

# МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ КВАНТОВОЙ ХИМИИ

*Б.И. Пякиля, ст.преподаватель  
Томский политехнический университет  
E-mail: [morphism@tpu.ru](mailto:morphism@tpu.ru)*

## Введение

В настоящее время атомистическое моделирование стало необходимым инструментом в задачах оптимизации взаимодействия лиганд-белок для создания новых лекарственных средств и поиска новых материалов для энергетики. Лежащие в основе моделирования методы квантовой механики, позволяют достаточно точно вычислять свойства органических и неорганических соединений. Однако, обычные, традиционные методы квантовой механики достаточно сложны для использования в связи со своей вычислительной сложностью и необходимые расчеты для промышленности могут занимать часы, дни, недели, что значительно сдерживает технологических прогресс.

Решением является использование методов машинного обучения, основывающихся на описании системы в терминах квантовой механики. Данные методы значительно ускоряют вычисление промышленно значимых свойств различных соединений и в то же время повышают качество оценки этих свойств по сравнению с существующими методами. В данной работе будут описаны преимущества методов машинного обучения в атомистическом моделировании для исследовательских задачах промышленности.

## Описание алгоритма

Процесс атомистического моделирования представляет собой вычислительный процесс оценки свойств веществ *ab initio*, т.е. без проведения экспериментов. Атомистическое моделирование основывается на использовании либо (полу-)эмпирических межатомных потенциалов, либо квантовомеханических моделей межатомного взаимодействия. Модели первой группы вычислительно эффективны, но получаемая ими точность иногда недостаточна. Модели же второй группы позволяют получить высокую точность, однако являются крайне затратными с точки зрения вычислительной сложности.

В основе методов второй группы лежит фундаментальная идея теории функционала плотности (DFT) [1]. Метод Кона-Шэма это стандартный метод теоретических исследований электронной структуры материалов. В этой теории, решение уравнения Кона-Шэма

$$\left[ -\frac{\nabla^2}{2} + V_{ion}(r) + \int dr' \frac{n(r')}{|r-r'|} + V_{xc}[n](r) \right] \varphi_i(r) = \varepsilon_i \varphi_i(r),$$

с плотностью  $n(r)$ , вычисляемое суммированием  $|\varphi_i(r)|^2$  для всех занятых состояний, позволяет получить полную энергию системы и распределение плотности взаимодействующей электронной системы под ионным потенциалом  $V_{ion}$ . Обменно-корреляционный потенциал  $V_{xc}[n]$  является функционалом плотности. Хорошо известны материалы, для которых точное DFT-описание еще не получено. Также некоторые современные функционалы критикуемы за смещение в сторону точного описания энергии, а не плотности, несмотря на то, что оба явления значимы.

Далее рассмотрим применение методов машинного обучения в задаче предсказания обменно-корреляционных потенциалов  $V_{xc}[n]$ .

Первые работы по применению методов машинного обучения к функционалам плотности проводились группой Burke [2, 3, 4], но в данной работе мы можем представить результаты [1], где коллектив авторов предлагает использовать нейросетевой алгоритм для предсказания  $V_{xc}$ . Для построения архитектуры нейросети, изображенной на рисунке 1, авторы использовали многослойный перспетрон, который имеет на входе признаки, описывающие различные небольшие молекулы (вода, аммиак, оксид азота). Данные признаки вычисляются на основе плотности  $n(r)$  и их полное математическое приводится в работе [1]. Используя данный метод машинного обучения, но не ограничиваясь только им, можно получить значительное ускорение проводимых вычислений, что крайне важно для проведения большего количества вычислительных экспериментов без потери качества предсказания. Однако важно помнить, что правильно подобранные экспертом или другим алгоритмом признаки, а также соответствующее разнообразие обучающих данных являются важнейшими элементами процесса обучения алгоритма.

Алгоритмы машинного обучения имеют отличную возможность поиска законов, лежащих в основе предоставленных данных, но имеют проблемы, связанные с переобучением, что выражается в

некачественном обобщении в случае новых данных, которые могут отличаться видом молекулы, что выражается как и в структуре, так и в конформации молекулы. Одним из решением данной проблемы является такая инженерия входных признаков и такое изменение архитектуры нейросети, которые смогли бы учитывать различные физические принципы (симметрия и пр.), лежащие в основе вычисления потенциалов молекул.

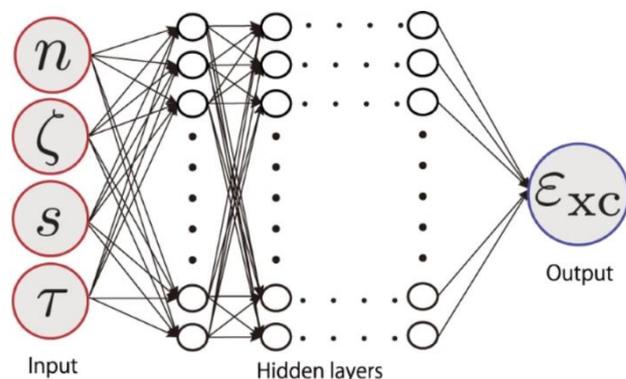


Рис. 1. Архитектура нейросети.

### Заключение

В работе описан пример использования метода машинного обучения для решения задачи квантовой химии, а именно метод, основанный на использовании нейросетевого алгоритма. В работе не упомянуты конкретные вычислительные результаты, сравнения с существующими решениями, а также дальнейшие пути развития алгоритма. Указанные недостатки будут устранены в последующих работах на данную тему.

### Список использованных источников

1. Nagai R. Completing density functional theory by machine learning hidden messages from molecules. *Comput Mater.* – 2020. – vol. 43.
2. Snyder J. C. Finding density functionals with machine learning. *Phys. Rev. Lett.* – 2012. – vol. 108.
3. Snyder J. C. Orbital-free bond breaking via machine learning. *Chem. Phys.* – 2013. – vol. 139.
4. Li L. Understanding machine-learned density functionals. *Int. Quantum Chem.* – 2016. – vol. 116. –P. 819-833.