Adapting explainable artificial intelligence to support drug design

This dissertation explores the application of explainability methods in computer-aided drug design, focusing on two use cases.

First, we investigate how atomic features used in representation impact the performance of graph neural networks. Our experiments involving Graph Convolutional Neural Networks trained using twelve representations confirm that the choice of atomic features results in an improved or reduced performance and that the optimal set of features is task-specific. We extend our analysis to more sophisticated graph models and show that the choice of atomic features remains important, although less so when rich bond information is available. Furthermore, we apply explainability methods to investigate how selected atomic features are used by Graph Convolutional Neural Networks and reveal a correlation between feature values and their importance scores.

In the second use case, we propose Explanation-driven optimisation (EDO), a transparent method to optimise drug candidates. EDO relies on explanations calculated by SHAP and derives a set of rules which guide the optimisation of molecules. We evaluate EDO on the task of increasing metabolic stability of drug candidates and demonstrate that it achieves promising results even when only a few modifications of the input molecule are allowed. Additionally, we show how rules derived by EDO can be analysed to deepen our understanding of metabolic processes. To the best of our knowledge, we are the first to demonstrate that explanations can be used for optimising molecular properties.

Our contributions include the preparation of datasets for predicting metabolic stability, the application of explainability methods to uncover the impact of atomic features used in the representation on the performance of graph neural networks, demonstration that explanations can be used for optimising molecular properties by using a method we proposed, a thorough evaluation of this method, and finally, presenting how rules generated by this method can be used for scientific discovery.

Agnieszka Wójcik