**Прогнозирование растворимости органических соединений в различных растворителях методами глубокого обучения**

*Сидорова Э. С.*

*Студент, 3 курс специалитета*

*Московский государственный университет имени М.В. Ломоносова,   
химический факультет, Москва, Россия*

*E-mail: elina\_sidorova*[*@internet.ru*](mailto:ivanov@yandex.ru)

Предсказание растворимости химических соединений является одной из ключевых задач в химической науке и промышленности, так как помогает снизить затраты на экспериментальные исследования, оптимизировать химические процессы [1]. Нейросеть, предложенная в данной работе, позволяет с точностью, превосходящей существующие модели, прогнозировать растворимость исходя из структурной формулы соединения и небольшого набора экспериментальных данных о растворителе [2].

В работе использовались данные о десятичном логарифме растворимости 130 молекул в 26 растворителях из базы данных BigSolDB [3]. В качестве нейросети была использована архитектура, которая содержит в себе два входа: графовая свертка использовалась для представления самой молекулы, и линейный вход был выбран для представления растворителя. Было проведено исследование использования различных теоретических и экспериментальных дескрипторов как входных данных для линейной части. В качестве теоретических дескрипторов выступали комбинации структурных дескрипторов из модуля rdkit [4], MACCSKeys [5], предсказанные другими моделями энергии Гиббса сольватации, физико-химические характеристики.

Лучшие результаты показали модели, где растворитель описывался с помощью вектора из восьми чисел, которые равны молярной массе, показателю преломления, температуре плавления, кипения и некоторым другим экспериментальным характеристикам, а также набором из MACCSKeys ключей. Результаты представлены в Табл.1.

Таблица 1. Метрики качества работы нейросети GCNN\_FCNN на тестовой выборке

|  |  |  |
| --- | --- | --- |
| Метрика качества | Экспериментальные дескрипторы | Экспериментальные дескрипторы + MACCSKeys |
| R2 | 0.941 | 0.944 |
| Среднеквадратичная ошибка | 0.565 | 0.548 |
| Средняя абсолютная ошибка | 0.377 | 0.343 |

**Литература**

1. Karthikeyan A., Priyakumar, U. Artificial intelligence: machine learning for chemical sciences // J. Chem. Sci. 2022. 134 (1)

2. Yu, J., Zhang, C., Cheng, Y. SolvBERT for solvation free energy and solubility prediction: a demonstration of an NLP model for predicting the properties of molecular complexes // Digital Discovery. 2023. V. 2. № 2. P. 409-421

3. Krasnov L., Mikhaylov S., Fedorov M.V., Sosnin S. BigSolDB: solubility dataset of compounds in organic solvents and water in a wide range of temperatures // Zenodo. 2022.

4. RDKit: Open-source cheminformatics. <https://www.rdkit.org>

5. Durant, J. L., Leland, B. A., Henry, D. R., Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. // Journal of chemical information and computer sciences. 2002. V. 42. № 6. P. 1273-1280