GENERAL PROGRAM - SIMBig 2022

20 min presentation

November 16, 2022

Hour	Title	Speaker
08h30 - 09h00	***Reception***	
09h00 - 09h20	**Welcoming to SIMBig2022**	Juan Antonio Lossio-Ventura Hugo Alatrista
Session 1		
09h20 - 10h20	**Title**: Taming Neural Language Models into Trustworthy Conversational Virtual Assistants **Abstract**: What if computers can truly converse with us in our native tongue? Computers will transform into effective, personalized assistants for everybody. Commercial chatbots today are notoriously brittle as they are hardcoded to handle a few possible choices of user inputs. Recently introduced large language neural models, such as GPT-3, are remarkably fluent, but they are prone to hallucinations, often producing incorrect statements. This talk describes how we can tame these neural models into robust, trustworthy, and cost-effective conversational agents.	**Dr. Monica Lam** *Stanford University USA* Stanford University
10h20 - 10h40	A Preliminary Analysis of Twitter's LGBTQ+ Discussions	Abu Naweem Khan, Rahat Ibn Rafiq
10h40 - 11h00	User-agnostic Model for Prediction of Retweets Based on Social Neighborhood Information	Pablo Gabriel Celayes, Martín Ariel Domínguez and Damián Barsotti
11h00 - 11h20	***Coffee Break***	
11h20 - 11h40	Segmentation and classification of pages for digitized documents of the Public Prosecutor's Office	Kevin Rivera Vergaray, Diana Mary Quintanilla Perez and Angel Jimy Espezua Chalco
11h40 - 12h40	**Title:** Unsupervised Machine Learning for Explainable Medicare Fraud Detection **Abstract:** A major concern in the US federal health care system is overbilling, waste and fraud by providers, who face incentives to misreport on their claims to receive higher reimbursements. In this talk, focusing on Medicare (the health insurance program for elderly adults and the disabled), I will present data-driven techniques to identify providers with spending patterns consistent with overbilling. Our approach is (i) fully unsupervised, avoiding laborious human labeling and (ii) explainable to end users, guiding the auditing process. Data from the Department of Justice on providers facing anti-fraud lawsuits validate our approach.	**Dr. Leman Akoglu** *Caneguie Mellon University, USA* Carnegie Mellon University
12h40 - 13h00	A Semantic Query Engine for Knowledge Rich Legal Digital Libraries	Hasan Jamil
13h00 - 14h20	***Lunch***	
Session 2		
14h20 - 14h40	Using features based on elongation to enhance sentiment analysis	Abderrahim Rafae, Mohammed Erritali, Younes Madani and Mathieu Roche
14h40 - 15h40	**Title:** "What if" AI with real-world data from prediction to intervention modeling to fairness and disparities **Abstract:** The rapid adoption of electronic health record (EHR) systems has made large collections of real-world data (RWD) that reflect the characteristics and outcomes of the patients being treated in real-world settings, available for research. The increasing availability of RWD combined with the advancements in artificial intelligence (AI), especially machine learning (ML) and deep learning (DL) offer untapped opportunities to generate real-world evidence (RWE) to answer a wide range of biomedical and clinical questions. This talk will focus on a number of applications of "AI" on real-world data and witness how we are moving away from simply building traditional prediction models to more careful study design to consider models that are "interventional." Furthermore, we will discuss how real-world data and AI can bring in both opportunities for addressing health disparities and improve health equity, e.g., through social determinants of health, but also challenges to be fair, due to a number of inherent limitations such as structural missingness and data quality issues.	**Dr. Jiang Bian** *University of Florida, USA* UF UNIVERSITY of FLORIDA
15h40 - 16h00	Image Generation from sketches and text-guided attribute edition	Dennis Sumiri and José Ochoa
16h00 - 16h20	***Coffee Break***	
16h20 - 16h40	Getting Quechua Closer to Final Users through Knowledge Graphs	Elwin Huaman, Jorge Luis Huaman and Wendi Huaman
16h40 - 17h00	ADRAS: Airborne Disease Risk Assessment System for Closed Environments	Wilber Rojas, Edwin Salcedo and Guillermo Sahonero
17h00 - 17h20	Predictive sentiment analysis model regarding the variation of the dollar exchange rate	Joaquin Rodolfo Bernabe Polo

