Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа прикладной математики и информатики
Кафедра интеллектуальных систем

Направление подготовки: 03.03.01 Прикладные математика и физика (бакалавриат)
Направленность (профиль) подготовки: Компьютерные технологии и интеллектуальный анализ данных

Нейросетевая аппроксимация плотности для построения вероятностно-метрического пространства

(бакалаврская работа)

Студент:
Вареник Наталия Викторовна

(подпись студента)
Научный руководитель:
Стрижов Вадим Викторович,
д-р физ.-мат. наук

(подпись научного руководителя)

Москва 2020

Содержание

Аннотация

Решается задача восстановления совместной плотности распределения взаимного расположения и угловой ориентации аминокислотного остатка и маленькой молекулы — лиганда. Взаимное расположение и угловая ориентация описываются расстоянием между аминокислотой и лигандом и двумя углами, которые представляют собой сферические координаты лиганда в локальной системе координат аминокислоты. В данной работе описывается нейросетевой метод оценки совместной плотности распределения, в качестве которого предлагается использовать монотонный многослойный перцептрон с сигмоидной функцией активации. Данная модель применяется для непрерывной аппроксимации эмпирической функции распределения пространственной молекулярной конфигурации аминокислоты и лиганда. В работе предлагается модификация перцептрона для аппроксимации неубывающей функции, которая сохраняет ее дифференцируемость и аналитический вид. Итоговая плотность распределения получается путем дифференцирования сети. Поскольку модифицированная нейронная сеть остается, как и классичекий перцептрон, композицией аналитических функций, производная может быть выражена с помощью аналитической формулы без необходимости использования численных методов дифференцирования. Возможность аналитического дифференцирования обладает значительным преимуществом при восстановлении плотности в пространстве большой размерности, так как в этом случае численные методы аппроксимации производных высшего порядка становятся неэффективными и нестабильными, поэтому неприменимы на практике.

Ключевые слова: *оценка совместной плотности распределения, многомерные данные, многослойный перцептрон, эмпирическая функция распределения, взаимное расположение, угловая ориентация, аминокислота, лиганд*

1 Введение

Методы и вероятностные модели искусственного интеллекта сделали большой шаг вперед в области вычислительной структурной биологии. Это стало возможным благодаря прогрессу в области машинного обучения и активного применения методов анализа данных в вычислительных экспериментах. В связи с тем, что число известных белковых структур значительно меньше числа выявленных белковых последовательностей, разработка надежного метода прогнозирования белковой структуры является важной задачей биоинформатики [1–3]. С этой целью возникает потребность в разработке вероятностно-метрической модели, которая описывают пространственные конфигурации контактов между аминокислотными остатками и стыковки белок-лиганд для прогнозирования множества функциональных состояний одного и того же белка, предсказания того, как белки взаимодействуют друг с другом, с небольшими молекулами, как они образуют сборки (докинг белка), и, в конечном счете, моделирования структуры белка. Все эти проблемы в настоящее время очень трудно решить без использования дополнительной информации из различных источников данных.

Для уточнения модели прогнозирования и оценки качества используются функции, называемые потенциалом [4], которые задают близость рассматриваемой пространственной конфигурации (взаимного расположения и угловой ориентации) к эталонному множеству конфигураций. В [5] описывается KORP — парный статистический потенциал для белков, использующий совместную плотность распределения взаимного расположения и угловой ориентации и показано, что он имеет лучшую точность различения конфигураций и является быстрым методом оценки качества моделирования белковых структур по сравнению с потенциалами, построенными из физических соображений. Использование статистической информации о структуре белка является основной причиной преимущества KORP в плане точности и времени работы.

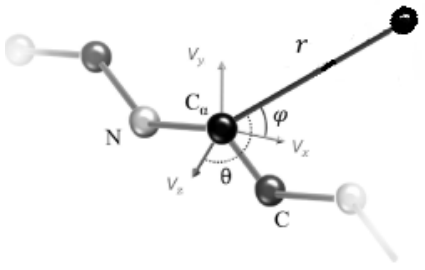

Рис. 1.1: Схематическое изображение, определяющее взаимную ориентацию и положение аминокислоты и лиганда. Относительная ориентация описывается двумя полярными углами φ и θ между радиус-вектором лиганда и осями V_x и V_z локальной системы координат аминокислоты.

