GENERAL PROGRAM - SIMBig 2023

Time in Mexico City, CDMX, Mexico (CST)
Central Time Zone
20 min presentation

December 13, 2023

Hour	Title	Speaker
08h30 - 09h20	***Reception***	
09h20 - 09h40	**Welcoming to SIMBig2023**	Hiram Calvo Juan Antonio Lossio-Ventura
Session 1		
9h40 - 10h40	**Title**: Generative Information Retrieval **Abstract**: I will discuss the historical paradigm of information retrieval systems, highlighting their limitations in integrating information from multiple documents into a direct answer. The advent of large language models (LLMs), empowered by deep neural networks, has challenged this paradigm by enabling models to memorize facts and relations. I will outline the emergence of generative information retrieval systems, addressing challenges such as grounding answers, attributing passages to primary sources, providing nuanced responses, avoiding bias, and navigating the evolving content ecosystem influenced by LLMs, raising questions about the treatment of generated content, differentiation from human-authored content, and the impact on business models.	**Dr. Marc Najork** *Google, USA* 
10h40 - 11h00	Predicting Course Performance on a Massive Open Online Course Platform: A Natural Language Processing Approach	Grant Alphenaar, Rahat Ibn Rafiq
11h00 - 11h20	***Coffee Break***	
11h20 - 11h40	Corporate Events Prediction Using Earning Call Transcripts	Zhaomin Xiao, Yachen Cui, Zhelu Mai, Jiancheng Li, Zhouer Xu
11h40 - 12h00	Comparing incremental learning approaches for a growing sign language dictionary	Joe N. Huamani Malca, Gissella Bejarano
12h00 - 12h20	Comparative Analysis of GPT Models for Detecting Cyberbulling in Social Media Plataforms Thread	Mohammad Shafiqul Islam, Rahat Ibn Rafiq
12h20 - 12h40	Prediction of Siderophores Partition Coefficient using Artificial Neural Network	Jesus Alvarado, Miquéias Silva, Moises Stevend Meza, Adriano Werhli, Karina Machado
12h40 - 13h00	Improvement of EduBPMN Transformation Rules from an Empirical Validation	Jorge Eduardo Díaz Suárez, José Ignacio Panach Navarrete
13h00 - 14h00	***Lunch***	
Session 2		
14h00 - 14h20	Analysis of Mexican women's decision-making power using machine learning strategies.	Paulina Aldape Bretado, Mariano de Jesús Gómez Espinoza, Juanita Hernández López, Azucena Yoloxóchitl Ríos Mercado, Alvaro Eduardo Cordero Franco
14h20 - 15h20	**Title**: Social Media Mining - A Bountiful Frontier in AI and Data Science **Abstract**: Unprecedented research opportunities emerge in AI and Data Science through the lens of social media data. We contemplate the role of social media mining in the rapid development of AI and Big Data.	**Dr. Huan Liu** *Arizona State University, USA* 
15h20 - 15h40	Multivariable-unistep Prediction of Travel Times in Public Transport Buses Using LSTM and Convolutional LSTM	Elizon F. Carcausto Mamani, Etson Ronaldao Rojas Cahuana, Harley Vera Olivera, Edwin Alvarez
15h40 - 16h00	***Coffee Break***	

16h00 - 17h00	**Title**: Deep Learning for Nonlinear System Modeling **Abstract**: Deep learning has demonstrated remarkable success in image and speech recognition, natural language processing, and various other domains. However, it has not yet been extensively utilized in modeling complex nonlinear systems. In this talk, we will explore several deep learning methods for nonlinear system identification. These methods include the use of long short-term memory networks (LSTMs) for modeling dynamic behaviors, convolutional neural networks (CNNs) for addressing spatial dependencies in nonlinear modeling, and transfer learning to handle limited data for specific nonlinear systems. Additionally, we will discuss the application in time series prediction. The efficacy of these deep learning methods will be validated through several benchmark examples.	**Dr. Wen Yu** *CINVESTAV-IPN, Mexico* 
17h00 - 17h20	Novel algorithm to predict electoral trends, case in Mexico.	Mauricio Flores Geronimo, Ulises Cruz Valencia, Daniela Grave Aragón, Renato García Gonzáles, Manuel Alejandro Guerrero Martínez

