

ФГАОУВО «Московский физико-технический институт  
(национальный исследовательский университет)»  
Физтех-школа прикладной математики и информатики  
Кафедра «Интеллектуальные системы»  
при Вычислительном центре им. А. А. Дородницына РАН

Северилов Павел Андреевич

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

03.03.01 – Прикладные математика и физика

Выпускная квалификационная работа бакалавра

**Научный руководитель:**  
д. ф.-м. н. Стрижов Вадим  
Викторович

Москва  
2020

# Содержание

<b>1 Введение</b>	<b>4</b>
<b>2 Постановка задачи</b>	<b>5</b>
2.1 CAD score . . . . .	5
2.2 Задача регрессии . . . . .	6
<b>3 Спектральный анализ</b>	<b>7</b>
3.1 Архитектура сети . . . . .	10
<b>4 Вычислительный эксперимент</b>	<b>10</b>
4.1 Матрица смежности . . . . .	12
4.2 Собственное пространство матриц смежности . . . . .	14
<b>5 Результаты</b>	<b>16</b>
<b>6 Заключение</b>	<b>17</b>
<b>Список литературы</b>	<b>18</b>

## **Аннотация**

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 с моделированных белковых структур. Она предсказывает близость с моделированной и реальной структуры белка в виде CAD<sub>score</sub>. 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 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, описан-

ного 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  $\mathfrak{D} = (\mathbf{X}, \mathbf{y})$  – 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<sub>score</sub> [9], LDDT, GDT. В данной работе выбран CAD<sub>score</sub>.

### 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<sub>(i,j)</sub> 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<sub>(i,j)</sub><sup>bounded</sup> = min (CAD<sub>(i,j)</sub>, T<sub>(i,j)</sub>). Таким образом: CAD<sub>score</sub> для всей

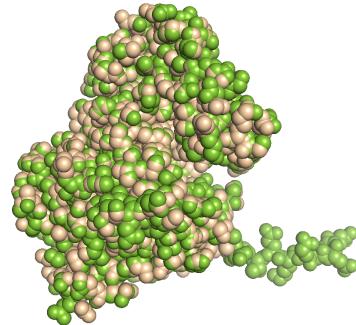


Рис. 1: Пересечение реальной и смоделированной структур

структуры определяется как

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

На рисунке 1 представлен пример пересечения реальной структуры T0861 (жёлтый) и её модели Atome2\_CBS\_TS4 (зелёный) при  $\text{CAD}_{\text{score}} = 0.829$

## 2.2 Задача регрессии

Рассматривается множество параметрических моделей  $\mathfrak{F}$ , взятых из класса графовых свёрточных нейронных сетей:  $\mathfrak{F} = \{\mathbf{f}_k: (\mathbf{w}, \mathbf{X}) \rightarrow \hat{\mathbf{y}} \mid k \in \mathfrak{K}\}$ , 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  $\text{CAD}_{\text{score}} y_i$  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  $\mathcal{L}(\mathbf{y}, \mathbf{X}, \mathbf{w}) = (\hat{\mathbf{y}} - \mathbf{y})^2$ , где  $\hat{\mathbf{y}} = \mathbf{f}(\mathbf{X}, \mathbf{w})$  –  $\text{CAD}_{\text{score}}$  предсказанный моделью  $\mathbf{f}$ ,  $\mathbf{y}$  – данный в выборке  $\text{CAD}_{\text{score}}$ .

Часто для оценивания модели используются коэффициенты корреляции Пирсона ( $R$ ), Спирмена ( $\rho$ ) [5–7]. Для каждой реальной структуры белка вычисляются коэффициенты корреляции Пирсона ( $R^{\text{target}}$ ), Спирмена ( $\rho^{\text{target}}$ ) между истинными и прогнозируемыми  $\text{CAD}_{\text{score}}$  для смоделированных 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  $\text{CAD}_{\text{score}}$  для смоделированных структур белка, соответствующих реальной структуре  $i$ . Тогда коэффи-

циенты 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  $\mathbf{g}_\theta$ . Соотношение 1 вычислительно 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  $\mathbf{g}_\theta$  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  $L^2\left([-1, 1], \frac{dx}{\sqrt{1-x^2}}\right)$

Представляя  $\mathbf{g}_\theta$  в виде

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

где  $\tilde{\Lambda} = 2\Lambda/\lambda_{\max} - \mathbf{I}_n \in [-1, 1]$ ,  $\lambda_{\max}$  – 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*:

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

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,  $\sigma$  – 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.

