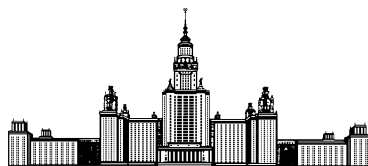


Московский государственный университет имени М.В. Ломоносова



Факультет Вычислительной Математики и Кибернетики
Кафедра Математических Методов Прогнозирования

Выпускная квалификационная работа

«Расстояния, параметризованные размерами, и их аппроксимация»

Выполнила:
студентка 417 группы
Петренко Дарья Павловна

Научный руководитель:
к.ф.-м.н.
Майсурадзе Арчил Ивериевич

Москва, 2020

Оглавление

1	Введение	2
2	Расстояния и задачи их аппроксимации	4
2.1	Основная система аксиом и её независимость	4
2.2	Усиления системы аксиом расстояния	7
2.3	Ослабления системы аксиом расстояния	9
2.4	Цели и задачи работы	9
2.5	Многомерное шкалирование	10
2.5.1	Модель Торгерсона	12
2.5.2	Взвешенная евклидова модель	16
2.5.3	Трехмодальная модель	18
3	Модели расстояний, параметризованных размерами, и их обучение	20
3.1	Модель пропорциональных расстояний	20
3.1.1	Метод 1	21
3.2	Модель пропорциональных конфигураций	22
3.3	Модель индивидуальных различий	23
3.3.1	Метод 2	25
3.4	Модель рядов	25
3.4.1	Метод 3	26
4	Эксперименты	29
4.1	Описание данных	29
4.2	Исследование метода 1	31
4.3	Исследование метода 2	34
4.4	Исследование метода 3	35
4.5	Сравнение и выводы	38
5	Заключение	39
	Список литературы	40

Глава 1

Введение

В числа, описывающие пару объектов, вкладывают много разных смыслов. Будем рассматривать интерпретацию, в которой число характеризует трудоемкость перехода от объекта к объекту. Примеры можно найти в транспортных сетях из различных предметных областей: такой трудоемкостью может являться время перемещения груза между городами или временная задержка при передаче сообщения между процессами в многопроцессорной системе.

Можно заметить, что для приведенных примеров трудоемкость перехода между объектами определяется не только свойствами самих объектов. В первом примере она зависит также от массы перевозимого груза, а во втором — от размера сообщения. Это наводит на мысль о том, что для описания такой трудоемкости недостаточно традиционного понятия расстояния. Нужно использовать функцию от трех аргументов, два из которых характеризуют объекты, а третий — размер груза. Назовем такую функцию расстоянием, параметризованным размером.

Большинство нормально функционирующих транспортных систем удовлетворяет набору очевидных требований. Например, время перемещения любого груза неотрицательно, а время перевозки груза из города до него же самого равно нулю при любом размере этого груза. Также для любой пары городов время перевозки груза между ними не зависит от того, какой из городов был стартовой точкой, а какой конечной. Также очевидно, что тяжелый груз требует больше времени для перемещения, чем легкий, а перевозка любого груза целиком занимает меньше времени, чем по частям.

Несоответствие транспортной системы каким-либо из описанных требований может говорить о наличии неисправностей в ее работе, поэтому важно уметь проверять их выполнение. При определении функции расстояния, параметризованного размером, потребуем, чтобы она удовлетворяла этим условиям. Формализация требований позволит легко проверить, выполняются ли они в реальных системах.

Часто при сборе информации о транспортных системах возникает слишком большое количество данных, иногда даже избыточное, которое сложно хранить и анализировать. Эту проблему можно решить, если ввести интерпретируемые модели расстояний, параметризованных размером, и аппроксимировать реальные данные с помощью этих моделей.

Дальнейшее изложение построено по следующему плану. В главе 2 вводится формальное определение расстояния, параметризованного размером, а также некоторые его сужения и обобщения. Также приводится обзор моделей расстояния и основанных на них методов многомерного шкалирования, дающий теоретическую базу, необходимую для дальнейшего исследования. В главе 3 вводятся конкретные семейства расстояний, пара-

метризованных размером. Для каждого из семейств приводятся достаточные условия на параметры, гарантирующие выполнение всех аксиом расстояния, параметризованного размером. Рассматриваются методы, позволяющие аппроксимировать исходные данные с использованием функций из этих семейств. В главе 4 приводятся результаты применения методов к реальным данным и сравнение методов. Глава 5 содержит основные выводы и результаты, полученные в ходе исследования.

Глава 2

Расстояния и задачи их аппроксимации

2.1 Основная система аксиом и её независимость

Расстояние принято определять как функцию на паре объектов, удовлетворяющую некоторому набору требований. Однако при решении различных задач требования к функции могут меняться, что приводит к существованию целого комплекса обобщений и сужений понятия расстояния. То же самое можно сказать и о расстоянии, параметризованном размером. В данном разделе вводится основное определение расстояния, параметризованного размером. В двух следующих разделах приведены некоторые его обобщения и сужения.

Напомним традиционное определение расстояния для пары объектов.

Определение 1 ([5]). Пусть X — произвольное множество. *Расстоянием* (distance) на X называется функция $\rho : X \times X \rightarrow \mathbb{R}$, удовлетворяющая следующей системе аксиом: $\forall x_1, x_2 \in X$

$$\text{D1. } \rho(x_1, x_1) = 0 \quad (\text{рефлексивность})$$

$$\text{D2. } \rho(x_1, x_2) = \rho(x_2, x_1) \quad (\text{симметричность})$$

$$\text{D3. } \rho(x_1, x_2) \geq 0 \quad (\text{неотрицательность})$$

Согласно [8], система аксиом называется *противоречивой*, если из нее логическим путем можно вывести как некоторое утверждение, так и утверждение, противоречащее первому. Система аксиом называется *непротиворечивой*, если она не является противоречивой.

Для системы аксиом совокупность объектов и отношений, удовлетворяющая всем требованиям этой системы, называется *моделью* этой системы аксиом. Система аксиом называется *совместной*, если она имеет модель. Совместность системы аксиом является достаточным условием ее непротиворечивости.

Указанные требования (D1)—(D3) непротиворечивы. Действительно, классические аксиомы метрики являются более строгими, при этом известно много разных метрик.

Для системы аксиом *независимость* аксиомы от остальных означает, что её замена на отрицание не приводит к противоречию, то есть она не является логическим следствием из множества остальных аксиом. Иными словами, аксиома независима от остальных аксиом в том и только том случае, если существует модель, при которой эта аксиома ложна, а все остальные аксиомы истинны. Построение такой модели является классическим методом

доказательства независимости. Если аксиома независима от всех остальных аксиом, то она тем более независима от любого меньшего их подмножества. Независимость системы аксиом — свойство системы, состоящее в том, что каждая аксиома является независимой от остальных.

Теорема 1. Система аксиом (D1)—(D3) является независимой.

Доказательство. Пусть $X = \{x_1, x_2\}$ — множество произвольных объектов.

- Пример: аксиома (D1) ложна, остальные истинны
 $\rho(x_1, x_1) = 1, \quad \rho(x_2, x_2) = 1, \quad \rho(x_1, x_2) = \rho(x_2, x_1) = 2$
- Пример: аксиома (D2) ложна, остальные истинны
 $\rho(x_1, x_2) = 1, \quad \rho(x_2, x_1) = 2, \quad \rho(x_1, x_1) = 0, \quad \rho(x_2, x_2) = 0$
- Пример: аксиома (D3) ложна, остальные истинны
 $\rho(x_1, x_2) = \rho(x_2, x_1) = -1, \quad \rho(x_1, x_1) = 0, \quad \rho(x_2, x_2) = 0$

□

Обобщим определение расстояния введением третьего аргумента, который будем называть *размером*. Его интуитивный смысл заключается в следующем: если измеряемое значение — время, за которое можно доехать из одного города в другой, то это время может зависеть не только от расстояния между городами в километрах, но и от массы перевозимого груза: чем тяжелее груз, тем дольше едет автомобиль.

Определение 2. Пусть X — произвольное множество. Пусть S — частично упорядоченное множество с операцией сложения. *Расстоянием, параметризованным размером*, будем называть функцию $\rho(x_1, x_2, s) : X \times X \times S \rightarrow \mathbb{R}$, удовлетворяющую следующей системе аксиом:

- аксиомы расстояния $\forall x_1, x_2 \in X, \forall s \in S$
- D1. $\rho(x_1, x_1, s) = 0$ (рефлексивность)
- D2. $\rho(x_1, x_2, s) = \rho(x_2, x_1, s)$ (симметричность)
- D3. $\rho(x_1, x_2, s) \geq 0$ (неотрицательность)
- аксиомы размера $\forall x_1, x_2 \in X, \forall s_1, s_2 \in S$
- S1. $s_1 \leq s_2 \implies \rho(x_1, x_2, s_1) \leq \rho(x_1, x_2, s_2)$ (монотонность)
- S2. $\rho(x_1, x_2, s_1 + s_2) \leq \rho(x_1, x_2, s_1) + \rho(x_1, x_2, s_2)$ (неделимость)

Аксиома (S1) означает, что для каждой фиксированной пары объектов функция должна быть возрастающей от размера. Аксиома (S2) означает, что для каждой фиксированной пары объектов функция должна быть субаддитивной от размера.

Аксиомы размера имеют естественную интерпретацию для транспортных систем. Монотонность говорит о том, что тяжелый груз требует для перевозки больше времени, чем легкий. Неделимость — о том, что большой груз невыгодно перевозить частями.

Очевидно, что для каждого фиксированного размера расстояние, параметризованное размером, является просто расстоянием. В математике есть традиционное понятие

мультиметрического пространства. В определении 2 требуется больше, чем просто «мультирасстояние»: на размерах есть порядок; аксиомы связывают между собой расстояния для разных размеров.

Предложение 2. Требования (D1)—(D3), (S1)—(S2) непротиворечивы.

Доказательство. Приведем пример, для которого все аксиомы (D1)—(D3), (S1)—(S2) выполнены. Возьмем $S = \{1, 2, 3\} \subset \mathbb{R}$, $X = \{x_1, x_2\}$ — множество произвольных объектов.

$$\rho(x_1, x_1, 1) = \rho(x_1, x_1, 2) = \rho(x_1, x_1, 3) = 0$$

$$\rho(x_2, x_2, 1) = \rho(x_2, x_2, 2) = \rho(x_2, x_2, 3) = 0$$

$$\rho(x_1, x_2, 1) = \rho(x_2, x_1, 1) = 2$$

$$\rho(x_1, x_2, 2) = \rho(x_2, x_1, 2) = 3$$

$$\rho(x_1, x_2, 3) = \rho(x_2, x_1, 3) = 4$$

□

Теорема 3. Система аксиом (D1)—(D3), (S1)—(S2) является независимой.

Доказательство. Возьмем $S = \{1, 2, 3\} \subset \mathbb{R}$, $X = \{x_1, x_2\}$ — множество произвольных объектов.

- Пример: (D1) ложна, остальные истинны

$$\rho(x_1, x_1, 1) = \rho(x_1, x_1, 2) = \rho(x_1, x_1, 3) = 1$$

$$\rho(x_2, x_2, 1) = \rho(x_2, x_2, 2) = \rho(x_2, x_2, 3) = 1$$

$$\rho(x_1, x_2, 1) = \rho(x_2, x_1, 1) = 2$$

$$\rho(x_1, x_2, 2) = \rho(x_2, x_1, 2) = 3$$

$$\rho(x_1, x_2, 3) = \rho(x_2, x_1, 3) = 4$$

- Пример: (D2) ложна, остальные истинны

$$\rho(x_1, x_1, 1) = \rho(x_1, x_1, 2) = \rho(x_1, x_1, 3) = 0$$

$$\rho(x_2, x_2, 1) = \rho(x_2, x_2, 2) = \rho(x_2, x_2, 3) = 0$$

$$\rho(x_1, x_2, 1) = 2, \quad \rho(x_2, x_1, 1) = 5$$

$$\rho(x_1, x_2, 2) = 3, \quad \rho(x_2, x_1, 2) = 6$$

$$\rho(x_1, x_2, 3) = 4, \quad \rho(x_2, x_1, 3) = 7$$

- Пример: (S1) ложна, остальные истинны

$$\rho(x_1, x_1, 1) = \rho(x_1, x_1, 2) = \rho(x_1, x_1, 3) = 0$$

$$\rho(x_2, x_2, 1) = \rho(x_2, x_2, 2) = \rho(x_2, x_2, 3) = 0$$

$$\rho(x_1, x_2, 1) = \rho(x_2, x_1, 1) = 3$$

$$\rho(x_1, x_2, 2) = \rho(x_2, x_1, 2) = 2$$

$$\rho(x_1, x_2, 3) = \rho(x_2, x_1, 3) = 1$$

- Пример: (S2) ложна, остальные истинны

$$\rho(x_1, x_1, 1) = \rho(x_1, x_1, 2) = \rho(x_1, x_1, 3) = 0$$

$$\rho(x_2, x_2, 1) = \rho(x_2, x_2, 2) = \rho(x_2, x_2, 3) = 0$$

$$\rho(x_1, x_2, 1) = \rho(x_2, x_1, 1) = 1$$

$$\rho(x_1, x_2, 2) = \rho(x_2, x_1, 2) = 2$$

$$\rho(x_1, x_2, 3) = \rho(x_2, x_1, 3) = 4$$

Пусть теперь $S = \{-1, -2, -3\} \subset \mathbb{R}$, $X = \{x_1, x_2\}$ — множество произвольных объектов.

- Пример: (D3) ложна, остальные истинны
 $\rho(x_1, x_1, -1) = \rho(x_1, x_1, -2) = \rho(x_1, x_1, -3) = 0$
 $\rho(x_2, x_2, -1) = \rho(x_2, x_2, -2) = \rho(x_2, x_2, -3) = 0$
 $\rho(x_1, x_2, -1) = \rho(x_2, x_1, -1) = -1$
 $\rho(x_1, x_2, -2) = \rho(x_2, x_1, -2) = -2$
 $\rho(x_1, x_2, -3) = \rho(x_2, x_1, -3) = -4$

□

2.2 Усиления системы аксиом расстояния

Приведенное выше определение расстояния можно усилить («сузить») путём введения разных новых аксиом [5]. Например, в такие сужения входят стандартные определения псевдометрики и метрики.