November 17, 2022

Hour	Title	Speaker
08h30 - 09h00	***Reception***	
Session 3		
09h00 - 09h20	Maximising Influence Spread in Complex Networks by Utilising Community-based Driver Nodes as Seeds	Abida Sadaf, Luke Mathieson Piotr Brodka, and Katarzyna Musial
09h20 - 10h20	**Title:** A/B testing for high-quality alternatives **Abstract:** A/B testing is widely used to tune search and recommendation algorithms, to compare product variants as efficiently and effectively as possible, and even to study animal behavior. With ongoing investment, due to diminishing returns, the items produced by the new candidate show smaller and smaller improvement in quality from the items produced by the current system. By formalizing this observation, we develop the first closed-form analytical expressions for the sample efficiency of a number of widely-used families of slate-based comparison tests. In empirical trials, these theoretical sample complexity results are shown to be predictive of real-world testing efficiency outcomes. These findings offer opportunities for both more cost-effective testing and a better analytical understanding of the problem.	**Dr. Andrew Tomkins** *Google, USA* 
10h20 - 10h40	Peak Anomaly Detection from Environmental Sensor-Generated Watershed Time Series Data	Byung Suk Lee, John Clay Kaufmann, James B. Shanley, Donna M. Rizzo and Ijaz Ul Haq
10h40 - 11h00	Gas Sensors and Machine Learning for Quality Evaluation of Grape Spirits (Pisco)	Renzo Bolivar, Edgar Sarmiento and Guina Sotomayor Alzamora
11h00 - 11h20	***Coffee Break***	
11h20 - 11h40	Optimal layer selection on DCNN using freezing and binary Search	Henry Herrera and José Herrera
11h40 - 12h40	**Title:** Deep Data Integration **Abstract:** We are witnessing the widespread adoption of deep learning techniques as avant-garde solutions to different computational problems in recent years. In data integration, the use of deep learning techniques has helped establish several state-of-the-art results in long standing problems, including information extraction, entity matching, data cleaning, and table understanding. In this talk, I will reflect on the strengths of deep learning and how that has helped move forward the needle in data integration. I will also discuss a few challenges associated with solutions based on deep learning techniques and describe some opportunities for the future work.	**Dr. Wang-Chiew Tan** *Facebook, USA* 
12h40 - 13h00	Using Unsupervised Learning to Profile Public Service Accessibility based on the Public Transport Infrastructure	Leibnitz Rojas-Bustamante, Crayla Alfaro, Ivan Molero, Dennis Aparicio and Miguel Nunez-Del-Prado
13h00 - 14h20	***Lunch***	
Session 4		
14h20 - 14h40	Multiple Scale Comparative Analysis of Classical, Dynamic and Intelligent Edge Detection Schemes	Zhengmao Ye and Hang Yin
14h40 - 15h40	**Title:** Machine Learning for Emotion Prediction, Ideology Detection and Polarization Analysis using COVID-19 Tweets **Abstract:** The adversarial impact of the Covid-19 pandemic has created a health crisis globally all over the world. This unprecedented crisis forced people to lockdown and changed almost every aspect of the regular activities of the people. Thus, the pandemic is also impacting everyone physically, mentally, and economically, and it, therefore, is paramount to analyze and understand emotional responses during the crisis affecting mental health. Negative emotional responses at fine-grained labels like anger and fear during the crisis might also lead to irreversible socio-economic damages. In this talk, I will discuss a neural network model trained using manually labeled data to detect various emotions at fine-grained labels in the Covid-19 tweets automatically. I will discuss about a manually labeled tweets dataset on COVID-19 emotional responses along with regular tweets data. A custom Q&A roBERTa model to extract phrases from the tweets that are primarily responsible for the corresponding emotions has been designed. None of the existing datasets and work currently provide the selected words or phrases denoting the reason for the corresponding emotions. Further, we propose a deep learning model leveraging the pre-trained BERT-base to detect the political ideology from the tweets for political polarization analysis. The experimental results show a considerable improvement in the accuracy of ideology detection when we use emotion as a feature.	**Dr. Madria Sankay** *Missouri University, USA* 
15h40 - 16h00	Soil organic carbon predicton using digital color sensor in Peru	Elida Montero, Alex Vásquez, Laura Alayo, Pedro Gutiérrez and Carlos Mestanza
16h00 - 16h20	***Coffee Break***	
16h20 - 17h20	**Title:** Causal inference in machine learning for actionable healthcare **Abstract:** Machine (deep) learning is increasingly becoming key to precision medicine—from identifying disease risks and taking preventive measures, to making diagnoses and personalizing treatment for individuals. Precision medicine, however, is not only about predicting risks and outcomes, but also about weighing interventions. Interventional clinical predictive models require the correct specification of cause and effect, and the calculation of so-called counterfactuals, that is, alternative scenarios. In biomedical research, observational studies are commonly affected by confounding and selection bias. Without robust assumptions, often requiring a priori domain knowledge, causal inference is not feasible. Data-driven prediction models are often mistakenly used to draw causal effects, but neither their parameters nor their predictions necessarily have a causal interpretation. Here we discuss how target trials, transportability and prediction invariance are linchpins to developing and testing intervention models.	**Dr. Yi Guo** *University of Florida, USA* 
17h00 - 17h20	Ideation of Computational Thinking Programs by Assembling Code Snippets from the Web	Hasan Jamil

