

Оценка качества прогнозирования структуры белка с использованием графовых свёрточных нейронных сетей*

Северилов П.А.¹

severilov.ra@phystech.edu

¹Московский физико-технический институт (МФТИ)

Решается задача оценки качества (QA – Quality Assessment) прогнозирования белковых структур. В работе показывается применимость к рассматриваемой задаче графовых свёрточных нейронных сетей, основанных на спектральной теории. Описание белковых структур представляется в виде графов. Спектральная теория графов определяет свёртку в нейронных сетях. Нейросеть в работе получает на вход матрицы координат атомов и матрицы смежности смоделированных белковых структур. Она предсказывает близость смоделированной и реальной структуры белка в виде CAD_{score}. Нейросеть обучается на наборах данных CASP7-CASP11 и тестируется на данных CASP12. На CASP12 достигается уровень ошибки MSE равный 0.051. Дополнительный анализ корреляционных коэффициентов Пирсона и Спирмена подтверждает применимость метода для различных белковых структур. Эксперименты в данной работе показывают новые направления в задаче QA.

Ключевые слова: белковые структуры, графы, графовые нейронные сети, свёрточные нейронные сети, спектральные свёртки.

1 Введение

Белки являются наиболее универсальными макромолекулами в живых системах и выполняют важнейшие функции практически во всех биологических процессах [1]. (?Понимание белковых структур и выполняемых задач помогают контролировать биологические процессы.?) Форма белковой структуры определяет выполняемые ей функции (?её функционал?) [1]. Но из имеющихся последовательностей аминокислот в белке трудно определить, в какую форму сворачивается структура. Идентификация структуры занимает большое количество времени и ресурсов, к тому же, не всегда возможна.

Каждые два года проводятся соревнования Critical Assessment of protein Structure Prediction (CASP [2]) по решению задачи предсказания структуры. Вычислительные методы, которые её решают состоят из двух этапов: генерация конформаций белка из их аминокислотных последовательностей и оценивание качества предсказания. В данной работе рассматривается только второй этап.

Белковая структура состоит из одной или нескольких цепочек более мелких молекул – аминокислотных остатков. Последовательность остатков $S = \{a_i\}_{i=1}^N$ представляет его первичную структуру, где a_i является одним из 22 типов аминокислот. Взаимодействия между соседними остатками и окружающей средой определяют, как цепочка будет сворачиваться в сложные структуры, которые представляют вторичную структуру и третичную структуру белка.

Поэтому для задач с участием белковых структур модель должна учитывать как пространственную информацию об атомах, третичную структуру, так и признаки в виде последовательностей аминокислот, первичную структуру белка. В работах [3, 4] для моделирования белков используются LSTM или 1D-CNN, которые представляют белки в виде

*Научный руководитель: В.В. Стрижов

последовательности с пространственными признаками. В работах [5, 6] моделируется пространственная структура белков с использованием 3D-CNN, но не учитывается структура последовательностей. На основе графов моделируются как последовательности, так и геометрические структуры белков.

В работе [7] графовые нейронные сети на основе алгоритма, описанного в [8], показывают результаты, превосходящие остальные современные методы. Основные результаты в этой области полагаются на рточные нейронные сети (CNN) [6].

Поэтому предлагается использование графовых свёрточных нейронных сетей.

2 Постановка задачи

Пусть $\mathcal{D} = (\mathbf{X}, \mathbf{y})$ – заданная выборка, где $\mathbf{X} \in \mathbb{R}^{m \times n \times 3}$ – тензор объект-признак, объекты $\mathbf{x}_i \in \mathbb{R}^{1 \times n_i \times 3}, i = \overline{1, m}$ – это молекулы, каждая из которых описана множеством 3-мерных координат всех ее атомов, а $\mathbf{y} = [y_1, \dots, y_m]^T \in \mathbb{R}^{m \times 1}$ – оценка близости предсказанной и реальной структуры белка. Оценка близости измеряется различными метриками: CAD_{score} [9], LDDT, GDT. В данной работе выбран CAD_{score}.

