

Оценка качества предсказания ИК-спектров полициклических ароматических углеводородов с использованием нейронных сетей

Б.Г. Бегларян, А.С. Закускин, Т.А. Лабутин

Химический факультет МГУ имени М. В. Ломоносова, 119991 Москва, Ленинские горы, 1с3

E-mail: babken_beglaryan@laser.chem.msu.ru

Полициклические ароматические углеводороды (ПАУ) играют ключевую роль в эволюции межзвездной среды. Поэтому для астрохимических исследований, основанных на наблюдательных данных, особый интерес представляет моделирование ИК-спектров этих молекул, при этом необходимо получение спектров для очень большого числа соединений. Традиционно используемая для таких задач теория функционала плотности (DFT) исключительно ресурсоемка, особенно для ПАУ большого размера. Это послужило основанием для применения нейронных сетей для ускорения предсказания ИК-спектров ПАУ. Целью данной работы являлась оценка погрешности предсказания ИК-спектров астрохимически значимых молекул ПАУ нейронными сетями. Оценка была произведена для данных из работы [1]. Также была осуществлена попытка повышения качества предсказания спектров ПАУ за счет рассмотрения улучшения структуры обучающего набора, а также оптимизации функции потерь и способов представления молекул.

Построение и обучение нейронной сети для предсказания ИК-спектров ПАУ проводилось в среде Python. Кодировка молекул была воспроизведена с помощью встроенной функции библиотеки RDKit, разбивающей молекулы на фрагменты согласно алгоритму Моргана. Нами была выполнена оценка качества предсказания ИК-спектров с применением различных способов кодировки. Среди рассмотренных функций потерь были как стандартные (RMSE), так и более специфичные для спектральных данных, например, EMD (Earth Movers's Distance). Исходный обучающий набор был дополнен молекулами из баз данных, представленных в работе [2]. При этом большее число этих молекул не относится к классу ПАУ. Значительного улучшения качества предсказания можно добиться путем селекции ароматических молекул с функциональными группами. Оценка погрешности качества предсказания показала, что нейронные сети успешно применимы для решения задачи моделирования ИК-спектров ПАУ. В качестве потенциального способа улучшения предсказания спектров стоит также рассмотреть возможность применения графовых нейронных сетей на стадии кодирования молекул.

Список литературы

[1] P. Kovács et al., *The Astrophysical Journal*, **2020**, 902(2).

[2] C. McGill et al., *Journal of Chemical Information and Modeling*. **2021**, 61(6), 2594-2609.