Определение 3. ([5]) Пусть X — произвольное множество. Функция $\rho : X \times X \rightarrow \mathbb{R}$ называется *псевдометрикой*, если она удовлетворяет определению 1 и справедливо $\forall x_1, x_2, x_3 \in X$

$$D4. \rho(x_1, x_2) \leq \rho(x_1, x_3) + \rho(x_2, x_3) \quad (\text{неравенство треугольника})$$

Известно, что система аксиом (D1)—(D4) уже не является независимой:

- неотрицательность можно вывести из остальных аксиом,
- симметричность можно вывести из остальных аксиом.

Определение 4. ([5]) Функция $\rho : X \times X \rightarrow \mathbb{R}$ называется *метрикой*, если она удовлетворяет определению 3 и справедливо $\forall x_1, x_2 \in X$

$$D5. \rho(x_1, x_2) = 0 \Leftrightarrow x_1 = x_2 \quad (\text{тождественность неразличимых})$$

При этом пара (X, ρ) называется *метрическим пространством*, а множество X называется *подлежащим множеством* метрического пространства. Элементы метрического пространства называются *точками*.

Условие (D5) независимо от (D1)—(D4). Действительно, существуют псевдометрики, не являющиеся метриками.

Аналогично тому, как путем добавления новых аксиом вводятся сужения расстояния, введем сужения расстояния, параметризованного размером. Снова пусть X — произвольное множество, S — частично упорядоченное множество с операцией сложения.

Определение 5. Функция $\rho : X \times X \times S \rightarrow \mathbb{R}$ называется *псевдометрикой, параметризованной размером*, если она удовлетворяет 2 и справедливо $\forall x_1, x_2, x_3 \in X, \forall s \in S$

$$D4. \rho(x_1, x_2, s) \leq \rho(x_1, x_3, s) + \rho(x_2, x_3, s) \quad (\text{неравенство треугольника})$$

Определение 6. Функция $\rho : X \times X \times S \rightarrow \mathbb{R}$ называется **метрикой, параметризованной размером**, если она удовлетворяет определению 5 и справедливо $\forall x_1, x_2 \in X, \forall s \in S$

$$D5. \rho(x_1, x_2, s) = 0 \Leftrightarrow x_1 = x_2 \quad (\text{тождественность неразличимых})$$

Предложение 4. Система аксиом (D1)—(D5), (S1)—(S2) непротиворечива.

Доказательство. Приведем пример, когда все аксиомы выполняются. Пусть $S = \{1, 2, 3\} \subset \mathbb{R}$, $X = \{x_1, x_2, x_3\}$ — множество произвольных элементов.

$$\rho(x_1, x_1, 1) = \rho(x_1, x_1, 2) = \rho(x_1, x_1, 3) = 0$$

$$\rho(x_2, x_2, 1) = \rho(x_2, x_2, 2) = \rho(x_2, x_2, 3) = 0$$

$$\rho(x_3, x_3, 1) = \rho(x_3, x_3, 2) = \rho(x_3, x_3, 3) = 0$$

$$\rho(x_1, x_2, 1) = \rho(x_2, x_1, 1) = 1$$

$$\rho(x_1, x_2, 2) = \rho(x_2, x_1, 2) = 2$$

$$\rho(x_1, x_2, 3) = \rho(x_2, x_1, 3) = 3$$

$$\rho(x_1, x_3, 1) = \rho(x_3, x_1, 1) = 2$$

$$\rho(x_1, x_3, 2) = \rho(x_3, x_1, 2) = 3$$

$$\rho(x_1, x_3, 3) = \rho(x_3, x_1, 3) = 4$$

$$\rho(x_2, x_3, 1) = \rho(x_3, x_2, 1) = 3$$

$$\rho(x_2, x_3, 2) = \rho(x_3, x_2, 2) = 4$$

$$\rho(x_2, x_3, 3) = \rho(x_3, x_2, 3) = 5$$

□

Предложение 5. Аксиома (S1) независима от системы аксиом (D1)—(D5), (S2). Аксиома (S2) независима от системы аксиом (D1)—(D5), (S1).

Доказательство. Пусть $S = \{1, 2, 3\} \subset \mathbb{R}$, $X = \{x_1, x_2, x_3\}$ — множество произвольных элементов.

- Пример: (S1) ложна, остальные истинны

$$\rho(x_1, x_1, 1) = \rho(x_1, x_1, 2) = \rho(x_1, x_1, 3) = 0$$

$$\rho(x_2, x_2, 1) = \rho(x_2, x_2, 2) = \rho(x_2, x_2, 3) = 0$$

$$\rho(x_3, x_3, 1) = \rho(x_3, x_3, 2) = \rho(x_3, x_3, 3) = 0$$

$$\rho(x_1, x_2, 1) = \rho(x_2, x_1, 1) = 5$$

$$\rho(x_1, x_2, 2) = \rho(x_2, x_1, 2) = 4$$

$$\rho(x_1, x_2, 3) = \rho(x_2, x_1, 3) = 3$$

$$\rho(x_1, x_3, 1) = \rho(x_3, x_1, 1) = 4$$

$$\rho(x_1, x_3, 2) = \rho(x_3, x_1, 2) = 3$$

$$\rho(x_1, x_3, 3) = \rho(x_3, x_1, 3) = 2$$

$$\rho(x_2, x_3, 1) = \rho(x_3, x_2, 1) = 3$$

$$\rho(x_2, x_3, 2) = \rho(x_3, x_2, 2) = 2$$

$$\rho(x_2, x_3, 3) = \rho(x_3, x_2, 3) = 1$$

- Пример: (S2) ложна, остальные истинны

$$\rho(x_1, x_1, 1) = \rho(x_1, x_1, 2) = \rho(x_1, x_1, 3) = 0$$

$$\rho(x_2, x_2, 1) = \rho(x_2, x_2, 2) = \rho(x_2, x_2, 3) = 0$$

$$\rho(x_3, x_3, 1) = \rho(x_3, x_3, 2) = \rho(x_3, x_3, 3) = 0$$

$$\rho(x_1, x_2, 1) = \rho(x_2, x_1, 1) = 1$$

$$\rho(x_1, x_2, 2) = \rho(x_2, x_1, 2) = 2$$

$$\rho(x_1, x_2, 3) = \rho(x_2, x_1, 3) = 4$$

$$\rho(x_1, x_3, 1) = \rho(x_3, x_1, 1) = 2$$

$$\rho(x_1, x_3, 2) = \rho(x_3, x_1, 2) = 3$$

$$\rho(x_1, x_3, 3) = \rho(x_3, x_1, 3) = 6$$

$$\begin{aligned}\rho(x_2, x_3, 1) &= \rho(x_3, x_2, 1) = 3 \\ \rho(x_2, x_3, 2) &= \rho(x_3, x_2, 2) = 4 \\ \rho(x_2, x_3, 3) &= \rho(x_3, x_2, 3) = 8\end{aligned}$$

□

2.3 Ослабления системы аксиом расстояния

Хотя определение 1 содержит только три простых условия, даже они в реальных транспортных сетях из различных предметных областей не выполняются. Например, при передаче сообщений в компьютерных сетях время передачи информации с узла на тот же самый узел больше нуля, что противоречит аксиоме рефлексивности. Симметричность, а также обе аксиомы размера из определения 2 также часто нарушаются. Получается, что гарантированно выполняется только неотрицательность.

Чтобы ослабить набор требований, введем расширение определения расстояния, параметризованного размером. Такие ослабления требований для обычных расстояний часто называют *различиями*. Соответственно, в данной работе будет рассматриваться *различие, параметризованное размером*. Снова пусть X — произвольное множество, S — частично упорядоченное множество с операцией сложения.

Определение 7. Функция $\rho : X \times X \times S \rightarrow \mathbb{R}$ называется *различием, параметризованным размером*, если $\forall x_1, x_2 \in X, \forall s \in S$ она удовлетворяет аксиомам

$$D1. \rho(x_1, x_1, s) = 0 \quad (\text{рефлексивность})$$

$$D3. \rho(x_1, x_2, s) \geq 0 \quad (\text{неотрицательность})$$

2.4 Цели и задачи работы

Введенные выше понятия позволяют сформулировать цели и задачи работы.

Пусть дан конечный набор объектов мощности N и конечный набор размеров мощности K . Под набором мы понимаем такую коллекцию, в которой каждый элемент имеет уникальный индекс от 1 до мощности набора, хотя значения элементов могут повторяться. Такие коллекции в теории программирования часто называют векторами. Ниже на объекты и размеры мы будем ссылаться по их индексам. В большинстве прикладных задач размеры пропорциональны своим индексам в наборе.

Для всех пар объектов и всех размеров измерено различие δ . Получается трехмерный тензор $\Delta \in \mathbb{R}^{N \times N \times K}$ такой, что элемент δ_{ijk} содержит измеренное различие между i -тым и j -тым объектами, параметризованное расстоянием с индексом k .

Отметим, что в реальных прикладных задачах в указанном тензоре различий некоторые элементы могут отсутствовать: теоретически они существуют, но измерения не проводились.

Цель данной работы состоит в том, чтобы разработать и исследовать методы «сжатия» этого тензора различий с возможностью последующего восстановления с потерями. При этом предполагается, что желательно гарантировать выполнение аксиом из определения 2.

Для этого в работе решаются следующие задачи:

- проводится систематизация формальных постановок,
- вводятся модели расстояний, параметризованных размерами,
- разрабатываются конкретные методы, основанные на этих моделях,
- выполняется реализация методов,
- проводятся эксперименты на реальных данных,
- проводится сравнение методов.

Предполагается, что различия, параметризованные размерами, будут аппроксимироваться расстояниями, параметризованными размерами. Необходимо отметить, что предложенная аксиоматизация расстояний, параметризованных размерами, является новой. Соответственно, все предлагаемые методы будут новыми, хотя и будут рассмотрены аналоги методов для традиционных расстояний.

2.5 Многомерное шкалирование

Рассматриваемые в данной работе постановки задач для расстояний, параметризованных размерами, в случае обычных расстояний относятся к классу задач многомерного шкалирования. Поэтому в данном разделе на основе [1], [2], [6] проводится обзор этого класса задач. Важной частью этих задач являются модели расстояний.

Везде далее будем использовать следующее обозначение для евклидовой метрики:

$$\text{eucl}(x, y) = \sqrt{\sum_{k=1}^K (x_k - y_k)^2} \quad (2.1)$$

Множество натуральных чисел $1, \dots, N$ обозначим через $[1, N]$.

Пусть дан конечный набор объектов мощности N . Под набором снова будем понимать коллекцию, в которой каждый элемент имеет уникальный индекс от 1 до мощности набора, при этом значения элементов могут повторяться. На объекты будем ссылаться по индексам.

Пусть для каждой пары объектов измерено различие. Полученные значения образуют двумерную матрицу $\Delta \in \mathbb{R}^{N \times N}$, в которой элемент δ_{ij} задает каким-то образом измеренное различие между i -тым и j -тым объектами.

В задаче многомерного шкалирования предполагается, что известна только матрица Δ . Требуется найти такие представления для исходных объектов в некотором метрическом пространстве X с расстоянием dist , чтобы расстояние между представлениями i -того и j -того объектов максимально «согласовалось» с элементами δ_{ij} матрицы Δ . Полученные представления $x_i, \dots, x_N \in X$ называют *точечной конфигурацией*.

В некоторых постановках различие для каждой пары объектов измеряется несколькими способами, и полученные значения образуют несколько двумерных матриц $\Delta_k \in \mathbb{R}^{N \times N}$, $k \in [1, K]$. При этом, в отличие от совокупности двумерных срезов тензора различий, параметризованных размерами, не предполагается наличие каких-либо отношений или операций на множестве матриц. Примером такой ситуации может послужить получение сведений об объектах от нескольких независимых экспертов. Тогда для каждой из матриц ищется своя точечная конфигурация в некотором метрическом пространстве.

В зависимости от свойств пространства исходных объектов и пространства, в котором ищется решение, от количества входных матриц различий, а также от трактовки понятия «согласования» можно выделить несколько подклассов задач многомерного шкалирования. Так, исходное пространство может быть евклидовым, иным метрическим или неметрическим. Матриц различий может быть одна или несколько. Пространство решений может быть евклидовым или иным метрическим. Можно требовать точного или приближенного сохранения значений матрицы или их рангов.

В частности, если требуется сохранение значений матрицы Δ , то говорят о метрическом многомерном шкалировании. Если требуется сохранение их рангов, то о неметрическом. Если входных матриц различий несколько, то говорят о многомерном шкалировании индивидуальных различий.

Для решения задач каждого из подклассов существуют различные методы. Конкретный метод многомерного шкалирования включает в себя описание модели расстояния (множества нужной размерности и семейства функций расстояния, определенных на этом множестве), в рамках которой будет осуществляться поиск аппроксимации исходных различий. Также указывается, какое решение считается оптимальным, и приводится конкретный алгоритм поиска этого решения. При необходимости описываются ограничения на исходное пространство, необходимые для корректной работы алгоритма.

Условие оптимальности можно задать при помощи функции потерь, которую в рамках терминологии многомерного шкалирования принято называть *функцией стресса*. Примером функции стресса для метрического многомерного шкалирования может служить норма Фробениуса между матрицей различий Δ и матрицей попарных расстояний точечной конфигурации

$$S(X) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} - \text{dist}(x_i, x_j))^2}$$

Согласно [6], в случае, когда исходное пространство евклидово, от матрицы Δ можно перейти к матрице Грама B для исходных объектов. Так как евклидова метрика инвариантна относительно сдвигов, то для устранения неоднозначности будем искать матрицу Грама в предположении, что центроид исходных представлений объектов лежит в начале координат.

Введем обозначение

$$A = \{a_{ij}\}, \quad a_{ij} = -\frac{1}{2}\delta_{ij}^2$$

Тогда искомая матрица вычисляется по формуле

$$B = H \times A \times H,$$

где H - центрирующая матрица:

$$H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T, \quad \mathbf{1} = (1, \dots, 1)^T - \text{вектор из } n \text{ единиц}$$

Тогда если решение ищется в евклидовом пространстве, то в качестве функции стресса можно рассматривать норму Фробениуса между матрицами Грама для исходных объектов и для построенной точечной конфигурации

$$S(X) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (b_{ij} - \langle x_i, x_j \rangle)^2},$$

где $\langle x_i, x_j \rangle$ - скалярное произведение между представлениями i -го и j -го объектов.