17h20 - 17h40	Smart doorbell with Telegram notification for multifamily dwellings	Gianfranco Perez Aquise and Frank E. Escobedo-Bailón

November 18, 2022

Hour	Title	Speaker
08h30 - 09h00	***Reception***	
Session 5		
09h00 - 09h20	Using features based on elongation to enhance sentiment analysis	Abderrahim Rafae, Mohammed Erritali, Younes Madani and Mathieu Roche
09h20 - 10h00	Performance analysis of machine learning algorithms for food fraud prediction	Joshep Douglas Estrella Condor and Félix Armando Fermín-Pérez
10h00 - 11h00	**Title:** Friends Don't Let Friends Deploy Black-Box Models: Glass-Box Learning to the Rescue! **Abstract:** I'll present a variety of case studies where glass-box learning methods uncover surprising statistics in data that would make deploying a black-box model trained on that data risky. Fortunately, the high-accuracy glass-box ML methods now available make it possible to detect and correct these problems before deployment, and to protect privacy and reduce bias, too.	**Dr. Rich Caruana** *Microsoft, USA* 
10h20 - 10h40	Poster session (short contributions)	
10h40 - 11h00	***Coffee break***	
11h00 - 12h00	Poster session (short contributions)	
12h00 - 13h00	**Title:** Ingesting Knowledge from Diverse Sources to Open Domain Social Conversations **Abstract:** Following the recent advancements in language modeling and availability of large natural language datasets, the last decade has been flourishing for conversational AI research. The progress also helped emphasize the importance of reasoning over a diverse set of external knowledge and task completion resources for forming relevant, informative, and accurate responses. In this talk, I will discuss our recent work on integrating knowledge to conversation responses from such a diverse set of resources, challenges associated with these, and progress we made so far.	**Dr. Hakkani-Tur Dilek** *Amazon, USA* 
13h00 - 13h20	Closure of SIMBig 2022	Juan Antonio Lossio-Ventura Hugo Alatrista