Метод «Гусеница»–SSA для анализа временных рядов с пропусками

Осипов Евгений Вадимович, 522-я группа

Санкт-Петербургский Государственный Университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель — к.ф.-м.н. **Н.Э. Голяндина**
Рецензент — к.ф.-м.н. **В.В. Некруткин**

Санкт-Петербург
2006г.

Постановка задачи

- Базовый метод «Гусеница»–SSA эффективно решает задачу выявления составляющих временного ряда:
  - трендовой,
  - гармонических,
  - шумовых.
- Метод применим для рядов без пропусков.
- Многие реальные временные ряды часто содержат пропуски.
- ***Задача*:**
  построить модифицированный алгоритм, который
  - **решает те же задачи,**
  - **заполняет пропуски.**

Понятия метода «Гусеница»–SSA

- $F_N = (f_0, \ldots, f_{N-1})$, L — длина окна;
- Вектора вложения $\{X_i\}_{i=1}^{K}$, $K = N - L + 1$,
  $X_i = (f_{i-1}, \ldots, f_{i+L-1})^{\mathrm{T}}$, $\mathbf{X} = [X_1 : \cdots : X_L]$;
- $\mathcal{L}^{(L)}(F_N) = \mathrm{span}(X_1, \ldots, X_L)$ — траекторное пространство,
  его базис $\{U_i\}$ — собственные вектора $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$,
  SVD матрицы $\mathbf{X}$: $\mathbf{X} = \sum \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$;
- Выделение $F_N^{(1)}$ из $F_N = F_N^{(1)} + F_N^{(2)}$:
  - выбор $\{i_1, \ldots, i_r\}$ и построение $\mathcal{L}^{(1)} = \mathrm{span}(U_{i_1}, \ldots, U_{i_r})$;
  - проектирование векторов $\{X_i\}_{i=1}^{K}$ на $\mathcal{L}^{(1)}$ $\longrightarrow$ $\widehat{\mathbf{X}}^{(1)} = [\widehat{X}_1^{(1)} : \cdots : \widehat{X}_L^{(1)}]$;
    $\widehat{X}_i^{(1)} = \mathbf{\Pi}^{(1)} X_i$;
  - диагональное усреднение: $\widehat{\mathbf{X}}^{(1)} \longrightarrow \widehat{F}_N^{(1)}$.

Понятия метода «Гусеница»–SSA

- F_N — ряд *конечного ранга* d, если $\dim \mathcal{L}^{(L)}(F_N) = d \quad \forall\, L$.

- *Любой ряд, являющийся линейной комбинацией произведений полиномов, экспонент и гармоник, является рядом конечного ранга.*

- Прогнозирование значений ряда в $\mathcal{L}^{(L)}(F_N)$, т.е. получение последней компоненты вектора вложения в виде линейной комбинации остальных.

- $F_N^{(1)}$ и $F_N^{(2)}$ *слабо разделимы*, если $\mathcal{L}^{(1)} \perp \mathcal{L}^{(2)}$ и $\mathcal{K}^{(1)} \perp \mathcal{K}^{(2)}$.

- Наличие разделимости дает возможность выделить $F_N^{(1)}$ из $F_N^{(1)} + F_N^{(2)}$.

Построение базиса $\mathcal{L}^{(L)}(F_N)$

$$F_{12} = (\star\ \star\ \star\ \star\ \star\ \star\ \cdot\ \cdot\ \cdot\ \cdot\ \star\ \star), \quad L = 5,\ d = 2$$

$$\mathbf{X} = \begin{pmatrix} \star & \star & \star & \star & \star & \star & \cdot & \cdot \\ \star & \star & \star & \star & \star & \cdot & \cdot & \cdot \\ \star & \star & \star & \star & \cdot & \cdot & \cdot & \cdot \\ \star & \star & \star & \cdot & \cdot & \cdot & \cdot & \star \\ \star & \star & \cdot & \cdot & \cdot & \cdot & \star & \star \end{pmatrix}.$$

Предложение

Если среди векторов вложения, не содержащих пропущенные значения, найдется хотя бы d линейно независимых, то траекторное пространство столбцов находится точно.

