

三维时间序列聚类的模拟实验

该部分主要通过模拟生成 3 类三维时间序列来验证模型聚类效果。在此我们提出一种新型的基于特征相关性的距离度量来刻画多维时间序列的距离。分析过程只要包括如下几部分：

- 随机三维序列的生成与数据预处理（包括断点查找与标准化）
- 相关性距离的度量方法
- 结果展示与可视化

一、生成随机序列

首先生成各类时间序列的特征序列，每类特征序列的表达式均由此生成：

$$X_t = f(t) + \varepsilon_t, \quad t = 1, 3, 5, \dots, 399$$

其中 ε_t 服从正态分布。均值为 0，方差为序列 $\{f(t)\}$ 方差的 0.6 倍。定义 $f(t)$ 表达式如下表所示：

	类 A	类 B	类 C
特征 1	$f(t) = \sin\left(\frac{t}{10}\right) + \ln t$	$f(t) = t^{0.5}$	$f(t) = \cos\left(\frac{t}{20}\right)$
特征 2	$f(t) = \exp\left(\frac{t}{150}\right) + \cos\left(\sqrt{\frac{t}{30}}\right)$	$f(t) = \frac{10}{1 + 5 \exp\left(\frac{t}{40}\right)} + \cos\left(\frac{t}{75}\right)$	$f(t) = \exp\left(\sin\left(\frac{t}{10}\right)\right)$
特征 3	$f(t) = -\sin\left(\frac{t}{50}\right) - \ln\left(\frac{t}{20}\right)$	$f(t) = \sin\left(\frac{t}{50}\right)$	$f(t) = -t^{1.5} + t^{0.5}$

生成的序列图如下图所示。其中 Aa 表示第 A 类第 a 个三位序列，Aa1,Aa2,Aa3 分别表示该三维时间序列的三个指标序列。

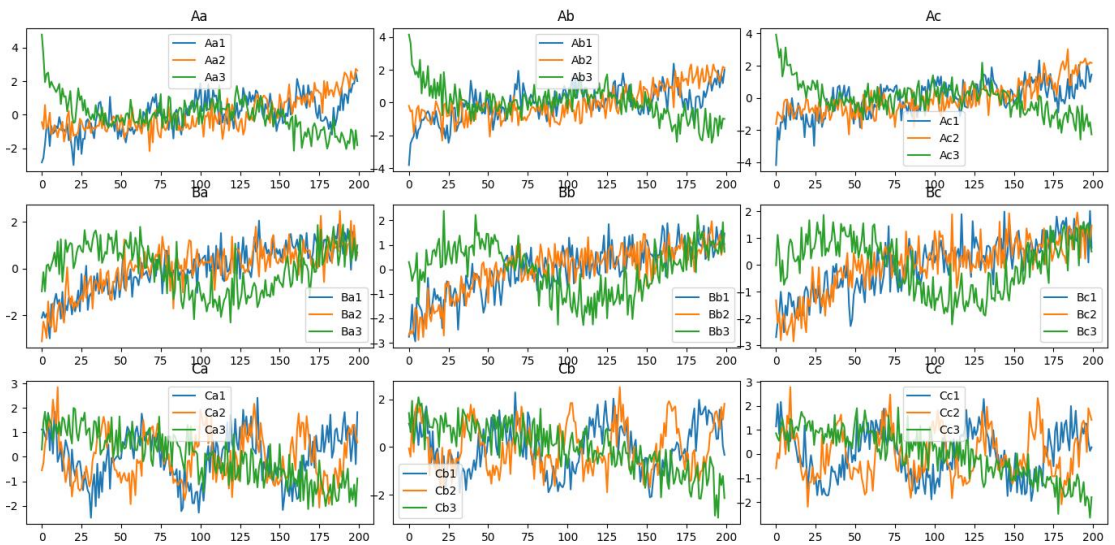


Figure 1

二、对原始序列进行断点查找

对原始三维时间序列的每一条特征序列都进行断点查找, 最终获得的断点序列如下图所示:

