5th Winter School on Data Analytics

November 20-22, 2020

Nizhny Novgorod, Russia

Laboratory of Algorithms and Technologies for Network Analysis of National Research University Higher School of Economics

Keldysh Institute of Applied Mathematics of Russian Academy of Science

Moscow Center for Fundamental and Applied Mathematics

School Lecturers.

Emilio Carrizosa, University of Sevilla, Spain
Mihalis Doumpos, Technical University of Crete, Crete
Mario Guarracino, University of Cassino and ICAR-CNR, Italy
Ilias Kotsireas, Wilfrid Laurier University, Canada
Harilaos Mertzanis, Abu Dhabi University, United Arab Emirates
Panos Pardalos, University of Florida, USA and HSE
Marcello Pelillo, Ca' Foscari University of Venice, Italy
Wolfram Wiesemann, Imperial College London, UK

Co-Chairs of the school

Panos M. Pardalos University of Florida and LATNA, HSE
Alexander I. Aptekarev, Keldysh Institute of Applied Mathematics of the Russian Academy of Science, Moscow Center for Fundamental and Applied Mathematics

Program Committee

Fuad Aleskerov, NRU HSE
Mikhail Batsyn, NRU HSE
Fedor Fomin, Bergen University, Norway and St Petersburg department of Steklov Mathematical Institute
Valery Kalyagin, NRU HSE
Yury Kochetov, Russian Academy of Sciences, Novosibirsk
Alexander Koldanov, NRU HSE
Valentina Kuskova, NRU HSE, Laboratory of Applied Network Research
Dmitriy Malyshev, NRU HSE
Oleg Prokopyev, University of Pittsburgh, USA
Andrey Raigorodskii, Moscow Institute of Physics and Technology, Moscow State University, Yandex
Nikolay Zolotykh, Lobachevsky State University, Nizhny Novgorod
Andrey Savchenko, NRU HSE

Day 1, Friday, November 20.

Zoom

15:00 -15:50 Marcello Pelillo,

Lecture 1. From Cliques to Equilibria: The Dominant-Set Approach to Clustering

16:00 -16:50 Marcello Pelillo,

Lecture 2. From Cliques to Equilibria: The Dominant-Set Approach to Clustering

17:00 – 17:50 Wolfram Wiesemann,

Lectures 1: Introduction to Robust Optimization

18:00 -18:50 Wolfram Wiesemann,

Lectures 2: Introduction to Robust Optimization

Day 2, Saturday, November 21.

Zoom

15:00 -15:50 Emilio Carrizosa,

Lectures 1: On Interpretability in Data Analytics

15:00 -16:50 Emilio Carrizosa,

Lectures 2: On Interpretability in Data Analytics

17:00 – 17:50 Mihalis Doumpos,

Data analytics and decision aiding models for credit risk modeling and banking

18:00 -18:50 Harilaos Mertzanis,

A data-driven approach to financial institution soundness around the world

Day 3, Sunday, November 22.

Zoom

15:00 -15:50 Panos Pardalos,

Recent Research on Support Vector Machines

16:00 -16:50 Mario Guarracino,

Lecture 1: Introduction to machine learning for biomedical data science

17:00 – 17:50 Mario Guarracino,

Lecture 2: Introduction to classification and clustering

18:00 -18:50 Ilias Kotsireas

Spectral Clustering

Emilio Carrizosa,
University of Sevilla, Spain

On Interpretability in Data Analytics

Abstract: An important challenge in Data Analytics is to make classification and prediction algorithms more interpretable, in the sense that we should know which attributes, and at which extent, contribute in the prediction.

A powerful tool to attain this aim is Mathematical Optimization, which allows us to pose in a natural way the multiobjective problem of optimizing the performance and, at the same time, the number of attributes or measurement costs.