Предложение

Если ряд F_N имеет ранг d, $e_1, e_L \notin \mathcal{L}^{(L)}(F_N)$, и ряд имеет $L + d - 1$ подряд идущих непропущенных значений, то траекторное пространство столбцов находится точно.

Проектирование векторов вложения

- $F_N = F_N^{(1)} + F_N^{(2)}$, разделимость ($\mathcal{L}^{(1)} \perp \mathcal{L}^{(2)}$), $\{R_i\}_{i=1}^d$ — базис $\mathcal{L}^{(1)}$;
- $X = \begin{pmatrix} X|_{\mathcal{I}\setminus\mathcal{P}} \\ \cdot \end{pmatrix} \xrightarrow{\mathrm{I}} \begin{pmatrix} X^{(1)}|_{\mathcal{I}\setminus\mathcal{P}} \\ \cdot \end{pmatrix} \xrightarrow{\mathrm{II}} \begin{pmatrix} X^{(1)}|_{\mathcal{I}\setminus\mathcal{P}} \\ X^{(1)}|_{\mathcal{P}} \end{pmatrix}$,
  $\mathcal{P}$ — множество индексов пропущенных значений в X, $\mathcal{I} = \{1, \ldots, L\}$;

Предложение

Пусть $\mathcal{L}^{(1)}|_{\mathcal{I}\setminus\mathcal{P}} \perp \mathcal{L}^{(2)}|_{\mathcal{I}\setminus\mathcal{P}}$.
Тогда $X^{(1)}|_{\mathcal{I}\setminus\mathcal{P}} = \mathbf{\Pi}^{(1)}_{\mathcal{I}\setminus\mathcal{P}}(X|_{\mathcal{I}\setminus\mathcal{P}})$ *и для* $\mathbf{R} = [R_1 : \ldots : R_d]$

- $\Pi^{(1)}_{\mathcal{I}\setminus\mathcal{P}} =$
  $$\mathbf{R}|_{\mathcal{I}\setminus\mathcal{P}}(\mathbf{R}|_{\mathcal{I}\setminus\mathcal{P}})^{\mathrm{T}} + \mathbf{R}|_{\mathcal{I}\setminus\mathcal{P}}(\mathbf{R}|_{\mathcal{P}})^{\mathrm{T}}(\mathbf{E}_{|\mathcal{P}|} - \mathbf{R}|_{\mathcal{P}}\mathbf{R^T}|_{\mathcal{P}})^{-1}\mathbf{R}|_{\mathcal{P}}(\mathbf{R}|_{\mathcal{I}\setminus\mathcal{P}})^{\mathrm{T}};$$

Предложение

Пусть $\mathrm{span}(e_i|\ i \in \mathcal{P}) \cap \mathcal{L}^{(1)} = \{0_L\}$.
Тогда $X^{(1)}|_{\mathcal{P}} = (\mathbf{E}_{|\mathcal{P}|} - \mathbf{R}|_{\mathcal{P}}\mathbf{R}^{\mathrm{T}}|_{\mathcal{P}})^{-1}\mathbf{R}|_{\mathcal{P}}\mathbf{R}^{\mathrm{T}}|_{\mathcal{I}\setminus\mathcal{P}}X^{(1)}|_{\mathcal{I}\setminus\mathcal{P}}$.

