

# ПРЕДСКАЗАНИЕ ЭНЕРГИИ СВЯЗЫВАНИЯ БЕЛКОВЫХ КОМПЛЕКСОВ С ПОМОЩЬЮ СВЁРТОЧНЫХ СЕТЕЙ

А. С. Рак

*Белорусский государственный университет, Минск;*

*kventinel@gmail.com;*

*науч. рук. – А. Ю. Хадарович, мл. науч. сотр. ГНУ «ОИПИ НАН Беларусь»*

Разработан алгоритм предсказания энергии связывания белковых комплексов с помощью свёрточных нейронных сетей. Алгоритм был протестирован на наборе белковых комплексов из базы данных. Данный алгоритм может использоваться в качестве одного из этапов разработки лекарственных средств.

**Ключевые слова:** свёрточные нейронные сети; белковые комплексы; энергия связывания.

Вероятность успеха на начальной фазе исследования лекарственных средств зависит от предсказаний или измерений близости кандидата (лиганда) и цели (белка). Пространство синтетически доступных лигандов очень велико, поэтому на данный момент невозможно исследовать всё это пространство. Задача состоит в тестировании как можно большего числа лигандов.

Внедрение моделей машинного обучения в качестве одного из этапов разработки лекарственных средств позволило улучшить виртуальный скрининг лекарств, а также другие физически обоснованные оценки малых молекул. Однако такие модели имеют фундаментальный недостаток: молекулы нужно представлять как вектор признаков фиксированных размеров.

Для исправления этого недостатка можно использовать нейронные сети. Гибкость глубоких нейронных сетей позволяет моделям получать признаки более высоких порядков из простого представления данных. Вместо выбора фиксированного параметра для образования признаков в атомных векторах, эти параметры оптимизируются одновременно с нейронной сетью, что позволяет модели принимать самостоятельные решения о признаках, важных для предсказания связывания лиганда.

## Атомарная свёрточная нейронная сеть (АСНС)

В данной сети используются две свёрточные операции: атомная свёртка и радиальный пулинг. Атомная свёртка использует матрицу расстояний до  $M$  ближайших соседей для извлечений признаков некоторой окрестности атома из данного представления молекулы.

## Конструкция матрицы расстояний и списка соседей

Матрица расстояний  $R$  и матрица атомных номеров  $Z$  строятся из атомных координат. На данном этапе используется лишь  $M$  ближайших соседей для того, что бы сложность этого этапа была  $O(NM)$ , а не  $O(N^2)$ . В данной работе используется  $M = 12$ . Таким образом, из начальной матрицы размера  $(N, 3)$  получается матрица расстояний размера  $(N, M)$ . Также используются типы ближайших атомов, для них матрица  $Z$  имеет размер  $(N, M)$ .

### Атомная свёртка.

Выход атомной свёртки конструируется из матрицы расстояний  $R$  и матрицы атомных типов  $Z$ . К матрице  $R$  применяется фильтр размера  $(1 \times 1)$  с шагом 1 и глубиной  $N_{at}$ , где  $N_{at}$  – это число уникальных атомных типов присутствующих в молекулярной системе. Ядром атомной свёрточной операции является функция:

$$(K * R)_{i,j}^a = R_{i,j} K_{i,j}^a$$

где

$$K_{i,j}^a = \begin{cases} 1, & Z_{i,j} = N_a \\ 0, & \text{иначе} \end{cases}$$

где  $N_a$  – это атомный номер из  $a = \{1, \dots, N_{at}\}$ . Атомарный конволюционный слой применяется к матрице расстояний до соседей и создаёт матрицу  $E$  размера  $(N, M, N_{at})$ .

### Слой радиального пулинга

Радиальный пулинг – это процесс, уменьшающий размер матрицы признаков, который принимает на вход тензор размера  $(N, M, N_{at})$ . Уменьшение размерности используется для предотвращения переобучения. В дополнение к этим свойствам после применения радиального пулинга получается тензор, который инвариантен к перестановкам соседей у атомов.

