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$(\sqrt{\lambda_i}, U_i, V_i)$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$\mathbf{X}_I = \mathbf{X}_{i_1} + ... + \mathbf{X}_{i_p}$。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$\tilde{\mathbb{X}}^{(k)} = \left(\tilde{x}_1^{(k)}, \ldots, \tilde{x}_N^{(k)}\right)$。最终，初始序列被分解为$m$成分：

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

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

SSA基础算法中的步骤需要一些说明。 在这个部分我们简要讨论所涉及的程序的含义。

2.1.2.1 嵌入 目录

嵌入是一个把一维时间序列$\mathbb{X} = (x_1, \ldots, x_N)$使用向量$X_i = (x_i, \ldots, x_{i+L-1})$转换为多维序列$X_1, \ldots, X_K$的过程，其中$K = N - L + 1$。嵌入的参数是窗口长度$L$。注意轨迹矩阵(2.1)具有明显的对称性。$\mathbb{X}^T$相当于是窗口长度为K的多维序列。

对于动力系统专家来说，一种常用的技术是获得所有成对的滞后向量$\mathbf{X}_i$和$\mathbf{X}_j$之间的经验，然后计算时间序列相关的维度。该维数与产生时间序列的动力系统的吸引子的分形维数有关。需要注意的是，在这种方法中，L必须相当小而K相当大。同样，在具有汉克尔矩阵结构的结构总最小二乘（Structural Total Least Squares）中，通常的做法是选择$L = r + 1$，其中r是近似矩阵的猜测秩。

在SSA中，窗口长度L应该足够大。特别是，L必须足够大，以便每个L滞后向量包含初始序列$\mathbb{X} = (x_1, \ldots, x_N)$的一个基础部分。$L$必须足够大时，才有可能将每个$L$滞后向量$\mathbf{X}_i$视为一个单独的序列。

2.1.2.2 SVD分解 目录

SVD分解可以使用不同的术语被描述，并且可以被用于不同的目的。让我们从SVD一般的特点开始讲解，这些特点对于理解SSA十分重要。

正如我们已经提到过的，SVD分解可以对任意的维度为$L*K$的矩阵$\mathbf{X}=[X_1:\dots.X_K]$进行分解，分解形式如下：

$$\mathbf{X}=\sum_{i=1}^{d}\sqrt{\lambda_i}U_iV_i^{\mathrm{T}} \tag{2.6}$$

其中$\lambda_i(i=1,...,L)$是矩阵$\mathbf{S}=\mathbf{X}\mathbf{X}^{\mathrm{T}}$的特征值，并且按照递减的顺序进行排列。$d=max\{i,suchthat\lambda_i>0\}=rank(\mathbf{X})$，$\{U_1,\dots,U_d\}$是$\mathbf{S}$的政教特征向量，并且$V_i=\mathbf{X}^{\mathrm{T}}U_i/\sqrt{\lambda_i}$。

标准的SVD术语称$\sqrt{\lambda_i}$为奇异值；U_i和V_i为矩阵$\mathbf{X}$的左奇异向量和右奇异向量。如果我们定义$\mathbf{X}_i=\sqrt{\lambda_i}U_iV_i^{\mathrm{T}}$，那么公式(2.6)可以被写作公式(2.2)的形式。

如果所有特征向量的重数为1，那么展开式(2.2)是唯一确定的。否则，如果至少有一个特征值的重数大于1，那么在选择相应的特征向量时就有了自由度。我们应该假设以某种方式选择特征向量并且选择是固定的。

公式(2.6)说明SVD分解具有以下的对称性：$V_1,\dots,V_d$构成了矩阵$\mathbf{X}^{\mathrm{T}}\mathbf{X}$的特征正交系统。注意到轨迹矩阵的行和列都是原时间序列的子序列，因此左右奇异向量都具有时间结构，所以这些奇异向量也可以看做时间序列。