Примером функции стресса в задачах неметрического многомерного шкалирования может служить следующая функция:

$$S(X) = \sqrt{\sum_{i=1}^N \sum_{j=1}^N (f(\delta_{ij}) - \text{dist}(x_i, x_j))^2},$$

где f - произвольная монотонная функция такая, что $\forall i, j, i', j' \in [1, N]$

$$\delta_{ij} < \delta_{i'j'} \Rightarrow f(\delta_{ij}) < f(\delta_{i'j'}) \quad (2.2)$$

Как правило, функция f не имеет формульного представления и задается поточечно. Методы неметрического шкалирования неприменимы для задач сжатия матрицы с последующим восстановлением значений, так как помимо построенной точечной конфигурации потребовалось бы хранение значений функции f в N^2 точек, что по объему занимаемой памяти равносильно хранению исходной матрицы Δ . Поэтому в данной работе методы неметрического многомерного шкалирования рассмотрены не будут.

В задачах многомерного шкалирования индивидуальных различий в качестве функции стресса может использоваться сумма функций стресса отдельных матриц.

Для измерения итогового качества аппроксимации исходной матрицы различий матрицей попарных расстояний построенной точечной конфигурации часто используется функция стресса Stress-1 [1], которая не зависит от абсолютных величин различий:

$$\text{Stress} - 1 = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^N (\delta_{ij} - \text{dist}(x_i, x_j))^2}{\sum_{i=1}^N \sum_{j=1}^N \text{dist}^2(x_i, x_j)}} \quad (2.3)$$

В случае, когда исходных матриц несколько, она имеет вид

$$\text{Stress} - 1 = \sqrt{\frac{\sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N (\delta_{ijk} - \text{dist}(x_i^k, x_j^k))^2}{\sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N \text{dist}^2(x_i^k, x_j^k)}} \quad (2.4)$$

Стоит отметить, что реальных задачах некоторые значения в матрице Δ могут отсутствовать: теоретически они существуют, но измерения не проводились. Возможность получения решения для матрицы с отсутствующими значениями является важной характеристикой методов.

Рассмотрим несколько моделей метрического многомерного шкалирования и примеры основанных на них методов.

2.5.1 Модель Торгерсона

Одной простейших моделей многомерного шкалирования является модель, предложенная Торгерсоном [2]. Она предполагает, что различия аппроксимируются расстояниями в многомерном евклидовом пространстве.

Классическое многомерное шкалирование

На модели Торгерсона основан метод классического многомерного шкалирования, подробно описанный в [2]. Метод применяется для одной матрицы различий Δ в предположении, что исходное пространство евклидово. Алгоритм может обеспечить как точное, так

и приближенное сохранение значений. При этом для приближенного сохранения следует задать размерность искомого представления. Пропуски в данных алгоритм не обрабатывает.

Как было сказано выше, в случае, когда исходное пространство евклидово, можно перейти от матрицы попарных различий Δ к матрице Грама B . Будем считать оптимальным решение, минимизирующее норму Фробениуса между матрицами Грама исходного представления объектов и построенной точечной конфигурации x_1, \dots, x_N

$$S(X) = \sqrt{\sum_{i=1}^N \sum_{j=1}^N (b_{ij} - \langle x_i, x_j \rangle)^2} \quad (2.5)$$

В случае, когда исходное пространство евклидово, матрица Грама B - симметричная неотрицательно определенная. Тогда все ее собственные значения неотрицательны, и существует ортонормированная система соответствующих им собственных векторов. То есть спектральное разложение матрицы B можно искать в следующем виде:

$$B = V \Lambda V^T, \quad (2.6)$$

где $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ - диагональная матрица неотрицательных собственных значений, $V = [v_1, \dots, v_n]$ - ортонормированная матрица соответствующих собственных векторов. Без потери общности будем считать, что $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$.

Пусть требуется точное сохранение значений исходной матрицы. Пусть B имеет ранг p . Тогда $\lambda_{p+1} = \dots = \lambda_n = 0$, и верна формула

$$B = V_1 \Lambda_1 V_1^T, \quad (2.7)$$

где $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_p)$, $V_1 = [v_1, \dots, v_p]$.

Тогда можно выбрать

$$X = V_1 \Lambda_1^{\frac{1}{2}}, \quad \Lambda_1^{\frac{1}{2}} = \text{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_p^{\frac{1}{2}}) \quad (2.8)$$

где i -й столбец матрицы X является представлением i -го объекта. Матрица Грама полученной точечной конфигурации будет в точности равна матрице B , и будет оптимальной в смысле выбранной функции стресса.

Пусть требуется приближенное сохранение значений исходной матрицы, при этом точечная конфигурация ищется в пространстве размерности L . Тогда обозначим

$$\Lambda_2 = \text{diag}(\lambda_1, \dots, \lambda_L), \quad V_2 = [v_1, \dots, v_L] \quad (2.9)$$

В [3] показано, что оптимальным в смысле функции стресса (2.5) решением будет

$$X = V_2 \Lambda_2^{\frac{1}{2}}, \quad \Lambda_2^{\frac{1}{2}} = \text{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_L^{\frac{1}{2}}) \quad (2.10)$$

Стоит отметить, что при невыполнении предположения о том, что исходное пространство евклидово, алгоритм неприменим. В [4] показано, что в этом случае матрица Грама B будет иметь отрицательные собственные значения, и попытки извлечь из них квадратный корень приведут к некорректному результату.

В силу того, что в реальных задачах предположение о евклидовости исходного пространства почти всегда не выполнено, классическое многомерное шкалирование используется в основном для получения начального приближения для других итерационных алгоритмов. При этом отрицательные собственные значения матрицы Грама приравниваются к нулю.

Метод SMACOF

Другим методом, основанным на модели Торгерсона, является SMACOF [2]. Метод также применяется для одной входной матрицы различий Δ , но не накладывает ограничений на исходное пространство. Метод обеспечивает приближенное сохранение значений входной матрицы. Также метод обеспечивает возможность присвоения веса w_{ij} каждой компоненте матрицы различий. Если некоторые значения в матрице отсутствуют, то можно заменить их на произвольные числа, а соответствующие веса приравнять к нулю. Тогда при оптимизации эти компоненты не будут учитываться.

Оптимальным считается решение, минимизирующее взвешенную сумму квадратов разностей компонент исходной матрицы различий и матрицы попарных расстояний построенной точечной конфигурации. То есть функция стресса имеет вид

$$S(X) = \sum_{i=1}^N \sum_{j=1}^N w_{ij} (\text{eucl}(x_i, x_j) - \delta_{ij})^2 \quad (2.11)$$

Для описания алгоритма минимизации функции стресса введем новое понятие. Пусть X, Z - произвольные множества, $f : X \rightarrow \mathbb{R}$, $g : X \times Z \rightarrow \mathbb{R}$.

Определение 8. Функция $g(x, \xi)$ называется **вариационной верхней оценкой** на функцию $f(x)$, если верно следующее:

- (1) $\forall x \in X, \quad \forall \xi \in Z \quad f(x) \leq g(x, \xi)$
- (2) $\forall x_0 \in X \quad \exists \xi \in Z : \quad f(x_0) = g(x_0, \xi)$

Выберем начальное приближение $x^{[0]}$. Тогда минимизация функции f может осуществляться итерационно следующим образом:

step 1: $\xi^{[t+1]} : g(x^{[t]}, \xi^{[t+1]}) = f(x^{[t]})$

step 2: $x^{[t+1]} = \arg \min_x g(x, \xi^{[t+1]})$

Так как

$$f(x^{[t]}) = g(x^{[t]}, \xi^{[t+1]}) \geq g(x^{[t+1]}, \xi^{[t+1]}) \geq f(x^{[t+1]})$$

то полученная последовательность приближений $\{x^{[t]}\}$ обеспечивает монотонное неубывание функции $f(x)$.

Найдем вариационную верхнюю оценку для функции стресса (2.11). Раскрыв скобки, можно представить функцию в следующем виде:

$$\begin{aligned} S(X) &= \sum_{i=1}^N \sum_{j=1}^N w_{ij} \delta_{ij}^2 + \sum_{i=1}^N \sum_{j=1}^N w_{ij} \text{eucl}^2(x_i, x_j) - 2 \sum_{i=1}^N \sum_{j=1}^N w_{ij} \delta_{ij} \text{eucl}(x_i, x_j) = \\ &= \eta_s^2 + \eta_d^2(X) - 2\rho(X) \end{aligned} \quad (2.12)$$

Обозначим $A_{ij} = (e_i - e_j)(e_i - e_j)^T$. Далее, можно преобразовать слагаемые:

$$\eta_d^2(X) = \sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - x_j)^T V (x_i - x_j) = \text{tr}(X^T V X),$$

$$\text{где } V = \sum_{i=1}^n \sum_{j=1}^n w_{ij} A_{ij} \quad (2.13)$$

$$\begin{aligned} \rho(X) &= \sum_{i=1}^N \sum_{j=1}^N \delta_{ij} \text{eucl}(x_i, x_j) = \sum_{i=1}^N \sum_{j=1}^N \frac{\delta_{ij}}{\text{eucl}(x_i, x_j)} \text{eucl}^2(x_i, x_j) = \\ &= \sum_{i=1}^N \sum_{j=1}^N \frac{\delta_{ij}}{\text{eucl}(x_i, x_j)} (x_i - x_j)^T (x_i - x_j) = \text{tr}(X^T B(X) X), \end{aligned}$$

$$\text{где } B(X) = \sum_{i=1}^N \sum_{j=1}^N s_{ij} A_{ij}, \quad s_{ij}(X) = \begin{cases} \frac{w_{ij} \delta_{ij}}{\text{eucl}(x_i, x_j)}, & \text{eucl}(x_i, x_j) \neq 0 \\ 0, & \text{eucl}(x_i, x_j) = 0 \end{cases} \quad (2.14)$$

Тогда функцию стресса можно представить в следующем виде:

$$S(X) = \eta_\delta^2 + \text{tr}(X^T V X) - 2 \text{tr}(X^T B(X) X) \quad (2.15)$$

Верхней вариационной оценкой для нее является функция

$$T(X, Y) = \eta_\delta^2 + \text{tr}(X^T X) - 2 \text{tr}(X^T B(Y) Y) \quad (2.16)$$

Покажем это. Пусть $\tilde{\rho}(X, Y) = \text{tr}(X^T B(Y) Y)$.

$$\begin{aligned} \rho(X) &= \sum_{i=1}^N \sum_{j=1}^N w_{ij} \delta_{ij} \sqrt{(x_i - x_j)^T (x_i - x_j)} = \\ &= \sum_{i=1}^N \sum_{j=1}^N \frac{w_{ij} \delta_{ij}}{\text{eucl}(y_i, y_j)} \sqrt{(x_i - x_j)^T (x_i - x_j) (y_i - y_j)^T (y_i - y_j)} = \\ &= \{ \text{неравенство Коши-Буняковского-Шварца} \} \geq \\ &\geq \sum_{i=1}^N \sum_{j=1}^N \frac{w_{ij} \delta_{ij}}{\text{eucl}(y_i, y_j)} (x_i - x_j)^T (y_i - y_j) = \tilde{\rho}(X, Y) \end{aligned}$$

Тогда

$$S(X) - T(X, Y) = -2\rho(X) + 2\tilde{\rho}(X, Y) = 2(\tilde{\rho}(X, Y) - \rho(X)) \leq 0 \quad (2.17)$$

Также легко видеть, что $T(X, X) = S(X)$. Следовательно, $T(X, Y)$ - верхняя вариационная оценка для $S(X)$ по определению.

$T(X, Y)$ - квадратичная функция относительно переменной X , следовательно, она достигает минимума по X в единственной точке, в которой производная равна 0.

$$\frac{\partial T(X, Y)}{\partial X} = 2VX - 2B(Y)Y = 0$$

Матрица V вырождена, так как сумма ее элементов в каждой строке и в каждом столбце равна 0. Следовательно, для нее не существует обратной. Будем считать приближенно, используя псевдообращение Мура-Пенроуза. Пусть V^+ - псевдообратная для V . Тогда

$$\hat{X} = V^+ B(Y) Y \quad (2.18)$$

Так как $T(X, X) = S(X)$, то $Y^{[t+1]} = X^{[t]}$. Следовательно, выбрав начальное приближение $X^{[0]}$, каждое следующее можно получить по следующей формуле:

$$X^{[t+1]} = V^+ B(X^{[t]}) X^{[t]} \quad (2.19)$$

Алгоритм гарантирует монотонное неубывание значения функции стресса при росте t .

Можно заметить, что диагональные элементы исходной матрицы различий не учитываются при оптимизации, а для каждой пары недиагональных элементов δ_{ij} , δ_{ji} учитывается только их сумма. Поэтому во время оптимизации в функции стресса можно суммировать только по $j < i$, а на вход подавать матрицу $\Delta + \Delta^T$ с весами $w + w^T$, при этом отсутствующие значения в матрице Δ заменив нулями.

2.5.2 Взвешенная евклидова модель

Описание модели приводится согласно [6]. Модель была разработана для решения задач многомерного шкалирования индивидуальных различий, когда на вход подается несколько матриц попарных различий $\Delta_1, \dots, \Delta_K$. Взвешенная евклидова модель является обобщением модели Торгерсона.

В евклидовом пространстве некоторой размерности L строится одна («общая», «групповая») конфигурация для объектов, представленная в виде матрицы $X = [x_1, \dots, x_N]^T$. Будем искать точеную конфигурацию X_k для k -той матрицы различия в виде

$$X_k = X \times \text{diag}(r_k), \quad (2.20)$$

где r_k - вектор весов размерности L . Согласно [6], эти веса иногда называют весами важности или весами характеристик.