На входе этого слоя получается тензор размера  $(N, M, N_{at})$ . Затем выбирается число  $N_r$  – число желаемых радиальных фильтров. Каждый радиальный фильтр объединяет часть тензора  $(1 \times M \times 1)$ , т.е. после применения всех радиальных фильтров получается тензор размера  $(N, N_{at}, N_r)$ . Каждый радиальный фильтр применяет следующие операции:

$$f_s(r_{i,j}) = \exp \frac{-(r_{i,j} - r_s)^2}{\sigma_s^2} f_c(r_{i,j})$$

$$f_c(r_{i,j}) = \begin{cases} \frac{1}{2} \cos\left(\frac{\pi r_{i,j}}{R_c}\right), & 0 < r_{i,j} < R_c \\ 0, & r_{i,j} \geq R_c \end{cases}$$

$$P_{i,n_a,n_r} = \beta_{n_r} \sum_{j=1}^M f_{n_r}(E_{i,j,n_a}) + b_{n_r}$$

где  $r_s$  и  $\sigma_s$  – это обучаемые параметры, а  $\beta$  и  $b$  – это константы.

## Полносвязный слой в атомарной конволовиционной нейронной сети

После получения из пулингового слоя тензора  $P$  размера  $(N, N_{at}, N_r)$  его координаты выравниваются для каждого атома, переводя его в матрицу размера  $(N, N_{at} * N_r)$ . На следующем шаге на вход полносвязного слоя передаются по отдельности векторы, соответствующие атомам обрабатываемой молекулы размера  $N_{at} * N_r$ . После этого результат суммируется по всем атомам.

## Обучение

Структура АСНС полностью дифференцируема по отношению к положению атомов. Энергия связывания комплекса вычисляется по следующей формуле:

$$\Delta G_{complex} = G_{complex} - G_{protein} - G_{ligand}$$

При обучении используется следующая функция ошибки:

$$L = (\Delta G_{complex} - \gamma_{complex})^2,$$

где  $\gamma_{complex}$  – это логарифм  $K_i$  (энзимная ингибиторная константа).

## Результаты

Для оценки полученных результатов используется коэффициент детерминации Пирсона:

$$R^2 = 1 - \frac{\delta^2}{\delta_y^2} = 1 - \frac{SS_{res}/n}{SS_{tot}/n} = 1 - \frac{SS_{res}}{SS_{tot}}$$

где  $SS_{res} = \sum_{i=1}^n (y_i - \bar{y}_i)^2$  – сумма квадратов остатков регрессии,  $y_i, \bar{y}_i$  – фактические и расчётные значения объясняемой переменной соответственно.

$SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$  и  $\bar{y}$  – среднее значение фактической переменной.

При использовании АЧС можно получить коэффициент детерминации Пирсона  $R^2 \approx 0.4$ .

## Вывод

В данной работе используется общая конволовационная операция в трёхмерном пространстве для изучения химических взаимодействий на атомном уровне, которая принимает в качестве входных данных атомные координаты в трёхмерном пространстве, и демонстрируется применение в структурном биоактивном предсказании. АЧС обучается предсказывать энергию связывания белково-лигандных комплексов через прямой подсчёт энергии, ассоциированной с комплексом, белком и лигандом. Определяются нековалентные взаимодействия, присутствующие в комплексе и отсутствующие в белке или лиганде, и модель обрабатывает силы взаимодействия, связанные с этими параметрами. Однако существенным недостатком такой сети является то, что для предсказания энергии связывания требуется структура не только лиганда и белка, но также и их комплекса, а это сильно сказывается на применимости такой архитектуры, ибо зачастую сразу известны только структуры белка и лигандов, а получение хороших предсказаний структуры их комплекса является сложной задачей.

АЧС также могут быть применимы к ряду приложений, непосредственно не связанных с обнаружением лекарственных средств. Например, они могут улучшить существующие нейронные сети на основе потенциалов для автоматизированного подбора потенциальной энергии поверхности (силовые поля), которые имеют широкое применение в теоретической химии для ускорения начальной симуляции молекулярной динамики многочастичных систем.

## Библиографические ссылки

1. *Gomes J. [et al.] Atomic Convolutional Networks for Predicting Protein-Ligand Binding Affinity* // Cornell University Library, 2017, arXiv:1703.10603.
2. *Krizhevsky. A. [et al.] ImageNet Classification with Deep Convolutional Neural Network* // NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012, Vol. 1, P. 1097–1105.
3. *Hechtlinger Y. [et al.] A Generalization of Convolutional Neural Networks to Graph-Structured Data»* // Cornell University Library, 2017, arXiv:1704.08165.