

Winter School on Data Analytics, November 17-18, 2022. Nizhny Novgorod

7th Winter School on Data Analytics

November 17-18, 2022

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

<https://nnov.hse.ru/en/latna/conferences/da2022>

Winter School on Data Analytics, November 17-18, 2022. Nizhny Novgorod

School Lecturers.

Roberto Battiti (University of Trento and HotelInCloud.com, Italy)

Dmitry Kiselev (Faculty of Computer Science, HSE Moscow)

Andrey Kuznetsov (Sber AI)

Pierre Miasnikof (University of Toronto, Canada)

Panos Pardalos (LATNA, HSE and University of Florida)

Alex Shestopaloff (Queen Mary University of London and a Fellow of the Alan Turing Institute, 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, Huawei Research Institute and 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

Dmitriy Malyshev, NRU HSE

Andrey Raigorodskii, Moscow Institute of Physics and Technology, Moscow State University, Yandex

Nikolay Zolotykh, Lobachevsky State University, Nizhny Novgorod

Andrey Savchenko, NRU HSE

The Winter School is organized in distant format, in Zoom

<https://zoom.us/j/97807586194?pwd=UGxmeGpDc3BvK3JFK1NrRjlJQzVldz09>

Conference ID: 978 0758 6194 Password: 465021

Winter School on Data Analytics, November 17-18, 2022. Nizhny Novgorod

Day 1, Thursday, November 17. Zoom

15:00 -15:50 Pierre Miasnikof

Lecture 1: Graph clustering quality

16:00 -16:50 Pierre Miasnikof

Lecture 2: Statistical testing of clusterability

17:00 -17:50 Roberto Battiti

Tourism and Hospitality: Relevant Problems for Machine Learning and Intelligent Optimization

18:00 -18:50 Panos Pardalos

AI and Data Analytics in Economics and Finance

Day 2, Friday, November 18. Zoom

13:00 -13:50 Alex Shestopaloff

Lecture 1: Bayesian methods for reduced rank regression models

14:00 -14:50 Alex Shestopaloff

Lecture 2: Developments in principal component regression

15:00 -15:50 Andrey Kuznetsov

Text2Image Generation using Diffusion Models

(диффузионные модели для генерации изображений по текстовым описаниям)

16:00 - 16:50 Dmitry Kisilev

Graph-based recommender systems

Roberto Battiti

University of Trento and HotelInCloud.com, Italy

Lecture: Tourism and Hospitality: Relevant Problems for Machine Learning and Intelligent Optimization

The tourism and hospitality sector are undergoing disruptive changes made possible by the connection between abundant data (searches, reservations, reviews, ...), models based on machine learning, and optimization tools to identify optimal or improving decisions. Through the cloud, this wave of innovation is percolating from international chains to individual hotels of medium and small size. The complexity and relevance of the sector are worth further investigation by competent researchers. In the talk we review the main applications of learning and intelligent optimization techniques to tourism and hospitality.

Dmitry Kiselev

Faculty of Computer Science, HSE Moscow

Lecture : Graph-based recommender systems

The seminar will describe the recommender systems problem and how it can be formalized using graphs. After, we will discuss the main concepts of graph neural network techniques and how they can be utilized to improve the quality of recommender systems from different perspectives.

Andrey Kuznetsov (Sber AI)

Lecture: Text2Image Generation using Diffusion Models (диффузионные модели для генерации изображений по текстовым описаниям)

Pierre Miasnikof

University of Toronto, Canada

Lecture 1: Graph clustering quality

Graph clustering, labeling vertices so they form densely connected subgraphs, is a fundamental task in the study of complex networks. Unfortunately, a formal definition of graph clusters does not exist, consequently the question of clustering quality remains an open question. Nevertheless, a quality assessment of clustering algorithms is essential in graph clustering (community detection). As stated in the literature, “(...) *running a clustering algorithm over a set of randomly generated data points will always produce clusters which, however, have little meaning*”.

In this lecture, I will introduce graph-clustering quality measures based on comparisons of global, intra- and inter-cluster densities, an accompanying statistical significance test and a step-by-step routine for clustering quality assessment.

Lecture 2: Statistical testing of clusterability

As mentioned earlier, graph clustering is a fundamental task in the study of complex networks. Unfortunately, not all graphs are clusterable. Attempting to cluster graphs that do not have a clustered structure, graphs that cannot be meaningfully summarized through clusters, is not only a waste of time, it also leads to misleading conclusions.

In this lecture, I will show a statistical test to determine if a graph meets the pre-requisite (necessary) conditions for clustering.

Panos Pardalos

University of Florida, USA and HSE

Lecture: AI and Data Analytics in Economics and Finance

Alex Shestopaloff

Queen Mary University of London and
a Fellow of the Alan Turing Institute, UK,

Lecture 1. Bayesian methods for reduced rank regression models.

Reduced rank regression models are multivariate regression models of correlated response vectors. An example would be the returns of a stock at different time horizons, say 1, 3, 5 days. In this lecture, I will consider some current developments in Bayesian reduced rank regression modeling, challenges, and potential approaches to overcoming them.

Lecture 2. Developments in principal component regression.

Principal component regression is a widely used method based on replacing a set of correlated predictor variables with the principal components of their covariance. I will consider some recent developments of PCA regression methods, including novel algorithmic techniques as well as their robustness properties and links to other techniques.