

Восстановление значений многомерных временных рядов с помощью тензорного разложения и выделения временного индекса

Тихонов Денис Максимович

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

2024 г.

Восстановление значений многомерных временных рядов с помощью тензорного разложения и выделения временного индекса

Цель работы: Построение алгоритма восстановления пропущенных или зашумленных значений многомерного временного ряда сложной структуры с выделение индекса временной составляющей.

Проблема: Классические тензорные методы восстановления зашумленных значений временных рядов не используют априорное знание об особенностях одного из индексов входных данных — времени.

Гипотеза: использование в методах априорного знания о временной составляющей повысит качество восстанавливаемого многомерного временного ряда.

Требуется: разработать алгоритм, основанный на классический методах тензорного разложения, учитывающий временной индекс с помощью маскирования данных.

Постановка задачи восстановления временных рядов

- ① Дан многомерный временной ряд одноканальных снимков МРТ

$$\underline{\mathbf{X}} \in \mathbb{R}^{l_1 \times l_2 \times l_3}$$

- ② Модель восстановления — модификация разложения Такера, представляющего тензор через $\underline{\mathbf{G}}$ *центральный* тензор и фактор-матрицы $\mathbf{U}^{(n)}$:

$$\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)}$$

- ③ Решаемая задача оптимизации:

$$\begin{aligned} & \underset{(R_1, \dots, R_N)}{\text{minimize}} \sum_m R_m, \\ & \text{s.t.} \quad \|\underline{\mathbf{Q}}_T * (\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}})\|_F^2 < \epsilon \end{aligned}$$

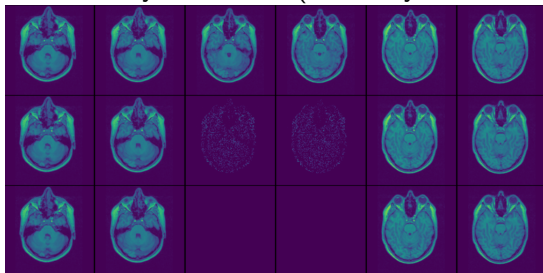
где (R_1, \dots, R_N) - ранги тензора в разложении Такера

Зашумленный временной ряд МРТ

- ① Дан многомерный временной ряд и тензор маскирования

$$\underline{\mathbf{X}} \in \mathbb{R}^{l_1 \times l_2 \times \dots \times l_N}, \underline{\mathbf{Q}}_T \in \{0, 1\}^{l_1 \times l_2 \times \dots \times l_N}$$

- ② Индексы l_1, l_3 обозначают координаты пикселей на снимке, индекс l_3 обозначает номер снимка или время, в которое снимок был сделан.
- ③ Предполагаем, что в моменты времени $l_3 \in 3, 4$ данные были сняты сильно зашумленными (или отсутствовали вовсе).



Разложение Такера для восстановления зашумленных тензоров

Input: тензоры $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \dots \times I_N}$, $\underline{\mathbf{Q}} \in \{0, 1\}^{I_1 \times I_2 \times \dots \times I_N}$.

Output: Разложение через тензор $\underline{\mathbf{G}}$ и фактор-матрицы $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$.

- 1 Случайная инициализация тензоров $\underline{\mathbf{G}}, \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(N)}$
- 2 repeat
- 3 $\hat{\underline{\mathbf{X}}} = \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \dots \times_N \mathbf{U}^{(N)}$
- 4 $\underline{\mathbf{Z}} = \underline{\mathbf{Q}} * \underline{\mathbf{X}} + (\underline{\mathbf{1}} - \underline{\mathbf{Q}}) * \hat{\underline{\mathbf{X}}}$
- 5 for $n = 1, \dots, N$ do:
- 6 $\underline{\mathbf{Y}} = \underline{\mathbf{Z}} \times_1 \mathbf{U}^{(1)} \dots \times_N \mathbf{U}^{(N)}$ исключая $\mathbf{U}^{(n)}$
- 7 $\mathbf{U}_{(n)}$ – матрица из R_n старших собств. векторов $\mathbf{Y}_{(n)}$
- 8 end for
- 9 $\underline{\mathbf{G}} = \underline{\mathbf{Z}} \times_1 \mathbf{U}^{(1)} \dots \times_N \mathbf{U}^{(N)}$
- 10 until разность норм $\|\underline{\mathbf{X}}\|_F - \|\hat{\underline{\mathbf{X}}}\|_F$ не перестанет убывать
- 11 return тензор $\underline{\mathbf{G}}$ и фактор-матрицы $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$

Модификация разложения для временных рядов

- 1 Для учета временной составляющей ввести дополнительный тензор маскирования применяемый к случайным временным срезам на каждой итерации

$$\underline{\mathbf{Q}}_T \in \{0, 1\}^{l_1 \times l_2 \times \dots \times l_N}.$$

В него так же включаются восстанавливаемые временные срезы.