Данная работа посвящена задаче аппроксимации совместной плотности распределения взамного расположения и угловой ориентации аминокислоты и лиганда. Взаимное расположение и угловая ориентация задаются расстоянием r между аминокислотным остатком и ли-

гандом и углами φ и θ, которые являются сферическими координатами лиганда в локальной системе координат аминокислоты, см. Рис. 1.1. Тогда задача состоит в восстановлении совместной плотности распределения $p_{a,b}(r, \theta, \varphi)$ для каждой пары аминокислоты и лиганда типов a и b соответственно. Локальная трехмерная система координат аминокислоты определяется из системы:

$$\begin{cases} \mathbf{V}_z = (\mathbf{r}_{CC_\alpha} + \mathbf{r}_{NC_\alpha})/|\mathbf{r}_{CC_\alpha} + \mathbf{r}_{NC_\alpha}|, \\ \mathbf{V}_y = (\mathbf{V}_z \times \mathbf{r}_{NC_\alpha})/|\mathbf{V}_z \times \mathbf{r}_{NC_\alpha}|, \\ \mathbf{V}_x = \mathbf{V}_y \times \mathbf{V}_z. \end{cases} \tag{1.1}$$

где $\mathbf{r}_{CC_\alpha} = \mathbf{r}_C - \mathbf{r}_{C_\alpha}$, $\mathbf{r}_{NC_\alpha} = \mathbf{r}_N - \mathbf{r}_{C_\alpha}$ – векторы из C_α в C, N.

Для восстановления плотности распределения предлагается использовать нейросетевой метод. Недавние разработки в области глубинного обучения предоставили новые методы восстановления плотности с ипользованием нейронных сетей. В [6] предлагается два метода восстановления плотности, метод стохастического обучения и метод гладкой интерполяции кумулятивной функции. Оба метода используют для обучения функцию распределения, посчитанную на обучающей выборке. Подбор параметров осуществляется методом обратного распространения ошибки. При этом функция ошибки содержит регуляризатор, который контролирует, чтобы модель была монотонно возрастающей, так как она моделирует истинную функцию распределения. Метод гладкой интерполяции в отличие от стохастического обучения кумулятивной функции работает не только для одномерных данных но и для многомерных, но для получения итоговой плотности использует численный метод дифференцировая, который с ростом порядка дифференцирования становится нестабильным и неэффективным. В [7] используется многослойный перцептрон для восстановления непосредственно плотности распределения. Поиск оптимальных параметров производится методом максимизации логарифма правдоподобия, но для того, чтобы полученная модель аппроксимировала истинную плотность нужно ее еще отнормировать. Для этого предлагается воспользоваться численным методом интергирования. В [8] описан метод нейросетевой авторегрессионной оценки функции распределения для многомерных биномиальных данных и его модификация для вещественного случая. Метод работает в строгом предположении, что совместная плотность распределения выражается через последовательные условные одномерные распределения. Его недостатком является то, что он сильно чувстителен к последовательному порядку переменных, что влияет на различную способность восстанавливать определенные функции плотности, поэтому на практике поиск оптимального последовательного порядка переменных является трудной задачей. В качестве решения проблемы предлагается комбинировать модели, полученные на упорядоченных последовательностях случай-

ных переменных, но данный подхой обладает низкой вычислительной эффективностью, так как для каждой упорядоченной последовательности случайных переменных нужно заново восстанавливать условные одномерные плотности распределения.

В данной работе предлагается реализовать нейросетевой метод восстановления плотности для получения полностью аналитической, а поэтому и легко дифференцируемой по носителю функции. Это позволит без затруднения находить максимумы плотности распределения, которые соответствуют устойчивым пространственным конфигурациям аминокислоты и лиганда. Устойчивые конфигурации в свою очередь соответствуют минимуму энергии взаимодействия пары аминокислота-лиганд.

