Прогнозирование сегмента временного ряда методом сингулярного анализа спектра

Гудков Денис

ФПМИ МФТИ

Цель работы

Рассказать о прогнозировании сегмента временного ряда методом сингулярного анализа спектра. Проанализировать исходную размерность матрицы Ганкеля и ее сниженную размерность.

Применение сингулярного разложения (SSA)

SSA позволяет:

- Прогнозировать значения временного ряда
- Выделять тренды и периодичность
- Удалять шум
- Разделять детерминистические и стохастические компоненты

Постановка задачи прогнозирования с помощью SSA

Пусть задан временной ряд:

$$X = \{x_1, x_2, \dots, x_N\}, \quad x_t \in \mathbb{R}.$$

Необходимо предсказать будущие значения:

$$x_{N+1}, x_{N+2}, \dots, x_{N+M}.$$

Построение матрицы Ганкеля

Выбираем длину окна L и формируем матрицу траекторий:

$$\mathrm{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \dots & x_N \end{pmatrix}, \quad K = N - L + 1.$$

Данная матрица называются матрицей Ганкеля. Рекомендации по выбору L:

- Слишком маленькое L: плохо выделяются сезонные колебания, шум может доминировать
- Слишком большое L: тяжёлые вычисления, важные колебания становятся менее четкими
- Сезонные компоненты: лучше выбирать $L \geq$ периода ряда

Сингулярное разложение

Пусть $\mathrm{S} = \mathrm{X}\mathrm{X}^{\mathrm{T}}$, а $\lambda_1, \ldots, \lambda_L$ — собственные значения матрицы S, $U_1, \ldots, U_L$ — ортонормированный базис из собственных векторов матрицы S, соответствующих этим собственным значениям. Пусть $d = \mathrm{rank} X$ и $V_i = \mathrm{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i}, \quad i = 1, \ldots, d$.
В этой записи SVD матрицы X можно записать как

$$\mathrm{X} = \mathrm{X}_1 + \cdots + \mathrm{X}_d, \tag{1}$$

где

$$\mathrm{X}_i = \sqrt{\lambda_i}\, U_i V_i^{\mathrm{T}}. \tag{2}$$

Векторы $\sqrt{\lambda_i} V_i = \mathrm{X}^{\mathrm{T}} U_i$ называются векторами главных компонент. Векторы U_i называются левыми сингулярными векторами матрицы X, а числа $\sqrt{\lambda_i}$ являются её сингулярными значениями.

Перегруппировка компонент

Множество индексов $\{1, \dots, d\}$ делится на m непересекающихся подмножеств $I_1, \dots, I_m$.
Пусть

$$I = \{i_1, \dots, i_p\}.$$

Тогда результирующая матрица X_I, соответствующая группе I, определяется как

$$\mathrm{X}_I = \mathrm{X}_{i_1} + \ldots + \mathrm{X}_{i_p}.$$

Результирующие матрицы вычисляются для групп

$$I = I_1, \dots, I_m,$$

и сгруппированное SVD-разложение X можно записать как

$$\mathrm{X} = \mathrm{X}_{I_1} + \ldots + \mathrm{X}_{I_m}.$$

Как выбираются компоненты для перегруппировки

1. **Анализ сингулярного спектра.** Смотрим на убывание $\sqrt{\lambda_i}$. Резкие скачки говорят о границах между трендом, сезонностью и шумом. Малые сингулярные значения убираются.
2. **Анализ левых сингулярных векторов.** Гладкие компоненты без колебаний соответствуют тренду, синусоидальные и идущие парами — сезонности, хаотичные — шуму. Пары компонент объединяются в одну группу.

Итоговая группировка состоит из компоненты тренда и сезонных компонент.