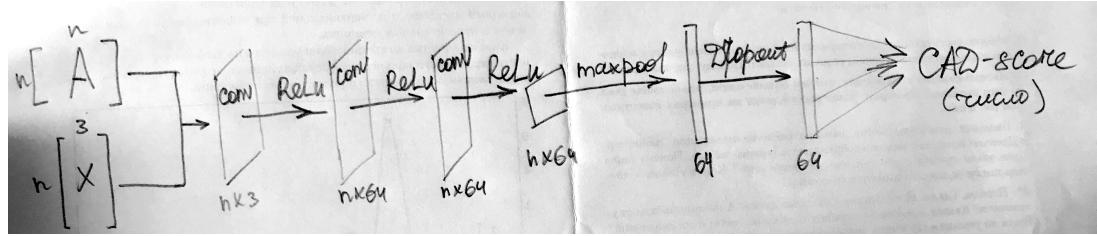


Рис. 2: Схематическое представление архитектуры свёрточной нейронной сети, использованной в данной работе

## 4 Вычислительный эксперимент

Данные берутся с соревнований CASP. Для реальной структуры белка берется еще смоделированная структура. Для них вычисляется CAD-

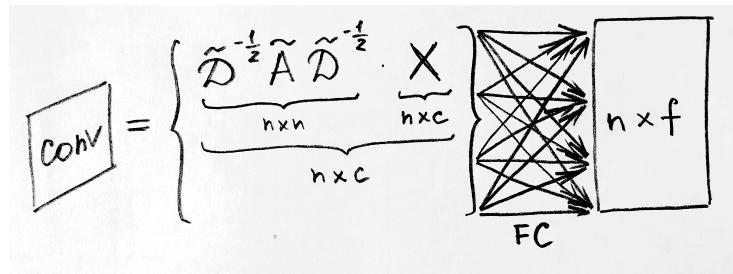


Рис. 3: Схема свёрточных слоев.  $X$  – подаваемая на вход слою матрица размера  $n \times c$ ,  $f$  – параметр свёртки. В модели взяты следующие значения  $f$  для свёрточных слоев: 64,64,64

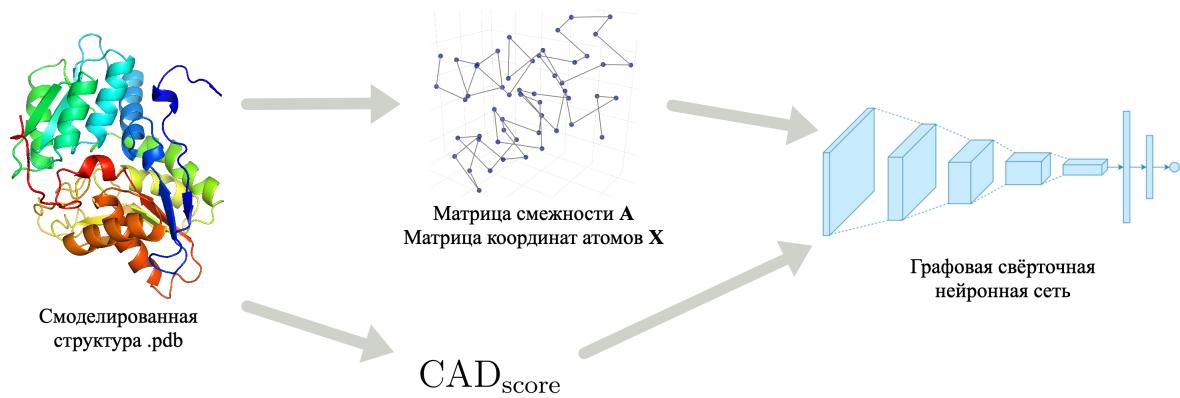


Рис. 4: Общая схема эксперимента

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	Train, Validation
CASP 8	122	34789	
CASP 9	117	34946	
CASP 10	103	26254	
CASP 11	83	16094	
CASP 12	40	6924	Test