Тогда расстояние между представлениями i -го и j -го объектов в k -той точечной конфигурации будет иметь вид

$$\text{eucl}(x_i \times \text{diag}(r_k), x_j \times \text{diag}(r_k)) = \sqrt{\sum_{l=1}^L r_{kl}^2 (x_{il} - x_{jl})^2} \quad (2.21)$$

Метод INDSCAL

На взвешенной евклидовой модели основан метод INDSCAL [2]. Метод применяется для K матриц различий $\Delta_1, \dots, \Delta_K$ в предположении, что исходное пространство евклидово. Метод обеспечивает приближенное сохранение значений матриц, пропуски в данных не обрабатывает.

Так как исходное пространство евклидово, то можно перейти от матриц попарных различий к матрицам Грама B_1, \dots, B_K . Оптимальным считается решение, минимизирующее сумму квадратов норм Фробениуса между матрицами Грама исходных представлений объектов и построенных точечных конфигураций.

$$S(X, r) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N \left(b_{ijk} - \sum_{l=1}^L r_{lk} x_{il} x_{jl} \right)^2 \quad (2.22)$$

Функционал квадратичный по r , но не является квадратичным по X . Введем переменные X^R и X^L , запишем функционал в следующем виде:

$$S(X, r) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N \left(b_{ijk} - \sum_{l=1}^L r_{lk} x_{il}^L x_{jl}^R \right)^2 \quad (2.23)$$

Полученный функционал квадратичный по всем переменным, и его можно оптимизировать методом наименьших квадратов по каждой из переменных при фиксированных значениях остальных. Авторами утверждается, что X^L и X^R сойдутся с точностью до домножения на диагональную матрицу с ненулевыми элементами на диагонали: $\hat{X}^R = \hat{X}^L \times \text{diag}(c)$, так как

$$\sum_{l=1}^L r_{kl} x_{il}^L x_{jl}^R = \sum_{l=1}^L \frac{r_{kl}}{c_l} (x_{il}^L * c_l) x_{jl}^R \quad \forall c \in \mathbb{R}^L$$

Поэтому на последней итерации нужно приравнять значение X^R к X^L и пересчитать r .

Обобщение метода SMACOF для взвешенной евклидовой модели

Существует обобщение рассмотренного выше метода SMACOF, основанное на взвешенной евклидовой модели [1]. Метод применяется для K матриц различий $\Delta_1, \dots, \Delta_K$, входное пространство может быть произвольным. Метод обеспечивает приближенное сохранение значений матриц и может обрабатывать пропуски в данных с использованием матрицы весов.

Оптимальным считается решение, минимизирующее функционал стресса

$$S(X, r) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N w_{ijk} \left(\delta_{ijk} - \sqrt{\sum_{l=1}^L r_{lk} (x_{il} - x_{jl})^2} \right)^2 \quad (2.24)$$

Обозначим $X_k = X \times \text{diag}(r_k)$. Перепишем функцию стресса в следующем виде:

$$S(X, r) = \sum_{k=1}^K \left(\sum_{i=1}^N \sum_{j=1}^N w_{ijk} \delta_{ijk}^2 + \text{tr}(X_k^T V_k X_k) - 2 \text{tr}(X_k^T B(X_k) X_k) \right)$$

Будем оптимизировать итерационно. Пусть $X_k^{[t]} = X^{[t]} \times r_k^{[t]}$ - приближение на t -том шаге итерационного процесса.

Пусть $Y = \{Y_1, \dots, Y_K\}$. Верхняя вариационная оценка для функции стресса имеет вид

$$T(X, Y, r) = \sum_{k=1}^K \left(\sum_{i=1}^N \sum_{j=1}^N w_{ijk} \delta_{ijk}^2 + \text{tr}(X_k^T V_k X_k) - 2 \text{tr}(X_k^T B(Y_k) Y_k) \right) \quad (2.25)$$

Пусть $\bar{X}_k = V_k^+ B(Y_k) Y_k$. Тогда

$$\begin{aligned} T(X, Y, r) &= \sum_{k=1}^K \left(\sum_{i=1}^N \sum_{j=1}^N w_{ijk} \delta_{ijk} + \text{tr}(X_k^T V_k X_k) - 2 \text{tr}(X_k^T V_k \bar{X}_k) \right) = \\ &= \sum_{k=1}^K \left(\sum_{i=1}^N \sum_{j=1}^N w_{ijk} \delta_{ijk}^2 + \text{tr}((X_k - \bar{X}_k)^T V_k (X_k - \bar{X}_k)) - \text{tr}(\bar{X}_k^T V_k \bar{X}_k) \right) \end{aligned}$$

При любом фиксированном Y от X , r зависит только второе слагаемое. Функционал квадратичный по этим переменным, и можно оптимизировать, например, методом наименьших квадратов по каждой из переменных при фиксированных значениях остальных. Так как $T(X, \{X \times \text{diag}(r_1), \dots, X \times \text{diag}(r_K)\}, r) = S(X, r)$, то $Y_k^{[t]} = X_k^{[t]}$. Тогда минимизировать можно следующим образом:

$$\bar{X}_k = V_k^+ B(X_k^{[t]}) X_k^{[t]} \quad (2.26)$$

$$X^{[t+1]}, r^{[t+1]} = \arg \min_{X, r} \sum_{k=1}^K \text{tr}((X \times \text{diag}(r_k) - \bar{X}_k)^T V_k (X \times \text{diag}(r_k) - \bar{X}_k)) \quad (2.27)$$

Стоит отметить, что не обязательно на каждой итерации доводить минимизацию по X, r до сходимости. Для любых $X^{[t+1]}$ и $r^{[t+1]}$ таких, что

$$\begin{aligned} \sum_{k=1}^K \text{tr}((X^{[t+1]} \times \text{diag}(r_k^{[t+1]}) - \bar{X}_k)^T V_k (X^{[t+1]} \times \text{diag}(r_k^{[t+1]}) - \bar{X}_k)) &\leq \\ &\leq \sum_{k=1}^K \text{tr}((X^{[t]} \times \text{diag}(r_k^{[t]}) - \bar{X}_k)^T V_k (X^{[t]} \times \text{diag}(r_k^{[t]}) - \bar{X}_k)) \end{aligned} \quad (2.28)$$

верно

$$\begin{aligned} S(X^{[t]}, r^{[t]}) = T(X^{[t]}, \{X_1^{[t]}, \dots, X_K^{[t]}\}, r^{[t]}) &\geq \{(2.28)\} \geq \\ &\geq T(X^{[t+1]}, \{X_1^{[t]}, \dots, X_K^{[t]}\}, r^{[t+1]}) \geq S(X^{[t+1]}, r^{[t+1]}) \end{aligned} \quad (2.29)$$

То есть значение функции стресса монотонно снижается.

2.5.3 Трехмодальная модель

Трехмодальная модель [6] используется в задачах многомерного шкалирования индивидуальных различий и является обобщением взвешенной евклидовой модели на случай произвольных матриц весов.

Аналогично взвешенной евклидовой модели в евклидовом пространстве некоторой размерности L строится групповая конфигурация для объектов представленная в виде матрицы $X = [x_1, \dots, x_N]^T$. Будем искать точечную конфигурацию X_k для k -той матрицы различий в виде

$$X_k = X \times r_k, \quad (2.30)$$

где $r_k \in \mathbb{R}^{L \times L}$ - матрица весов. Тогда расстояние между представлениями i -го и j -го объектов в k -той точечной конфигурации будет иметь вид

$$\text{eucl}(x_i \times r_k, x_j \times r_k) \quad (2.31)$$

Обобщение метода SMACOF для трехмодальной модели

Существует еще одно обобщение метода SMACOF, основанное уже на трехмодальной модели [2]. Метод применяется для K матриц различий $\Delta_1, \dots, \Delta_K$, входное пространство может быть произвольным. Метод обеспечивает приближенное сохранение значений матриц и может обрабатывать пропуски в данных с использованием матрицы весов.

Оптимальным считается решение, минимизирующее функционал стресса

$$S(X, r) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N w_{ijk} (\delta_{ijk} - \text{eucl}(x_i \times r_k, x_j \times r_k))^2 \quad (2.32)$$

Алгоритм оптимизации функции стресса аналогичен алгоритму обобщения метода SMACOF для взвешенной евклидовой модели.

Глава 3

Модели расстояний, параметризованных размерами, и их обучение

3.1 Модель пропорциональных расстояний

Пусть S - частично упорядоченное множество размеров, на котором определена операция сложения. Пусть X — множество, на котором задано расстояние dist . Пусть r — функция из размеров в вещественные числа. Для $x_1, x_2 \in X$, $s \in S$ рассмотрим функцию

$$\rho(x_1, x_2, s) = r(s) \text{dist}(x_1, x_2) \quad (3.1)$$

Теорема 6. Если dist является расстоянием, а функция $r(s)$ неотрицательна, монотонна и субаддитивна, т.е. выполнено

$$\forall s \in S \quad r(s) \geq 0, \quad (3.2)$$

$$\forall s_1, s_2 \in S \quad s_1 \leq s_2 \Rightarrow r(s_1) \leq r(s_2), \quad (3.3)$$

$$\forall s_1, s_2 \in S \quad r(s_1 + s_2) \leq r(s_1) + r(s_2), \quad (3.4)$$

то функция (3.1) удовлетворяет всем аксиомам расстояния, параметризованного размером. Если dist является псевдометрикой, то ρ является псевдометрикой, параметризованной размером. Если dist является метрикой и дополнительно r положительна, то ρ является метрикой, параметризованной размером.

Доказательство. В [7] показано, что умножение на неотрицательное число расстояния (псевдометрики) гарантирует сохранение аксиом расстояния (псевдометрики). Умножение на положительное число метрики гарантирует сохранение аксиом метрики.

Покажем выполнение аксиомы монотонности (S1). В силу (3.3) и неотрицательности функции dist для любой пары объектов $x_1, x_2 \in X$ при $s_1, s_2 \in S$, $s_1 \leq s_2$ выполнено

$$\begin{aligned} \rho(x_1, x_2, s_2) - \rho(x_1, x_2, s_1) &= r(s_2) \text{dist}(x_1, x_2) - r(s_1) \text{dist}(x_1, x_2) = \\ &= (r(s_2) - r(s_1)) \text{dist}(x_1, x_2) \geq 0 \end{aligned}$$

Покажем выполнение аксиомы неделимости (S2). В силу (3.4) и неотрицательности dist выполнено $\forall x_1, x_2 \in X$, $\forall s_1, s_2 \in S$

$$\begin{aligned} \rho(x_1, x_2, s_1) + \rho(x_1, x_2, s_2) &= r(s_1) \text{dist}(x_1, x_2) + r(s_2) \text{dist}(x_1, x_2) = \\ &= (r(s_1) + r(s_2)) \text{dist}(x_1, x_2) \geq r(s_1 + s_2) \text{dist}(x_1, x_2) = \rho(x, y, s_1 + s_2) \end{aligned}$$

□

В методах сжатия трехмерного тензора, основанных на модели пропорциональных расстояний, решение ищется в следующем виде: строится единая «групповая» конфигурация $X = [x_1, \dots, x_N]^T$ в метрическом пространстве заданной размерности L , а также вектор весов $r \in \mathbb{R}^K$. Для обеспечения выполнения аксиом расстояния, параметризованного размером, можно потребовать, чтобы вектор r удовлетворял достаточным условиям теоремы 6. То есть для случая, когда размеры пропорциональны индексам входного тензора, достаточно потребовать

$$r_k \geq 0 \quad \forall k \in [1, K] \quad (3.5)$$

$$r_{k_2} - r_{k_1} \geq 0 \quad \forall k_1 \leq k_2 \in [1, K] \quad (3.6)$$

$$r_{k_1} + r_{k_2} - r_{k_1+k_2} \geq 0 \quad \forall k_1, k_2 \in [1, K], k_1 + k_2 \leq K \quad (3.7)$$

3.1.1 Метод 1

Метод основан на модели пропорциональных расстояний с функцией $\text{dist} = \text{eucl}$. Он применяется для трехмерного тензора различий $\Delta \in \mathbb{R}^{N \times N \times K}$ для произвольного входного пространства. Метод обеспечивает приближенное сохранение значений входного тензора и обрабатывает пропуски в данных с использованием матрицы весов.

Оптимальным считается решение, минимизирующее функцию стресса

$$S(r, X) = \sum_{k=1}^s \sum_{i=1}^n \sum_{j=1}^n w_{ijk} (r_k \text{eucl}(x_i, x_j) - \delta_{ijk})^2 \quad (3.8)$$

и удовлетворяющее условиям (3.5) — (3.7).

Будем оптимизировать методом покоординатного спуска по X и по r , минимизируя по одной группе переменных при фиксированном значении второй.

Формулу (3.8) можно записать в следующем виде:

$$\begin{aligned} S(r, X) &= \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N w_{ijk} (r_k \text{eucl}(x_i, x_j) - \delta_{ijk})^2 = \\ &= \sum_{k=1}^s \sum_{i=1}^n \sum_{j=1}^n w_{ijk} \delta_{ijk} + \sum_{k=1}^s r_k^2 \sum_{i=1}^n \sum_{j=1}^n w_{ijk} \text{eucl}^2(x_i, x_j) - 2 \sum_{k=1}^s r_k \sum_{i=1}^n \sum_{j=1}^n w_{ijk} \delta_{ijk} \text{eucl}(x_i, x_j) \end{aligned}$$

Зафиксируем X . С учетом требований (3.5) — (3.7) получаем задачу минимизации квадратичной по r функции с линейными ограничениями. Оптимальное значение параметров можно найти с помощью соответствующих методов условной оптимизации.

