

Сферические свёрточные нейронные сети для QSAR предсказаний*

Вареник Н.В., Попова М.С., Стрижов В.В
varenik.nv@phystech.edu

Задача прогнозирования молекулярных свойств, например, биологической активности или растворимости на основе атомной структуры называется QSAR (Qualitative Structure Activity Relationships) предсказание. Это классическая задача в области разработки лекарств. Несмотря на то, что множество алгоритмов, таких как квантильная регрессия, нейронные сети на основе радиально-базисных функций являются приемлемыми решениями, все еще есть необходимость в более точной модели. В работе была выбрана модель сферических свёрточных нейронных сетей, изначально предложенная Taso S. Cohen et. al. для распознавания 3D-форм и положена под тщательное изучение в контексте QSAR предсказаний. Результаты модели сравниваются с результатами более общих моделей, таких как свёрточные нейронные сети, рекуррентные нейронные сети, свёрточные нейронные сети с подающим на вход графом и случайный лес.

Ключевые слова: *QSAR предсказание, сферические свёрточные нейронные сети, разработка лекарств.*

1 Введение

Идея QSAR (Qualitative Structure Activity Relationships) заключается в том, чтобы связать 2D или 3D структурное представление молекулы с её биологическими или химическими свойствами. Эта задача очень важна в сфере разработки лекарственных препаратов. Цель нашего исследования - создать точный инструмент прогнозирования QSAR. Было несколько попыток решить эту задачу. Изначально, было предложено использовать графическое представление молекулы для вычисления индекса Винера и терминального индекса Винера [6], которые коррелируют с такими понятиями как критическая точка [4], вязкость [3], но они не имеют четкой связи с растворимостью или активностью, которые особо важны в разработке лекарств. Машинное обучение, как развивающаяся наука дает возможность используя ее различные методы, такие как случайный лес, квантильная и самосогласованная регрессии, нейронные сети постепенно улучшать качество прогнозирования в различных отраслях задачи нахождения QSAR. Также рассматривался вопрос рационального деления выборки на обучающую и тестовую [2]. Был сделан вывод, что оптимальный размер обучающей и тестовой выборки следует устанавливать на основе конкретного набора данных и типа используемых дескрипторов. Стоит отметить модель нейронных сетей, предложенную в 2014 году и активно используемую в наши дни, так как она дает довольно неплохие результаты [5], в основе которой лежат радиальные базисные функции и самосогласованная регрессия. Метод, предложенный в данной статье основан на сферических свёрточных нейронных сетях [1]. Они обладают уникальной особенностью, такой как возможность проектирования сферического сигнала без искажений. Их разработчик Taso et. al. тестировал сферические свёртки в различных задачах, в том числе и для предсказания энергии распыления из молекулярной геометрии. Модель дала отличные результаты, поэтому возник интерес в её применении к задаче QSAR предсказаний.

*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В.В. Задачу поставил: Эксперт И.О. Консультант: Консультант И.О.

Основным недостатком предлагаемой модели является её сложность, связанная с большим числом ее параметров. Однако, ожидается, что данная модель станет универсальным решением нашей задачи. Результаты модели сравниваются с обычными моделями, такими как CNN, RNN, CNN с подающимся на вход графом и RF на данных, взятых из Benchmark Datasets.

Литература

- [1] Taco Cohen, Mario Geiger, Jonas Koehler, and Max Welling. Spherical cnns. In *International Conference on Learning Representations*, March 2018.
- [2] Alexander Golbraikh and Alexander Tropsha. Predictive qsar modeling based on diversity sampling of experimental datasets for the training and test set selection. *Journal of Computer-Aided Molecular Design*, 16(5-6):357–369, 2002.
- [3] D. H. Rouvray and B. C. Crafford. The dependence of physical-chemical properties on topological factors. *South African Journal of Science*, 72:47, September 1976.
- [4] Leonard I. Stiel and George Thodos. The normal boiling points and critical constants of saturated aliphatic hydrocarbons. *AIChE Journal*, 8:527–529, September 1962.
- [5] Alexey V. Zakharov, Megan L. Peach, Markus Sitzmann, and Marc C. Nicklaus. A new approach to radial basis function approximation and its application to qsar. *Journal of Chemical Information and Modeling*, 54(3):713–719, 2014.
- [6] 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 (CloudTech?15)*, March 2015.