Development of Hybrid Machine Learning Tools with the Use of Graph Theory Methods

Doctoral School – Information and Biomedical Technologies – Polish Academy of Sciences

Supervisor: dr hab. inż. Marek Gągolewski <gagolews at ibspan.waw.pl>

Department of Stochastic Methods, Systems Research Institute, Polish Academy of Sciences
Deakin University, School of IT, Melbourne, VIC, Australia

Project Description

Traditionally, most of the research work in the field of machine learning concerns building solutions for the processing of structured data, where each observation in a dataset is represented by a feature vector in some multidimensional space. Nowadays, however, there is a growing interest, with the help of complex network analysis, to extract a graph structure from the underlying binary relations that we can discover in the data. A typical example is a group of social network users, where in addition to their individual characteristics, we might have access to some relations describing how they interact with each other, e.g., who is a friend of whom. The desire to use such graph structures to improve existing or come up with new machine learning solutions has led to a rapid increase in the popularity of graph representation learning, which is now one of the most popular topics in artificial intelligence [1].

The aim of this project is to advance the research in graph-based machine learning by developing new tools for, amongst others:

- minimum spanning tree-based algorithms and their applicability for unsupervised learning [2, 3, 4],
- extraction of the underlying graph structure from structured data and the investigation of its properties in connection with graph machine learning tools [5, 6],
- graph-based dimensionality reduction and clustering validation methods [7, 8].

It is also important to note that graph data is mostly sparse, which requires new approaches to the storing and processing of such datasets.

References

[1] W. L. Hamilton, R. Ying, and J. Leskovec. Representation learning on graphs: Methods and applications. *IEEE Data Eng. Bull.*, 40:52–74, 2017.

[2] M. Gagolewski, M. Bartoszuk, and A. Cena. Genie: A new, fast, and outlier-resistant hierarchical clustering algorithm. *Information Sciences*, 363:8–23, 2016.

[3] O. Grygorash, Y. Zhou, and Z. Jorgensen. Minimum spanning tree based clustering algorithms. pages 73–81, 2006.

[4] D. Xu and Y. Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2, 8 2015.

[5] Z. Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521:452–459, 2015.

[6] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P.S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32:4–24, 2021.

[7] E. Kosman, J. Oren, and D. Castro. *LSP : Acceleration and Regularization of Graph Neural Networks via Locality Sensitive Pruning of Graphs.* 11 2021.

[8] M. Gagolewski, M. Bartoszuk, and A. Cena. Are cluster validity measures (in)valid? *Information Sciences*, 581:620–636, 2021.