第25卷 第2期
1999年3月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 25, No. 2
Mar., 1999

研究简报

# 成组技术与相似性系数[1)]

裘聿皇
（中国科学院自动化研究所 北京 100080）

**关键词** 成组技术，相似性系数，聚类分析.

# GROUP TECHNOLOGY AND THE SIMILARITY COEFFICIENT

QIU Yuhuang
(*Institute of Automation, Chinese Academy of Sciences, Beijing* 100080)

**Key words** Group technology, similarity coefficient, cluster analysis.

## 1 引言

把聚类分析用于成组技术，把机床划分成机床单元，分类以后，可以得到机床的合理的物理布局和逻辑布局.

成组技术的建模方法之一是所谓矩阵表示方法，建立机床-零件的关联矩阵 $A=[a_{ij}]$. 矩阵的行号代表机床号，列号表示零件号. 矩阵的元素只取0或1. $a_{ij}=1$，表示机床 $i$，加工零件 $j$. 否则 $a_{ij}=0$.

作者在六五攻关项目：京津地区生态-经济区划，也曾用(0,1)矩阵建模，提出新的聚类分析方法，进行计算机辅助区划[2]. 并用这种聚类分析方法，对我国草兔进行了聚类研究[3]. 现在作出补充修改并应用于成组技术研究.

## 2 数学描述

### 2.1 相似性系数的定义

设 $A$ 是如前已建立的 $m\times n$ 关联矩阵. $A$ 的转置矩阵记为 $A^{\mathrm{T}}$. 令

$$B^{(1)} = AA^{\mathrm{T}}. \tag{1}$$

以 $\boldsymbol{x}_i$ 记矩阵 $A$ 的第 $i$ 行行向量($i=1,2,\cdots,m$)，则 $B^{(1)}$的元素 $b_{ij}^{(1)}$($i,j=1,2,\cdots,m,i\neq j$)，表明了 $\boldsymbol{x}_i\cap\boldsymbol{x}_j$ 含“1”的元素的个数. 令

$$B^{(0)} = \overline{A}\,\overline{A}^{\mathrm{T}}, \tag{2}$$

其中 $\overline{A}$ 的每一个元素是 $A$ 的元素求反，则 $B^{(0)}$的元素 $b_{ij}^{(0)}$($i,j=1,2,\cdots,m,i\neq j$)，表明了 $\boldsymbol{x}_i\cap\boldsymbol{x}_j$ 含“0”的元素的个数. 令

1) 国家自然科学基金资助项目(编号 69681002, 69635030)，本文初稿曾在1998年中国控制会议上宣读.

收稿日期 1997-06-03 收修改稿日期 1998-04-20

$$D = B^{(1)} + B^{(0)}, \tag{3}$$

则 $D$ 的元素 $d_{ij}(i,j=1,2,\cdots,m,i\neq j)$，表明了 $\boldsymbol{x}_i \cap \boldsymbol{x}_j$ 共同“0”与共同“1”的个数之和. 显然 $d_{ii}=n,(i=1,2,\cdots,m)$.

我们定义相似性系数为

$$S_{ij} = d_{ij}/n, \tag{4}$$

它刻划了机床 $i$ 与机床 $j$ 加工零件工况的相似性. 计算式(4)时，由于矩阵的对称性，只要计算 $B^{(1)},B^{(0)},D$ 的主对角线右边的元素，即可得到 $S_{ij}$.

### 2.2 半格 $\varepsilon=\langle \boldsymbol{E},\oplus\rangle$ 及 $\boldsymbol{T}_1,\boldsymbol{T}_2$映射

令三元素集合 $E=\{0,1,\delta\}$($\delta$ 在计算机中用0,1以外的数表示)，“$\oplus$”为“加法”运算，规则如下：

$$a \oplus a = a, \quad a \in E, \tag{5}$$

$$a \oplus b = \delta, \quad a,b \in E, \quad a \neq b. \tag{6}$$

显然 $E$ 上的$\oplus$运算满足结合律及交换律，$\varepsilon=\langle E,\oplus\rangle$是一个半格(Semilattice)[4]，可以用来模拟三值逻辑. 这里1代表“是”，0代表“否”，$\delta$ 代表“不知道”，“不确定”等. $\delta$ 在不同的场合起着不同的作用.

对于其元素属于 $E$ 的向量 $\boldsymbol{x}_i,\boldsymbol{x}_j$(文中的向量都用行向量表示)，$\oplus$定义如下：

$$\boldsymbol{x}_i \oplus \boldsymbol{x}_j = (x_{i1} \oplus x_{j1}, x_{i2} \oplus x_{j2}, \cdots, x_{in} \oplus x_{jn}), \tag{7}$$

其中 $\boldsymbol{x}_{ik},\boldsymbol{x}_{jk}(k=1,2,\cdots,n)$为 $\boldsymbol{x}_i,\boldsymbol{x}_j$ 的第 $k$ 个分量.