Оптимизировать по X будем, аналогично алгоритму SMACOF, с использованием верхней вариационной оценки. Перепишем функцию стресса следующим образом:

$$\begin{aligned} S(r, X) &= \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N w_{ijk} \delta_{ijk} + \left(\sum_{k=1}^s r_k^2 \right) \sum_{i=1}^N \sum_{j=1}^N w_{ijk} \text{eucl}^2(x_i, x_j) - \\ &\quad - 2 \sum_{i=1}^n \sum_{j=1}^n \left(\sum_{k=1}^s r_k w_{ijk} \delta_{ijk} \right) \text{eucl}(x_i, x_j) \end{aligned}$$

Становится ясно, что формулы для вычисления итерационных приближений $X^{[t]}$ подвергнутся следующим изменениям:

$$X^{[t+1]} = V_r^+ B_r(X^{[t]}) X^{[t]}$$

$$\text{где } V_r = \sum_{i=1}^n \sum_{j=1}^n \left(\sum_{k=1}^s r_k^2 w_{ijk} \right) A_{ij}$$

$$B_r(X) = \sum_{i=1}^n \sum_{j=1}^n s_{ij}^r A_{ij}, \quad s_{ij}^r(X) = \begin{cases} \frac{\sum_{k=1}^s r_k w_{ijk} \delta_{kij}}{\text{eucl}(x_i, x_j)}, & \text{eucl}(x_i, x_j) \neq 0 \\ 0, & \text{eucl}(x_i, x_j) = 0 \end{cases} \quad (3.9)$$

Также стоит отметить, что так как методы условной оптимизации имеют большую вычислительную сложность, то в случаях, когда лишь незначительная часть входных данных противоречит аксиомам размера, первые несколько итераций алгоритма можно сделать без учета ограничений на r .

3.2 Модель пропорциональных конфигураций

Пусть снова S - частично упорядоченное множество размеров с операцией сложения, X - множество, на котором задано расстояние dist . Пусть r - функция из размеров в вещественные числа, а для всех точек $x \in X$ определена операция умножения на скаляр. Тогда для $x_1, x_2 \in X$, $s \in S$ рассмотрим функцию

$$\rho(x_1, x_2, s) = \text{dist}(r(s)x_1, r(s)x_2) \quad (3.10)$$

Можно утверждать, что если dist является расстоянием (псевдометрикой), то для $\forall r$ функция ρ удовлетворяет всем аксиомам расстояния (псевдометрики), так как умножение на скаляр переводит одинаковые точки в одинаковые, а разные — в одинаковые (равные нулю при $r(s) = 0$) или разные.

Если dist — метрика, а $\forall s \in S$ верно $r(s) \neq 0$, то функция ρ удовлетворяет аксиомам метрики. Действительно, умножение на ненулевой скаляр одинаковые точки переводит в одинаковые, а разные — в разные.

Ограничения, которые необходимо наложить на функцию r , чтобы обеспечить выполнение аксиом размера, зависят от конкретного вида функции dist . Рассмотрим частный случай.

Предложение 7. Для случая $\text{dist} = \text{eucl}$ модель пропорциональных конфигураций эквивалентна модели пропорциональных расстояний.

Доказательство. Для $\forall x_1, x_2 \in X$, $s \in S$

$$\begin{aligned} \rho(x_1, x_2, s) &= \text{eucl}(r(s)x_1, r(s)x_2) = \sqrt{\sum_{l=1}^L (r(s)x_{1l} - r(s)x_{2l})^2} = \\ &= \sqrt{r(s)^2 \sum_{l=1}^L (x_{1l} - x_{2l})^2} = |r(s)| \text{eucl}(x_1, x_2) \end{aligned}$$

□

3.3 Модель индивидуальных различий

Пусть теперь X — множество, на котором определено расстояние dist такое, что его элементы — векторы размерности L . Пусть S — частично упорядоченное множество размеров с операцией сложения. Пусть $r : S \rightarrow \mathbb{R}^L$ — функция из размеров в векторы вещественных чисел размерности L , а для всех элементов X определена операция матричного умножения на вещественную матрицу $\in \mathbb{R}^{L \times L}$. Тогда $\forall x_1, x_2 \in X, \forall s \in S$ рассмотрим функцию

$$\rho(x_1, x_2, s) = \text{dist} \left(x_1 \times \text{diag}(r(s)), x_2 \times \text{diag}(r(s)) \right) \quad (3.11)$$

Можно утверждать, что если dist является расстоянием (псевдометрикой), то для $\forall r$ функция ρ удовлетворяет аксиомам расстояния (псевдометрики). Умножение вектора на диагональную матрицу равносильно умножению i -той его компоненты на i -тый диагональный элемент $\forall i \in [1, L]$. Такое преобразование переводит одинаковые элементы множества X в одинаковые, а разные — в одинаковые (с равными нулю компонентами при $r_i(s) = 0$) или в разные.

Если dist — метрика, то если выполнено

$$r_i(s) \neq 0 \quad \forall s \in S, \forall i \in [1, L],$$

то ρ удовлетворяет аксиомам метрики. Умножение каждой компоненты вектора на ненулевой скаляр переводит одинаковые векторы в одинаковые, а разные — в разные.

Ограничения, которые необходимо наложить на функцию r , чтобы обеспечить выполнение аксиом размера, зависят от функции dist . Рассмотрим частный случай.

Теорема 8. Если в модели индивидуальных различий $\text{dist} = \text{eucl}$, а r удовлетворяет аксиомам неотрицательности, монотонности и субаддитивности (суммирование и сравнение вещественных векторов $r(s)$ покомпонентное): $\forall s_1, s_2 \in S$

$$r(s_1) \geq 0 \quad (3.12)$$

$$s_1 \leq s_2 \Rightarrow r(s_1) \leq r(s_2) \quad (3.13)$$

$$r(s_1 + s_2) \leq r(s_1) + r(s_2) \quad (3.14)$$

то функция ρ удовлетворяет аксиомам размера $(S1) - (S2)$.

Доказательство. Покажем выполнение аксиомы монотонности $(S1)$. Пусть $x_1, x_2 \in X, s_1, s_2 \in S$ — произвольные элементы. Верно, что

$$s_1 \leq s_2 \Rightarrow \{(3.13)\} \Rightarrow r(s_1) \leq r(s_2) \Rightarrow \{(3.12)\} \Rightarrow r^2(s_1) \leq r^2(s_2) \quad (3.15)$$

Также в силу того, что функция eucl неотрицательна, верно

$$\begin{aligned} \rho(x_1, x_2, s_1) \leq \rho(x_1, x_2, s_2) &\Leftrightarrow \text{eucl} \left(x_1 \times r(s_1), x_2 \times r(s_1) \right) \leq \text{eucl} \left(x_1 \times r(s_2), x_2 \times r(s_2) \right) \Leftrightarrow \\ &\Leftrightarrow \rho^2(x_1, x_2, s_1) \leq \rho^2(x_1, x_2, s_2) \end{aligned} \quad (3.16)$$

Тогда $\forall x_1, x_2 \in X, \forall s_1, s_2 \in S, s_1 \leq s_2$

$$\begin{aligned} \rho^2(x_1, x_2, s_1) - \rho^2(x_1, x_2, s_2) &= \sum_{l=1}^L r_l^2(s_1) (x_{1l} - x_{2l})^2 - \sum_{l=1}^L r_l^2(s_2) (x_{1l} - x_{2l})^2 = \\ &= \sum_{l=1}^L (r_l^2(s_1) - r_l^2(s_2)) (x_{1l} - x_{2l})^2 \leq \{(3.15)\} \leq 0 \end{aligned}$$

А, согласно (3.16), верно

$$\rho^2(x_1, x_2, s_1) - \rho^2(x_1, x_2, s_2) \leq 0 \Leftrightarrow \rho(x_1, x_2, s_1) - \rho(x_1, x_2, s_2) \leq 0$$

Теперь покажем выполнение аксиомы (S2). $\forall x_1, x_2 \in X, \forall s_1, s_2 \in S$

$$\begin{aligned} \rho^2(x_1, x_2, s_1 + s_2) &= \sum_{l=1}^L r_l^2(s_1 + s_2) (x_{1l} - x_{2l})^2 \leq \{(3.12), (3.14)\} \leq \sum_{l=1}^L (r_l(s_1) + r_l(s_2))^2 (x_{1l} - x_{2l})^2 = \\ &= \sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2)) (x_{1l} - x_{2l})^2 + 2 \sum_{l=1}^L r_l(s_1) r_l(s_2) (x_{1l} - x_{2l})^2 \leq \{(3.12)\} \leq \\ &\leq \sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2)) (x_{1l} - x_{2l})^2 \end{aligned}$$

В силу того, что оба выражения неотрицательны, верно

$$\begin{aligned} \rho^2(x_1, x_2, s_1 + s_2) &\leq \sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2)) (x_{1l} - x_{2l})^2 \Leftrightarrow \\ &\Leftrightarrow \rho(x_1, x_2, s_1 + s_2) \leq \sqrt{\sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2)) (x_{1l} - x_{2l})^2} \end{aligned}$$

Так как квадратный корень из суммы неотрицательных элементов не превосходит сумму корней этих элементов, то

$$\begin{aligned} \rho(x_1, x_2, s_1 + s_2) &\leq \sqrt{\sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2)) (x_{1l} - x_{2l})^2} \leq \\ &\leq \sqrt{\sum_{l=1}^L r_l^2(s_1) (x_{1l} - x_{2l})^2} + \sqrt{\sum_{l=1}^L r_l^2(s_2) (x_{1l} - x_{2l})^2} = \rho(x_1, x_2, s_1) + \rho(x_1, x_2, s_2) \end{aligned}$$

□

В методах сжатия трехмерного тензора различий, основанных на модели индивидуальных различий, решение ищется в следующем виде: строится единая «групповая» конфигурация $X = [x_1, \dots, x_N]^T$ в метрическом пространстве заданной размерности L , а также набор векторов весов $r = [r_1, \dots, r_K]^T \in \mathbb{R}^{K \times L}$. В случае, когда $\text{dist} = \text{eucl}$, для обеспечения выполнения аксиом расстояния, параметризованного размером, можно потребовать, чтобы векторы r удовлетворяли достаточным условиям теоремы 8. То есть для случая, когда размеры пропорциональны индексам входного тензора, достаточно потребовать

$$r_k \geq 0 \quad \forall k \in [1, K] \quad (3.17)$$

$$r_{k_2} - r_{k_1} \geq 0 \quad \forall k_1 \leq k_2 \in [1, K] \quad (3.18)$$

$$r_{k_1} + r_{k_2} - r_{k_1 + k_2} \geq 0 \quad \forall k_1, k_2 \in [1, K], k_1 + k_2 \leq K \quad (3.19)$$

Арифметические операции и сравнения поэлементные.

3.3.1 Метод 2

Метод основан на модели индивидуальных различий с функцией $\text{dist} = \text{eucl}$. Он применяется для трехмерного тензора различий $\Delta \in \mathbb{R}^{N \times N \times K}$, входное пространство произвольное. Метод обеспечивает приближенное сохранение значений тензора и обрабатывает пропуски в данных с использованием матрицы весов.

Оптимальным считается решение, минимизирующее функцию стресса

$$S(X, r) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N w_{ijk} \left(\text{eucl}(x_i \times \text{diag}(r_k), x_j \times \text{diag}(r_k)) - \delta_{ijk} \right)^2 \quad (3.20)$$

и удовлетворяющее условиям (3.17) — (3.19).

Для оптимизации будем использовать модифицированный алгоритм обобщения метода SMACOF. Процесс оптимизации итерационный, и на каждой итерации, аналогично (2.26) и (2.27), приближение параметров вычисляется по формулам

$$\begin{aligned} \bar{X}_k &= V_k^+ B(X_k^{[t]}) X_k^{[t]} \\ X^{[t+1]}, r^{[t+1]} &= \arg \min_{X, r} \sum_{k=1}^K \text{tr}((X \times \text{diag}(r_k) - \bar{X}_k)^T V_k (X \times \text{diag}(r_k) - \bar{X}_k)) \end{aligned}$$

Только теперь минимизация квадратичного функционала (2.27) по параметру r условная, с линейными ограничениями (3.17) — (3.19). Оптимальное значение параметра можно найти с помощью соответствующих методов условной оптимизации. Стоит отметить, что так как методы условной оптимизации имеют большую вычислительную сложность, то в случаях, когда лишь незначительная часть входных данных противоречит аксиомам размера, первые несколько итераций алгоритма можно сделать без учета ограничений.

3.4 Модель рядов

Пусть X — произвольное множество на котором определена функция различия α , S — частично упорядоченное множество с операцией сложения. Пусть $r : X \times S \in \mathbb{R}$. Для $x_1, x_2 \in X$, $s \in S$ рассмотрим функцию

$$\rho(x_1, x_2, s) = \alpha(x_i, x_j) r(x_i, s) + \alpha(x_j, x_i) r(x_j, s) \quad (3.21)$$

Теорема 9. Если функция α неотрицательна и рефлексивна, а r неотрицательна, монотонна и субаддитивна, то есть

$$\alpha(x_i, x_j) \geq 0 \quad (3.22)$$

$$\alpha(x_i, x_i) = 0 \quad (3.23)$$

$$r(x_i, s) \geq 0 \quad (3.24)$$

$$s_1 \leq s_2 \Rightarrow r(x_i, s_1) \leq r(x_i, s_2) \quad (3.25)$$

$$r(x_1, s_1 + s_2) \leq r(x_1, s_1) + r(x_1, s_2) \quad (3.26)$$

то функция (3.21) является расстоянием, параметризованным размером.

Доказательство. Покажем, что выполнены аксиомы расстояния. Аксиома рефлексивности (D1): $\forall x_1 \in X, s_1 \in S$

$$\rho(x_1, x_1, s_1) = \alpha(x_1, x_1)r(x_1, s_1) + \alpha(x_1, x_1)r_1, s_1) = \{(3.23)\} = 0 \quad (3.27)$$

Аксиома симметричности (D2): $\forall x_1, x_2 \in X, s_1 \in S$

$$\rho(x_1, x_2, s_1) = \alpha(x_1, x_2)r(x_1, s_1) + \alpha(x_2, x_1)r(x_2, s_1) = \rho(x_2, x_1, s_1) \quad (3.28)$$

Аксиома неотрицательности (D3) выполнена в силу неотрицательности функций α и r .