2.1 CAD score

Обозначим через G множество всех пар элементов последовательности аминокислот (остатков) (i, j) , имеющих ненулевую площадь контакта $T_{(i,j)}$ в реальной структуре. Затем для каждой пары остатков $(i, j) \in G$ вычисляется площадь контакта $M_{(i,j)}$ смоделированной структуры.

Для каждой пары остатков $(i, j) \in G$ определяется разность площадей контакта CAD_(i,j) как абсолютная разница площадей контакта между остатками i и j в реальной T и смоделированной структуре M :

$$\text{CAD}_{(i,j)} = |T_{(i,j)} - M_{(i,j)}|.$$

Для вычислительной стабильности берется ограниченный CAD: CAD_(i,j)^{bounded} = min (CAD_(i,j), T_(i,j)). Таким образом: CAD_{score} для всей структуры определяется как

$$\text{CAD}_{\text{score}} = 1 - \frac{\sum_{(i,j) \in G} \text{CAD}_{(i,j)}^{\text{bounded}}}{\sum_{(i,j) \in G} T_{(i,j)}}.$$

2.2 Задача регрессии

Рассматривается множество параметрических моделей \mathfrak{F} , взятых из класса графовых свёрточных нейронных сетей: $\mathfrak{F} = \{f_k: (\mathbf{w}, \mathbf{X}) \rightarrow \hat{\mathbf{y}} \mid k \in \mathfrak{K}\}$, где $\mathbf{w} \in \mathbb{W}$ – параметры модели, а $\hat{\mathbf{y}} \in \mathbb{R}^{m \times 1}$ – вектор оценок предсказаний CAD-scores.

Решается задача регрессии для предсказания численного значения CAD_{score} y_i белка на основе его смоделированной пространственной структуры \mathbf{x}_i .

Параметры модели $\mathbf{w} \in \mathbb{W}$ подбираются в соответствии с минимизацией функции ошибки на обучении. Определим функцию ошибки $\mathfrak{L}(\mathbf{y}, \mathbf{X}, \mathbf{w}) = (\hat{\mathbf{y}} - \mathbf{y})^2$, где $\hat{\mathbf{y}} = f(\mathbf{X}, \mathbf{w})$ – CAD_{score} предсказанный моделью f , \mathbf{y} – данный в выборке CAD_{score}.

Часто для оценивания модели используются коэффициенты корреляции Пирсона (R), Спирмена (ρ) [5–7]. Для каждой реальной структуры белка вычисляются коэффициенты корреляции Пирсона (R^{target}), Спирмена (ρ^{target}) между истинными и прогнозируемыми

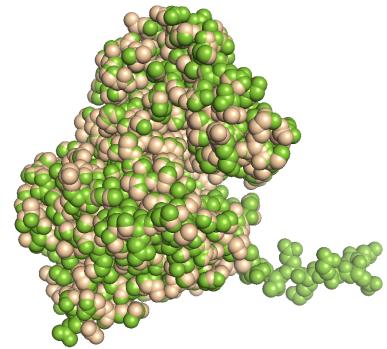


Рис. 1 Пересечение реальной структуры T0861 (жёлтый) и её модели Atome2_CBS_TS4 (зелёный) при CAD_{score} = 0.829

CAD_{score} для сгенерированных структур, соответствующих данной реальной структуре белка. Затем коэффициенты корреляции усредняются по всем реальным структурам (m штук). Обозначим \mathbf{y}_i и $\hat{\mathbf{y}}_i$ соответственно вектор истинных значений и вектор предсказаний CAD_{score} для сгенерированных структур белка, соответствующих реальной структуре i . Тогда коэффициенты корреляции записываются:

$$R = R(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{m} \sum_{i=1}^m R_i^{target} = \frac{1}{m} \sum_{i=1}^m \text{PEARSON}(\mathbf{y}_i, \hat{\mathbf{y}}_i)$$

$$\rho = \rho(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{m} \sum_{i=1}^m \rho_i^{target} = \frac{1}{m} \sum_{i=1}^m \text{SPEARMAN}(\mathbf{y}_i, \hat{\mathbf{y}}_i)$$

3 Спектральный анализ

Для обобщения свёрточных нейронных сетей на графы необходимо определить свёрточные фильтры на графах. Существует два известных подхода: пространственный и спектральный [10, 11]. Как показано в [12] пространственный подход не имеет общего математического определения трансляции на графах, в то время как спектральный метод имеет хорошее математическое обоснование. Поэтому рассматривается спектральная теория графов.

