Тензорный метод SSA в задаче декомпозиции многомерных временных рядов

Сёмкин Кирилл Вадим Стрижов

Московский Физико-Технический Институт

2023

Цель исследования

Проблема

Классические методы декомпозиции временных рядов не позволяют учесть взаимосвязь и общую природу порождения набора временных рядов.

Цель

Предложить метод, позволяющий выделить общую для набора сигналов структуру, и на её основании произвести разложение на аддитивные компоненты.

Решение

Использовать гипотезу порождения рядов общей динамической системой и идею метода SSA. Состыковать траекторные матрицы в тензор, применить каноническое тензорное разложение, на её основе получить декомпозицию типа SVD для каждой матрицы, далее разложить каждый ряд на компоненты.

Литература

- Stephan Rabanser and Oleksandr Shchur and Stephan Günnemann *Introduction to Tensor Decompositions and their Applications in Machine Learning*. arXiv, 2017.
- D. Stepanov and N. Golyandin *SSA-based approaches to analysis and forecast of multidimensional time series*. 2005.
- Kouchaki, Samaneh and Sanei, Saeid *Tensor based singular spectrum analysis for nonstationary source separation*. MLSP, 2013.
- Fu, Hang and Sun, Genyun and Zhang, Aizhu and Shao *Tensor Singular Spectrum Analysis for 3-D Feature Extraction in Hyperspectral Images*. IEEE Transactions on Geoscience and Remote Sensing, 2023.

Постановка задачи

Пусть имеем набор временных рядов $\{x_i(t_j)\}_{i=1}^{m}$, где сетка по времени $t_j \in \overrightarrow{1, N}$. Стоит задача разложения временных рядов на аддитивные компоненты:

$$x_1(t) = f_1(t) + f_2(t) + \ldots + f_{n_1}(t)$$
$$x_2(t) = g_1(t) + g_2(t) + \ldots + g_{n_2}(t)$$
$$\ldots$$

Проблема

Способов разложения бесконечно много, нужно конкретизировать требования на компоненты $f_i(t), g_i(t)$ и т.д.

Метод SSA

Имеем один временной ряд. Предполагаем его порождение некоторой (скрытой) динамической системой, используем теорему Такенса для её восстановления.

Выбираем длину окна L, строим *вектора задержек* $x_k = (x(t_k)\ x(t_{k+1})\ \dots\ x(t_{k+L-1}))$, собираем их по столбцам в ганкелеву матрицу T (*траекторная матрица*).

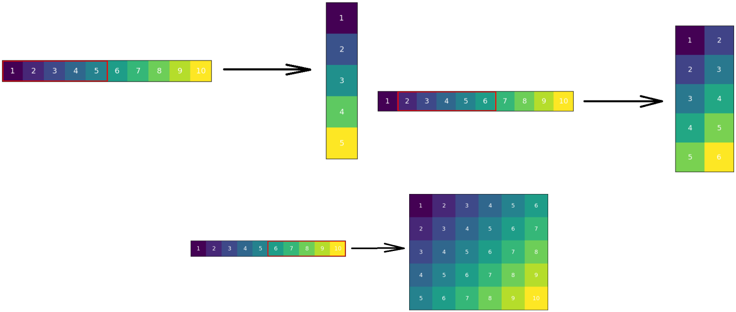


Рис. 1: Построение траекторной матрицы

Метод SSA

Далее к матрице применяется SVD-разложение $T = \sum^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i$

Предполагается существование разбиение факторов на 2 (или более) группы так, что $T = X_1 + X_2$, причём X_1, X_2 - также траекторные матрицы некоторых сигналов $f_1(t), f_2(t) \Rightarrow x(t) = f_1(t) + f_2(t)$

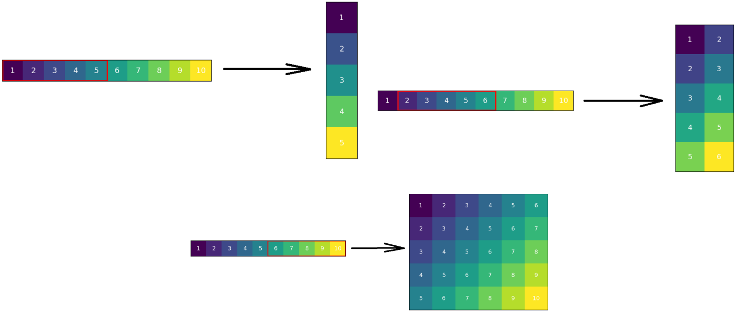


Рис. 1: Построение траекторной матрицы

Метод mSSA

Имеем набор временных рядов. Строим траекторные матрицы для каждого ряда $T_1, T_2, \ldots, T_m$, конкатенируем все матрицы в одну $T = [T_1 \;\; T_2 \ldots T_m]$, которую далее раскладываем с помощью SVD