Таблица 1: Наборы данных

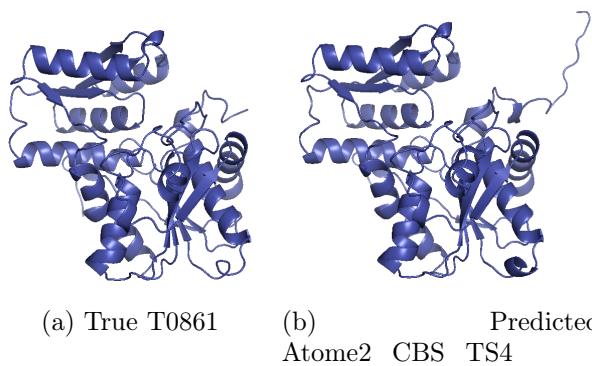


Рис. 5: Пример реальной и смоделированной структуры белка

## 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  $r_{\min} = 0.01\text{\AA}$ ,  $r_{\max} = (0.6 \cdot (\rho_{\text{atom1}} + \rho_{\text{atom2}}))^2$ ,  $\rho_{\text{atom}}$  – радиус атома (максимально возможное  $r_{\max} = 5.76$  – при  $\rho_{\text{atom1}} = \rho_{\text{atom2}} = 2.0$ ).

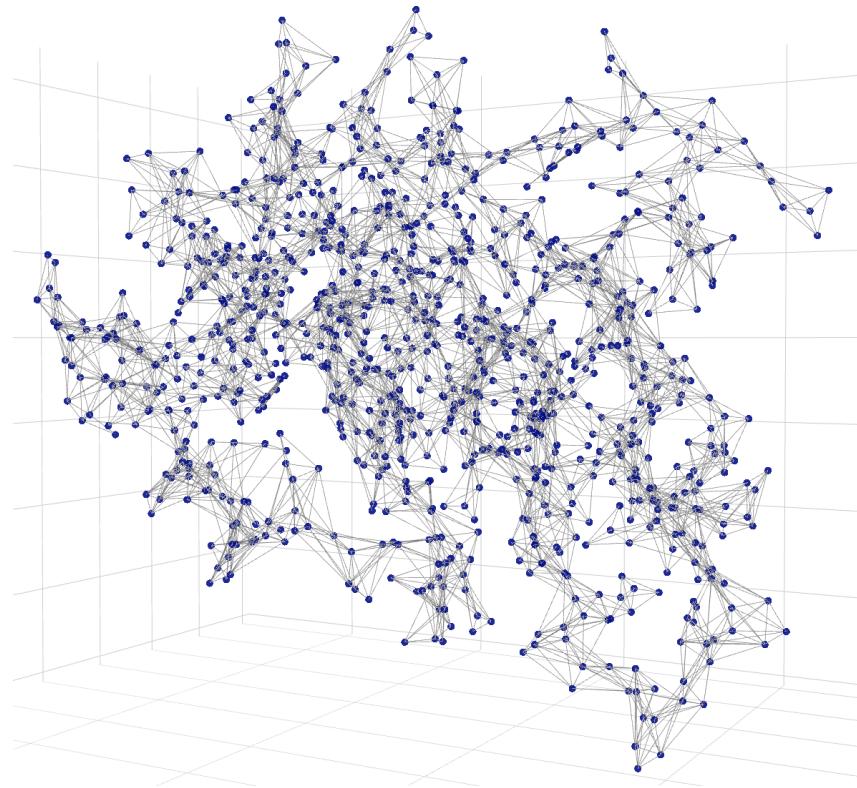


Рис. 6: Трехмерное представление с помощью координат и полученной матрицы смежности

По попарным расстояниям между атомами на Рис. 7 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.