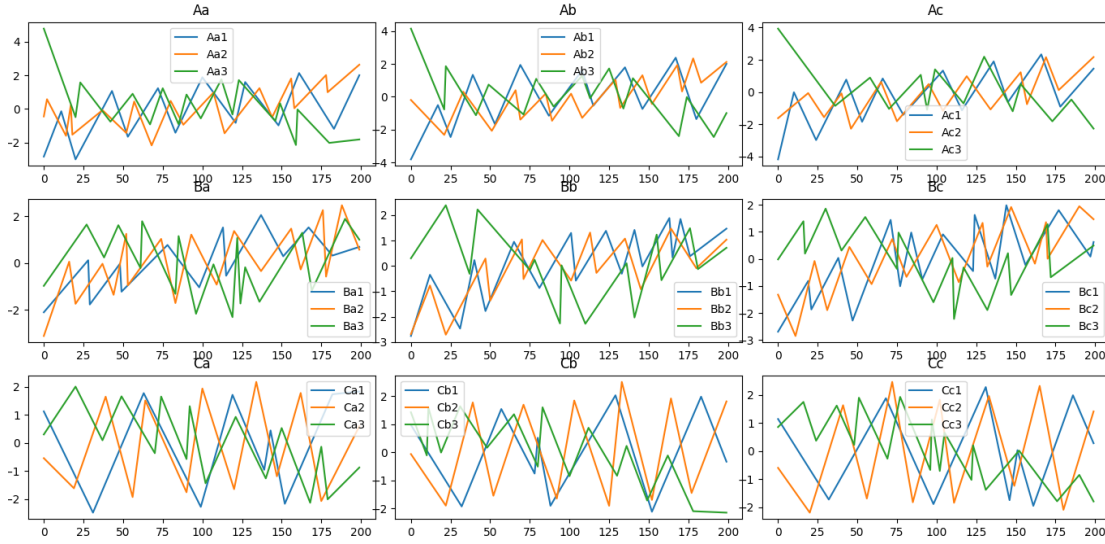


Figure 2

多维时间序列的距离度量主要采用基于相关系数的距离度量。假设现有两组三维时间序列:

$$X_t = (X_{1t} \quad X_{2t} \quad X_{3t})' = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ X_{31} & X_{31} & \cdots & X_{3n} \end{pmatrix}$$

$$Y_t = (Y_{1t} \quad Y_{2t} \quad Y_{3t})' = \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ Y_{31} & Y_{31} & \cdots & Y_{3n} \end{pmatrix}$$

因指标间的值差异较大, 故先对原始时间序列进行标准化。标准化的公式如下所示:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_i}{S_i}, \quad i = 1, 2, 3, \quad j = 1, 2, \dots, n$$

其中:

$$\bar{x}_i = \frac{\sum_{j=1}^n x_{ij}}{n}$$

$$S_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}$$

现考虑多维时间序列不同维度间的距离。

记

$$d_{ij} = x_{ij}^* - y_{ij}^*, \quad i = 1, 2, 3, \quad j = 1, 2, \dots, n$$

则第 i 个维度间的距离为:

