Теплицев вариант анализа сингулярного спектра

Потешкин Егор Павлович, гр.20.Б04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Санкт-Петербург, 2023

Постановка задачи

Временной ряд $\mathsf{X} = (x_1, \ldots, x_N)$ — последовательность наблюдений, упорядоченных по времени.

Примеры: биржевой курс, замеры температуры в течении нескольких лет.

Дано: временной ряд состоит из тренда, сезонности (неслучайные состовляющие) и шума (случайная состовляющая): $\mathsf{X} = \mathsf{T} + \mathsf{S} + \mathsf{R}$.

Проблема: Как выделить неслучайные компоненты?

Решение: метод SSA.

Задача: Выделить сигнал как можно точнее.

Обозначения: оператор вложения и ганкелизации

$\mathsf{X} = (x_1, \ldots, x_N)$. Зафиксируем *длину окна* L, $1 < L < N$.

Оператор вложения $\mathcal{T}$:

$$\mathcal{T}(\mathsf{X}) = \mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix}$$

— *траекторная матрица*.

Оператор ганкелизации $\mathcal{H}$ — усреднение матрицы по побочным диагоналям.

SSA: алгоритм

Входные данные: временной ряд $\mathsf{X} = (x_1, \ldots, x_N)$.
Параметр: длина окна L.
Результат: m восстановленных временных рядов.

Рис.: Алгоритм SSA

MSSA: алгоритм

MSSA — общение SSA на многомерный случай, когда X — набор из D временных рядов (каналов). Также этот метод дает преимущество по сравнению с SSA, если сигналы имеют в большой степени одинаковую структуру.

Пусть $\mathsf{X} = \{\mathsf{X}^{(d)}\}_{d=1}^{D}$ — D-канальный временной ряд с длинами $N_1, \ldots, N_D$. Тогда требуется определить только шаг построения траекторной матрицы:

- *Вложение*: составная траекторная матрица $L \times K$

$$\mathbf{X} = [\mathcal{T}(\mathsf{X}^{(\mathbf{1})}) : \ldots : \mathcal{T}(\mathsf{X}^{(\mathbf{D})})] = [\mathbf{X}^{(\mathbf{1})} : \ldots : \mathbf{X}^{(\mathbf{D})}],$$

где $K = \sum_{i=1}^{D} K_i$, $K_i = N_i - L + 1$.

Стационарный случай

Если X — стационарный временной ряд, то можно улучшить базовый метод, используя вместо сингулярного тёплицево разложение матрицы $\mathbf{X}$, не изменяя при этом остальные этапы.

Тёплицев вариант MSSA был специально разработан для анализа стационарных временных рядов, поскольку имеет преимущество перед стандартным MSSA.

Существует 2 метода Toeplitz MSSA:

1. Метод Sum.
2. Метод Block.

Тёплицев MSSA: обозначение

Пусть X — D-канальный временной ряд с одинаковыми длинами $N_d = N\ \forall d = 1, \ldots, D$. Зафиксируем L.

Определим матрицу $\mathbf{T}^L_{l,k}$ с элементами

$$\{\mathbf{T}^L_{l,k}\}_{ij} = \frac{1}{\widetilde{N}_{i,j}} \sum_{n=\max(1,1+i-j)}^{\min(N,N+i-j)} x^{(l)}_n x^{(k)}_{n+j-i},\ 1 \leqslant i,j \leqslant L,$$

где $\widetilde{N}_{i,j} = \min(N, N+i-j) - \max(1, 1+i-j) + 1$.

Тёплицев MSSA: метод Sum

1. Построить $\mathbf{T}_{\mathsf{Sum}} = \sum_{i=1}^{D} \mathbf{T}_{i,i}^{L}$.
2. Найти ортонормированные собственные векторы $H_1, \ldots, H_L$ матрицы $\mathbf{T}_{\mathsf{Sum}}$ и получить разложение

$$\mathbf{X} = \sum_{i=1}^{L} H_i Z_i^{\mathrm{T}},$$

где $Z_i = \mathbf{X}^{\mathbf{T}} H_i$.

Трудоемкость построения матрицы: $\mathcal{O}(DN^2)$.