In this course we will illustrate the use of Mathematical Optimization strategies in different problems, such as dimensionality reduction (sparse PCA), sparse linear models with performance constraints, cost-sensitive Support Vector Machines with performance constraints or functional data, sparse classification and regression trees, interpretable clustering, etc., with special focus on the methods developed by the research group in Optimization in IMUS, the Institute of Mathematics of the University of Seville.

Mihalis Doumpos,
Technical University of Crete, Crete

Data analytics and decision aiding models for credit risk modeling and banking

Abstract: Over the past few decades, the financial sector has been under continuous changes in all areas of designing, implementing, and managing the services and products provided to consumers, firms, and investors. One of the most notable changes involves the extensive use of analytical modeling techniques for financial decision making and risk management. These modeling techniques originate from various disciplines, such as statistics, operations research, and computer science. This talk focuses on the applications of data analytics, based on machine learning and decision aiding methods, with emphasis on credit risk modeling and banking. We present the context for building descriptive and predictive models in these areas, the existing challenges, and overview established and state-of-the-art data analytic methodologies used this domain. Results from illustrative applications are also presented.

Mario Guarracino,
University of Cassino and ICAR-CNR, Italy

Lecture 1: Introduction to machine learning for biomedical data science

Abstract: In this lecture we will analyze some recent advancements in the field of data science and how machine learning is providing effective solutions to many problems in biomedicine. Examples will cover the analysis of experimental data types produced by biomedical equipments to solve different problems.

Lecture 2: Introduction to classification and clustering

Abstract: We will review basic machine learning concepts and algorithms. We will explore simple clustering and classification algorithms such as Hierarchical Clustering, K-means, Nearest neighbors, Support Vector Machines, Generalized Eigenvalue Classifiers, Decision trees and Random Forest. Simple examples will be provided to effectively explore pros and cons of these algorithms.

Ilias Kotsireas,
Wilfrid Laurier University, Canada

Spectral Clustering

Abstract: Spectral clustering has become one of the most popular modern clustering algorithms. It is simple to implement, can be solved efficiently by standard linear algebra software, and very often outperforms traditional clustering algorithms such as the k-means algorithm. We will describe the necessary concepts needed to define spectral clustering, including the similarity matrix and various graph Laplacians. We will also touch briefly on the perturbation approach to spectral clustering and the eigengap heuristic for the number of clusters. We will illustrate spectral clustering with specific examples, so as to solidify our understanding of the algorithm and the related concepts.

Harilaos Mertzanis,
Abu Dhabi University, United Arab Emirates

A data-driven approach to financial institution soundness around the world

Abstract: We use a fully data-driven approach and information provided by the IMF's financial soundness indicators to measure the soundness of a country's financial system around the world. Given the nature of the measurement problem, we apply principal component analysis (PCA) to deal with the presence of strong cross-sectional dependence in the data due to unobserved common factors. Using this comprehensive sample and various statistical methods, we produce a data-driven measure of financial soundness that provides policy makers and financial institutions with a tool that is easy to implement and update.

Panos Pardalos,
University of Florida, USA and HSE

Recent Research on Support Vector Machines

Abstract: In this lecture we are going to discuss recent developments for an important class of algorithms in data sciences, the Support Vector Machines. Some results have been published in the papers below:

1. "Ramp-loss nonparallel support vector regression: Robust, sparse and scalable approximation" (Tang, Long, Tian, Yingjie, Yang, Chunyan, Panos M.Pardalos), Knowledge-Based Systems, Volume 147, 1 May 2018, Pages 55-67.
2. "A novel perspective on multiclass classification: Regular simplex support vector machine" (Tang Long, Tian Yingjie, Panos M. Pardalos) Information Sciences, Vol. 480, pp 324-338 (April 2019).
3. "Structural improved regular simplex support vector machine for multiclass classification" (Long Tang, Yingjie Tian, Wenjun Li, Panos M Pardalos), Applied Soft Computing, Volume 91, 106235 (June 2020)
https://www.sciencedirect.com/science/article/abs/pii/S1568494620301757
4. "A Survey of Support Vector Machines with Uncertainties" (Ximing Wang, Panos M. Pardalos) Annals of Data Science, (2014) 1 (3-4), pp. 293 - 309.