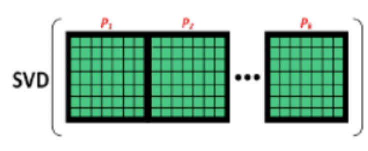


Рис. 2: Метод mSSA

$$T = \sum_{i}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i \Leftrightarrow \begin{cases} T_1 = \sum_{i}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^1 \\ T_2 = \sum_{i}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^2 \\ \ldots \\ T_m = \sum_{i}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^m \end{cases}$$

Метод tSSA

Теперь состыкуем матрицы по третьему измерению, получим траекторный тензор T. Применим CPD-разложение:

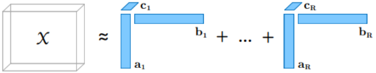


Рис. 3: Иллюстрация CPD разложения

С математической точки зрения это выглядит так:

$$\mathrm{T} = \sum_{i}^{r} \mathrm{a}_i \otimes \mathrm{b}_i \otimes \mathrm{c}_i \Leftrightarrow \begin{cases} T_1 = \sum_{i}^{r} \mathrm{c}_i[1] \cdot \mathrm{a}_i \mathrm{b}_i \\ T_2 = \sum_{i}^{r} \mathrm{c}_i[2] \cdot \mathrm{a}_i \mathrm{b}_i \\ \dots \\ T_m = \sum_{i}^{r} \mathrm{c}_i[m] \cdot \mathrm{a}_i \mathrm{b}_i \end{cases}$$

Вычислительный эксперимент

Цель вычислительного эксперимента: сравнить работу методов mSSA и tSSA на синтетических (набор синусов с разными фазами) и на реальных временных рядах (акселерометрия).

Метрики качества:

- интерпретируемость компонент разложения
- расхождение исходных рядов от суммы полученных декомпозиций

Синтетические данные

Сгенерирован набор $x_i(t) = \sin(w_i \cdot t) + \varepsilon_i$, где $w_i \in \{1, 2, 3\}$ - частоты, ε_i - небольшой гауссовский шум.