Тёплицев MSSA: метод Block

1. Построить $\mathbf{T}_{\mathsf{Block}} = \begin{pmatrix} \mathbf{T}^K_{1,1} & \mathbf{T}^K_{1,2} & \cdots & \mathbf{T}^K_{1,D} \\ \mathbf{T}^K_{2,1} & \mathbf{T}^K_{2,2} & \cdots & \mathbf{T}^K_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}^K_{D,1} & \mathbf{T}^K_{D,D} & \cdots & \mathbf{T}^K_{D,D} \end{pmatrix}$, где $K = N - L + 1$.
2. Найти ортонормированные собственные векторы $Q_1, \ldots, Q_{DK}$ матрицы $\mathbf{T}_{\mathsf{Block}}$ и получить разложение

$$\mathbf{X} = \sum_{i=1}^{DK} (\mathbf{X}Q_i)Q_i^{\mathrm{T}} = \sum_{i=1}^{DK} P_i Q_i^{\mathrm{T}}.$$

Трудоемкость построения матрицы: $\mathcal{O}(D^2 N^2)$.

Численное исследование

Дано: $(\mathsf{F}^{(1)}, \mathsf{F}^{(2)}) = (\mathsf{S}^{(1)}, \mathsf{S}^{(2)}) + (\mathsf{R}^{(1)}, \mathsf{R}^{(2)})$, $N = 71$.

Задача: проверить точность базового и модифицированных методов для разных значений параметра L.

Рассмотрим 3 случая:

1. Косинусы с одинаковыми периодами:

$$s_n^{(1)} = 30\cos(2\pi n/12), \quad s_n^{(2)} = 20\cos(2\pi n/12), \quad n = 1, \ldots, N.$$

2. Косинусы с разными периодами:

$$s_n^{(1)} = 30\cos(2\pi n/12), \quad s_n^{(2)} = 20\cos(2\pi n/8), \quad n = 1, \ldots, N.$$

3. Полиномы первой степени (нестационарные ряды):

$$s_n^{(1)} = 6n, \quad s_n^{(2)} = 4n, \quad n = 1, \ldots, N.$$

Численное исследование. Результаты

Таблица: MSE восстановления сигнала.

Случай 1	$L = 12$	$L = 24$	$L = 36$	$L = 48$	$L = 60$
SSA	3.25	**2.01**	**2.00**	**2.01**	3.25
Toeplitz SSA	3.2	1.87	1.63	**1.59**	1.67
MSSA	3.18	1.83	1.59	**1.47**	2.00
Sum	3.17	1.75	1.44	**1.32**	**1.33**
Block	1.39	**1.26**	**1.25**	1.33	1.97
Случай 2	$L = 12$	$L = 24$	$L = 36$	$L = 48$	$L = 60$
SSA	3.25	**2.01**	**2.00**	**2.01**	3.25
Toeplitz SSA	3.2	1.87	1.63	**1.59**	1.67
MSSA	6.91	3.77	3.07	**2.88**	3.84
Sum	6.88	3.65	2.64	2.37	**2.27**
Block	4.47	3.67	**3.22**	**3.23**	3.8
Случай 3	$L = 12$	$L = 24$	$L = 36$	$L = 48$	$L = 60$
SSA	3.34	2.02	**1.87**	2.02	3.34
Toeplitz SSA	**5.08**	19.3	63.63	169.91	383.92
MSSA	3.3	1.9	1.59	**1.52**	2.09
Sum	**4.58**	14.48	46.39	123.06	277.61
Block	278.23	123.47	46.38	14.06	**3.25**

Выводы

1. Sum и Block версии Toeplitz MSSA для стационарного ряда точнее выделяют сигнал, чем Basic MSSA. Если в сигналах разных каналов присутствует одна и та же частота, то Block чуть лучше Sum. Но если частоты разные, то Block существенно хуже Sum.
2. В случае нестационарных рядов (например, в которых присутствует тренд) оба метода показывают плохие результаты.
3. Применять Block и Sum Toeplitz MSSA рекомендуется к временным рядам, заранее выделив из них тренд.
4. Рекомендуется использовать длину окна $L \gg (N+1)/2$ для метода Sum и $L \approx (N+1)/2$ для метода Block.