Теперь покажем, что выполнены аксиомы размера. Аксиома монотонности (S1): $\forall x_1, x_2 \in X, s_1, s_2 \in S, s_1 \leq s_2$

$$\begin{aligned} \rho(x_1, x_2, s_1) - \rho(x_1, x_2, s_2) &= \\ &= \alpha(x_1, x_2)r(x_1, s_1) + \alpha(x_2, x_1)r(x_2, s_1) - \alpha(x_1, x_2)r(x_1, s_2) - \alpha(x_2, x_1)r(x_2, s_2) = \\ &= \alpha(x_1, x_2)(r(x_1, s_1) - r(x_1, s_2)) + \alpha(x_2, x_1)(r(x_2, s_1) - r(x_2, s_2)) \leq \{(3.22), (3.25)\} \leq 0 \end{aligned}$$

Аксиома неделимости (S2): $\forall x_1, x_2 \in X, s_1, s_1 \in S$

$$\begin{aligned} \rho(x_1, x_2, s_1) + \rho(x_1, x_2, s_2) &= \\ &= \alpha(x_1, x_2)(r(x_1, s_1) + r(x_1, s_2)) + \alpha(x_2, x_1)(r(x_2, s_1) + r(x_2, s_2)) \geq \{(3.22), (3.26)\} \geq \\ &\geq \alpha(x_1, x_2)r(x_1, s_1 + s_2) + \alpha(x_2, x_1)r(x_2, s_1 + s_2) = \rho(x_1, x_2, s_1 + s_2) \end{aligned}$$

□

В методах сжатия трехмерного тензора различий, основанных на модели рядов, решение ищется в следующем виде: строится матрица попарных различий между объектами $\alpha \in \mathbb{R}^{N \times N}$ и матрица $r \in \mathbb{R}^{N \times K}$. Для обеспечения выполнения аксиом расстояния можно потребовать, чтобы α и r удовлетворяли достаточным условиям теоремы 9. Для случая, когда размеры пропорциональны индексам, достаточно потребовать

$$\alpha_{ij} \geq 0 \quad \forall i, j \in [1, N] \quad (3.29)$$

$$\alpha_{ii} = 0 \quad \forall i \in [1, N] \quad (3.30)$$

$$r_{ik} \geq 0 \quad \forall i \in [1, N], \forall k \in [1, K] \quad (3.31)$$

$$r_{i,k_2} - r_{i,k_1} \geq 0 \quad \forall i \in [1, N], \forall k_1, k_2 \in [1, K], k_1 \leq k_2 \quad (3.32)$$

$$r_{i,k_1} + r_{i,k_2} - r_{i,k_1+k_2} \geq 0 \quad \forall i \in [1, N], \forall k_1, k_2 \in [1, K], k_1 + k_2 \leq K \quad (3.33)$$

3.4.1 Метод 3

Метод основан на модели рядов. Он применяется для трехмерного тензора различий $\Delta \in \mathbb{R}^{N \times N \times K}$, входное пространство произвольное. Метод обеспечивает приближенное сохранение значений тензора, пропуски в данных не обрабатывает.

Оптимальным считается решение, минимизирующее функцию стресса

$$S(r, \alpha) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N (\alpha_{ij}r_{ik} + \alpha_{ji}r_{jk} - \delta_{ijk})^2 \quad (3.34)$$

и удовлетворяющее условиям (3.29) — (3.33).

Функционал квадратичный по каждой из переменных при фиксированных значениях остальных. Раскроем скобки и преобразуем полученные слагаемые, чтобы понять, на какие группы можно разбить переменные для оптимизации.

$$\begin{aligned}
S_d^2(r, \alpha) &= \sum_{k=1}^s \sum_{i=1}^n \sum_{j=1}^n \delta_{kij}^2 + \sum_{k=1}^s \sum_{i=1}^n \sum_{j=1}^n (\alpha_{ij} r_{ki} + \alpha_{ji} r_{kj})^2 - 2 \sum_{k=1}^s \sum_{i=1}^n \sum_{j=1}^n \delta_{kij} (\alpha_{ij} r_{ki} + \alpha_{ji} r_{kj}) = \\
&= \sum_{k=1}^s \sum_{i=1}^n \sum_{j=1}^n \delta_{kij}^2 + \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij}^2 \sum_{k=1}^s r_{ki}^2 + \sum_{i=1}^n \sum_{j=1}^n \alpha_{ji}^2 \sum_{k=1}^s r_{kj}^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} \alpha_{ji} \sum_{k=1}^s r_{ki} r_{kj} - \\
&\quad - 2 \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} \sum_{k=1}^s \delta_{kij} r_{ki} - 2 \sum_{i=1}^n \sum_{j=1}^n \alpha_{ji} \sum_{k=1}^s \delta_{kij} r_{kj} = \\
&= \sum_{k=1}^s \sum_{i=1}^n \sum_{j=1}^n \delta_{kij}^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij}^2 \sum_{k=1}^s r_{ki}^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} \alpha_{ji} \sum_{k=1}^s r_{ki} r_{kj} - 2 \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} \sum_{k=1}^s (\delta_{kij} + \delta_{kji}) r_{ki}
\end{aligned}$$

Легко видеть, что итерационный процесс можно вести последовательно по следующим группам переменных при фиксированных значениях остальных:

- $\{r_{ki} \mid k = 1, \dots, s\}, \forall i = 1, \dots, n$
- $\{\alpha_{ij} \mid i, j = 1, \dots, n, i < j\}$
- $\{\alpha_{ij} \mid i, j = 1, \dots, n, i > j\}$
- $\{\alpha_{ii} \mid i = 1, \dots, n\}$

С учетом ограничений (3.29) — (3.33) получаем

$$\alpha_{ii} = 0 \quad \forall i \in [1, N] \quad (3.35)$$

$$\begin{aligned}
\sum_{i=1}^N \sum_{j=1}^N \left(2\alpha_{ij}^2 \sum_{k=1}^s r_{ki}^2 + \alpha_{ij} \left(4\alpha_{ji} \sum_{k=1}^s r_{ki} r_{kj} - 2 \sum_{k=1}^s (\delta_{kij} + \delta_{kji}) r_{ki} \right) \right) \rightarrow \min_{\alpha_{ij}, i < j} \\
\text{при условии } \alpha_{ij} \geq 0 \quad (3.36)
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^N \sum_{j=1}^N \left(2\alpha_{ij}^2 \sum_{k=1}^s r_{ki}^2 + \alpha_{ij} \left(4\alpha_{ji} \sum_{k=1}^s r_{ki} r_{kj} - 2 \sum_{k=1}^s (\delta_{kij} + \delta_{kji}) r_{ki} \right) \right) \rightarrow \min_{\alpha_{ij}, i > j} \\
\text{при условии } \alpha_{ij} \geq 0 \quad (3.37)
\end{aligned}$$

Для всех $i \in [1, N]$

$$\begin{aligned}
\sum_{k=1}^K \left(2r_{ki}^2 \left(\sum_{j=1}^n \alpha_{ij}^2 + \alpha_{ii}^2 \right) + 4r_{ki} \left(\sum_{\substack{j=1 \\ j \neq i}}^n r_j \alpha_{ij} \alpha_{ji} - 2 \sum_{j=1}^n \alpha_{ij} (\delta_{kij} + \delta_{kji}) \right) \right) \rightarrow \min_{r_{k,i}, k \in [1, K]} \\
\text{при условиях (3.31) — (3.33)} \quad (3.38)
\end{aligned}$$

Задачи минимизации квадратичного функционала с линейными ограничениями решаются с помощью соответствующих методов условной оптимизации. Так как методы условной оптимизации имеют большую вычислительную сложность, то в случаях, когда лишь незначительная часть входных данных противоречит аксиомам размера, первые несколько итераций алгоритма можно сделать без учета ограничений на параметры.

Глава 4

Эксперименты

4.1 Описание данных

Эксперименты проводились на данных, собранных в рамках тестирования коммуникаций многопроцессорной системы "Ломоносов" с помощью программного средства *network_test2*. Средство позволяет получить данные по задержкам передачи сообщения для конкретных пар процессов при разных длинах сообщений. В ходе тестирования для каждой упорядоченной пары процессов и каждой длины сообщения проводилось несколько измерений, после чего вычислялся ряд характеристик полученного эмпирического распределения задержки: минимальное значение, медиана, среднее значение, стандартное отклонение. Для исследования был выбран трехмерный тензор медиан.

Тестирование проводилось на 78 процессах для длин сообщений в диапазоне от 0 до 10000 байтов с шагом в 100 байтов. То есть получен тензор различий, параметризованных размером, $\Delta \in \mathbb{R}^{78 \times 78 \times 100}$, где δ_{ijk} соответствует медиане эмпирического распределения длины задержки при отправке сообщения длины, пропорциональной k , процессом с индексом i процессу с индексом j . На рисунке 4.1 приведена визуализация срезов полученного тензора при фиксированном индексе размера. Можно заметить, что двумерные срезы для различных длин сообщений имеют схожую структуру. Это позволяет предположить, что данные хорошо описываются моделями пропорциональных расстояний, пропорциональных конфигураций и индивидуальных различий.

На рисунке 4.2 приведена визуализация двумерных срезов тензора для различных процессов-получателей и различных длин сообщений при фиксированном индексе процесса-отправителя. Теперь зафиксируем 15 случайных пар индексов процесса-отправителя и процесса-получателя. На рисунке 4.3 приведена визуализация полученных одномерных срезов тензора для различных длин сообщений. Видно, что значение задержек передачи сообщения имеет тенденцию увеличения с ростом длины сообщения.

Чтобы измерить, какая доля входных данных противоречит аксиомам расстояния, параметризованного размером, введем следующие показатели:

$$MP = \frac{\sum_{i=1}^N \sum_{j=1}^N \sum_{k_1 < k_2} \mathbb{I}[\delta_{ijk_1} > \delta_{ijk_2}]}{\sum_{i=1}^N \sum_{j=1}^N \sum_{k_1 < k_2} 1} \quad (4.1)$$

$$IP = \frac{\sum_{i=1}^N \sum_{j=1}^N \sum_{k_1 < k_2} \mathbb{I}[k_1 + k_2 \leq K] \mathbb{I}[\delta_{ijk_1} + \delta_{ijk_2} < \delta_{ijk_1+k_2}]}{\sum_{i=1}^N \sum_{j=1}^N \sum_{k_1 < k_2} \mathbb{I}[k_1 + k_2 \leq K]} \quad (4.2)$$

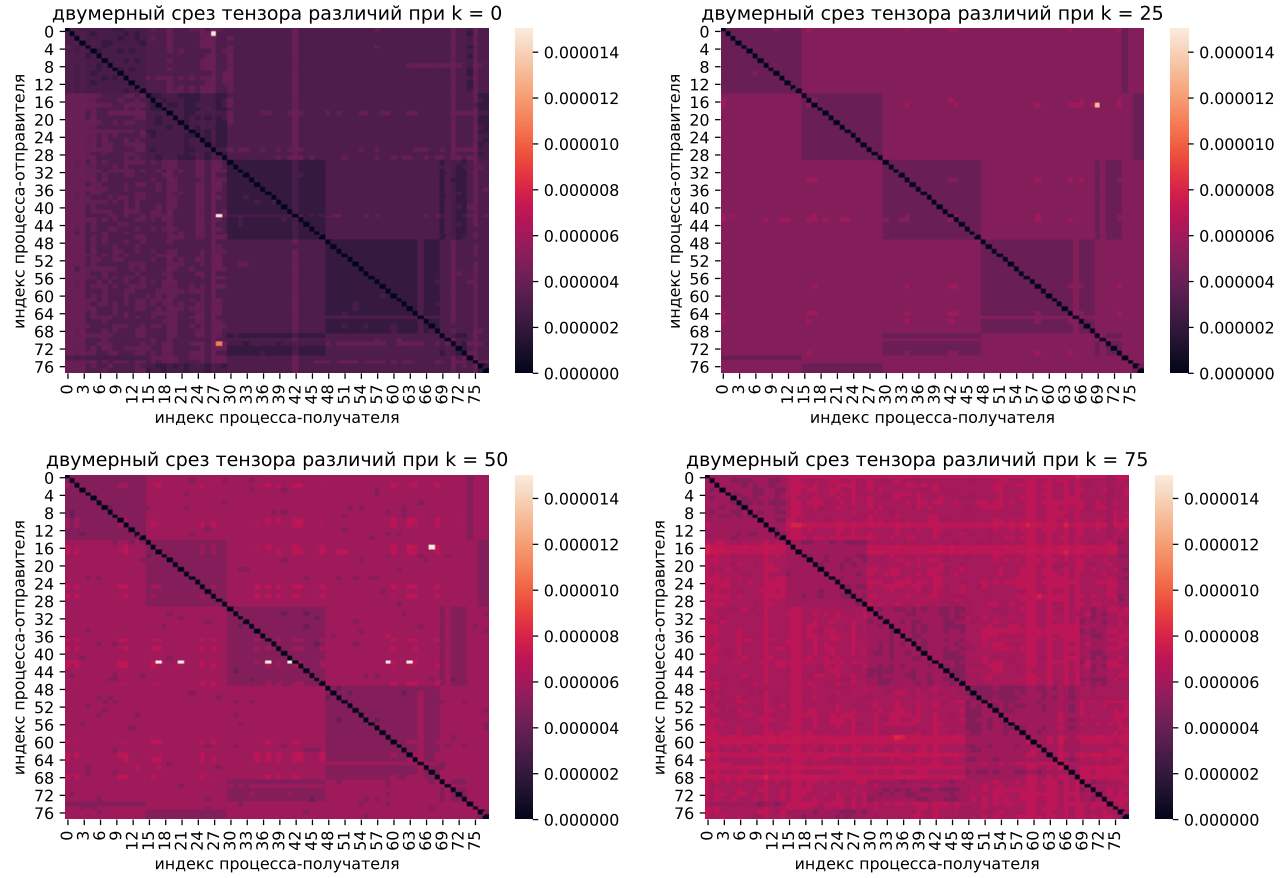


Рис. 4.1: Визуализация срезов тензора Δ при фиксированном индексе размера

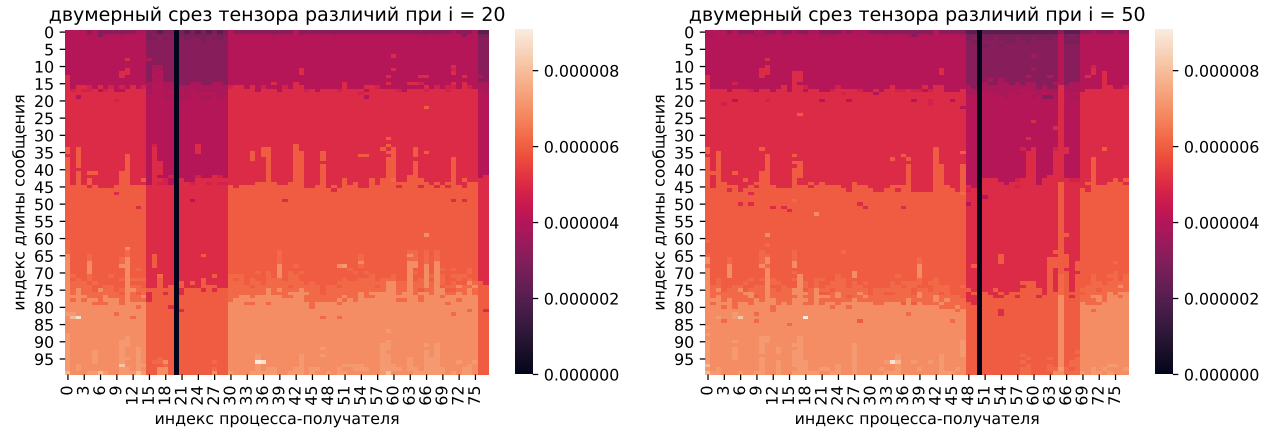


Рис. 4.2: Визуализация срезов тензора Δ при фиксированном индексе процесса-отправителя

$$RP = \frac{\sum_{k=1}^K \sum_{i=1}^N \mathbb{I}[\delta_{iik}! = 0]}{NK} \quad (4.3)$$

$$SP = \frac{\sum_{k=1}^K \sum_{i=1}^N \sum_{j < i} \mathbb{I}[\delta_{ijk}! = \delta_{jik}]}{0.5KN(N-1)} \quad (4.4)$$