Для оценки плотности распределения производится непрерывная аппроксимация функции распределения пространственной конфигурации, которая оценивается по имеющей выборке. Аппроксимация осуществляется с использованием монотонного перцептрона. После чего итоговая плотность получается дифференцированием модели, соответствующей наилучшей аппроксимации. В качестве решения исследуется метод [9], который является улучшенной модификацией метода гладкой интерполяции функции распределения, так как в нем устраняется необходимость в подборе метапараметра, оказывающего сильное влияние на результат восстановления функции распределения.

2 Постановка задачи

2.1 Задача восстановления плотности

Для каждой пары аминокислоты и лиганда (a, b) задана выборка:

$$\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n}, \tag{2.1}$$

где n – число найденных в базе данных пространственных конфигураций пары (a, b). Элемент выборки представляет из себя реализацию трехкомпонентного случайного вектора $\mathbf{x} = [r, \theta, \varphi]^\mathsf{T} \in \Omega$, где r – расстояние между аминокислотой и лигандом. Углы θ и φ – сферические координаты лиганда в локальной системе координат аминокислоты, $\Omega = [4\text{Å}, 20\text{Å}] \times [0, \pi] \times [0, 2\pi] \subset \mathbb{R}^3$.

Для пары (a, b) необходимо построить модель $\hat{p}^{a,b}(\mathbf{x}) = \hat{p}^{a,b}(\mathbf{x}|\mathbf{w}, \mathbf{z})$, которая аппроксимирует истинную плотность распределения пространственной конфигурации и является дифференцируемой по носителю Ω, где $\mathbf{w}$ – параметры модели, $\mathbf{z}$ – структурные параметры модели.

2.2 Восстановление плотности как задача регрессии

Для каждой пары (a, b) введем вектор ответов $\mathbf{y} = [y_1, \ldots, y_n]^\mathsf{T}$, где y_i является значением эмпирической функции распределения пространсвенной конфигурации аминокислоты и лиганда в точке $\mathbf{x}_i$:

$$y_i = \frac{1}{n} \sum_{j=1, j\neq i}^{n} \Theta(\mathbf{x}_j - \mathbf{x}_i), \tag{2.2}$$

где $\Theta(\mathbf{x})$ – тета-функция Хевисайда, равная 1 если $x_d \geq 0 \ \forall d = \overline{1, 3}$, иначе 0.

Рассмотрим множество параметрических моделей взятых из класса нейросетей прямой связи:

$$\mathfrak{F} = \{\mathbf{f}^{a,b} : (\mathbf{X}, \mathbf{w}, \mathbf{z}) \mapsto \mathbf{y}\}, \tag{2.3}$$

где $\mathbf{w} \in \mathbb{R}$ – параметры модели, $\mathbf{z} \in \mathbb{N}$ – структурные параметры.

Получаем регрессионную задачу, состоящую в гладкой, непрерывной аппроксимации эмпирической функции распределения пространсвенной конфигурации пары аминокислота-лиганд.

Определим функцию потерь:

$$S(\mathbf{y}, \mathbf{f}^{a,b}, \mathbf{X}, \mathbf{w}, \mathbf{z}) = ||\mathbf{y} - \mathbf{f}^{a,b}(\mathbf{X}, \mathbf{w}, \mathbf{z})||_2^2. \tag{2.4}$$

Параметры модели $\mathbf{w} \in \mathbb{R}$ подбираются в соответствии с минимизацией функции потерь методом обратного распространения

$$\mathbf{w}^* = \arg\min_{\mathbf{w}\in\mathbb{W}} S(\mathbf{w} \mid \mathbf{y}, \mathbf{f}^{a,b}, \mathbf{X}, \mathbf{z}). \tag{2.5}$$

Подобрав параметры получим непрерывную, аналитическую функцию, аппроксимирующую функцию распределения пространсвенной конфигурации аминокислоты и лиганда.