Элементы аминокислотной последовательности рассматриваются 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. 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} .

Определение 1. Графовый Лапласиан [13] – матрица $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$, где \mathbf{A} – матрица смежности графа \mathbf{G} , \mathbf{D} – диагональная матрица степеней вершин, $\mathbf{D}_{ii} = \sum_j (\mathbf{A}_{ij})$, \mathbf{I}_n – единичная матрица.

Матрица \mathbf{L} является вещественной симметричной положительной полуопределенной, поэтому может быть представлена в виде $\mathbf{L} = \mathbf{U} \Lambda \mathbf{U}^\top$, где $\mathbf{U} = [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{n-1}] \in \mathbb{R}^{n \times n}$ – это матрица собственных векторов, упорядоченных по собственным значениям, $\Lambda \in \mathbb{R}^{n \times n}$ – диагональная матрица собственных значений (спектр), $\Lambda_{ii} = \lambda_i$. Спектральное разложение Лапласиана позволяет определить преобразование Фурье для графов: собственные векторы соответствуют модам Фурье, а собственные значения – частотам.

Определение 2. Графовое преобразование Фурье [14] для сигнала $\mathbf{x} \in \mathbb{R}^n$ задается $\mathcal{F}(\mathbf{x}) = \mathbf{U}^\top \mathbf{x} \equiv \hat{\mathbf{x}} \in \mathbb{R}^n$, а обратное графовое преобразование Фурье: $\mathcal{F}^{-1}(\hat{\mathbf{x}}) = \mathbf{U} \hat{\mathbf{x}}$, где \mathbf{x} – вектор признаков всех вершин.

Данное преобразование является ключевым в определении графовой свёртки. Оно проецирует входной графовый сигнал на ортонормированное пространство, где базис формируется собственными векторами графового Лапласиана. Элементы преобразованного сигнала $\hat{\mathbf{x}}$ являются координатами сигнала в новом пространстве, так что входной сигнал может быть представлен как $\mathbf{x} = \sum_i \hat{x}_i \mathbf{u}_i$, что является обратным графовым преобразованием Фурье.

Теорема 1. (Теорема о свёртках) [15] Преобразование Фурье свёртки двух сигналов является покомпонентным произведением их преобразований Фурье, т.е.

$$\mathcal{F}(\mathbf{f} * \mathbf{g}) = \mathcal{F}(\mathbf{f}) \odot \mathcal{F}(\mathbf{g})$$

Следуя из теоремы 1, спектральная свёртка на графах определяется для сигнала \mathbf{x} и фильтра $\mathbf{g} \in \mathbb{R}^n$ как

$$\mathbf{x} * \mathbf{g} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{x}) \odot \mathcal{F}(\mathbf{g})) = \mathbf{U} (\mathbf{U}^\top \mathbf{x} \odot \mathbf{U}^\top \mathbf{g}) = \mathbf{U} \mathbf{g}_\theta \mathbf{U}^\top \mathbf{x}, \quad (1)$$

где $\mathbf{g}_\theta = \text{diag}(\mathbf{U}^\top \mathbf{g})$ – спектральные коэффициенты фильтра.

Спектральные методы отличаются выбором фильтра \mathbf{g}_θ . Соотношение 1 вычислиительно дорогое, т.к. спектральное разложение требует $O(n^3)$ операций, а перемножение с матрицей собственных векторов \mathbf{U} требует $O(n^2)$ операций. Chebyshev Spectral CNN (ChebNet) [16] обходит эти проблемы аппроксимацией \mathbf{g}_θ с помощью полиномов Чебышева $\mathbf{T}_k(\mathbf{x})$, убирая необходимость считать собственные векторы Лапласиана \mathbf{L} .