Marcello Pelillo
Ca' Foscari University of Venice, Italy

Lectures 1, 2. From Cliques to Equilibria: The Dominant-Set Approach to Clustering

Abstract: These lectures will provide an overview of "dominant sets," a graph-theoretic notion of a cluster which generalizes the concept of a maximal clique to edge-weighted graphs and has intriguing connections with optimization theory, (evolutionary) game theory, and dynamical systems theory. The idea is general and can be applied to weighted graphs, digraphs and hypergraphs alike. After introducing the main properties of dominant sets, along with some generalizations, I will discuss a few recent applications in the field of computer vision, including interactive image segmentation, geo-localization, multi-camera tracking, and person re-identification.

Bio Sketch

Marcello Pelillo is a Full Professor of Computer Science at Ca' Foscari University of Venice, Italy, where he leads the Computer Vision and Pattern Recognition group. He

has directed the European Centre for Living Technology (ECLT) and has held visiting research/teaching positions in various institutions such as Yale University, McGill University, the University of Vienna, York University (UK), National ICT Australia (NICTA), Wuhan University, Huazhong University of Science and Technology, and South China University of Technology. He has been General Chair for ICCV 2017, Program Chair for ICPR 2020, and has served in various roles in the organization of the main conferences of his research areas. He is the Specialty Chief Editor of Frontiers in Computer Vision and serves, or has served, on the Editorial Boards of several journals, including IEEE Transactions on Pattern Analysis and Machine Intelligence, IET Computer Vision, Pattern Recognition, and Brain Informatics. He also serves on the Advisory Board of the International Journal of Machine Learning and Cybernetics. Prof. Pelillo has been elected a Fellow of the IEEE and a Fellow of the IAPR, and is an IEEE SMC Distinguished Lecturer. His Erdös number is 2.

Wolfram Wiesemann,

Imperial College London, UK

Lectures 1, 2: Introduction to Robust Optimization

Abstract: Traditionally, uncertainty-affected decision problems are solved by modelling the uncertain problem data as random variables and subsequently discretizing the outcomes of these random variables. Although this is a very natural approach, it has several shortcomings: it requires the exact specification of the underlying stochastic process (which is rarely available in practice), and it results in a curse of dimensionality for dynamic (multi-stage) problems, which implies that the computation times grow exponentially with problem size. In this lecture, we review the rapidly growing literature on robust and distributionally robust optimization, which aims to alleviate the aforementioned shortcomings. A robust optimization problem specifies an uncertainty set that contains all possible values for the uncertain problem parameters, and it seeks the best decision in view of the worst parameter realization. A distributionally robust optimization problem, on the other hand, specifies an ambiguity set that contains all possible probability distributions that could govern the uncertain problem parameters, and it seeks the best decision in view of the worst probability distribution. Topics covered include the reformulation and solution of static and dynamic (distributionally) robust optimization problems as well as discrete robust optimization.

Note: To participate, a working knowledge of linear programming is required.

Bio sketch: Wolfram Wiesemann is professor of Analytics and Operations and Academic Director of the MSc Business Analytics (on-campus) at Imperial College

Business School, London, where he also serves as an Associate Director of the Centre for Process Systems Engineering as well as a Fellow of the Imperial Business Analytics centre. Wolfram has been a visiting researcher at the Institute of Statistics and Mathematics at Vienna University of Economics and Business, Austria, in 2010; the Computer-Aided Systems Laboratory at Princeton University, USA, in 2011; and the Industrial Engineering and Operations Research Department at Columbia University, USA, in 2012. Wiesemann’s research interests revolve around the methodological aspects of decision-making under uncertainty, as well as applications in operations management, energy, and finance.