若只进行一次聚类，以上准备就够了. 若是系统聚类从而形成聚类体系，则需要二元集合 $H=\{0,1\}\subset E$ 和两种不同的映射.

设 $T_1,T_2$是 $E$ 到 $H$ 上的两个映射. 其规则如下：

$$T_1 0 = 0, \quad T_1 1 = 1, \quad T_1 \delta = 0; \tag{8}$$

$$T_2 0 = 1, \quad T_2 1 = 0, \quad T_2 \delta = 0. \tag{9}$$

对于其元素 $a_{ij}$属于 $E$ 的矩阵 $A$，则 $T_k A$ 的每个元为

$$T_k a_{ij}, \tag{10}$$

其中 $i=1,\cdots,m,j=1,\cdots,n,k=1,2$.

### 2.3 聚类

本文的(4)式给出了相似性系数，由于分母是固定的 $n$，故实际聚类只要考虑 $d_{ij}$就行了. 本文的系统聚类分两部分.

第一部分是在 $D$ 阵上进行. 设立阈值 $\theta$，一般可取 $\theta=[n/2]$作为初值.

1) 令 $\eta=n$.

2) 在 $D$ 阵的主对角线右边寻找 $d_{ij}=\eta$. 若找不到，则转入5)；若找到这样的 $d_{ij}$，则进入下一步.

3) 机床 $i$ 与机床 $j$ 形成机床单元，$\boldsymbol{x}_i$ 与 $\boldsymbol{x}_j$ 合成一类 $\boldsymbol{H}_i$. 若同一行有 $d_{ij}=d_{ik}=\eta$，则 $\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_k$ 合成类 $\boldsymbol{H}_i$，其余类推. 以下都以两个集合合成一类为例.

4) $D$ 的 $i$ 行，$i$ 列，$j$ 行，$j$ 列全置零.(若手工进行，即把它们划去).

5) $\eta \leftarrow \eta-1$.

6) 若 $\eta=\theta$，则进入7)，否则转2).

7) 若 $D$ 还有非全零行 $i$，机床 $i$ 不与别的机床形成机床单元.

第二部分是把属于同一类的集合加以合并(或类与类合并)，产生类的代表：$\boldsymbol{x}_i \oplus \boldsymbol{x}_j$ 代替 $A$ 阵中的第 $i$ 行，第 $j$ 行. 若 $\boldsymbol{x}_i$ 自成一类，则不进行$\oplus$运算.

若类 $\boldsymbol{H}_p$ 与类 $\boldsymbol{H}_q$ 对应列(同一列)都是 $\delta$，且合并成类 $\boldsymbol{H}_p$ 与类 $\boldsymbol{H}_q$ 以前该列各自为1的元素个数相同，则将式(9)中的 $T_2\delta=0$修改为 $T_2\delta=1$.

随后，用 $T_1A$ 代替 $A$，用 $T_2A$ 代替 $\overline{A}$. 利用式(1)，(2)，(3)再次形成新的一轮 $B^{(1)}$，$B^{(0)}$，$D$. 若 $\boldsymbol{x}_i$，$\boldsymbol{x}_j$ 应归为一类，且 $j>i$，则把 $D$ 的 $j$ 行，$j$ 列置零，再在 $D$ 中进行新的一轮聚类，且 $\theta \leftarrow \theta-1$(把阈值减小，以便把更多类聚在一起)，进行第一部分工作.

把上述过程反复进行，直到 $\theta=1$，把所有机床聚为一类.

### 2.4 单元间的调动总数

当零件既要在机床单元 MC-$p$ 中加工，又要在 MC-$q$ 中加工，这时该零件便是瓶颈零件.

两个机床单元 MC-$p$，MC-$q$ 的单元间调动总数(ICM)可用下列方法求得.

设聚类结果 MC-$p$ 的类代表为 $\boldsymbol{H}_p$，MC-$q$ 的类代表为 $\boldsymbol{H}_q$.

a. 令 $\delta=1$，得 $\boldsymbol{H}'_p$，$\boldsymbol{H}'_q$；

b. 令 $\boldsymbol{\rho}=(\rho_1^{pq},\cdots,\rho_n^{pq})=\boldsymbol{H}'_p \oplus \boldsymbol{H}'_q|_{\delta=0}$；

c. 单元间的调动数为 $\mathrm{ICM}_{p,q}=\sum\limits_{k=1}^{n}\rho_k^{pq}$. (11)

## 3 例子分析

**例**[1].

$$A=\begin{array}{c}\text{零 件 号}\\ \begin{array}{ccccc}1&2&3&4&5\end{array}\\ \begin{bmatrix}1&1&0&0&1\\1&1&0&0&0\\0&0&1&1&1\\0&0&1&1&0\end{bmatrix}\end{array}\begin{array}{c}\\ \\1\\2\\3\\4\end{array}\ \text{机床号} \quad . \tag{12}$$

本文用这个例子来比较(4)式与文献[1]相似性系数，知两种相似性系数都有 $s_{14}=s_{23}=0/5=0$.