Синтетические данные

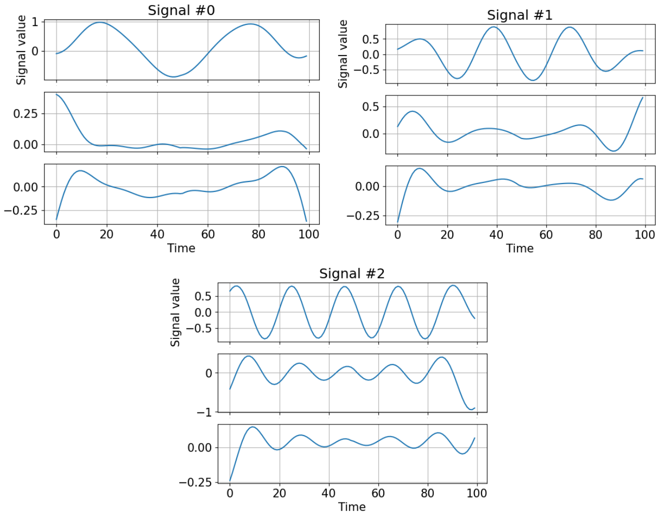


Рис. 4: Метод mSSA. Разложение рядов на компоненты

Синтетические данные

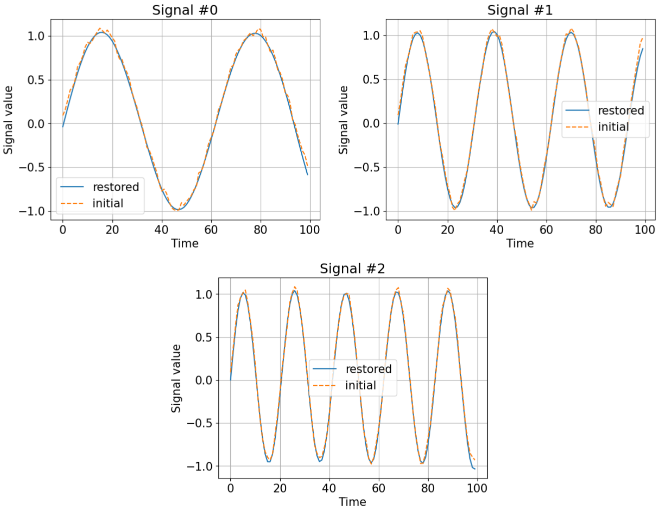


Рис. 4: Метод mSSA. Сопоставление истинного ряда и его аппроксимации

Синтетические данные

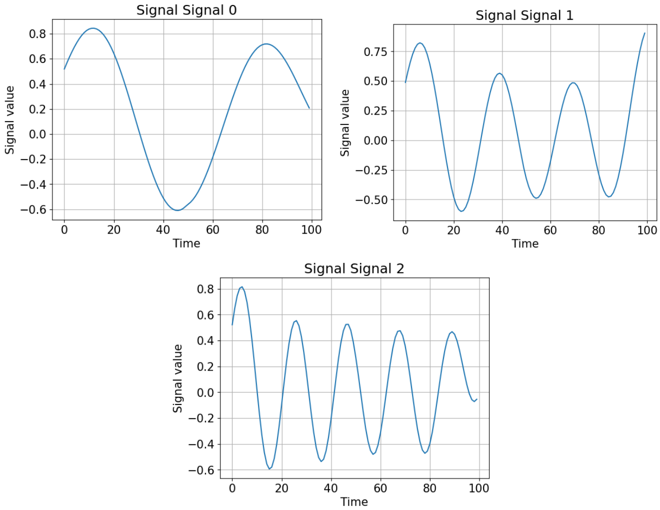


Рис. 5: Метод tSSA. Разложение рядов на компоненты

Синтетические данные

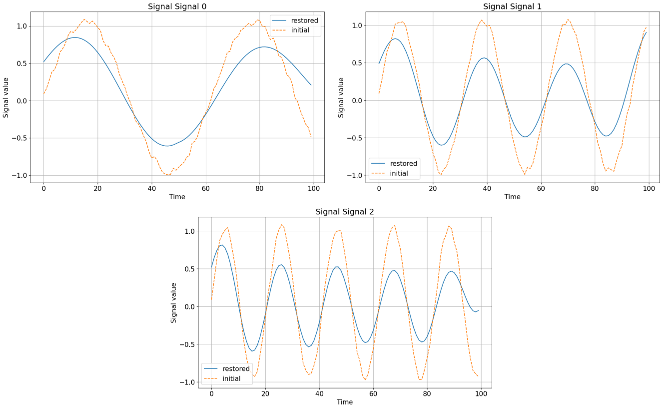
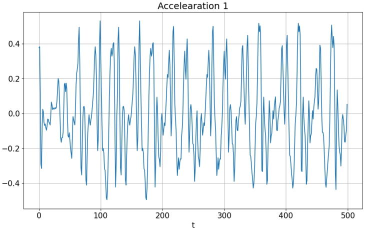
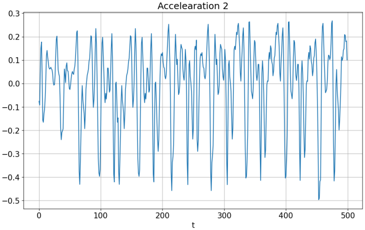


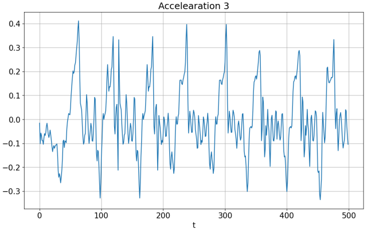
Рис. 5: Метод tSSA. Сопоставление истинного ряда и его аппроксимации

Акселерометрия

Имеем три временных ряда — компоненты вектора ускорения идущего человека.