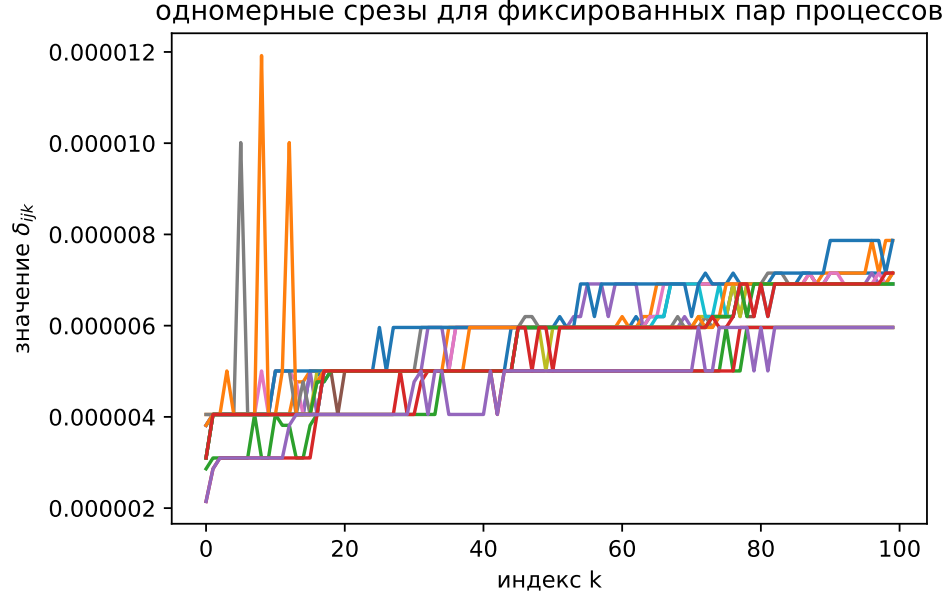


Рис. 4.3: Визуализация срезов Δ при фиксированных индексах процесса-отправителя и процесса-получателя

$$NNP = \frac{\sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N \mathbb{I}[\delta_{iik} < 0]}{N^2 K} \quad (4.5)$$

Для наших данных их значения составляют $MP = 0.008$, $IP = 0.0005$, $RP = 0$, $SP = 0.181$, $NNP = 0$. То есть введенные аксиомы расстояния, параметризованного размером, действительно выполняются для большей части реальных данных.

4.2 Исследование метода 1

Метод был реализован на языке Python с использованием библиотек `pumpy` и `scipy`. Так как условная оптимизация имеет большую вычислительную сложность, то сначала выполнялась безусловная минимизация функции стресса. Итерационный процесс останавливался при

$$S(X^{[t-1]}, r^{[t-1]}) - S(X^{[t]}, r^{[t]}) < \epsilon, \quad \epsilon = 1e-12 \quad (4.6)$$

Затем полученные значения параметров использовались в качестве начального приближения для условной оптимизации. Оптимизация производилась с использованием метода `SLSQP`, реализованного в библиотеке `scipy.optimize` (необходимо вызвать функцию `scipy.optimize.minimize` с параметром `method = "SLSQP"`). Метод вызывался с параметром `tol = 1e-4`, в качестве начального приближения на каждой итерации использовались вычисленные аналитически значения параметров, минимизирующие квадратичный функционал без учета ограничений. Итерационный процесс останавливался при выполнении условия .

Эксперименты были проведены для нескольких значений гиперпараметра L . В качестве метрик качества аппроксимации исходного тензора были использованы итоговое значение функции стресса и значение функции `Stress-1`. Чтобы измерить, насколько зна-

чение параметра r отклоняется от достаточных условий теоремы 6, введем показатели

$$rNNP = \frac{\sum_{k=1}^K r_k \geq 0}{K} \quad (4.7)$$

$$rMP = \frac{\sum_{k_1 < k_2} \mathbb{I}[r_{k_1} > r_{k_2}]}{\sum_{k_1 < k_2} 1} \quad (4.8)$$

$$rSP = \frac{\sum_{k_1 < k_2} \mathbb{I}[k_1 + k_2 \leq K] \mathbb{I}[r_{k_1} + r_{k_2} < r_{k_1+k_2}]}{\sum_{k_1 < k_2} \mathbb{I}[k_1 + k_2 \leq K]} \quad (4.9)$$

Используемая модель расстояния гарантирует выполнение аксиом расстояния, поэтому для проверки аксиом расстояния, параметризованного размером, для полученного тензора будем измерять только показатели MP и IP .

Значения описанных показателей для результатов безусловной и условной оптимизации функции стресса метода 1 приведены в таблице 4.1. На рисунке 4.4 приведены графики зависимости значений параметра r_k от индекса k после безусловной оптимизации и после завершения алгоритма при $L = 20$.

L	оптимизация	Stress	Stress-1	rNNP	rMP	rSP	MP	IP
L = 1	безусловная	6.327e-06	0.704	0	0.141	0	0.008	0
	условная	6.327e-06	0.704	0	0	0	0	0
L = 2	безусловная	2.675e-06	0.404	0	0.162	0	0.008	0
	условная	2.675e-06	0.404	0	0	0	0	0
L = 3	безусловная	1.590e-06	0.301	0	0.141	0	0.008	0
	условная	1.590e-06	0.301	0	0	0	0	0
L = 4	безусловная	1.101e-06	0.247	0	0.162	0	0.009	0
	условная	1.101e-06	0.247	0	0	0	0	0
L = 5	безусловная	8.340e-07	0.214	0	0.162	0	0.009	0
	условная	8.340e-07	0.213	0	0	0	0	0
L = 6	безусловная	6.741e-07	0.191	0	0.162	0	0.008	0
	условная	6.741e-07	0.191	0	0	0	0	0
L = 8	безусловная	4.860e-07	0.162	0	0.162	0	0.008	0
	условная	4.861e-07	0.162	0	0	0	0	0
L = 10	безусловная	3.823e-07	0.143	0	0.172	0	0.009	0
	условная	3.824e-07	0.143	0	0	0	0	0
L = 15	безусловная	2.647e-07	0.119	0	0.172	0	0.009	0
	условная	2.647e-07	0.119	0	0	0	0	0
L = 20	безусловная	2.1325e-07	0.106	0	0.182	0	0.009	0
	условная	2.132e-07	0.106	0	0	0	0	0

Таблица 4.1: Значения показателей после безусловной и после условной оптимизации функции стресса метода 1 при различных значениях гиперпараметра L

Эксперименты показывают, что для использованных данных качество аппроксимации тензора различий после условной оптимизации незначительно ниже, чем после безусловной.

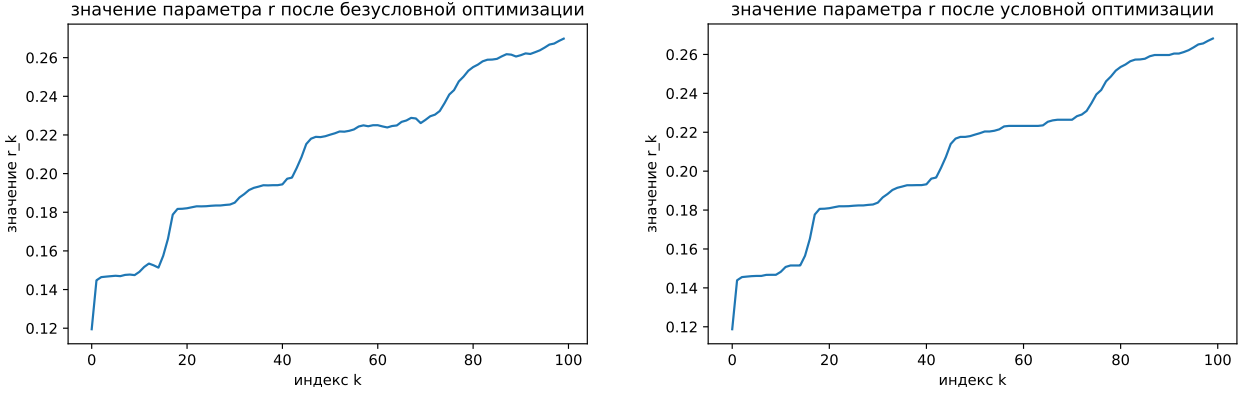


Рис. 4.4: Зависимость значения параметра r_k от индекса k после безусловной и после условной оптимизации функции стресса метода 1 при $L = 20$

При этом результат безусловной оптимизации функции стресса не удовлетворяет некоторым из достаточных условий теоремы 6, а также некоторым из аксиом расстояния, параметризованного размерам. В то время как результат условной оптимизации удовлетворяет и достаточным условиям теоремы, и всем аксиомам расстояния, параметризованного размером, то есть разработанный метод работает корректно.

На рисунке 4.5 приведен график зависимости Stress-1 от значения гиперпараметра L . График показывает, что с незначительной потерей точности аппроксимации можно приблизить входной тензор моделью с небольшим значением L .

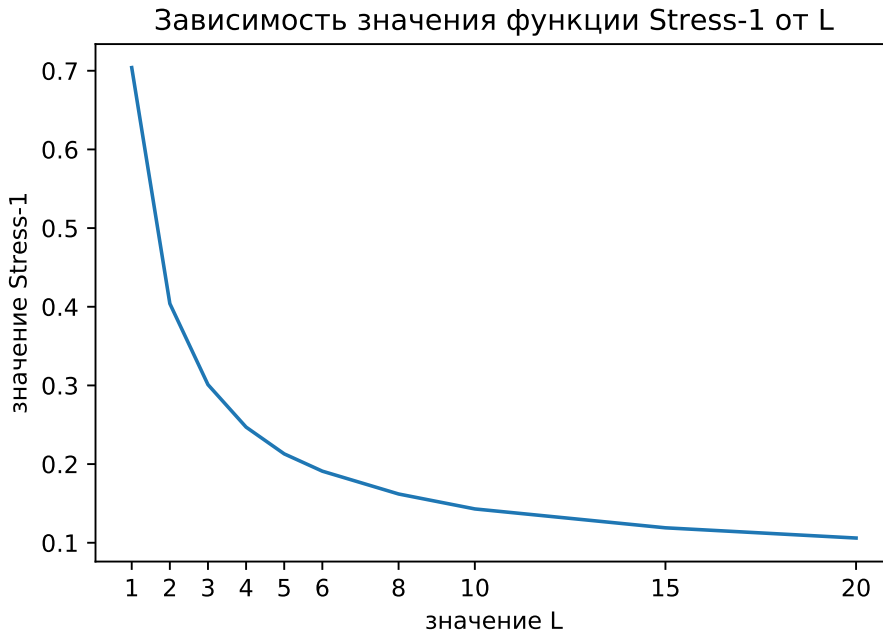


Рис. 4.5: Зависимость значения функции Stress-1 от значения гиперпараметра L для результатов метода 1

4.3 Исследование метода 2

Алгоритм был реализован на языке Python с использованием библиотек `numpy` и `scipy`. Аналогично экспериментам с методом 1, сначала выполнялась безусловная минимизация функции стресса. Итерационный процесс останавливался при выполнении условия (4.2).

Затем полученные значения параметров использовались в качестве начального приближения для условной оптимизации. Оптимизация производилась с использованием метода SLSQP, реализованного в библиотеке `scipy.optimize`. Метод вызывался с параметром $tol = 1e - 4$, в качестве начального приближения на каждой итерации использовались вычисленные аналитически значения параметров, минимизирующие квадратичный функционал без учета ограничений. Итерационный процесс останавливался при выполнении (4.2).

Для метода 2 показатели, измеряющие, насколько не выполняются достаточные условия для параметра r , имеют вид

$$rNNP = \frac{\sum_{k=1}^K \sum_{l=1}^L r_{kl} \geq 0}{KL} \quad (4.10)$$

$$rMP = \frac{\sum_{k_1 < k_2} \sum_{l=1}^L \mathbb{I}[r_{k_1 l} > r_{k_2 l}]}{\sum_{k_1 < k_2} L} \quad (4.11)$$

$$rSP = \frac{\sum_{k_1 < k_2} \sum_{l=1}^L \mathbb{I}[k_1 + k_2 \leq K] \mathbb{I}[r_{k_1 l} + r_{k_2 l} < r_{k_1 + k_2 l}]}{\sum_{k_1 < k_2} \mathbb{I}[k_1 + k_2 \leq K] L} \quad (4.12)$$

Используемая модель расстояния гарантирует выполнение аксиом расстояния, поэтому для проверки аксиом расстояния, параметризованного размером, для полученного тензора будем измерять только показатели MP и IP . Также были измерены итоговое значение функции стресса и значение функции Stress-1.