December 14, 2023

Hour	Title	Speaker
09h00 - 09h20	***Reception***	
Session 3		
09h20 - 09h40	A Bayesian Machine Learning Model for Predicting Uncontrolled Type 2 Diabetes Mellitus	Adriana Robles-Cabrera, Christopher Stephens, Joana Itzel Ríos-Barba, Areli Monserrat Rios-Vazquez, Romel Calero-Ramos
09h40 - 10h40	**Title**: Frontiers of Foundation Models for Time Series Modeling and Analysis **Abstract**: Recent development in deep learning has spurred research advances in time series modeling and analysis. Practical applications of time series raise a series of new challenges, such as multi-resolution, multimodal, missing value, distributeness, and interpretability. In this talk, I will discuss possible paths to foundation models for time series data and future directions for time series research.	**Dr. Yan Liu** *University of Southern California, USA* 
10h40 - 11h00	Laws of Federated Learning For Bioinformatics	Prakash Chourasia, Taslim Murad, Zahra Tayebi, Sarwan Ali, Imdad Ullah Khan, Murray D. Patterson
11h00 - 11h20	***Coffee Break***	
11h20 - 11h40	Beyond Accuracy: Measuring Representation Capacity of Embeddings to Preserve Structural and Contextual Information	Sarwan Ali
11h40 - 12h00	Visualizing Software Test Requirements Using NLP and Human-in-the-Loop Approach	S. M. Azizul Hakim, Rahat Ibn Rafiq, Michael Lingg
12h00 - 13h00	**Title**: Machine Learning and Big Data Analytics for Sustainable Food-Energy-Water Nexus **Abstract**: The critical resources of food, energy, and water (FEW) are pivotal for human survival. Achieving sustainable access to these resources aligns with the United Nations' Sustainable Development Goals for 2030. Technological advancements have generated vast FEW-related data, attracting data scientists to address FEW nexus challenges. Notable challenges include predicting wind energy production, analyzing land irrigation efficiency, monitoring water quality, detecting plant diseases, and advancing precision agriculture, etc. These challenges find resolution through the application of machine learning techniques and Big Data analytics. Associated themes encompass spatio-temporal data analysis, handling high-dimensional data, addressing data imbalances, managing the complexity of heterogeneous data sources, and integrating disparate data sets through data fusion strategies. This presentation embarks on an exploration of the FEW systems and the concomitant data science challenges it engenders. Subsequently, it shares preliminary findings from ongoing research endeavors, specifically focusing on the long-term prediction of time series within wind energy systems, even in the presence of missing data points.	**Dr. Xiaoou Li** *CINVESTAV-IPN, Mexico* 

13h00 - 14h00	***Lunch***	
Session 4		
14h00 - 14h20	Expanding Chemical Representation with k-mers and Fragment-based Fingerprints for Molecular Fingerprinting	Sarwan Ali, Prakash Chourasia, Murray D. Patterson
14h20 - 15h20	**Title**: Responsible Thinking: laying the operational foundation **Abstract**: In this talk I will chart out a practical and operational foundation for what I coin as responsible thinking. I believe that this is critical for building modern technologies that are at the intersection of AI and society.	**Dr. Mona Diab** *Caneguie Mellon University, USA* Carnegie Mellon University
15h20 - 15h40	***Coffee Break***	
15h40 - 17h20	**Track on Diversity and Inclusion** *(hybrid session)* ***Zoom Link***: *https://cnrs.zoom.us/j/97680757116?pwd=M042YnhpdHVCNVI0dUtJVXh5blk5Zz09* *Passcode: sqq5ci* ***Detailed program***: *http://vargas-solar.com/dei-track-simbig-2023/*	
15h40 – 16h00	Building a Diverse Future: A Practical Guide to Structuring Gender-Inclusive STEM	Genoveva Vargas-Solar, CNRS, LIRIS, France
16h00 – 16h20	Diversity and Inclusion-Aware Data Science	Barbara Catania, University of Genova, Italy
16h20 – 17h20	**Panel**: Practices and methodologies for designing data science research projects: careers, applications, and responsible actions ***Moderators***: *Barbara Catania, Genoveva Vargas-Solar, and Ester Zumpano.* ***Panelists***: *Luciana Ayciriex, Sagrario Castillo Camporro, and Greta Temporin.*	
17h20 - 17h40	Recommendations for the Development of Augmented Reality Video Games for Children with ADHD	Augusto Gerardo Morante Castañeda, Omar Jesus Cahuana Rios, Jorge Eduardo Díaz Suárez
17h40 - 18h00	Parallel EvoCluster: An Open-Source Parallel Nature-Inspired Optimization Clustering	Edwin Alvarez Mamani, Milagros Yarahuaman Rojas, Raul Huillca Huallparimachi