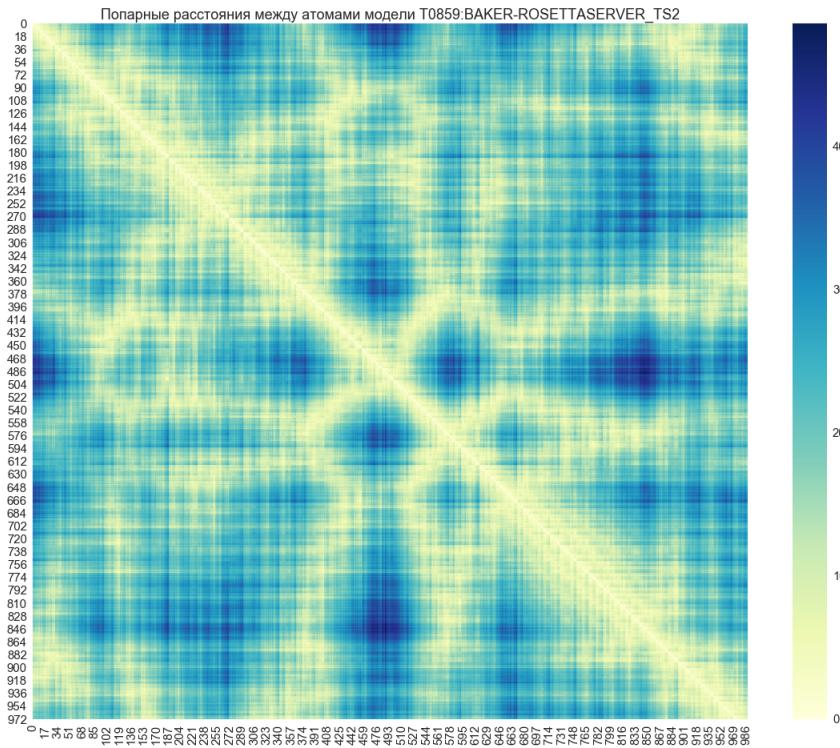


Рис. 7: попарные расстояния между атомами модели T0859\_BAKER-ROSETTASERVER\_TS2 (CASP12)