Значения описанных показателей для результатов безусловной и условной оптимизации функции стресса метода 2 приведены в таблице 4.2. На рисунке 4.6 приведена визуализация значений параметра r после безусловной и после условной оптимизации.

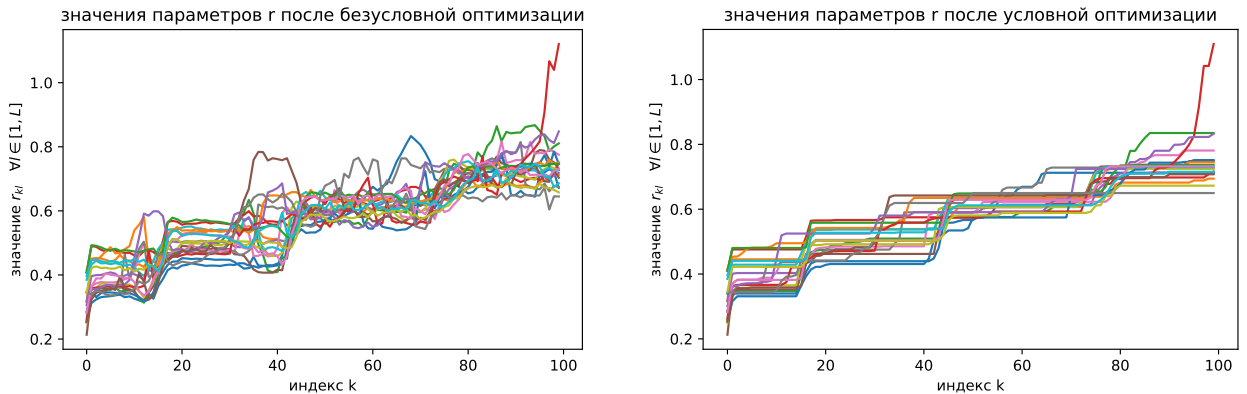


Рис. 4.6: Зависимость значений параметров $r_{kl} \forall l \in [1, L]$ от индекса k после безусловной и после условной оптимизации функции стресса метода 2 при $L = 20$

Эксперименты показывают, что для использованных данных качество аппроксимации тензора различий после условной оптимизации незначительно ниже, чем после безусловной. При этом результат безусловной оптимизации функции стресса не удовлетворяет

L	оптимизация	Stress	Stress-1	rNNP	rMP	rSP	MP	IP
L = 2	безусловная	6.271e-06	0.699	0	0.172	0	0.008	0
	условная	6.271e-06	0.699	0	0	0	0	0
L = 2	безусловная	2.725e-06	0.378	0	0.217	0	0.010	0
	условная	2.701e-06	0.406	0	0	0	0	0
L = 3	безусловная	1.585e-06	0.301	0	0.246	0	0.0127	0
	условная	1.588e-06	0.301	0	0	0	0	0
L = 4	безусловная	1.095e-06	0.247	0	0.240	0	0.015	0
	условная	1.096e-06	0.247	0	0	0	0	0
L = 5	безусловная	8.300e-07	0.213	0	0.299	0	0.015	0
	условная	8.297e-07	0.213	0	0	0	0	0
L = 6	безусловная	6.671e-07	0.190	0	0.286	0	0.017	0
	условная	6.676e-07	0.191	0	0	0	0	0
L = 8	безусловная	4.768e-07	0.160	0	0.346	0	0.024	0
	условная	4.761e-07	0.160	0	0	0	0	0
L = 10	безусловная	3.710e-07	0.141	0	0.372	0	0.031	0
	условная	3.719e-07	0.141	0	0	0	0	0
L = 15	безусловная	2.479e-07	0.115	0	0.413	0.00013	0.043	0
	условная	2.495e-07	0.115	0	0	0	0	0
L = 20	безусловная	1.929e-07	0.101	0	0.432	0.00064	0.051	0
	условная	1.967e-07	0.102	0	0	0	0	0

Таблица 4.2: Значения показателей для результатов безусловной и условной оптимизации функции стресса метода 2 при различных значениях гиперпараметра L

некоторым из достаточных условий теоремы 8, а также некоторым из аксиом расстояния, параметризованного размерам. В то время как результат условной оптимизации удовлетворяет и достаточным условиям теоремы, и всем аксиомам расстояния, параметризованного размером, то есть разработанный метод работает корректно.

На рисунке 4.7 приведен график зависимости Stress-1 от значения гиперпараметра L . График показывает, что с незначительной потерей точности аппроксимации можно приблизить входной тензор моделью с небольшим значением L .

4.4 Исследование метода 3

Алгоритм был реализован на языке Python с использованием библиотек `numpy` и `scipy`. Сначала выполнялась безусловная минимизация функции стресса, итерационный процесс останавливался при выполнении условия

$$\sum_{i=1}^N \sum_{j=1}^N (\alpha_{ij}^{[t-1]} - \alpha_{ij}^{[t]})^2 < \epsilon, \quad \epsilon = 1e - 13 \quad (4.13)$$

Затем полученные значения параметров использовались в качестве начального приближения для условной оптимизации. Оптимизация производилась с использованием метода SLSQP, реализованного в библиотеке `scipy.optimize`. При минимизации по r_i метод вызывался с параметром $tol = 1e - 15$, при минимизации по α_{ij} - с параметром $tol = 1e - 3$.



Рис. 4.7: Зависимость значения Stress-1 от значения гиперпараметра L для результатов метода 2

При этом в качестве начального приближения во всех случаях использовались вычисленные аналитически значения параметров, минимизирующие квадратичный функционал без учета условий. Итерационный процесс останавливался при выполнении (4.13).

Для метода 3 показатели, измеряющие, насколько не выполняются достаточные условия для параметра r , имеют вид

$$rNNP = \frac{\sum_{i=1}^N \sum_{k=1}^K r_{ik} \geq 0}{NK} \quad (4.14)$$

$$rMP = \frac{\sum_{i=1}^N \sum_{k_1 < k_2} \mathbb{I}[r_{ik_1} > r_{ik_2}]}{\sum_{k_1 < k_2} N} \quad (4.15)$$

$$rSP = \frac{\sum_{i=1}^N \sum_{k_1 < k_2} \mathbb{I}[k_1 + k_2 \leq K] \mathbb{I}[r_{ik_1} + r_{ik_2} < r_{ik_1+k_2}]}{\sum_{i=1}^N \sum_{k_1 < k_2} \mathbb{I}[k_1 + k_2 \leq K] N} \quad (4.16)$$

Введем также показатели, измеряющие, насколько не выполняются достаточные условия для параметра α :

$$\alpha SP = \frac{\sum_{i=1}^N \alpha_{ii} \neq 0}{N} \quad (4.17)$$

$$\alpha NNP = \frac{\sum_{i=1}^N \sum_{j=1}^N \alpha_{ij} \geq 0}{N^2} \quad (4.18)$$

Также были измерены итоговое значение функции стресса и значение функции Stress-1. Полученные значения для результатов безусловной и условной оптимизации функции стресса метода 3 приведены в таблице 4.3.

Используемая модель расстояния гарантирует выполнение только аксиомы симметричности, поэтому для проверки аксиом расстояния, параметризованного размером, для

оптимизация	Stress	Stress-1	rNNP	rMP	rSP	αSP	αNNP
безусловная	9.055e-08	0.069	0	0.394	0.00031	0	0.00016
условная	9.811e-08	0.072	0	0	0	0	0

Таблица 4.3: Значения показателей для результатов безусловной и условной оптимизации функции стресса метода 3

полученного тензора будем измерять показатели MP, IP, RP, NNP . Полученные значения для результатов безусловной и условной оптимизации функции стресса метода 3 приведены в таблице 4.4.

оптимизация	MP	IP	RP	NNP
безусловная	0.0534	2.735e-05	0	0
условная	0	0	0	0

Таблица 4.4: Значения показателей для результатов безусловной и условной оптимизации функции стресса метода 3

На рисунке 4.8 приведена визуализация значений параметра r после безусловной оптимизации и после условной оптимизации.

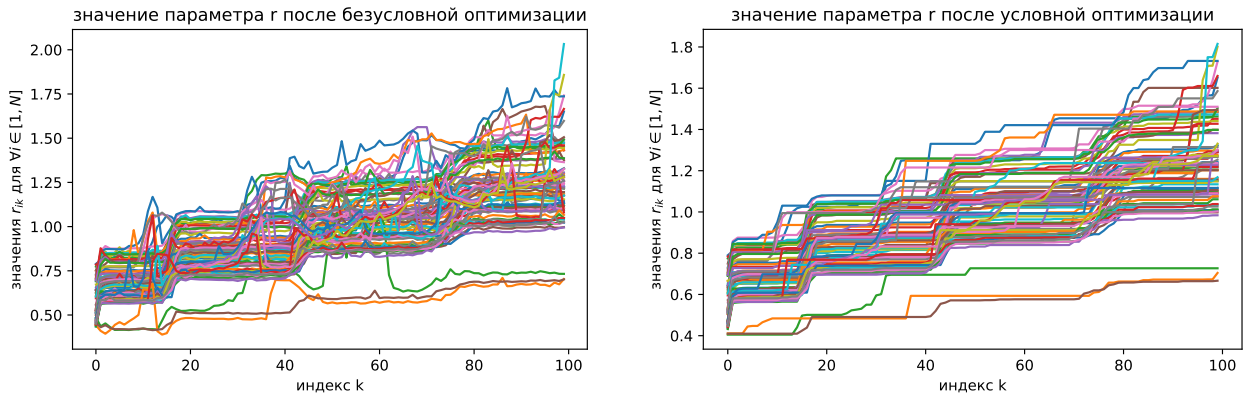


Рис. 4.8: Зависимость значений параметров $r_{ik} \forall i \in [1, N]$ от индекса k после безусловной оптимизации и после условной оптимизации функции стресса метода 3

Эксперименты показывают, что для использованных данных качество аппроксимации тензора различий после условной оптимизации незначительно ниже, чем после безусловной. При этом результат безусловной оптимизации функции стресса не удовлетворяет некоторым из достаточных условий теоремы 9, а также некоторым из аксиом расстояния, параметризованного размерам. В то время как результат условной оптимизации удовлетворяет и достаточным условиям теоремы, и всем аксиомам расстояния, параметризованного размером, то есть разработанный метод работает корректно.

4.5 Сравнение и выводы

Сравнение результатов безусловной и условной оптимизации функций стресса методов показывает, что для реальных данных, использованных для проведения экспериментов, требование обеспечения введенных в теоремах достаточных условий не приводит к значительному снижению качества аппроксимации.

Также эксперименты подтверждают, что предложенные методы, действительно, обеспечивают выполнение этих условий и, соответственно, выполнение для построенной аппроксимации всех аксиом расстояния, параметризованного размером. Стоит отметить, что для результатов безусловной оптимизации функций стресса некоторые из достаточных условий теорем, а также некоторые из аксиом расстояния, параметризованного размером, не выполняются.

Для методов 1 и 2 графики зависимости функции стресса от гиперпараметра L показывают, что реальные данные можно достаточно точно аппроксимировать точечной конфигурацией в пространстве небольшой размерности. В таблице 4.5 приведены значения функции Stress-1 для результатов методов 1 и 2 при различных значениях гиперпараметра L . Значения функции для метода 2 незначительно превосходят значения для метода 1, хотя модель расстояния, параметризованного размером, на которой основан метод 2, содержит намного больше параметров. Можно сделать вывод, что модель пропорциональных расстояний достаточно хорошо описывает данные, и использование своего веса для каждой координаты векторов групповой конфигурации излишне.

метод \ L	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 8	L = 10	L = 15	L = 20
метод 1	0.704	0.404	0.301	0.247	0.213	0.191	0.162	0.143	0.119	0.106
метод 2	0.699	0.406	0.301	0.247	0.213	0.191	0.160	0.141	0.115	0.102

Таблица 4.5: Значения функции Stress-1 для результатов методов 1 и 2 при различных значениях гиперпараметра L

Модель рядов содержит больше всего параметров, поэтому неудивительно, что основанный на ней метод дает наилучшее качество аппроксимации. Такой метод обеспечивает менее эффективное сжатие данных, чем методы 1 и 2, но позволяет получить для каждого объекта интерпретируемый вектор (ряд), описывающий изменение свойств этого объекта в зависимости от размера.

Глава 5

Заключение

В данной работе было введено формальное определение расстояния, параметризованного размером, а также некоторые его сужения и обобщения. Были рассмотрены конкретные модели таких расстояний и приведены достаточные условия на параметры этих моделей, обеспечивающие выполнение аксиом расстояния, параметризованного размером.

Также были разработаны методы сжатия трехмерного тензора различий, параметризованных размером, основанные на этих моделях. Предложенные методы гарантируют выполнение для аппроксимирующего тензора всех аксиом расстояния, параметризованного размером.

Методы были реализованы на языке Python и протестированы на реальных данных. Был выполнен анализ результатов экспериментов и проведено сравнение методов.

Литература

- [1] *Borg I., Groenen P. J. F.* Modern multidimensional scaling: Theory and applications. — Springer Science & Business Media, 2005.
- [2] *Cox T. F., Cox M. A.* Multidimensional Scaling. — New York: Chapman and Hall/CRC, 2000.
- [3] *Gower J. C.* A Q-technique for the calculation of canonical variates // *Biometrika*. — 1966. — Pp. 588–590.
- [4] *Mardia K. V., Kent J. T., Bibby J. M.* Multivariate analysis. — Probability and mathematical statistics. Academic Press Inc., 1979.
- [5] *Деза Е. И., Деза М. М.* Энциклопедический словарь расстояний. — Наука, 2008.
- [6] *Дэйвисон М. Л.* Многомерное шкалирование: методы наглядного представления данных. — Финансы и статистика, 1988.
- [7] *Майсурадзе А. И.* Гомогенные и ранговые базисы в пространствах метрических конфигураций // *Журнал вычислительной математики и математической физики*. — 2006. — Vol. 46, no. 2. — Pp. 344–360.
- [8] *Успенский В. А.* Что такое аксиоматический метод. — Ижевск: Научно-издательский центр «Регулярная и хаотическая динамика», 2001.