Тогда плотность распределения получается путем дифференцирования

$$\hat{p}^{a,b}(\mathbf{x}|\mathbf{w}, \mathbf{z}) = \frac{\partial^3}{\partial r \partial\theta \partial\varphi} f^{a,b}(\mathbf{x}, \mathbf{w}, \mathbf{z}). \tag{2.6}$$

3 Восстановление плотности распределения

3.1 Монотонный многослойный перцептрон как универсальный аппроксиматор монотонной функции

Основа решения задачи восстановления плотности распределения заключается в непрерывной аппроксимации эмпирической функции распределения, которая при достаточно большом размере выборки по закону больших чисел является адекватной оценкой теоретической функции распределения. Для аппроксимации функции распределения предлагается использовать многослойный перцептрон, так как по универсальной теореме об аппроксимации [10] многослойный перцептрон с сигмоидными функциями активации является универсальным аппроксиматором любой непрерывной функции. Но поскольку, функция распределения обладает свойством неубывания, то нужно контролировать, что перцептрон также является неубывающим. Следующая теорема позволяет решить этот вопрос.

Теорема 1 *(Bernhard Lang, [11])*

Полносвязный многослойный перцептрон вида:

$$f(\mathbf{x}) = b^{(3)} + \sum_{l=1}^{h_2} w_l^{(3)} \sigma \left(b_l^{(2)} + \sum_{i=1}^{h_1} W_{li}^{(2)} \sigma \left(b_i^{(1)} + \sum_{j=1}^{d} W_{ij}^{(1)} x_j \right) \right), \tag{3.1}$$

где d–размерность входа, h_1 – размерностью первого скрытого слоя, h_2 – размерность второго, обеспечивает монотонно возрастающее поведение по отношению к входу, если

$$w_l, W_{li}, W_{ij} \geq 0, \ i = \overline{1, h_1}, \ l = \overline{1, h_2} \tag{3.2}$$

Доказательство.

В силу достаточного условия возрастания функции многослойный перцептрон обеспечивает монотонное возрастание по отношению к входным переменным $x_j \in \mathbf{x}$, если:

$$\frac{\partial f}{\partial x_j} = \sum_{l=1}^{h_2} w_l^{(3)} \underbrace{\left(1 - \sigma^2 \left(b_l^{(2)} + \sum_{i=1}^{h_1} W_{li}^{(2)} \sigma \left(b_i^{(1)} + \sum_{j=1}^{d} W_{ij}^{(1)} x_j \right) \right) \right)}_{\geq 0} \cdot \tag{3.3}$$

$$\cdot \sum_{h=1}^{h_1} W_{li}^{(2)} \underbrace{\left(1 - \sigma^2 \left(b_i^{(1)} + \sum_{j=1}^{d} W_{ij}^{(1)} x_j \right) \right)}_{\geq 0} W_{ij}^{(1)} \geq 0,$$

где $1 - \sigma^2(x)$ – производная сигмоидной функции 1-го порядка, которая принимает неотрицательные значения, поскольку $\sigma(x)$ является возрастающей функцией.

Получаем, что монотонное возрастание многослойного перцептрона (3.1) равносильно условию:

$$w_l \cdot W_{li} \cdot W_{ij} \geq 0,\ i = \overline{1, h_1},\ l = \overline{1, h_2}. \tag{3.4}$$

Это неравенство выполняется, если потребовать:

$$w_l, W_{li}, W_{ij} \geq 0,\ i = \overline{1, h_1},\ l = \overline{1, h_2}. \tag{3.5}$$

■

Исходя из данной теоремы, для получения неубывающего перцептрона необходимо контролировать выполнение условия неотрицательности весов всех его слоев. Таким образом удастся устранить случаи, когда полученная модель в результате оптимизации функции потерь является функцией, которая по определению не может быть функцией распределения.

3.2 Метод гладкой интерполяции функции распределения

Пусть дана выборка $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbf{x}_i \in \mathbb{R}^d$, $p(\mathbf{x})$ – плотность распределения в точке $\mathbf{x}$. По выборке оценивается функция функция распределения $y(\mathbf{x})$ одним из следующих способов.

На носителе Ω распределения выборки определить равномерную сетку и в каждом узле сетки посчитать эмпирическую функцию распределения как

$$y(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \Theta(\mathbf{x} - \mathbf{x}_i), \tag{3.6}$$

где $\Theta(\mathbf{x}) = 1$, если $x_i \geq 0\ \forall i = 1, \ldots, d$, иначе 0.