SVD分解具有很多特征，其中一个如下：在所有秩$r<d$的矩阵$\mathbf{X}^{(r)}$中，矩阵$\sum_{i=1}^{r}\mathbf{X}_i$给出了对轨迹矩阵$\mathbf{X}$的最优估计，即$\left\|\mathbf{X}-\mathbf{X}^{(r)}\right\|_{\mathrm{F}}$最小。其中$F$表示矩阵的Frobenius范数，一个矩阵$\mathbf{Y}$的Frobenius范数为$\|\mathbf{Y}\|_F=\sqrt{\langle\mathbf{Y},\mathbf{Y}\rangle_F}$。其中两个矩阵$\mathbf{Y}=\{y_{ij}\}_{i,j=1}^{q,s}$和$\mathbf{Z}=\{z_{ij}\}_{i,j=1}^{q,s}$的内积定义为：

$$\langle\mathbf{Y},\mathbf{Z}\rangle_{\mathrm{F}}=\sum_{i,j=1}^{q,s}y_{ij}z_{ij}$$

对于向量而言，Frobenius范数等同于传统的欧几里得范数。

注意到$\|\mathbf{X}\|_{\mathrm{F}}^2=\sum_{i=1}^{d}\lambda_i$并且$\lambda_i=\|\mathbf{X}_i\|_{\mathrm{F}}^2$，因此我们考虑将$\lambda_i/\|\mathbf{X}\|_{\mathrm{F}}^2$作为度量矩阵$\mathbf{X}_i$在整个轨迹矩阵$\mathbf{X}$中所占的比重。最终可以使用$\sum_{i=1}^{r}\lambda_i/\|\mathbf{X}\|_{\mathrm{F}}^2$作为轨迹矩阵的最优近似的特征指标。而且，如果$\lambda_r\neq\lambda_{r+1}$，可以把$\sum_{i=r+1}^{d}\lambda_i$作为轨迹矩阵与其最优估计之间的误差距离。

现在我们考虑把轨迹矩阵$\mathbf{X}$看做是一系列L延时的向量。用$\mathcal{X}^{(L)}\subset\mathrm{R}^L$表示向量$X_1,\dots,X_K$组成的线性空间。我们把这个空间称作时间序列$\mathbb{X}$的L-轨迹空间。为了强调序列$\mathbb{X}$的作用，我们把这个空间记作$\mathcal{X}^{(L)}(\mathbb{X})$。公式(2.6)表明$\mathscr{U}=(U_1,\dots,U_d)$是这个轨迹空间的一组正交基向量。

令$Z_i = \sqrt{\lambda_i} V_i, i = 1, \ldots, d$，公式(2.6)可以被写作$\mathbf{X} = \sum_{i=1}^{d} U_i Z_i^{\mathrm{T}}$，并且对于延时向量$X_j$我们有$X_j = \sum_{i=1}^{d} z_{ji} U_i$，其中$z_{ji}$是向量$Z_i$中的元素。这意味着向量$Z_i$是由向量$X_j$在基向量$\mathscr{U}$下的第$i$个成分组成的。

让我们考虑转置的轨迹矩阵$\mathbf{X}^T$。引入$Y_i = \sqrt{\lambda_i} U_i$，我们可以得到展开式$\mathbf{X}^{\mathrm{T}} = \sum_{i=1}^{d} V_i Y_i^{\mathrm{T}}$，该公式对应着K-延时向量在正交基$V_1, \ldots, V_d$下的表示。至此，SVD分解引出了关于轨迹矩阵$\mathbf{X}$的两种几何描述。

上述关于SVD特性的描述可以用L延时向量的多变量几何语言进行描述，如下：令$r < d$，那么在整个r维子空间$\mathcal{L}_r$中，由$U_1, \ldots, U_r$张成的子空间可以最好得估计原空间。也就是说，最小化$\sum_{i=1}^{K} \operatorname{dist}^2(X_i, \mathcal{L}_r)$得到结果是$\mathcal{L}_r = \operatorname{span}(U_1, \ldots, U_r)$，而$\sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{d} \lambda_i$是这些延时向量的$r$维最佳估计的特征。

SVD的另一个特征与特征向量$U_1, \ldots, U_d$有关。具体地，第一个特征向量U_1决定了延时向量的投影变化最大的方向。后续的每一个特征向量都与之前的特征向量正交，并且延时向量在该方向上的投影变化也是最大的。因此，很自然地将第i个特征向量所表示的方向称作第i个主方向。注意到，矩阵$\mathbf{X}_i = U_i Z_i^{\mathrm{T}}$是延时向量在第$i$个主方向投影后构成的。