Определение 3. Полиномы Чебышева $\mathbf{T}_k(\mathbf{x})$ k -ого порядка задаются рекуррентным соотношением $\mathbf{T}_k(\mathbf{x}) = 2\mathbf{x} \cdot \mathbf{T}_{k-1}(\mathbf{x}) - \mathbf{T}_{k-2}(\mathbf{x})$, $\mathbf{T}_0(\mathbf{x}) = 1$, $\mathbf{T}_1(\mathbf{x}) = \mathbf{x}$. Образуют ортогональный базис в $L^2([-1, 1], \frac{dx}{\sqrt{1-x^2}})$

Представляя \mathbf{g}_θ в виде

$$\mathbf{g}_\theta = \sum_{k=0}^{K-1} \theta_k \mathbf{T}_k(\tilde{\Lambda}),$$

где $\tilde{\Lambda} = 2\Lambda/\lambda_{\max} - \mathbf{I}_n \in [-1, 1]$, λ_{\max} – максимальное собственное число \mathbf{L} , а также замечая, что

$$(\mathbf{U} \Lambda \mathbf{U}^\top)^k = \mathbf{U} \Lambda^k \mathbf{U}^\top$$

(собственные векторы образуют ортонормированный базис $\mathbf{U}^\top \mathbf{U} = \mathbf{I}$), получаем:

$$\mathbf{U} \mathbf{g}_\theta \mathbf{U}^\top \mathbf{x} = \mathbf{U} \left(\sum_{i=0}^{K-1} \theta_i \mathbf{T}_i(\tilde{\Lambda}) \right) \mathbf{U}^\top \mathbf{x} = \sum_{k=0}^{K-1} \theta_k \mathbf{T}_k(\tilde{\Lambda}) \mathbf{x}, \quad (2)$$

где $\tilde{\Lambda} = 2\Lambda/\lambda_{\max} - \mathbf{I}_n$.

Graph Convolutional Network (GCN) [17] используют первое приближение ChebNet. Предполагая $\lambda_{\max} \approx 2$ и $K = 1$, соотношение (2) упрощается до

$$\mathbf{x} * \mathbf{g} \approx \tilde{\theta}_0 \mathbf{x} + \tilde{\theta}_1 (\mathbf{L} - \mathbf{I}_n) \mathbf{x} = \tilde{\theta}_0 \mathbf{x} - \tilde{\theta}_1 \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{x}. \quad (3)$$

Приняв $\theta = \tilde{\theta}_0 = -\tilde{\theta}_1$, получаем:

$$\mathbf{x} * \mathbf{g} \approx \theta \left(\mathbf{I}_n + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \right) \mathbf{x}. \quad (4)$$

Оператор в скобках может привести к вычислительной нестабильности и взрыву или затуханию градиентов, т.к. собственные значения данного оператора $\in [0, 2]$. Для решения проблемы в [17] предлагается *трюк перенормировки*:

$$\mathbf{I}_n + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \rightarrow \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}, \text{ где } \tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n, \tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}.$$

Дан граф \mathbf{G} и матрица с информацией об узлах $\mathbf{X} \in \mathbb{R}^{n \times c}$ (n – число узлов и c – число признаков в каждом узле). Исходя из (4) и применяя трюк перенормировки, определяется слой свёртки графа таким образом:

$$\mathbf{Z} = \sigma \left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X} \mathbf{W} \right),$$

где $\mathbf{W} \in \mathbb{R}^{c \times t}$ – матрица параметров свёртки с t фильтрами, σ – нелинейная функция активации, а $\mathbf{Z} \in \mathbb{R}^{n \times t}$ – выходная матрица.

3.1 Архитектура сети

Основа архитектуры берется из существующей модели GCN [17], где изменяется выходной слой для работы с задачей регрессии и параметры свёрток.