Группы пропусков и их заполнение

- Пропуски $\quad\longrightarrow\quad$ *группы* (внутренние, левые/правые, сплошные);
  $$L = 3, \quad F_{19} = (\star\ \star\ \underbrace{\cdot\ \star\ \cdot}_{(1)}\ \star\ \star\ \star\ \star\ \underbrace{\cdot\ \cdot\ \cdot}_{(2)}\ \star\ \star\ \star\ \star\ \underbrace{\cdot\ \cdot\ \cdot}_{(3)})$$
  (1) — левая, (2) — внутренняя сплошная, (3) — правая сплошная;

- Группа пропусков $\quad\longleftrightarrow\quad$ *набор векторов вложения группы пропусков*
  $$(2): \quad \begin{pmatrix} \star & \star & \cdot & \cdot & \cdot \\ \star & \cdot & \cdot & \cdot & \star \\ \cdot & \cdot & \cdot & \star & \star \end{pmatrix};$$

- Разные способы заполнения пропусков
  $$\begin{pmatrix} \star & \star & {\color{red}\blacktriangle} & {\color{green}\blacktriangle} & {\color{blue}\blacktriangle} \\ \star & {\color{red}\blacktriangle} & {\color{green}\blacktriangle} & {\color{blue}\blacktriangle} & \star \\ {\color{red}\blacktriangle} & {\color{green}\blacktriangle} & {\color{blue}\blacktriangle} & \star & \star \end{pmatrix}.$$

Заполнение группы пропусков

- $$\begin{pmatrix} f_{i_1-3} & f_{i_1-2} & f_{i_1-1} & \cdot & \cdot \\ f_{i_1-2} & f_{i_1-1} & \cdot & \cdot & f_{i_1+2} \\ f_{i_1-1} & \cdot & \cdot & f_{i_1+2} & f_{i_1+3} \\ \cdot & \cdot & f_{i_1+2} & f_{i_1+3} & f_{i_1+4} \end{pmatrix};$$
- $X\big|_{\mathcal{P}} = \left(\mathbf{E}_{|\mathcal{P}|} - \mathbf{R}\big|_{\mathcal{P}}\mathbf{R}^{\mathrm{T}}\big|_{\mathcal{P}}\right)^{-1}\mathbf{R}\big|_{\mathcal{P}}\mathbf{R}^{\mathrm{T}}\big|_{\mathcal{I}\setminus\mathcal{P}}X\big|_{\mathcal{I}\setminus\mathcal{P}};$
- Применение формулы к каждому вектору *набора векторов вложения группы* — *«Одновременное восстановление пропусков»*;
- Другой способ — *«Последовательное восстановление слева (справа)»* на основе формулы при $\mathcal{P} = \{L\}$ ($\mathcal{P} = \{1\}$).
$$\begin{pmatrix} f_{i_1-3} & f_{i_1-2} & f_{i_1-1} & L_1 & L_2 \\ f_{i_1-2} & f_{i_1-1} & L_1 & L_2 & f_{i_1+2} \\ f_{i_1-1} & L_1 & L_2 & f_{i_1+2} & f_{i_1+3} \\ L_1 & L_2 & f_{i_1+2} & f_{i_1+3} & f_{i_1+4} \end{pmatrix}.$$

Построение модифицированного алгоритма

Два способа построения модифицированного алгоритма:

- Формальная замена скалярного произведения на операцию " * ":

  $A = (a_1, ..., a_m)^\mathrm{T}$ — с множеством индексов пропусков $\mathcal{A}$,
  $B = (b_1, ..., b_m)^\mathrm{T}$ — с множеством индексов пропусков $\mathcal{B}$,

  $$A^\mathrm{T} * B = \gamma \sum_{k:k\notin\mathcal{A}\cup\mathcal{B}} a_k b_k, \quad \text{где } \gamma = \frac{m}{m - |\mathcal{A}\cup\mathcal{B}|};$$

- Использование предложенных способов заполнения пропусков в выбранном подпространстве.

Программа

- Возможности программы позволяют обрабатывать реальные временные ряды;
- Реализовано много способов восстановления составляющих ряда с заполнением пропусков, что позволяет работать с различным расположением пропусков;
- Возможность применять различные методы к разным группам пропусков реализована в виде задания приоритетов методов.

Пример: ряд Airpass

Объемы пассажироперевозок по месяцам, отсутствует 1 год, $L = 36$

Пример: ряд Airpass

Тренд и главная часть сезонной составляющей