Или вычислить эмпирическую функцию распределения исходя только из точек выборки

$$y(\mathbf{x}_m) = \frac{1}{n-1} \sum_{i=1, i \neq m}^{n} \Theta(\mathbf{x}_m - \mathbf{x}_i). \tag{3.7}$$

Для аппроксимации эмпирической функции распределения используется многослойный перцептрон

$$f(\mathbf{x}) = \sum_{i=1}^{h} w_i^{(2)} \sigma\left(\mathbf{W}_{i*}^{(1)} \mathbf{x} + b_i^{(1)}\right) + b^{(2)} = \sum_{i=1}^{h} w_i^{(2)} \sigma\left(\sum_{j=1}^{d} W_{ij}^{(1)} x_j + b_i^{(1)}\right) + b^{(2)}. \tag{3.8}$$

Параметры перцептрона подбираются методом обратного распространения ошибки,

где ошибка состоит из слагаемого, описывающего среднеквадратичное отклонение прогноза от истинного значения и слагаемого, штрафующего модель за нарушение условия неубывания. После подбора оптимальных параметров плотность распределения получается дифференцированием модели

$$\hat{p}(\mathbf{x}) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} f(\mathbf{x}). \tag{3.9}$$

3.3 Недостатки метода гладкой интерполяции функции распределения

Чтобы обеспечить валидность модели для аппроксимации функции распределения в методе гладкой интерполяции функции распределения к квадратичной функции потерь добавляется регуляризатор, штрафующий модель за нарушение условия неубывания:

$$L(\mathbf{X}, \mathbf{w}) = \sum_{i=1}^{n} (f(\mathbf{x}_i) - y(\mathbf{x}_i))^2 + \lambda \sum_{k=1}^{n} \Theta(f(\mathbf{x}_k) - f(\mathbf{x}_k + \delta 1_d))[f(\mathbf{x}_k) - f(\mathbf{x}_k + \delta 1_d)]^2. \tag{3.10}$$

где λ – положительная константа, метапараметр, δ – положительное маленькое число, 1_d – единичный вектор размерности d.

Подбор метарпараметра осуществляется перебором по сетке, что приводит к дополнительным вычислительным затратам. При этом слишком большое значение λ может привести к сильно сглаженной аппроксимации функции распределения, что исказит получаемую плотность при дифференцировании. Слишком маленькое значение λ не может гарантировать неубывание модели по отношению к входу на носителе выборки Ω, где точек обучающей выборки либо очень мало, либо вообще нет.

Кроме этого, для метода гладкой интерполяции не было предложено эффективного способа вычисления производной (3.9), кроме численного дифференцирования.

3.4 Улучшенная модификация метода гладкой интерполяции функции распределения

В отличие от метода гладкой интерполяции для аппроксимации эмпирической функции распределения предлагается использовать многослойный перцептрон с неотрицательными весами всех слоев и сигмоидными функциями активации. По теореме 1 такая модель является универсальным аппроксиматором любой неубывающей функции.

Чтобы сохранить аналитический вид модели, удобный для дифференцирования предлагается обеспечивать неотрицательность весов за счет их экспоненцирования.

Тогда модель, используемая для аппроксимации функции распределения примет следующий вид

$$f(\mathbf{x}) = \sum_{i=1}^{h} e^{w_i^{(2)}} \sigma\left(\sum_{j=1}^{d} e^{W_{ij}^{(1)}} x_j + b_i^{(1)}\right) + b^{(2)}. \tag{3.11}$$

Получаем гладкую, определенную на всем носителе выборки и дифференцируемую по нему функцию, имеющую аналитический вид, что позволяет упростить ее дифференцирование.

3.5 Получение плотности

Поскольку, полученная функция для оценки функции распределения имеет аналитический вид, то ее производная также может быть получена аналитически. Для перцептрона с одним скрытым слоем формула принимает следующий вид:

$$p(\mathbf{x}) = \frac{\partial^d}{\partial x_1, \ldots, \partial x_d} f(\mathbf{x}) = \sum_{i=1}^{h} e^{w_i^{(2)}} \prod_{j=1}^{d} e^{W_{ij}^{(1)}} \sigma^{(d)}\left(\sum_{j=1}^{d} e^{W_{ij}^{(1)}} x_j + b_i^{(1)}\right) = \tag{3.12}$$

$$= \sum_{i=1}^{h} e^{w_i^{(2)} + \sum\limits_{j=1}^{d} W_{ij}^{(1)}} \sigma^{(d)}\left(\sum_{j=1}^{d} e^{W_{ij}^{(1)}} x_j + b_i^{(1)}\right),$$

где $\sigma^{(d)}$ – производная сигмоидной функции d-го порядка.