December 15, 2023

Hour	Title	Speaker
08h30 - 09h00	***Reception***	
Session 5 ***(online session)*** **Zoom link:** *cic-ipn-mx.zoom.us/my/simbig*		
09h20 - 09h40	The Study of Human Action Recognition in Videos with Long Short-Term Memory Model	Hassan Khan, Sammra Habib, Amna Qasim, Nisar Hussain, Gull Mehak, Asghar Afzal, Muhammad Usman, Zain Burraq
09h40 - 10h00	User-agnostic Model for Retweets Prediction Based on Graph-Embedding Representation of Social Neighborhood Information	Martin Dominguez, Pablo G Celayes, Damian Barsotti
10h00 - 10h20	Item Response Theory In Sample Reweighting To Build Fairer Classifiers	Diego Minatel, Nicolas Roque Santos, Vinícius Ferreira, Mariana Curi, Alneu A. Lopes
10h20 - 10h40	Analyzing Sentiments and Topics on Twitter towards Rising Cost of Living	Yanyi LI, Nian Ran, Yifu Chen, Renhua Zhou, Riza Batista-Navarro
10h40 - 11h00	A Framework to Transform Metadata and Document-level Tabular Spatial Information and Measurement to Marine Geology Gazetteer	Muhammad Asif Suryani, Christian Beth, Klaus Wallmann, Matthias Renz
11h00 - 11h20	Reviewer 2 Must be Stoped: Transformer-based Approaches for Predicting Paper Aceptance	Lukas Rimkus, Jonas Verbickas, Riza Batista-Navarro
11h20 - 11h40	Reinforcement learning & biologically inspired artificial neural networks	Ariel Fiuri, Martin Dominguez, Francisco Tamarit
11h40 - 12h00	Design optimization for high-performance computing using FPGA	Murat Isik, Kayode Inadagbo, Hakan Aktaş

12h00 - 13h00	**Title**: Recent advances in score-based diffusion models **Abstract**: Generative models are typically based on explicit representations of probability distributions (e.g., autoregressive) or implicit sampling procedures (e.g., GANs). I will present an alternative approach based on modeling directly the vector field of gradients of the data distribution (scores) which underlies recent score-based diffusion models. This framework allows flexible architectures, requires no sampling during training or the use of adversarial training methods. Additionally, score-based diffusion generative models enable exact likelihood evaluation through connections with neural ODEs, achieving state-of-the-art sample quality and excellent likelihoods on image datasets. I will discuss numerical and distillation methods to accelerate sampling and their application to inverse problem solving.	**Dr. Stefano Ermon** *Stanford University, USA*  Stanford University
13h00 - 13h20	CollabVR: A Social VR Architecture for social interaction between college students	Diego Johnson, Brayan Mamani, Cesar Salas
13h20 - 13h40	CLPSafe: Mobile application for avoid cloned of license plates using Deep Learning	Diego Sanchez, Jhon Silva, Cesar Salas
13h40 - 14h00	Closing of SIMBig 2023	Hiram Calvo & Hugo Alatrista