

Maximizing regularized DR-submodular functions on a convex set

Zhiyuan Dong (School of Mathematics and Statistics, Shandong Normal University), Yihan Wang (School of Mathematics and Statistics, Shandong Normal University) and Yang Zhou (School of Mathematics and Statistics, Shandong Normal University)

Tue-1A 11:25-11:55

In this paper we concentrate on maximizing a regularized DR-submodular function $F: [0,1]^n \to \mathbb R$, which is composed of DR-submodular function G and linear function L, on a convex set. To solve the problem of maximizing the regularized monotone DR-submodular function, we propose an approximation algorithm with an approximation ratio of $(1-\frac{1}{e},1)$. When the DR-submodular function is non-monotonic, we propose two approximation algorithms of $(\frac{1}{4},1)$ and $(\frac{1}{4},\log2)$ respectively for the non-negative and non-positive linear function L, which have good approximation guarantee in continuous-time and discrete-time. At the end, we also propose an approximation algorithm of $(\frac{1}{4},\frac{1}{2})$ for maximizing the regularized non-monotone DR-submodular function in the case of unconstrained linear part L.

Keywords: DR-submodular, regularized problem, convex, approximation algorithm

Modeling the change of public opinion caused by opinion posts' dissemination in social networks

Liman Du (University of Chinese Academy of Science), Wenguo Yang (University of Chinese Academy of Sciences) and Suixiang Gao (University of Chinese Academy of Sciences)

Wed-1A 10:15-10:35

The various extensively used social platforms make it possible for the public to facilitate quick and easy sharing of information as well as express their views and opinions in real time, which can lead to the rapid outbreak of online public opinion. To simulate the change of online public

opinion resulting from opinion posts' dissemination in social networks, this study propose a more realistic and novel model named as Agent-Based Opinion Dissemination model. It provides the additional advantage to incorporate the effect of real-world characteristics from the perspective of outer interventions and individuals' attributes. The experimental results show that our model can simulate the evolution of public opinion, thus providing a clearer explanation of the law and causes of public opinion's evolution from the perspective of individual and media behavior.
Keywords: Social Networks, Online Public Opinion, Information Dissemination, Agent-Based

Scalable Deep Metric Learning on Attributed Graphs

Xiang Li (The Ohio State University), Gagan Agrawal (University of Georgia), Ruoming Jin (Kent State University) and Rajiv Ramnath (The Ohio State University)

Wed-1A 10:35-10:55

We consider the problem of constructing embeddings of large attributed graphs and supporting multiple downstream learning tasks. We develop a graph embedding method, which is based on extending deep metric and unbiased contrastive learning techniques to 1) work with attributed graphs, 2) enabling a mini-batch based approach, and 3) achieving scalability. Based on a multi-class tuplet loss function, we present two algorithms – DMT for semi-supervised learning and DMAT-i for the unsupervised case. Analyzing our methods, we provide a generalization bound for the downstream node classification task and for the first time relate tuplet loss to contrastive learning. Through extensive experiments, we show high scalability of representation construction, and in applying the method for three downstream tasks (node clustering, node classification, and link prediction) better consistency over any single existing method.
Keywords: Attributed Graph, Deep Metric Learning, Graph Embedding, Graph Convolutional Network, Scalability

Evolutionary Parameter Optimization: A Novel Control Strategy for Chaotic Environments

Andrey Musayev (ITMO University) and Dmitry Grigoryev (SPbU)

Tue-1A 10:45-11:05

The efficient management of dynamic systems interacting with unstable immersions is crucial across various fields. These include the stabilization of turbulent flows, signal generation in radio engineering, and asset management optimization in capital markets. The primary challenge stems from the inherent unpredictability of deterministic chaos models, which introduces additional uncertainty. Numerical methods are the only feasible approach to evaluate the effectiveness of control strategies. This study focuses on developing empirical algorithms to identify and forecast local trends, aiming ultimately to establish extrapolation prediction techniques. The research specifically targets speculative trading within currency markets, characterized by stochastic chaos. Unlike physical and technical problems, currency markets are purely informational and lack inertia. As a result, traditional prediction algorithms that rely on reactive control strategies have

proven ineffective. This study seeks to address this efficiency gap by exploring control strategies that sequentially optimize evolutionary parameters while approximating the model structure of observation series.

Keywords: Stationarity, Evolutionary optimization, Asset allocation, Process dynamics prediction

An Approach for Web Content Classification with FastText

Huong Hoang Luong (FPT University), Lê Thị Thu Lan (Can Tho University) and Hai Thanh Nguyen (Can Tho University)

Mon-1B 16:45-17:05

Nowadays, with the Internet infrastructure and nearly global access, the amount and diversity of data are increasing rapidly. Many tasks require information retrieval and data collection for machine learning, research, and survey reports in various fields such as meteorology, science, geography, literature, and more. However, manual data collection and classification can be time-consuming and prone to errors. Additionally, AI assistants used for drafting or writing can sometimes be corrected regarding writing style and inappropriate language for the given context. Faced with these needs, In this article, Vietnamese documents are classified using the TF-IDF method, TF-IDF combined with SVD, and FastText at three levels: word level, n-gram level, and character level. For this approach, 15 categories were gathered from various online news sources. The dataset was preprocessed and trained using machine learning models such as SVM, Naive Bayes, Neural Network, and Random Forest to find the most effective method. The Random Forest combined with the FastText method was highly evaluated, achieving a success rate of 82% when measured against essential evaluation criteria of accuracy, precision, and F1 score.

Keywords: Content classification, FastText, TF-IDF

The Impact of News, Expert and Public Opinion on Painting Prices

Taisia Pimenova (Center of Econometrics and Business Analytics, SPbU), Valeria Kolycheva (St. Petersburg State University (Faculty of Economics & CEBA)), Alexander Semenov (University of Florida), Dmitry Grigoryev (SPbU) and Arsenii Pimenov (St. Petersburg State University)

Wed-1A 10:55-11:15

Online reviews play a crucial role in determining the price of goods, and the art market is no exception. However, due to the varying levels of expertise and influence among participants, it is essential to analyze each group separately. In this work, we explore the influence of experts and public sentiment on painting prices.

To investigate this, we employ a hedonic regression model with fixed effects for artists. Our dataset comprises 18,100 sold paintings at Sotheby's, Christie's and Phillips auction houses. Additionally, we examine the relationship between review sentiment and buyers' investment intentions, as purchasing paintings as an investment is a widespread practice in the secondary art market. Our results indicate that negative opinion from various sources and positive public opinion significantly influence the price of paintings in line with their respective valence. Furthermore, we observe that negative opinion from experts and news publications has a more

pronounced effect on pricing. Our findings contribute to empirical research on the price determinants of artworks and examining online reviews from different sources.
Keywords: art auctions, hedonic model, experts, online reviews

Table of Contents

CSoNet 2023

The International Conference on Computational Data and Social Networks (CSoNet) 2023