Производная сигмоидной функции активации обладает некоторым свойством замкнутости в том плане, что ее n-я производная является полиномом $(n+1)$-ой степени от $\sigma(x)$. В [12,13] описывается эффективный алгоритм подбора коэффициентов этого полинома.

4 Вычислительный эксперимент

Используемая база данных содержит двадцать аминокислотных остатков. Аминокислотные остатки взаимодействуют с лигандами, их сорок. В базе имеется пять признаков: тип аминокислотного остатка, тип лиганда, расстояние и 2 угла θ и ϕ. Ниже приведет график встречающихся в базе данных пространственных конфигураций, в котором наибольшие скопления точек демонстрируют, что статистически аминокислотный остаток и лиганд наиболее вероятно занимают данную пространственную конфигурацию.

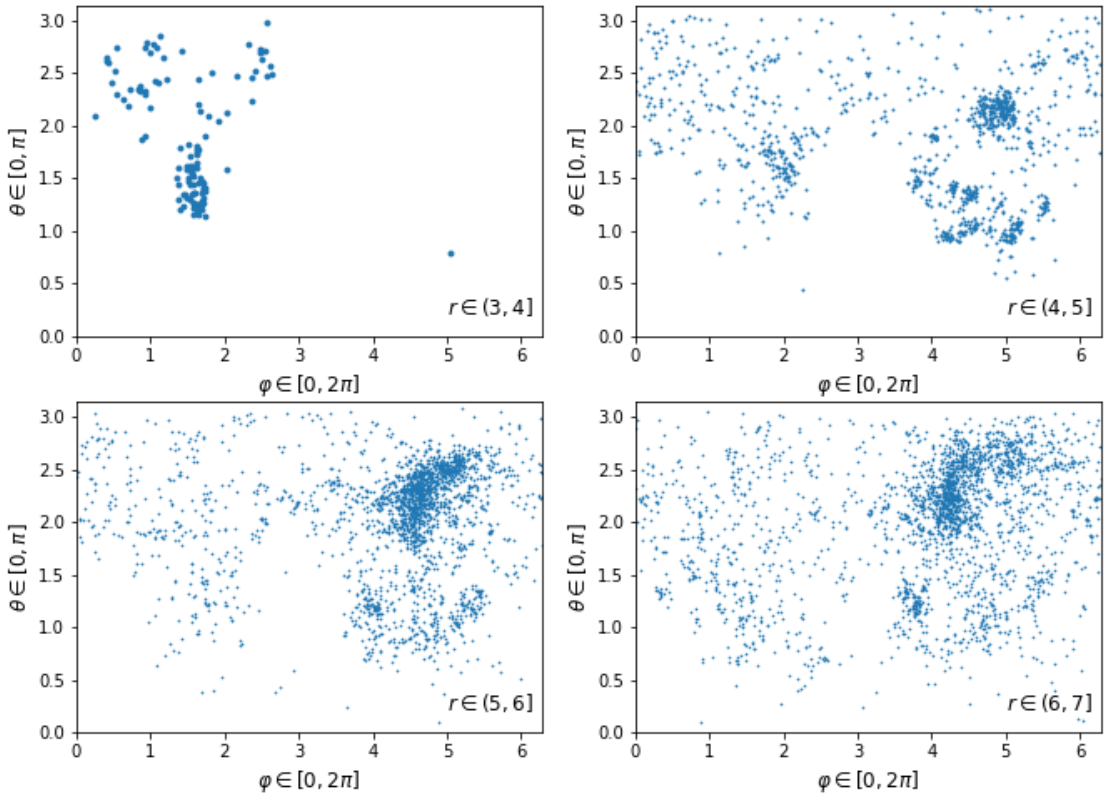

Рис. 4.1: График встречающихся в базе данных пространственных конфигураций аминокислотного пары (C_ar6x, ALA).

