

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОЛУЧЕНИЯ И ПРЕДСКАЗАНИЯ ДАННЫХ САМОДИФФУЗИИ ЧИСТЫХ ВЕЩЕСТВ И СРАВНЕНИЕ С РЕЗУЛЬТАТАМИ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ.

Бараблина Ю.Е. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Новотельнова А.В. (ИТМО)

Введение. Изучение коэффициента самодиффузии в жидкостях является важной задачей в химической физике, материаловедении и молекулярном моделировании. Традиционные методы молекулярной динамики требуют значительных вычислительных ресурсов и времени, что делает их применение для анализа больших объемов данных затруднительным. В связи с этим актуальным становится использование методов машинного обучения и нейронных сетей, которые позволяют быстро и эффективно предсказывать коэффициенты самодиффузии на основе имеющихся данных.

Использование машинного обучения для решения таких задач имеет ряд преимуществ:

1. Скорость анализа: Методы машинного обучения позволяют обрабатывать большие объемы данных за короткое время [1].
2. Минимизация экспериментальных затрат: Предсказание свойств жидкостей возможно без проведения дорогостоящих экспериментов [2].
3. Гибкость моделей: Нейронные сети и методы машинного обучения могут быть адаптированы для работы с различными типами данных и задач [3].

Целью данной работы является сравнение эффективности методов машинного обучения (Random Forest, Gradient Boosting) и нейронных сетей для предсказания коэффициента самодиффузии на основе данных молекулярного моделирования.

Основная часть. Для решения задачи предсказания коэффициента самодиффузии были использованы следующие методы:

1. Random Forest: Ансамблевый метод, основанный на построении множества решающих деревьев.
2. Gradient Boosting: Метод, который строит последовательность моделей, каждая из которых корректирует ошибки предыдущей.
3. Нейронная сеть: Многослойная модель, способная обучаться сложным нелинейным зависимостям.

Параметры модели Random Forest: количество деревьев – 100, среднеквадратичная ошибка – 0.046, коэффициент детерминации – 0.905; Gradient Boosting: количество деревьев – 100, среднеквадратичная ошибка – 0.047, коэффициент детерминации – 0.904; Нейронная сеть: архитектура: два скрытых слоя (64 и 32 нейрона), среднеквадратичная ошибка – 0.674, коэффициент детерминации – (-0.385)

Выводы. На рисунке 1 представлено сравнение предсказанных значений коэффициента самодиффузии с данными, полученными в результате молекулярного моделирования. Видно, что методы Random Forest и Gradient Boosting демонстрируют высокую точность ($R^2 > 0.9$), в то время как нейронная сеть показала низкую эффективность ($R^2 < 0$).

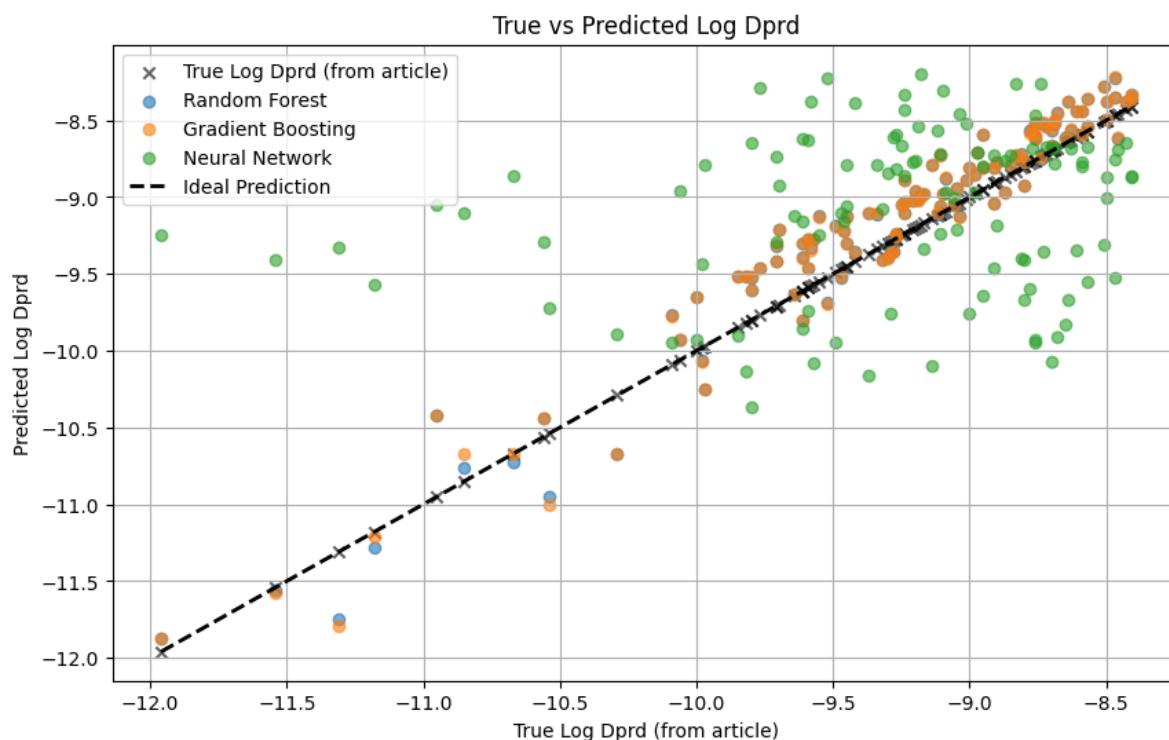


Рисунок 1. Сравнение предсказанных значений коэффициента самодиффузии с данными молекулярного моделирования.

Список использованных источников:

1. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, 559(7715), 547-555. DOI:10.1038/s41586-018-0337-2
2. Schütt, K. T., Gastegger, M., Tkatchenko, A., & Müller, K.-R. (2019). Quantum-chemical insights from deep tensor neural networks. *Nature Communications*, 10(1), 1-10. DOI:10.1038/s41467-019-12875-2
3. Noé, F., Tkatchenko, A., Müller, K.-R., & Clementi, C. (2020). Machine learning for molecular simulation. *Annual Review of Physical Chemistry*, 71, 361-390. DOI:10.1146/annurev-physchem-042018-052331