## 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.

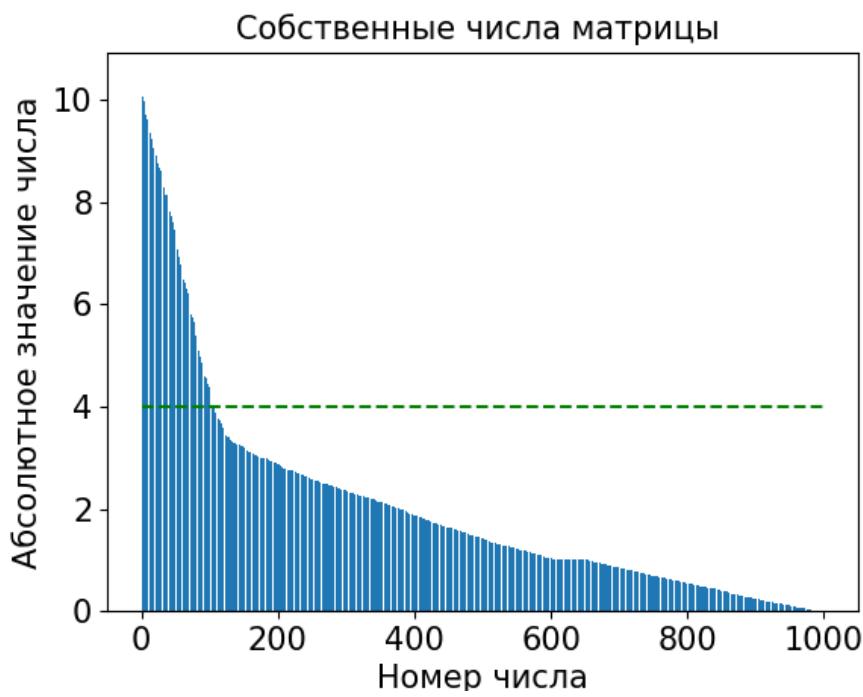


Рис. 8: Пример собственных чисел одной из матриц вместе с порогом 4

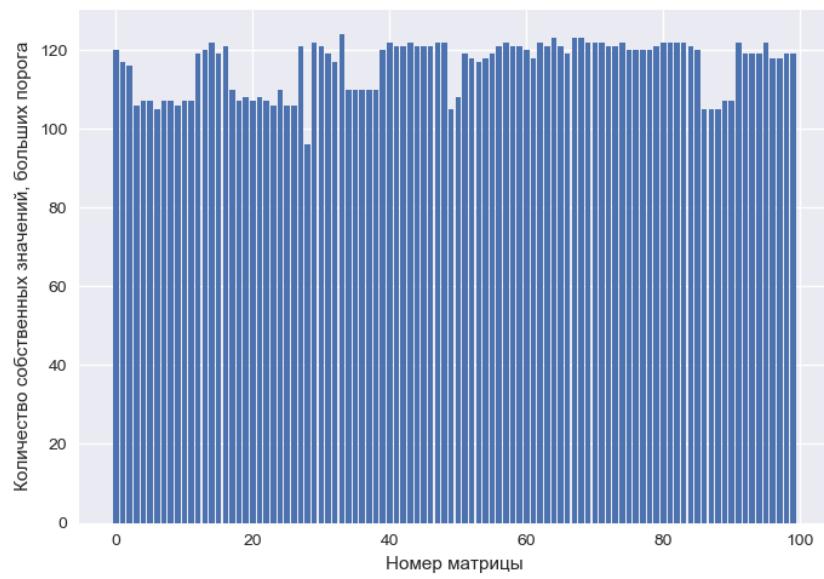


Рис. 9: Количество собственных значений матриц, больших порога 4

## 5 Результаты

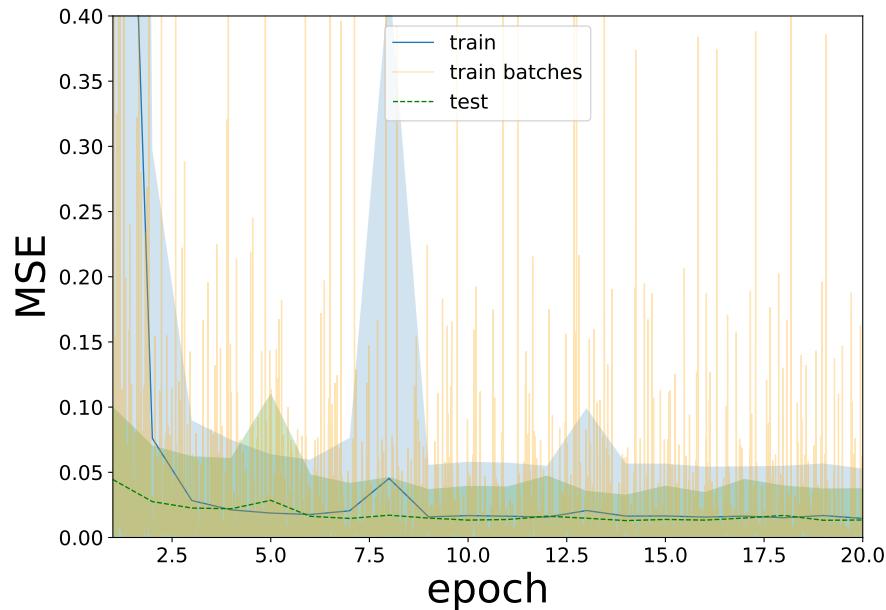


Рис. 10: График MSE ошибки GCN на обучающей и тестовой выборке

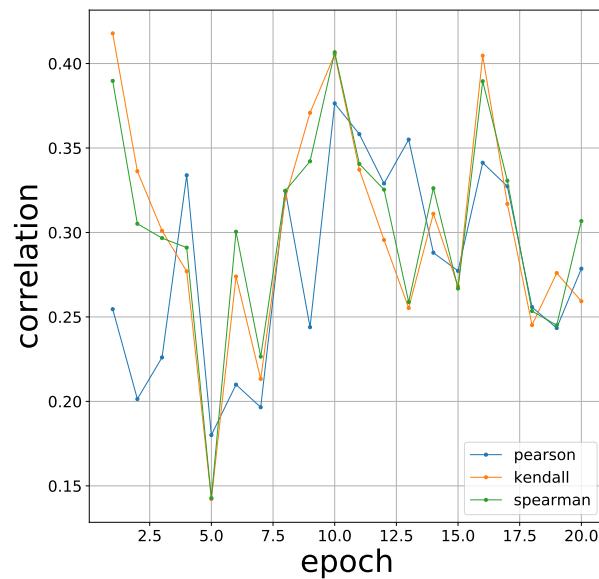


Рис. 11: Корреляция Пирсона, Кендалла, Спирмена

Сравнение с существующими методами QA:

Method	Spearmann $\rho$	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
<b>SpectralQA (МОЯ)</b>	<b>0</b>	<b>0</b>	<b>0</b>

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.

