POLISH NATIONAL AGENCY FOR ACADEMIC EXCHANGE

STER PROGRAMME

name of the unit:

ANALYSIS OF BIG DATA WITH OUTLIER DETECTION

Institute of Information Technology, Lodz University of Technology

symbol:

I-71

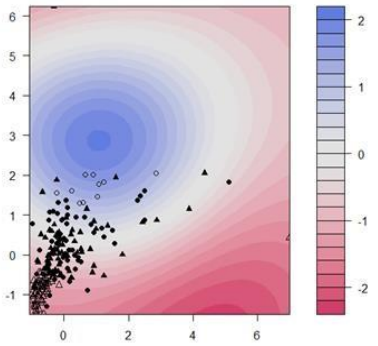
http://it.p.lodz.pl

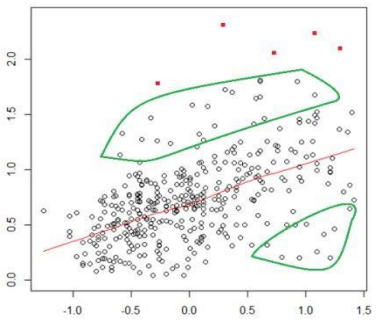
head of the unit:	potential promoters:	contact person:
Piotr Szczepaniak, MSc, PhD, DSc, Full Professor	Agnieszka Duraj, MSc, PhD, DSc Łukasz Chomątek, MSc, PhD	Agnieszka Duraj, MSc, PhD, DSc tel: (+48 42) 631-27-96 (+48 42) 631-39-54 agnieszka.duraj@p.lodz.pl

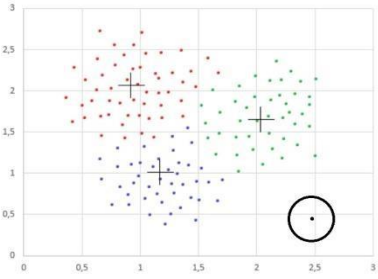
scope of activities:

Intelligent quantitative and qualitative analysis of big data with outlier detection.
The general goal of the research is improvement of known methods of data analysis and knowledge extraction, as well as development of new ones. The goal is achieved with the use of classic approaches and methods of the artificial intelligence. Particular attention is paid to evolutionary algorithms, their novel variants, and fuzzy sets applications.

graphic material







Presenta ctivities:

The research is focused on the following issues:

- separation of data for native, foreign, and outlier ones;
- variants of linguistic summarization applied to outlier detection;
- innovative approach to case-based reasoning;
- development of evolutionary algorithms, multi-objective in particular;
- multi-objective approach to outlier detection;
- hierarchical methods;
- outlier detection in data steams;
- innovations in methods based on statistics, distance, and density;
- verification of methods on real-world data;
- extraction and generalization of knowledge; consideration of context.
  The definition of an outlier often requires the cumulative application of several different criteria (e.g. low cardinality and distance from dominant "typical" patterns). For this reason, it is natural to develop and use multi-criteria optimization methods, here evolution algorithms.

1. A. Duraj, P. S. Szczepaniak: *Linguistic Summaries Using Interval-Valued Fuzzy Representation of Imprecise Information-An Innovative Tool for Detecting Outliers*. International Conference on Computational Science - ICCS, pp.500--513, Springer, 2021.
2. A. Duraj, P. S. Szczepaniak: *Outlier Detection in Data Streams—A Comparative Study of Selected Methods*. 25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES, Elsevier, 2021.
3. A. Duraj, P. S. Szczepaniak, L. Chomatek: *Intelligent Detection of Information Outliers Using Linguistic Summaries with Non-monotonic Quantifiers*. Springer Nature Switzerland AG 2020 M.-J. Lesot et al. (Eds.): IPMU 2020, CCIS 1239, pp. 787–799, 2020. https://doi.org/10.1007/978-3-030-50153-2_58.

The portfolio of research groups was created as part of the Programme "STER" – Internationalisation of doctoral schools" as part of the realization of the project "Curriculum for advanced doctoral education & taining – CADET Academy of Lodz University of Technology".

4. P.S. Szczepaniak, A. Duraj (2018): *Case-Based Reasoning – the Search for Similar Solutions and Identification of Outliers*. Complexity (ID 9280787; open access)	

Future activities:

1. Development of effective methods of detecting anomalies in data sets or patterns.
2. Development in the field of multi-criteria optimization methods, which consists in:
- Adaptation of optimization algorithms (in this case genetic) to classification tasks (exception - not exception).
- Development of dedicated genetic operators for the exception detection problem.
- Defining the method of selecting the components of the objective functions used in the multi-criteria optimization task.

Keywords: intelligent data analysis, outlier detection

List of internship proposal in this research team:

List of attachments: