DOI 10.18699/SBB-2022-199

A Bayesian approach to lipophilicity prediction

Pyakillya B.I.*, Goncharov V.I.*
Tomsk Polytechnic University, School of Computer Science & Robotics, Tomsk, Russia
** morphism@tpu.ru*

Key words: machine learning, Bayesian theorem, lipophilicity, drug design

Motivation and Aim: It is well-known that lipophilicity affects the absorption dynamics of drugs and chemical substances in total. Accurate prediction of lipophilicity is a key problem in drug discovery and other related biochemical or physicochemical problems [1]. Current rise in popularity of machine learning shed some light on this problem and gave state of the art models like graph neural networks or transformer-based. Unfortunately, most machine learning methods cannot afford full interpretation of its predictions to research community e.g., like models' uncertainty – which tends to be unknown. Overfitting is still a big issue in machine learning tasks, and it also must be taken into account based on models' uncertainty. Therefore, we proposed to use Bayesian approach to overcome this problem and consider the uncertainty of model's behavior [2].
Methods and Algorithms: We utilized the Bayesian approach in the neural networks design and used it in the problem of lipophilicity prediction. We used PubChem library as the training dataset and verified the Bayesian neural networks results in comparison to conventional multi-layer perceptrons. We proposed to use model's uncertainty as a standard deviation, which could be easily calculated in Bayesian sampling process.
Results: The connection between prediction error and uncertainty values were found, which gives a lot of information to computational chemistry groups and could be used as a guide to make the training process more valuable. In addition, complexity of Bayesian neural networks was found to be significantly lower than in the case of conventional neural networks. This complexity benefit comes without drop in prediction accuracy.
Conclusion: The results obtained suggest to use the Bayesian approach in the cheminformatics problems like property prediction tasks. The approach gives the ability to consider model's uncertainty and take into account how much effort a cheminformatics researcher should put to make the prediction accuracy higher. Moreover, lower amount of training parameters makes the training of Bayesian Neural networks more favorable in comparison to conventional models, which tend to overfit to the training data.

References

1. Arnott J.A. et al The influence of lipophilicity in drug discovery and design. *Expert Opin Drug Discov.* 2012;10:863-875.
2. Kendall A. et al. What uncertainties do we need in Bayesian deep learning for computer vision? In: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017 Long Beach, California, USA. 2017:5580-5590.