$$p_i = \sqrt{\frac{\sum_{j=1}^n d_{ij}^2}{n}}, \quad i = 1, 2, 3$$

则两组时间序列间的距离 P_{xy} 为：

$$P_{xy} = \sum_{i=1}^3 p_i \cdot r_{ii}^{-\frac{1}{2}}$$

其中 r_{ii} 表示第 i 个维度下两时间序列的 Pearson 相关系数。

故基于断点序列计算 9 个三维序列之间的相关性距离，得到如下距离矩阵。其中底纹颜色越深代表距离越接近。

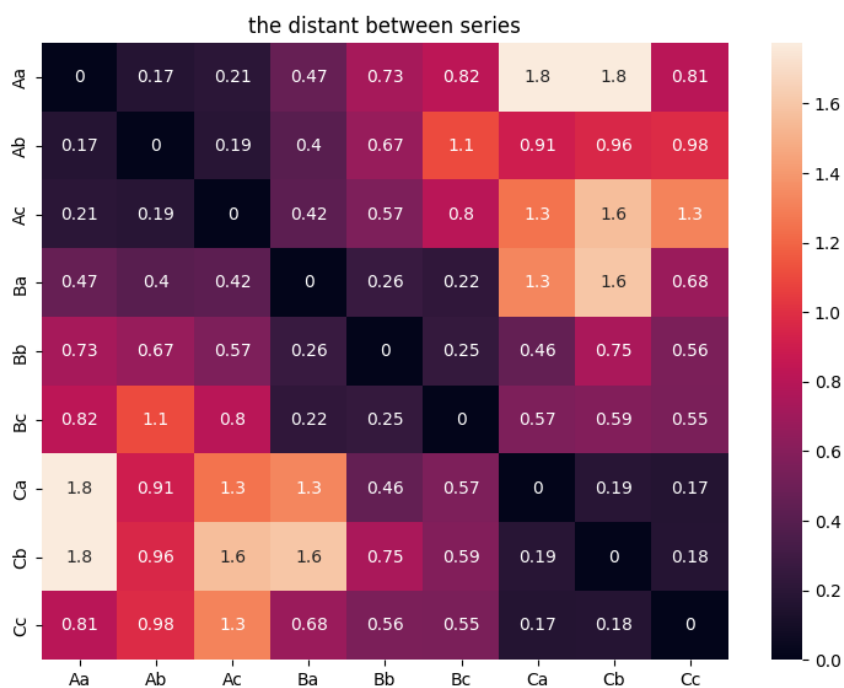


Figure 3

三、基于相关距离进行聚类

运用层次聚类法对上述距离矩阵进行聚类，得到的树状图如下所示：

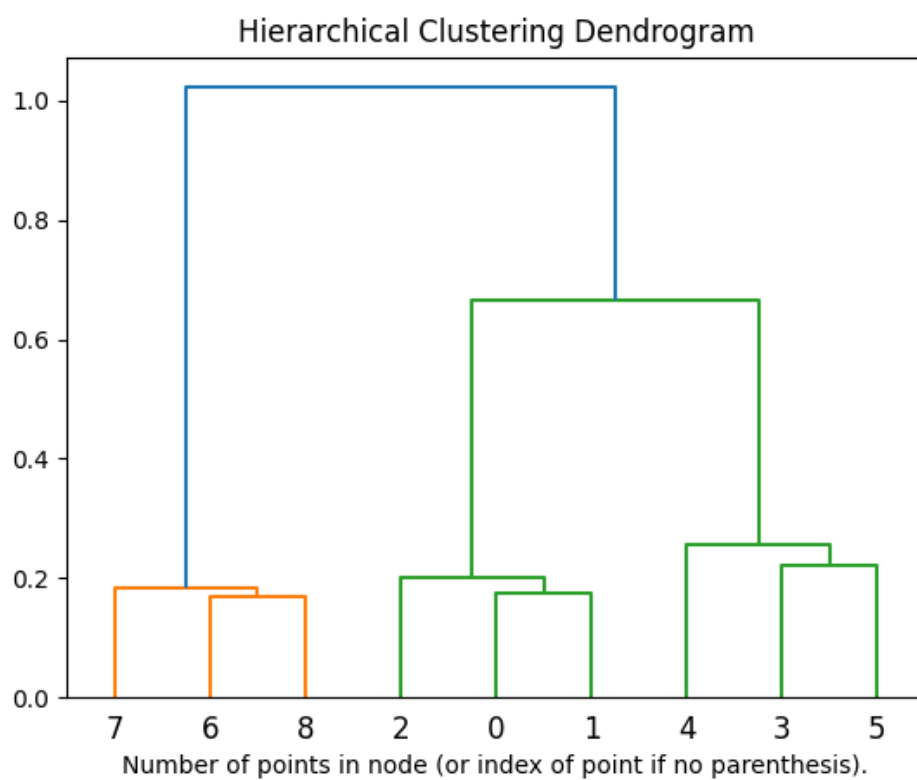


Figure 4

图中横轴序号依次对应下述序列：

0	1	2	3	4	5	6	7	8
Aa	Ab	Ac	Ba	Bb	Bc	Ca	Cb	Cc

从结果中能够看出，基于相关性进行聚类的思路可行。