- 2 Для выборка оптимальных рангов разложения Такера, предлагается начинать итеративный алгоритм с малых значений $R_1 = R_2 = \dots = R_N = 1$
- 3 На каждой итерации выбирать моду для повышения ранга основываясь на остатках по выбранной моде

$$\|\hat{\underline{\mathbf{X}}} \times_1 \mathbf{U}^{T(1)} \dots \times_N \mathbf{U}^{T(N)}\|_F,$$

исключая $\mathbf{U}^{T(n)}$ из произведения

Модифицированное разложение Такера

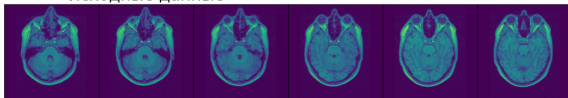
Input: тензоры $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \dots \times I_N}$, $\underline{\mathbf{Q}} \in \{0, 1\}^{I_1 \times I_2 \times \dots \times I_N}$.

Output: Разложение через тензор $\underline{\mathbf{G}}$ и фактор-матрицы $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$.

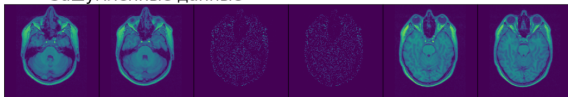
- 1 Случайная инициализация тензоров $\underline{\mathbf{G}}, \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(N)}$
- 2 **repeat**
- 3 Случайный временной тензор маскирования $\underline{\mathbf{Q}}_T$
- 4 Получить приближения классическим алгоритмом Такера $\underline{\mathbf{G}}, \mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$, используя дополнительное маскирование $\underline{\mathbf{Q}}_T$
- 5 Выбрать моду для повышения ранга
- 6 $\hat{m} = \underset{m}{\operatorname{argmax}} \|\hat{\underline{\mathbf{X}}} \times_1 \mathbf{U}^{\top(1)} \dots \times_N \mathbf{U}^{\top(N)}\|_F$, исключая $\mathbf{U}^{\top(m)}$ из произведения
- 7 Обновить ранг моды $R_{\hat{m}}$
- 8 **until** разность норм $\|\underline{\mathbf{X}}\|_F - \|\hat{\underline{\mathbf{X}}}\|_F$ не перестанет убывать
- 9 **return** тензор $\underline{\mathbf{G}}$ и фактор-матрицы $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$

Вычислительный эксперимент

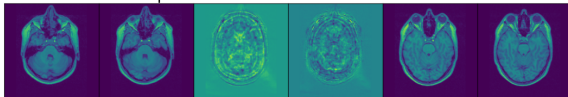
Исходные данные



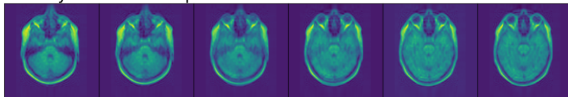
Зашумленные данные



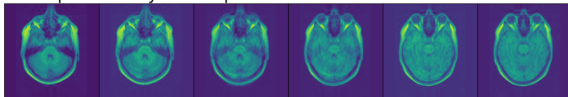
Каноническое разложение



Случайное маскирование



Маскирование с учетом времени



Сравнение методов на МРТ-изображениях головы человека

Результаты работы

Модель	Средний SSIM
CP-разложение	0.32
Случайное маскирование	0.54
Маскирование с учетом времени	0.58

SSIM (Structural similarity index measure) исходных и восстановленных МРТ-изображениях головы человека

Результаты:

- 1 Предложен модифицированный алгоритм восстановления тензорных временных рядов, основанный на разложении Такера;
- 2 Показано, что для задачи с временной структурой предложенный подход повышает точность восстанавливаемых данных.