UDK 543.087.9

AUTOMATED MACHINE LEARNING AND MACHINE LEARNING PIPELINES FOR CHEMICAL SYSTEMS ANALYSIS

Aliev T.A. (ITMO University)
Scientific supervisor – Prof. Dr. Skorb E.V
(ITMO University)

Machine leaning (ML) is a great tool for chemical systems analysis, e.g. numerical data [1]. However, there is a challenge to choose ML algorithm [2]. Data scientists try to use advanced models or combine them basing on a previous research. This approach cannot be applied for novel systems' analysis, and new instruments are required.

Automated machine learning (AutoML) helps to choose the effective ML algorithm according to the stated task and optimize hyperparameters automatically [3]. AutoML also provides the opportunity to use several models step by step to get better prediction. Several AutoML frameworks were varied for electrochemical data analysis to solve multiclassification and linear regression task. AutoML helps to choose algorithms that are preliminary effective for the stated task and open the way for further hyperparameters' optimization by a user.

LightAutoML called Lama was applied for linear regression task. FEDOT was used to solve multiclassification task. API Fedot was used to make ML pipeline instead of AutoML to increase the accuracy. This approach helps to combine ML pipeline with different specific algorithms according to the user's decision and automatically optimizes hyperparameters.

References:

1. A. Keith J. et al. Combining Machine Learning and Computational Chemistry for Predictive Insights into Chemical Systems // Chem Rev. American Chemical Society, 2021. Vol. 121, № 16. P. 9816–9872.
2. Lickert H. et al. Selection of Suitable Machine Learning Algorithms for Classification Tasks in Reverse Logistics // Procedia CIRP. Elsevier, 2021. Vol. 96. P. 272–277.
3. Xin D. et al. Whither AutoML? Understanding the Role of Automation in Machine Learning Workflows // Conference on Human Factors in Computing Systems - Proceedings. Association for Computing Machinery, 2021.