Расстояния, параметризованные размерами, и их аппроксимация

Петренко Дарья Павловна

Кафедра математических методов прогнозирования

e-mail: daria_petrenko@bk.ru

Научный руководитель — к.ф.-м.н. доц. Майсурадзе Арчил Ивериевич

Во многих прикладных областях при содержательной постановке и последующей формализации задач используется понятие расстояния. Однако для адекватного представления сложных ситуаций традиционного расстояния недостаточно и требуется использовать более богатые множества функций. В данной работе предлагается и исследуется теоретически и эмпирически одно из таких множеств — расстояния, параметризованные размером.

Определение 1. *Пусть X — произвольное множество. Пусть S — частично упорядоченное множество с операцией сложения.* Расстоянием, параметризованным размером*, будем называть функцию $\rho(x_1, x_2, s) : X \times X \times S \to \mathbb{R}$, 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. $\rho(x_1, x_2, s) \geqslant 0$ *(неотрицательность)*

- *аксиомы размера* $\forall x_1, x_2 \in X$, $\forall s_1, s_2 \in S$

S1. $s_1 \leqslant s_2 \Longrightarrow \rho(x_1, x_2, s_1) \leqslant \rho(x_1, x_2, s_2)$ *(монотонность)*

S2. $\rho(x_1, x_2, s_1 + s_2) \leqslant \rho(x_1, x_2, s_1) + \rho(x_1, x_2, s_2)$ *(неделимость)*

В работе доказана непротиворечивость и независимость системы аксиом (D1)–(D3), (S1), (S2). Также рассмотрен ряд сужений и расширений указанного множества функций, полученных добавлением новых аксиом или ослаблением требований соответственно. В частности, от *различия, параметризованного размером*, требуется только рефлексивность и неотрицательность.

Далее в работе ставится и решается различными способами следующая задача аппроксимации тензора различий. Рассмотрим конечный набор объектов мощности N и конечный набор размеров мощности K, причем объекты характеризуются только своими индексами в наборе, а размеры, для удобства дальнейших формулировок, — неотрицательные числа, пропорциональные своим индексам. Пусть для каждой пары объектов и каждого размера из рассматриваемых наборов измерено различие, параметризованное размером— получим трехмерный тензор $\Delta \in \mathbb{R}^{N \times N \times K}$, где элемент тензора δ_{ijk} содержит различие между i-тым и j-тым объектами, параметризованное k-тым размером. Требуется разработать модели и методы аппроксимации тензора

[*] Термин «аксиома» является традиционным для условий на функции расстояния.

Δ, причем различия δ_{ijk}, параметризованные размерами, аппроксимируются расстояниями ρ_{ijk}, параметризованными размерами. Точность аппроксимации оценивается с помощью взвешенного среднеквадратического отклонения $Stress = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{K} w_{ijk}(\delta_{ijk} - \rho_{ijk})^2$, где w_{ijk} — заданные веса.

В случае обычных расстояний задача относится к классу задач многомерного шкалирования, поэтому работа была построена в соответствии с идеологией современного многомерного шкалирования [1, 2]. При этом предложенные в работе модели расстояний, параметризованных размером, постановки задач аппроксимации и методы их решения являются оригинальными.

Для решения поставленной задачи в работе предложены 4 модели расстояний, параметризованных размером. Для каждой из моделей найдены достаточные условия на параметры («теоремы о достаточных условиях»), обеспечивающие выполнение всех аксиом из определения 1.

Задачи аппроксимации ставились как задачи минимизации взвешенного среднеквадратического отклонения $Stress$ исходного тензора от его аппроксимации в рамках выбранной модели, обеспечивая выполнение условий теоремы для соответствующей модели. Так как модели представлены в общем виде, то для постановки задачи аппроксимации для каждой из моделей был выбран конкретный частный случай. Частные случаи двух моделей совпали, поэтому было получено 3 различых задачи условной оптимизации.

Для решения каждой из задач в работе предложен свой метод условной оптимизации. Идея методов 1 и 2 состоит в итерационной условной минимизации вариационной верхней оценки функции стресса, идея метода 3 — в итерационной условной минимизации квадратичной функции стресса методом наименьших квадратов.

Все предложенные методы были реализованы и протестированы на реальных данных о задержках передачи сообщений между процессами в многопроцессорной системе «Ломоносов». Эксперименты показали хорошее качество аппроксимации для моделей с малым числом параметров (то есть высокую степень сжатия данных), а также сопоставимость потерь с безусловными постановками задач (в которых аксиомы из определения 1 игнорируются).

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.