



DISTINGUISHED LECTURE SERIES

Learning classifiers from concept drifting and imbalanced data streams

Dr. Jerzy Maciej Stefanowski

Poznań University of Technology, Poland

Thursday, May 23, 2024 10:00am
Auditorium 106, at IIS new Building



Abstract

In many real-world applications, incoming data are often available as data streams characterized by huge volumes of instances and rapid arrival rates, which often requires quick response. Besides special computational requirements, another challenge is the non-stationary characteristics of such data, where the data and target concepts change over time in a phenomenon of concept drift. In this talk, the different types of drifts will first be characterized. Then, we will present various, own proposals for ensembles of incremental predictive models (such as AUE, OAUE) and discuss experimental evaluation of their adaptation to various types of drifts. The task of learning classifiers from streams becomes even more challenging in the presence of additional data complexities, in particular imbalances between cardinalities of target classes. Most existing work focuses on designing new incremental algorithms for dealing with the global imbalance ratio only and does not consider other data complexities (which are known to be difficulty factors for learning classifiers from static data). As the interactions between concept drifts and such local data difficulty factors is not sufficiently investigated in concept drifting data streams, in this talk we will present a new categorization of concept drifts and local data difficulty factors for imbalanced data streams. Then, we will summarize results of our recent comprehensive experimental study with representative on-line classifiers applied to various synthetic and real-world imbalanced data streams. The results show differences in existing classifiers' reactions to such factors and drifts. Combinations of multiple difficulty factors are the most challenging for many classifiers, which are not able to recover from these drifts. Therefore, a specialized generalization of online bagging ensemble, which should handle some of these combined difficulties, will be presented. At the end of the talk, we will address the challenges of explaining the predictions of such streaming models and their reaction to changes in evolving data. This is related to the recent few works on adaptation of XAI paradigms in this context.

Biography

Jerzy Stefanowski is a full professor at Poznan University of Technology Institute of Computing Science, and leads the university's Machine Learning Lab. He is a Member of the Polish Academy of Sciences since 2021, and the Vice-President of the Polish Artificial Intelligence Society.

His research interests include data mining, machine learning, artificial intelligence and intelligent decision support. Major results are concerned with: explainable Artificial Intelligence, counterfactual or prototype explanations of machine learning black box models, induction of various types of rules, multiple classifiers - ensembles, mining class imbalanced data, incremental learning from evolving data streams and concept drift detection, data preprocessing, descriptive clustering of documents and medical applications of data mining. He is the author and co-author of over 230 research papers and 2 books. For more information, please refer to his personal webpage (<http://www.cs.put.poznan.pl/jstefanowski/>).

For more information:
<http://www.iis.sinica.edu.tw/>

