Литература

Литература Meryam Zeryouh, Mohamed El Marraki, and Mohamed Essalih. Some tools of qsar/qspr and drug development: Wiener and terminal wiener indices. In *Proceedings of 2015 International Conference on Cloud Computing Technologies and Applications (CloudTechвъ™15)*, March 2015.

[2] Meryam et. al., Some tools of QSAR/QSPR and drug development: Wiener and Terminal Wiener indices, 2015 International Conference on Cloud Technologies and Applications, 2015.

Spherical CNN for QSAR prediction*

Dorokhin S.[1]***, Popova M.***[2]
[1]MIPT, [2]University of North Carolina

Abstract: The task of predicting molecular properties e.g. biological activity or solubility based on the atomic structure is called QSAR (quantitative structure-activity relationship) prediction. It is a classical problem in drug design. Despite various algorithms (quantile regression, radial basis function neural networks) are an acceptable solution, there is still a need for more precise models. A model originally developed for 3D shapes recognition was chosen and put under carefull examination in context of QSAR forecasting. This model is Spherical CNN, first suggested by Taco S. Cohen et. al., who managed to demonstrate that this NN performs well in several applications, including atomization energy prediction. The implemented model is compared with common CNN, RNN, graph CNN and Random Forest.

Keywords: *QSAR prediction, Spherical CNN, drug design*

Introduction

The idea of QSAR (Qualitive Structure Activity Realtionships) is to associate 2D or 3D structural representaion of a molecule with its biological or chemical properties. This research is aimed at building an accurate QSAR prediction tool. There were several attempts to solve the problem. Meryam Zeryouh et. al. [2] used graph representaion and suggested formulas for calculation of Wiener indices of complicated graphs. Wiener indices do correlate with such properties as critical point (ref) or viscosity (ref), but there is still no distinct relation to solubility or target activity, which are of extreme importance in drug design. Nupur S Munjal et. al.(ref) performed a non-linear multi-colinearity regression analysis to build model which predicts paclitaxel solubility. However, their model is designed for a specific compound and thus lacks universality. Fatima Adilova and Alisher Ikramov (ref) analyzed Matched Molecular Pairs (MMP) method in context of QSAR prediction. They managed to demonstrate that such an approach is inappropiate for QSAR modelling.

The method suggested in this article is based on Spherical Convolution Neural Networks. This concept was introduced by Taco S. Cohen et. al. (ref), who defined the correlation of two signals in SO(3) rotation group and the generalized convolution. The unique feature of the CNN suggested in their article is that the abovementioned convolution allows to create a distortion-free projection of a spherical signal. Taco et. al. tested the CNN in various tasks, including prediction of atomization energies from molecular geometry. The model yielded excellent results and this makes applying it to QSAR prediction an interesting challenge.

The main drawback of the suggested model is its complexity resulting in significant number of

---

parameters (around 1M) and huge memory and time resources required. However, the resulting model is expcted to be a universal solution. It is compared with conventional CNN, RNN, graph CNN and Random Forest.