## Список литературы

- [1] Berg J.M., Tymoczko J.L., Stryer L. Biochemistry, Fifth Edition. — W.H. Freeman, 2002. — ISBN: 9780716730514. — URL: <https://books.google.ru/books?id=uDFqAAAAMAAJ>.
- [2] Protein Structure Prediction Center. — <http://predictioncenter.org/>.
- [3] Hurtado David, Uziela Karolis, Elofsson Arne. Deep transfer learning in the assessment of the quality of protein models. — 2018. — 04.
- [4] AngularQA: Protein Model Quality Assessment with LSTM Networks / Matthew Conover, Max Staples, Dong Si et al. // Computational and Mathematical Biophysics. — 2019. — 01. — Vol. 7. — P. 1–9.
- [5] Deep convolutional networks for quality assessment of protein folds / Georgy Derevyanko, Sergei Grudinin, Y. Bengio, Guillaume Lamoureux // Bioinformatics (Oxford, England). — 2018. — 01. — Vol. 34.
- [6] Pagès Guillaume, Charmettant Benoit, Grudinin Sergei. Protein model quality assessment using 3D oriented convolutional neural networks // Bioinformatics. — 2019. — 02. — Vol. 35, no. 18. — P. 3313–3319. — <http://oup.prod.sis.lan/bioinformatics/article-pdf/35/18/3313/30024731/btz122.pdf>.
- [7] GraphQA: Protein Model Quality Assessment using Graph Convolutional Network / Federico Baldassarre, David Menéndez Hurtado, Arne Elofsson, Hossein Azizpour. — 2019.
- [8] Relational inductive biases, deep learning, and graph networks / Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst et al. // ArXiv. — 2018. — Vol. abs/1806.01261.
- [9] Olechnovic Kliment, Kulberkytė Eleonora, Venclovas Ceslovas. CAD-score: a new contact area difference-based function for evaluation of protein structural models. // Proteins. — 2013. — Vol. 81 1. — P. 149–62.

- [10] A Comprehensive Survey on Graph Neural Networks / Zonghan Wu, Shirui Pan, Fengwen Chen et al. // CoRR. — 2019. — Vol. abs/1901.00596. — 1901.00596.
- [11] Graph Neural Networks: A Review of Methods and Applications / Jie Zhou, Ganqu Cui, Zhengyan Zhang et al. // CoRR. — 2018. — Vol. abs/1812.08434. — 1812.08434.
- [12] Spectral networks and locally connected networks on graphs / Joan Bruna, Wojciech Zaremba, Arthur Szlam, Yann Lecun // International Conference on Learning Representations (ICLR2014), CBLS, April 2014. — 2014.
- [13] Chung F. R. K. Spectral Graph Theory. — American Mathematical Society, 1997.
- [14] The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains. / David I. Shuman, Sunil K. Narang, Pascal Frossard et al. // IEEE Signal Process. Mag. — 2013. — Vol. 30, no. 3. — P. 83–98.
- [15] Mallat Stphane. A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way. — 3rd edition. — USA : Academic Press, Inc., 2008. — ISBN: 0123743702.
- [16] Defferrard Michaël, Bresson Xavier, Van gheynst Pierre. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering // Advances in Neural Information Processing Systems 29 / Ed. by D. D. Lee, M. Sugiyama, U. V. Luxburg et al. — Curran Associates, Inc., 2016. — P. 3844–3852. — URL: <http://papers.nips.cc/paper/6081-convolutional-neural-networks-on-graphs-with-fast-localized-spectral-filtering.pdf>.
- [17] Kipf Thomas N., Welling Max. Semi-Supervised Classification with Graph Convolutional Networks // arXiv:1609.02907 [cs, stat]. — 2017. — Feb. — arXiv: 1609.02907. URL: <http://arxiv.org/abs/1609.02907> (online; accessed: 2019-12-10).
- [18] An End-to-End Deep Learning Architecture for Graph Classification / Muhan Zhang, Zhicheng Cui, Marion Neumann, Yixin Chen. — 2018.

[19] R.Evans J.Jumper J.Kirkpatrick L.Sifre T.F.G.Green C.Qin A.Zidek  
A.Nelson A.Bridgland H.Penedones S.Petersen K.Simonyan S.Crossan  
D.T.Jones D.Silver K.Kavukcuoglu D.Hassabis A.W.Senior.  
De novo structure prediction with deep-learning based scoring // Thirteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstracts) 1-4. — 2018. — Dec. — URL:  
<https://deepmind.com/blog/article/alphafold>.