Для каждой пары аминокислоты и лиганда на носителе выборки Ω вводится равномерная сетка мелкостью 250 и в каждом узле сетки вычисляется оценка функции распределения по формуле (3.6).

Для непрерывной аппроксимации полученной эмпирической функции распределения используется монотонный 6-слойный перцептрон с размерностями скрытых пространств 10. Подбор оптимальных параметров осуществляется в соответствии с минимизацией MSELoss. После чего плотность распределения пространственной молекулярной конфигурации получается путем дифференцирования модели.

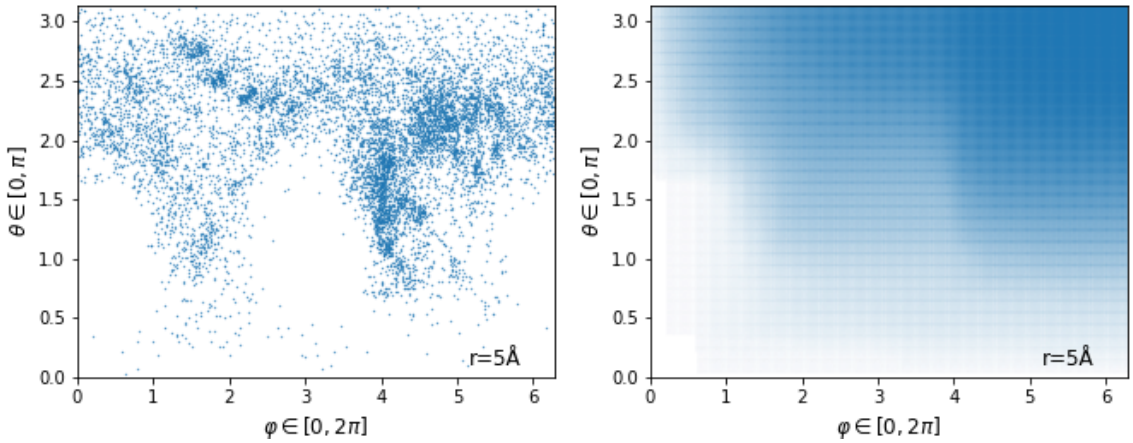

Рис. 4.2: Посчитанная функция распределения для (C_ar6x, ALA).

Ниже приведены результаты восстановления плотности.

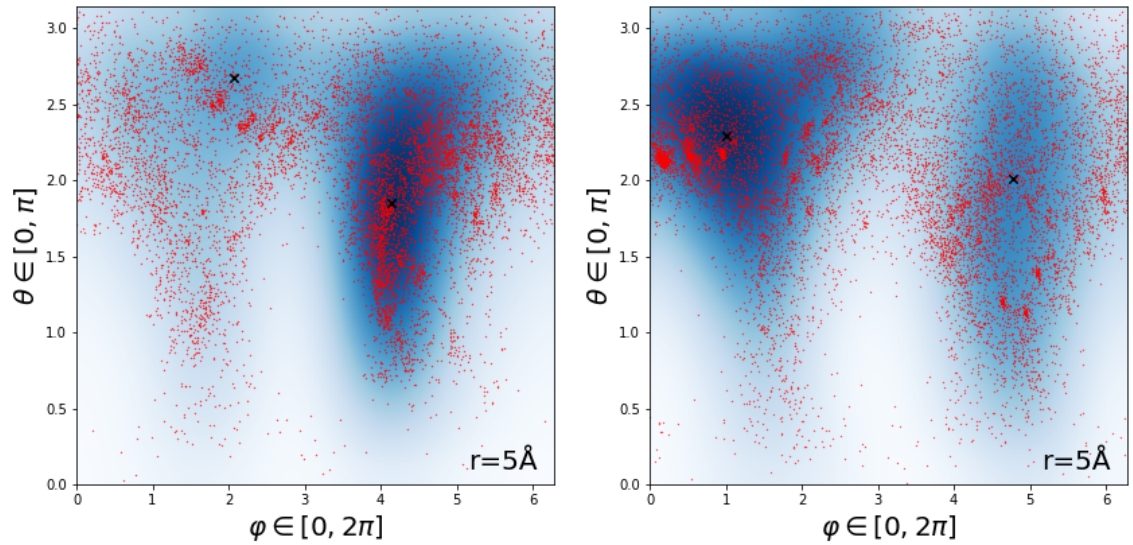

Рис. 4.3: Полученная плотность для пар (C_ar6x, ALA) и (C_ar6x, CYS).