根据定义，由(4)式得 $s_{12}=s_{34}=4/5=0.8$. 在文献[1]中 $s_{12}=s_{34}=2/3=0.75$(译者已指出应为0.67). $A$ 中的行向量 $X_1$，$X_2$前4个元素相同，都为1，1，0，0. 在5个元素中有4个相同，本文取0.8更合理.

由(4)式得 $s_{13}=s_{24}=1/5=0.2$. 在文献[1]中 $s_{13}=1/4=0.25$，$s_{24}=0/5=0$. 实际上，$X_1$，$X_3$第5个元素都是1，有1/5元素相同. $X_2$，$X_4$第5个元素都是0，也有1/5元素相同. 由聚类知机床1与机床2为一个机床单元，机床3与机床4为另一个机床单元. 在文献[1]中 $s_{13}=0.25$，$s_{24}=0$，两数相差太大，在逻辑上难于接受，这正是 McAuley 在文献[5]中所承认的缺点. 而本文是 $s_{13}=s_{24}=0.2$，正好相等，合理性很明显.

造成不同的计算结果的原因之一是：我们让两台机床都不加工某一零件，也作为一种相同的工况. 因此本文的相似性系数定义考虑得更全面、直观而合理.

文献[1]中的相似性系数另一个不合理的情况是：若两台机器都加工所有零件，则使

相似性系数的分母为0. 文献[1]的表达方式与文献[5]不同，根据 McAuley 的思想[5]，可得 $s_{13}=1/5=0.2$，其余相似性系数同文献[1]，[5]求得的结果相同.

$D$ 的主对角线右边最大元为 $d_{12}=d_{34}=4$，故机床1与机床2聚为一类成 MC-1，机床3与机床4聚为一类成 MC－2.

然后再对 MC-1与 MC-2聚类，由新的一轮 $D$ 中的 $d_{13}=1$，得集合 MC-1类和 MC-2类的相似性系数为1/5＝0. 2. 而在文献[1]中为0. 25. 这是由于把机床1与机床3之间工况的相似性系数替代了机床单元类之间的工况的相似性系数. 显然是不太合理的. 集合 MC-1，MC-2第五个元素都是$\begin{pmatrix}1\\0\end{pmatrix}$，因此本文的相似性系数较合理.

在聚类体系的基础上根据准则值(the value of the criteria)[5]，挑选最好的机床聚类(分组)，使机器得到合理安排，具体情况可见文献[5]或其他文献.

下面求单元间调动总数，由聚类可得机床单元 MC-1＝{1，2}，MC-2＝{3，4}，及两个零件族 $PF$-1＝{1，2}，$PF$-2＝{3，4}1). 形成局部可分解，而零件5便成为"瓶颈"零件.

我们可得 MC-1，MC-2的代表为 $\boldsymbol{H}_1=(1,1,0,0,\delta)$，$\boldsymbol{H}_2=(0,0,1,1,\delta)$.

a) 令 $\delta=1$得 $\boldsymbol{H}'_1=(1,1,0,0,1)$， $\boldsymbol{H}'_2=(0,0,1,1,1)$.

b) $\boldsymbol{\rho}^{1,2}=\boldsymbol{H}'_1\oplus\boldsymbol{H}'_2|_{\delta=0}=(0,0,0,0,1)$.

c) 得单元间调动数为 ICM＝0＋0＋0＋0＋1＝1.

与 Seifoddini，Wolfe 方法所得结果相同.

# 4 结论

本文提出了新的相似性系数及利用矩阵计算所有相似性系数的方法. 例子说明本文的相似性系数更合理. 新的聚类方法比 McAuley 所运用的单交连聚类分析方法也更合理更方便，MoAuley 是用单交连准则逐个聚类，而本文可以几个同时聚类在一起(参见文献[3]的例子)，更符合机床安排的实际需要.

## 参 考 文 献

1 Andrew Kusiak. Intelligent Manufacturing Systems. Prentice Hall, Inc. 1990(杨静宇，陆际联译，1993，清华大学出版社)

2 裘聿皇. 计算机辅助区划——一种新的聚类方法. 计算技术与自动化，1990，**9**(1)：1—6

3 裘聿皇. 我国草兔的聚类研究. 兽类学报，1989，**9**(3)：168—172

4 Garrett Birkhoff. Lattice Theory. New York：1948

5 John McAuley. Machine grouping for efficient prodution. *The production Engineer*, 1972，**51**(1)：53—57

**裘聿皇** 1942年生，1966年毕业于华东师范大学数学系. 现为中国科学院自动化研究所研究员、博士生导师. 感兴趣的领域有：系统理论与应用、$H^\infty$控制理论、遗传算法等各种算法及计算机软件等.

1) 按文献[1]的(8. 3)式，零件号排列为(1，3，2，4，5)，因此对文献[1]来说，应是 $PF$-1＝{1，3}，$PF$-2＝{2，4}才对. 故文献[1]有误. 本文零件号排列同列号.