Сравнение исходной и перегруппированной матриц Ганкеля

После SVD и перегруппировки компонент:

$$X = X_{I_1} + X_{I_2}$$

- X_{I_1} — тренд
- X_{I_2} — сезонность (пары гармоник)

Исходная матрица X содержит весь ряд, включая шум, а перегруппированная матрицавыделяет только значимые структурные компоненты. Снижение размерности упрощает анализ и повышает точность прогноза

Построение линейного рекуррентного соотношения

После перегруппировки компонент SSA формируем восстановленный ряд:

$$\mathsf{X}_{\text{recon}} = \sum_{i \in I} \mathsf{X}_i$$

Линейное рекуррентное соотношение для группы компонент имеет вид:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_{L-1} y_{t-L+1}, \quad t = L, \ldots, N$$

Коэффициенты $a_1, \ldots, a_{L-1}$ вычисляются через левые сингулярные векторы U_i группы I:

$$A = \frac{1}{1 - \nu^2} \sum_{i \in I} \pi(U_i)\, \underline{U_i}^\top,$$

где $\underline{U_i} = (u_{i1}, \ldots, u_{i,L-1})^\top$, $\pi(U_i) = u_{iL}$, $\nu^2 = \sum_{i \in I} u_{iL}^2$.
Источник: N. Golyandina, A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*, Springer, 2013.

Итоговый прогноз временного ряда

После построения линейного рекуррентного соотношения:

1. Для каждой группы I_j строится прогноз $\hat{y}^{(I_j)}_{N+t}$:

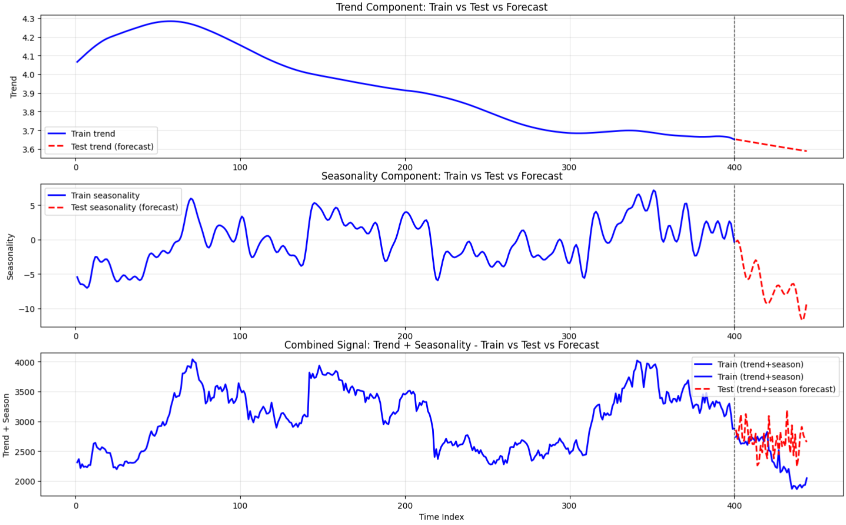
$$\hat{y}^{(I_j)}_{N+t} = a^{(j)}_1 \, y^{(I_j)}_{N+t-1} + \dots + a^{(j)}_{L-1} \, y^{(I_j)}_{N+t-L+1}, \quad t = 1, \dots, h$$

2. Итоговый прогноз ряда получается суммированием прогнозов по группам:

$$\hat{x}_{N+t} = \sum_{j=1}^{m} \hat{y}^{(I_j)}_{N+t}, \quad t = 1, \dots, h$$


Сингулярный спектр
$\sqrt{\lambda_i}$
800
700
600
500
400
300
200
100
0
0
25
50
75
100
125
150
175
200


Кумулятивная дисперсия SSA
Кумулятивная доля дисперсии
1.0
0.9
0.8
0.7
0.6
0
25
50
75
100
125
150
175
200
Компонента


Trend Component: Train vs Test vs Forecast
Trend
Train trend
Test trend (forecast)
Seasonality Component: Train vs Test vs Forecast
Seasonality
Train seasonality
Test seasonality (forecast)
Combined Signal: Trend + Seasonality - Train vs Test vs Forecast
Trend + Season
Train (trend+season)
Train (trend+season)
Test (trend+season forecast)
Time Index