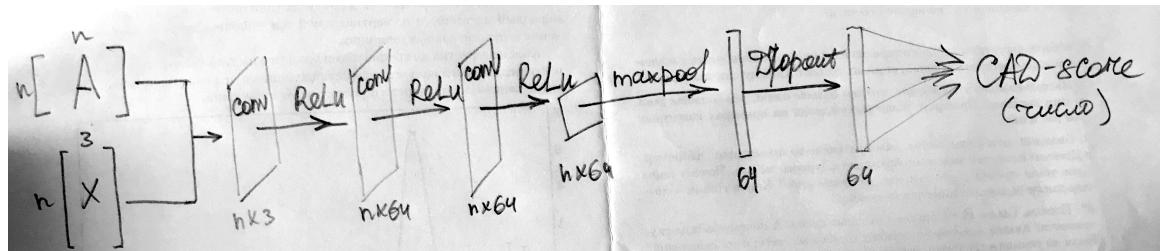


Рис. 2 Схематическое представление архитектуры свёрточной нейронной сети, использованной в данной работе

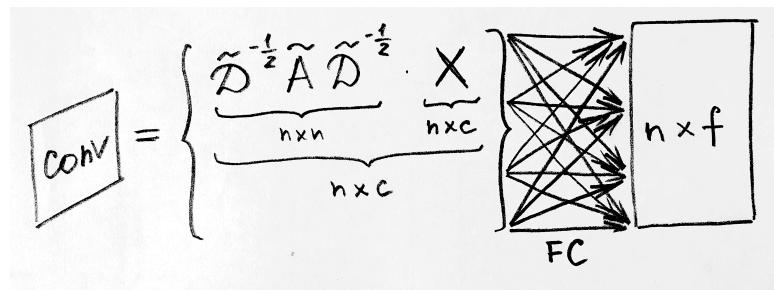


Рис. 3 Схема свёрточных слоев. X – подаваемая на вход слою матрица размера $n \times c$, f – параметр свёртки. В модели взяты следующие значения f для свёрточных слоев: 64, 64, 64

4 Вычислительный эксперимент

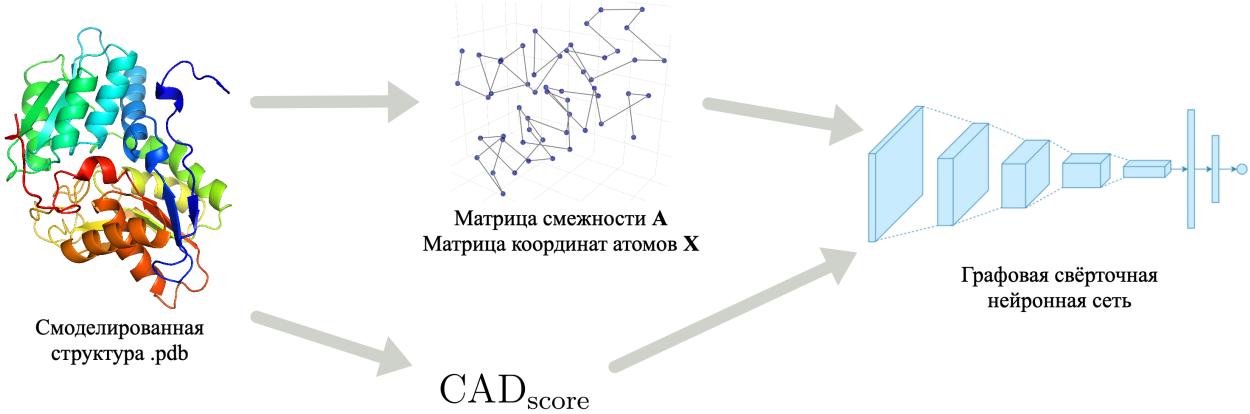


Рис. 4 Общая схема эксперимента

Данные берутся с соревнований CASP. Для реальной структуры белка берется еще смоделированная структура. Для них вычисляется CAD-score. Модель на тесте предсказывает CAD-score для смоделированной структуры, не имея возможности напрямую вычислить CAD-score по реальной структуре.

