目录

第二章 SSA基础 目录

2.1 基础算法 目录

2.1.1 算法描述 目录

考虑一个长度为N的时间序列$\mathbb{X} = \mathbb{X}_N = (x_1, \ldots, x_N)$。假设$N > 2$且$\mathbb{X}$是一个非零序列。用$L(1 < L < N)$表示窗口长度，并且$K = N - L + 1$。下面将讲述SSA的基础算法，其中包含两个阶段：分解和重构。

2.1.1.1 第一阶段：分解 目录

第一步：嵌入

我们将原始序列映射为长度为L的$K = N - L + 1$个滞后向量，称其为L-滞后向量。

$$X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}} \quad (1 \le i \le K)$$

由L-滞后向量组成L-轨迹矩阵：

$$\mathbf{X} = [X_1 : \ldots : X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots x_K \\ x_2 & x_3 & x_4 & \ldots x_{K+1} \\ x_3 & x_4 & x_5 & \ldots x_{K+2} \\ \vdots & \vdots & \vdots & \ddots \\ x_L & x_{L+1} & x_{L+2} & \ldots x_N \end{pmatrix} \tag{2.1}$$

滞后向量X_i是**轨迹矩阵**$\mathbf{X}$的列。$\mathbf{X}$的行和列都是原序列的子序列。$\mathbf{X}$中的位于(i,j)的元素$x_{ij} = x_{i+j-1}$，因此$\mathbf{X}$中反对角线的元素相等。（因此轨迹矩阵也被称称作**汉克尔矩阵**）
公式(2.1)定义了一个从时间序列到轨迹矩阵的一对一的映射。

第二步：奇异值分解
在这一步，我们对$\mathbf{X}$进行SVD分解。令$\mathbf{S} = \mathbf{X}\mathbf{X}^T$并且用$\lambda_1, \ldots, \lambda_L$表示$\mathbf{S}$的特征值，其中这些特征值按照从大到小的顺序排列。用$U_1, \ldots, U_L$表示对应的特征向量。
用d表示矩阵$\mathbf{X}$的秩，一般情况下，如果时间序列是真实世界中获取到的，$d = L^* = minL, K$。记$V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i} (i = 1, \ldots, d)$。轨迹矩阵$\mathbf{X}$可以被分解为：

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d \tag{2.2}$$

其中$\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$。
矩阵$\mathbf{X}_i$的秩为1，这样的矩阵被称作初等矩阵（elementary matrices），
集合$\left(\sqrt{\lambda_i}, U_i, V_i\right)$被称作三特征（eigentriple）简记作ET。

2.1.1.2 第二阶段：重构 目录

第三步：三特征分组
一旦得到(2.2)的分解结果，就把d个分解结果分配到m个不相交的子集中$I_1, \ldots, I_m$。令$I = \{i_1, \ldots, i_p\}$。
分组I对应的分组结果$\mathbf{X}_I = \mathbf{X}_{i_1} + \ldots + \mathbf{X}_{i_p}$。
最终分组后的结果记为

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m} \tag{2.3}$$

挑选集合$I_1, \ldots, I_m$的过程称作三特征分组，如果$m = d$，即$I_j = \{j\}, j = 1, \ldots, d$，则该分组成为初等分组。

第四步：对角平均
在这一步，我们需要把矩阵$\mathbf{X}_{I_j}$还原为时间序列。

令$\mathbf{Y}$表示$L*K$的一个矩阵，记$L^*=min(L,K),K^*=max(L,K)$，$N=L+K-1$。
令$y^*_{ij}=y_{ij}$如果$L<K$，否则$y^*_{ij}=y_{ji}$。
使用以下公式将矩阵$\mathbf{Y}$还原为时间序列：

$$y_k=\begin{cases}\frac{1}{k}\sum_{m=1}^{k}y^*_{m,k-m+1} & \text{for } 1\le k<L^*,\\ \frac{1}{L^*}\sum_{m=1}^{L^*}y^*_{m,k-m+1} & \text{for } L^*\le k\le K^*,\\ \frac{1}{N-k+1}\sum_{m=k-K^*+1}^{N-K^*+1}y^*_{m,k-m+1} & \text{for } K^*<k\le N.\end{cases}\tag{2.4}$$

该公式对应着求取矩阵反对角线元素的平均值。

对$\mathbf{X}_{I_k}$使用对角平均可以得到重构序列$\tilde{\mathrm{X}}^{(k)}=\left(\tilde{x}^{(k)}_1,\ldots,\tilde{x}^{(k)}_N\right)$。最终，初始序列被分解为$m$成分：

$$x_n=\sum_{k=1}^{m}\tilde{x}^{(k)}_n\quad(n=1,2,\ldots,N)\tag{2.5}$$

由初等分组得到的序列重构结果被称作初等重构序列（elementary reconstructed series）。

Remark 2.1 基础SSA算法可以自然扩展到复数时间序列：唯一的区别在于矩阵的转置要替换为复数的共轭转置。

2.1.2 对SSA基础算法中四个步骤的分析 目录