Акселерометрия

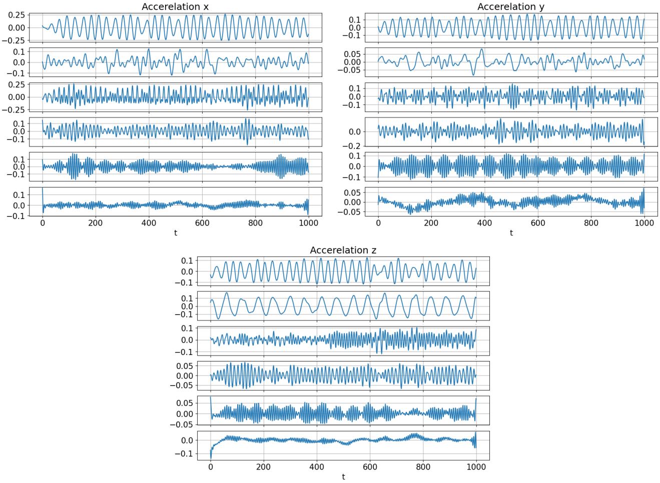


Рис. 6: Метод mSSA. Разложение рядов на компоненты

Акселерометрия

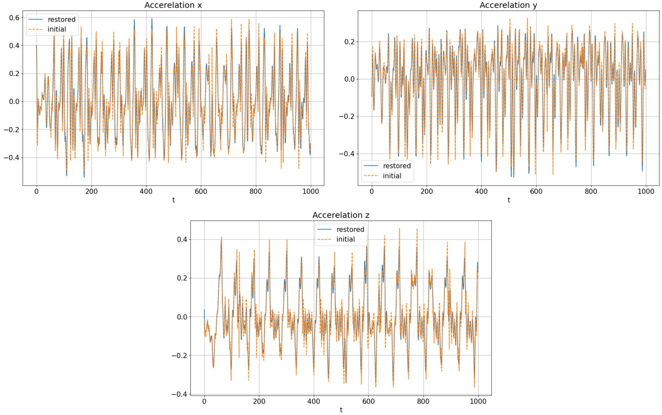


Рис. 6: Метод mSSA. Сопоставление истинного ряда и его аппроксимации

Акселерометрия

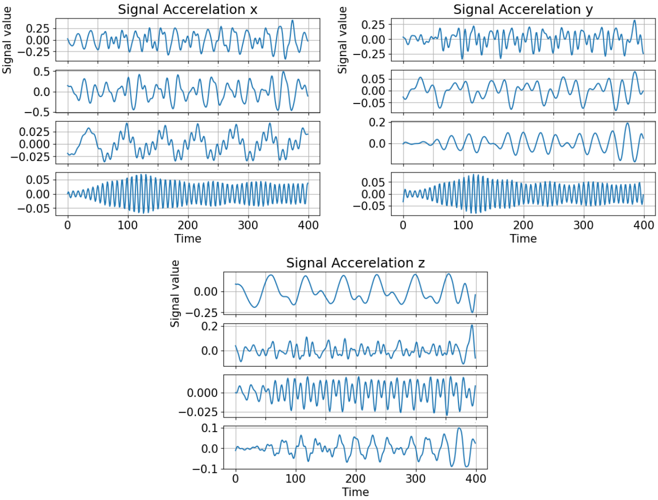


Рис. 7: Метод tSSA. Разложение рядов на компоненты

Акселерометрия

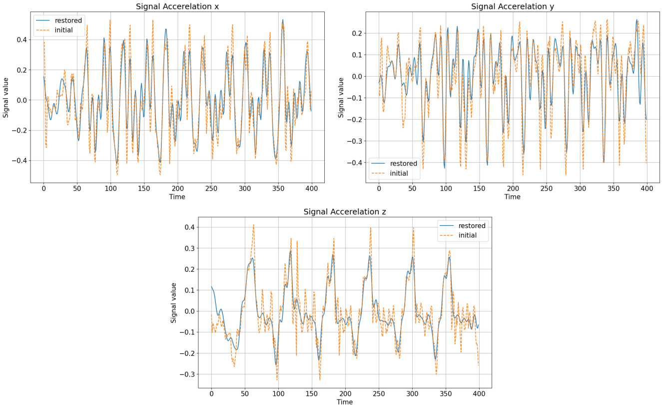


Рис. 7: Метод tSSA. Сопоставление истинного ряда и его аппроксимации

Анализ ошибки

- оба метода смогли вычленить сложные составляющие рядов, в том числе модулированные сигналы
- точность аппроксимации tSSA пока получается ниже, чем у mSSA
- тем не менее tSSA намного гибче в плане выбора компонент для каждого сигнала

Дальнейшая работа

Далее планируется внести поправки в алгоритм tSSA, продолжить его математический анализ, усложнить рассматриваемую модель, провести больше экспериментов.