Синим отмечена восстановленная плотность распределения пространственной конфигурации. Красным – встречающиеся в базе данных пространственные конфигурации пары. Черным – найденные экстремумы.

Таким образом, найденные экстремумы позволяют установить устойчивые пространственные конфигурации пары, соответствующие минимуму энергии взаимодействия, которые записываются в специальный каталог в порядке убывания значения экстремума и даются экспертам для проверки адекватности работы метода. Кроме этого полученная плотность используется для ранжирования эталонного множества белков состоящих из одного набора аминокислот. С этой целью строится вероятностно-метрического пространство, элементом которого является вектор из

расстояний между плотностями распределений всех пар аминокислота-лиганд наблюдаемого белка и белка из эталонного множества.

5 Заключение

В данной работе исследовался нейросетевой метод восстановления плотности распределения пространственной молекулярной конфигурации пары аминокислота-лиганд. В ходе исследования был проведен анализ существующих методов и их недостатков, а также предложена модификация многослойного перцептрона, являющаяся универсальным аппроксиматором произвольной неубывающей функции. Эта модификация позволяет получить определенную на всем носителе выборки, аналитически дифференцируемую функцию, которая является непрерывной аппроксимацией эмпирической функции распределения. Также в работе был предложен метод вычисления плотности распределения из полученной аппроксимации функции распределения. В результате эксперимента была проведена реализация описанного метода и показана адекватность его работы на исследуемой выборке.

Список литературы

[1] Andriy Kryshtafovych, Torsten Schwede, Maya Topf, Krzysztof Fidelis, and John Moult. Critical assessment of methods of protein structure prediction (casp) - round xiii. *Proteins*, 2019.

[2] Petr Popov and Sergei Grudinin. Knowledge of native protein-protein interfaces is sufficient to construct predictive models for the selection of binding candidates. *Journal of chemical information and modeling*, 55 10:2242–55, 2015.

[3] Maria Kadukova and Sergei Grudinin. Convex-pl: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization. *Journal of computer-aided molecular design*, 31(10):943—958, October 2017.

[4] Armando D. Solis. Deriving high-resolution protein backbone structure propensities from all crystal data using the information maximization device. *PLoS ONE*, 9, 2014.

[5] José Ramón López-Blanco and Pablo Chacón. Korp: knowledge-based 6d potential for fast protein and loop modeling. *Bioinformatics*, 2019.

[6] M. Magdon-Ismail and A. Atiya. Density estimation and random variate generation using multilayer networks. *IEEE Transactions on Neural Networks*, 13(3):497–520, 2002.

[7] Aristidis Likas. Probability density estimation using artificial neural networks. *Computer Physics Communications*, 135(2):167 – 175, 2001.

[8] Benigno Uria, Marc-Alexandre Côté, Karol Gregor, Iain Murray, and Hugo Larochelle. Neural autoregressive distribution estimation. *CoRR*, abs/1605.02226, 2016.

[9] Shengdong Zhang. From cdf to pdf - a density estimation method for high dimensional data. *CoRR*, abs/1804.05316, 2018.

[10] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314, December 1989.

[11] Bernhard Lang. Monotonic multi-layer perceptron networks as universal approximators. In *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - Volume Part II*, ICANN'05, page 31–37, Berlin, Heidelberg, 2005. Springer-Verlag.

[12] Feng Qi and Bai-Ni Guo. An explicit formula for derivative polynomials of the tangent function. *Acta Universitatis Sapientiae, Mathematica*, 9(2):348 – 359, 2017.

[13] Ali A. Minai and Ronald D. Williams. On the derivatives of the sigmoid. *Neural Networks*, 6(6):845 – 853, 1993.