Набор	Реальные структуры	Модели структур	Разбиение
CASP 7	95	19591	
CASP 8	122	34789	
CASP 9	117	34946	Train, Validation
CASP 10	103	26254	
CASP 11	83	16094	
CASP 12	40	6924	Test

Таблица 1 Наборы данных

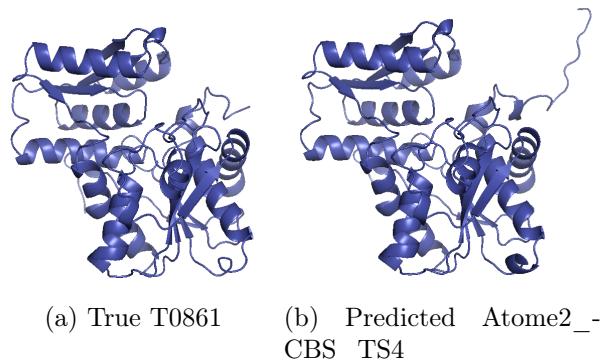


Рис. 5 Пример реальной и смоделированной структуры белка

4.1 Матрица смежности

Т.к. данные о белках не содержат информации о соединениях между атомами, т.е. нет матрицы смежности, построены соединения A по следующим правилам:

- не соединяются водород с водородом,
- атом не соединяется с водородом, если расстояние между ними $\geq 1.21 \text{ \AA}$,
- не соединяются атомы, которые находятся далеко в последовательности (номера остатков отличаются больше, чем на 1),
- не соединяются атомы, создающие дисульфидные связи,

- соединяются атомы, расстояние между которыми $r \in (r_{\min}, r_{\max}]$, где $r_{\min} = 0.01\text{\AA}$, $r_{\max} = (0.6 \cdot (\rho_{\text{atom1}} + \rho_{\text{atom2}}))^2$, ρ_{atom} – радиус атома (максимально возможное $r_{\max} = 5.76$ – при $\rho_{\text{atom1}} = \rho_{\text{atom2}} = 2.0$).

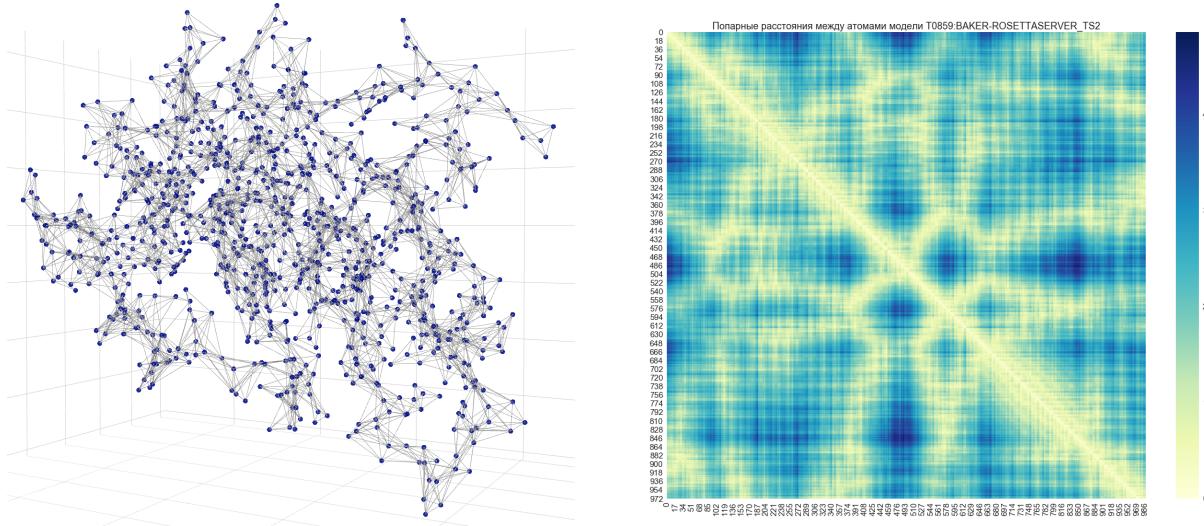


Рис. 6 Трехмерное представление с помощью координат и полученной матрицы смежности и попарные расстояния между атомами модели T0859 BAKER-ROSETTASERVER_TS2 (CASP12)

По попарным расстояниям между атомами на Рис. 6 видно, что могут иметь соединения атомы, обозначенные самым светлым желтым, т.к. максимально возможное расстояние между атомами, при котором они могут иметь соединение по представленным правилам составления матрицы смежности равно 5.76 . Т.е. матрица смежности будет сильно разреженной.