这种对于SVD分解的理解角度引出了以下的术语。我们称向量U_i为第i个主特征向量，向量V_i和$Z_i = \sqrt{\lambda_i} V_i$被称作第$i$个音字向量和第$i$个主成分。

Remark2.2：SSA基础算法中所使用的SVD分解类似于经典的多变量分析中的主成分分析（PCA）和平稳时间序列分析中的KL变换。然而，SSA中的SVD方法利用了轨迹矩阵具有汉克尔结构的特点。事实上，轨迹矩阵的行和列都是原始序列的子序列并具有相同的时域意义，而PCA和KL变换并非如此。

Remark 2.3：一般地，轨迹空间中的任何一组正交基$P_1, \ldots, P_d$都可以替代通过SVD分解得到的正交基$U_1, \ldots, U_d$。在这种情况下，公式(2.2)可以被替换为$\mathbf{X}_i = P_i Q_i^{\mathrm{T}}$，其中$Q_i = \mathbf{X}^{\mathrm{T}} P_i$。一种可以替换正交基的例子是，在拓普利兹SSA中的自相关矩阵中的正交基，见章节2.5.3。另一个例子可以在独立成分分析（ICA）和因子分析中看到，见章节2.5.4。

关于SVD进一步的讨论以及使用见章节2.5.7。

2.1.2.3 分组 目录

分组这一步骤是对公式(2.2)中的$\mathbf{X}_i$j进行分组分配。假设$m = 2$，$I_1 = I = \{i_1, ..., i_r\}$并且$I_2 = \{1, \ldots, d\} \backslash I$，其中$1 \le i_1 < ... < i_r \le d$。

分组的目标是对时间序列的附加成分进行划分。假设时间序列$\mathbb{X}$是两个时间序列$\mathbb{X}^{(1)}$和$\mathbb{X}^{(2)}$的和；也就是说$x_i = x_i^{(1)} + x_i^{(2)}, i = 1, ..., N$。固定窗口长度$L$并且用$\mathbf{X}, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}$分别表示时间序列$\mathbb{X}, \mathbb{X}^{(1)}, \mathbb{X}^{(2)}$的$L$的轨迹矩阵。考虑轨迹矩阵$\mathbb{X}$的SVD分解。如果存在一组下标$I \subset \{1, \ldots, d\}$，使得$\mathbf{X}^{(1)} = \sum_{i \in I} \mathbf{X}_i$并且$\mathbf{X}^{(2)} = \sum_{i \notin I} \mathbf{X}_i$，那么我们称$\mathbb{X}^{(1)}$和$\mathbb{X}^{(2)}$是（弱）可分的。在这个例子中，$\mathbb{X}^{(1)}$在整个时间序列中的贡献值为对应的特征值的比重$\sum_{i \in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$。

(这里由两个小段没有翻译，感觉不是很必要)

分组的目的是找到一组分组方案$I_1,...,I_m$，它们对应的轨迹矩阵$\mathbf{X}_{I_1},...,\mathbf{X}_{I_m}$具有汉克尔矩阵的结构。（主要注意公式(2.2)中对轨迹矩阵分解以后，分解后的成分可能不具有汉克尔矩阵的结构，而分组这一步骤正是为了让各个成分具有汉克尔矩阵的结构）

现在让我们从多变量几何的角度看分组这一步骤。令$\mathbf{X}=[X_1:...:X_K]$表示时间序列$\mathbb{X}$的轨迹矩阵，序列$\mathbb{X}^{(1)}$和$\mathbb{X}^{(2)}$是一组分解方式并且各自对应下标集合I和$\{1,...,d\}\backslash I$。这意味着由基向量$U_1,...,U_d$张成的轨迹空间$\mathcal{X}^{(L)}$被分为了两个部分，即$\mathcal{X}^{(L,1)}=\mathrm{span}\,(U_i,i\in I)$，$\mathcal{X}^{(L,2)}=\mathrm{span}\,(U_i,i\notin I)$。两个序列的可分离性意味着矩阵$\mathbf{X}_I$的列是延时向量$X_1,...,X_K$在特征子空间$\mathcal{X}^{(L,2)}$上的投影，恰好是序列$\mathbb{X}^{(1)}$的轨迹矩阵。