

Оценка качества структуры белка с использованием графовых нейронных сетей.*

Северилов П.А.¹

severilov.pa@phystech.edu

¹Московский физико-технический институт (МФТИ)

Оценка качества белковой модели (ИЛИ предсказание структуры белковой модели) является важной и пока открытой проблемой в структурной биоинформатике (биологии). В работе исследуется применимость графовых нейронных сетей в связке со сверточными к данной задаче

Ключевые слова: *GCN, графовые нейросети*.

1 Введение

Белки присутствуют практически в каждом биологическом процессе. Понимание их структуры и выполняемых задач помогают контролировать эти процессы. Белки спонтанным образом принимают форму в различных средах — форма диктует функционал. Но из имеющихся последовательностей аминокислот в белке трудно определить, в какую форму произойдет сворачивание. Идентификация структуры занимает большое количество времени и ресурсов, к тому же, не всегда возможна.

Вычислительные методы, которые решают задачу предсказания структуры белка в основном состоят из двух этапов: генерация конформаций белка из их аминокислотных последовательностей и оценивание предсказания. В данной работе рассматривается только ПЕРВУЮ ИЛИ ВТОРУЮ задачу. Данная проблема является крайне важной. Это подтверждается тем, что каждые два года проводится соревнование Critical Assessment of protein Structure Prediction (CASP) по решению этой задачи.

До недавнего времени лучшими методами (для контроля качества одной модели ИЛИ предсказания структуры белка) считались объединение подходов, основанных на функциях, предназначенных для узкого класса белков. С развитием нейронных сетей и ростом количества известных белковых структур, методы глубинного обучения превзошли эти результаты.

Основные результаты в этой области полагаются на сверточные нейронные сети (CNN) [1]. Т.к. имеющиеся данные представляют собой трехмерные координаты атомов, то это побудило нас к использованию графовых нейронных сетей и добавлению их к уже имеющимся архитектурам.

2 Постановка задачи

генерация конформаций белка из их аминокислотных последовательностей ИЛИ оценивание предсказания???

3 Теоретическая часть

3.1 Слой свертки графа

Дан граф \mathbf{A} и матрица с информацией об узлах $\mathbf{X} \in \mathbb{R}^{n \times c}$. Тогда слой свертки графа может быть представлен в следующей форме:

*Научный руководитель: В.В. Стрижов

$$\mathbf{Z} = f\left(\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{A}}\mathbf{X}\mathbf{W}\right),$$

где $\tilde{\mathbf{A}} = \mathbf{G} + \mathbf{I}$ – матрица смежности графа с добавлением петель, $\tilde{\mathbf{D}}$ это его диагональная матрица степеней вершин, где $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$, $\mathbf{W} \in \mathbb{R}^{c \times c'}$ – матрица параметров свертки обучаемого графа, f – нелинейная функция активации, а $\mathbf{Z} \in \mathbb{R}^{n \times c'}$ – выходная матрица.

4 Представление белков в виде графов

Интуитивное представление белка с помощью графов может быть сформировано, рассматривая элементы аминокислотной последовательности как отдельные узлы, чьи связи (ребра) описывают пространственные отношения между ними.

В общем случае граф \mathbf{G} определяется набором (\mathbf{V}, \mathbf{A}) , где $\mathbf{V} \in \mathbb{R}^{n \times c}$ определяет вершины или узлы графа, n – число узлов и c – число признаков в каждом узле. Матрица смежности $\mathbf{A} \in \mathbb{R}^{n \times n}$ определяет соединения между n узлами (ребра), где \mathbf{A}_{ij} – сила связи между узлами i и j . Используя это определение графа, белковые структуры можно определить как графы, признаки элементов аминокислотной последовательности которых закодированы в элементах \mathbf{V} узлов, а пространственная близость между элементами закодирована в матрице смежности \mathbf{A} .

5 Вычислительный эксперимент

6 Результаты

7 Связанные работы

Литература

- [1] Guillaume Pagès, Benoit Charmettant, and Sergei Grudinin. Protein model quality assessment using 3D oriented convolutional neural networks. *Bioinformatics*, 35(18):3313–3319, 02 2019.
- [2] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. *CoRR*, abs/1901.00596, 2019.
- [3] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, and Maosong Sun. Graph neural networks: A review of methods and applications. *CoRR*, abs/1812.08434, 2018.