4.2 Собственное пространство матриц смежности

Для каждой полученной матрицы смежности A производится сингулярное разложение для получения сингулярных чисел матрицы. Для оценки числа необходимых главных компонент матрицы используется правило сломанной трости.

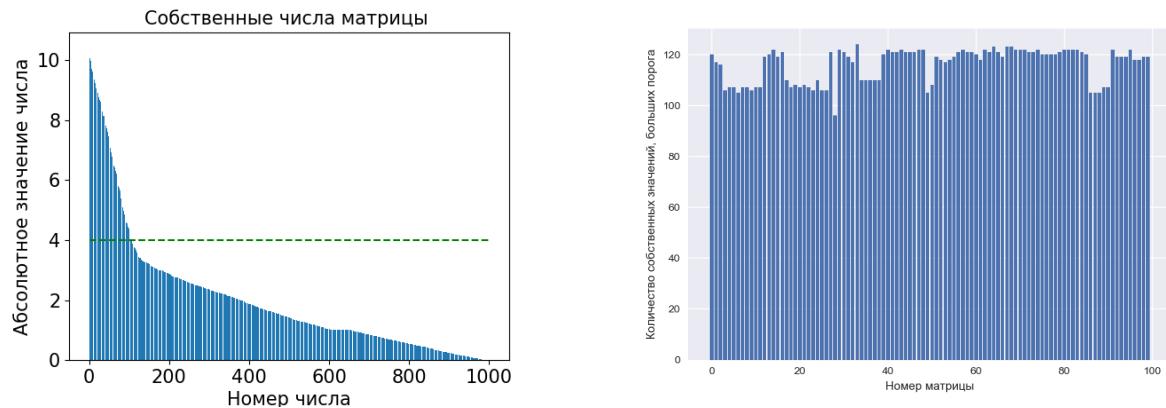


Рис. 7 Пример собственных чисел одной из **Рис. 8** Количество собственных значений матриц, больших порога 4

5 Результаты

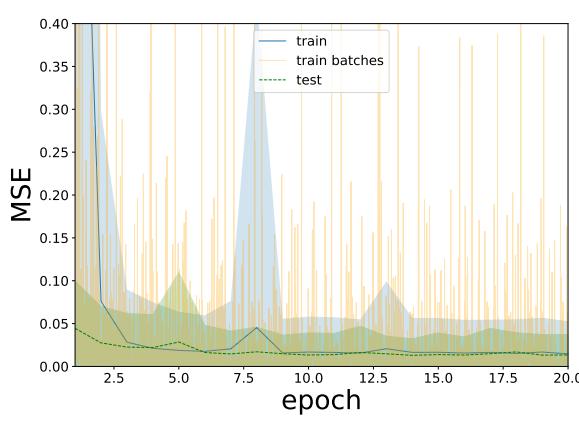


Рис. 9 График MSE ошибки GCN на обучающей и тестовой выборке

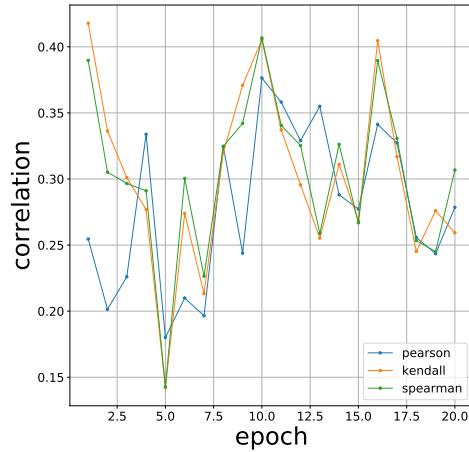


Рис. 10 Корреляция Пирсона, Кендалла, Спирмена

Сравнение с существующими методами QA:

Method	Spearmann ρ	Pearson R	z-score
ProQ3D	0.801	0.750	1.670
VoroMQA	0.803	0.766	1.410
SBROD	0.685	0.762	1.282
Ornate	0.828	0.781	1.780
SpectralQA (МОЯ)	0	0	0

Таблица 2 Сравнение корреляции Пирсона, Спирмена и z-score существующих современных алгоритмов с моделью SpectralQA на данных CASP12

НЕСКОЛЬКО СЛОВ О РЕЗУЛЬТАТАХ ИЗ ТАБЛИЦЫ

6 Заключение

Впервые для задачи оценки качества прогнозирования структуры белка применены графовые свёрточные нейронные сети, в которых свёртки определены на основе спектральной теории графов. В качестве улучшения, можно в основе архитектуры сети использовать другие существующие улучшения спектральных свёрток: CayleyNet, Adaptive Graph Convolution Network (AGCN), GRAPH WAVELET NEURAL NETWORK. Также предлагается учитывать в данных дополнительные химические свойства атомов и в матрице смежности учитывать не только наличие связи, но и расстояния между атомами при наличии связи.

Литература

- [1] J. Berg, J. Tymoczko, and L. Stryer, *Biochemistry, Fifth Edition*. W.H. Freeman, 2002.
- [2] “Protein structure prediction center.” <http://predictioncenter.org/>.
- [3] D. Hurtado, K. Uziela, and A. Elofsson, “Deep transfer learning in the assessment of the quality of protein models,” 04 2018.
- [4] M. Conover, M. Staples, D. Si, M. Sun, and R. Cao, “Angularqa: Protein model quality assessment with lstm networks,” *Computational and Mathematical Biophysics*, vol. 7, pp. 1–9, 01 2019.
- [5] G. Derevyanko, S. Grudinin, Y. Bengio, and G. Lamoureux, “Deep convolutional networks for quality assessment of protein folds,” *Bioinformatics (Oxford, England)*, vol. 34, 01 2018.
- [6] G. Pagès, B. Charmettant, and S. Grudinin, “Protein model quality assessment using 3D oriented convolutional neural networks,” *Bioinformatics*, vol. 35, pp. 3313–3319, 02 2019.
- [7] F. Baldassarre, D. M. Hurtado, A. Elofsson, and H. Azizpour, “Graphqa: Protein model quality assessment using graph convolutional network,” 2019.
- [8] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. F. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, Çağlar Gülcühre, H. F. Song, A. J. Ballard, J. Gilmer, G. E. Dahl, A. Vaswani, K. R. Allen, C. Nash, V. Langston, C. Dyer, N. M. O. Heess, D. Wierstra, P. Kohli, M. M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu, “Relational inductive biases, deep learning, and graph networks,” *ArXiv*, vol. abs/1806.01261, 2018.
- [9] K. Olechnovic, E. Kulberkytė, and C. Venclavas, “Cad-score: a new contact area difference-based function for evaluation of protein structural models.,” *Proteins*, vol. 81 1, pp. 149–62, 2013.
- [10] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A comprehensive survey on graph neural networks,” *CoRR*, vol. abs/1901.00596, 2019.
- [11] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, and M. Sun, “Graph neural networks: A review of methods and applications,” *CoRR*, vol. abs/1812.08434, 2018.
- [12] J. Bruna, W. Zaremba, A. Szlam, and Y. Lecun, “Spectral networks and locally connected networks on graphs,” in *International Conference on Learning Representations (ICLR2014), CBLS, April 2014*, 2014.
- [13] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.
- [14] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains.,” *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.
- [15] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. USA: Academic Press, Inc., 3rd ed., 2008.
- [16] M. Defferrard, X. Bresson, and P. Van gheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” in *Advances in Neural Information Processing Systems 29* (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, eds.), pp. 3844–3852, Curran Associates, Inc., 2016.
- [17] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” *arXiv:1609.02907 [cs, stat]*, Feb. 2017. arXiv: 1609.02907.
- [18] A. Oliveira and R. J. Sassi, “Behavioral malware detection using deep graph convolutional neural networks,” Nov 2019.
- [19] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, “An end-to-end deep learning architecture for graph classification,” 2018.

- [20] J. L. T. C. A. A. H. S. K. S. D. D. K. D. A. R.Evans, J.Jumper, “De novo structure prediction with deep-learning based scoring,” *Thirteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstracts) 1-4*, Dec 2018.