第 6 章　模糊关系与模糊聚类分析

现实世界中，事物之间存在着这样或那样的关系。一些是界限非常明确的关系，如“父子关系”、“大小关系”等等，可以简单地用“是”与“否”或者“1”与“0”来刻画。而更多的则是界限不明显的关系，如“朋友关系”、“相近关系”、“相像关系”等。对于这类关系再用简单的“是”与“否”或者“1”与“0”来刻画显然是不合适的，必须用模糊集理论来进行描述，这就是模糊关系。

模糊关系是模糊集理论最重要的内容之一，其应用范围十分广泛，几乎遍及模糊数学的所有应用领域。本章主要介绍模糊关系的基本理论及其在聚类分析中的应用，内容主要包括：模糊矩阵，模糊关系，模糊聚类分析。

6.1　模糊矩阵

如同数值矩阵在经典数学中的作用一样，模糊矩阵是建立模糊数学方法的重要工具之一。同时，当论域有限时，模糊关系可以用模糊矩阵来表示。

6.1.1　模糊矩阵的定义及其运算

定义 1　设 $R = (r_{ij})_{m\times n}$ 是一个 $m\times n$ 的矩阵，如果

$$r_{ij}\in[0, 1]，1\le i\le m, 1\le j\le n \tag{6.1}$$

则称 R 是**模糊矩阵**，通常用 $M_{m\times n}$ 来表示所有 $m\times n$ 的模糊矩阵构成的集合。

定义 2　设 $R, S\in M_{m\times n}$ 且 $R = (r_{ij})_{m\times n}$，$S = (s_{ij})_{m\times n}$，则

(1) R 与 S **相等**定义为：$R = S \Leftrightarrow r_{ij} = s_{ij}, \forall i, j$；

(2) R **包含于** S 定义为：$R\subseteq S \Leftrightarrow r_{ij} \le s_{ij}, \forall i, j$；

(3) R 与 S 的**交**定义为：$R\cap S = (r_{ij} \wedge s_{ij})_{m\times n}$；

(4) R 与 S 的**并**定义为：$R\cup S = (r_{ij} \vee s_{ij})_{m\times n}$；

(5) R 的**余**定义为：$R^C = (1-r_{ij})_{m\times n}$。

例 1　设 $R, S, T\in M_{2\times 2}$ 且

$$R=\begin{pmatrix}0.4 & 0.7\\ 0.5 & 0.1\end{pmatrix}，S=\begin{pmatrix}0.9 & 0.4\\ 0.2 & 0.7\end{pmatrix}，T=\begin{pmatrix}0.9 & 0.6\\ 0.4 & 0.8\end{pmatrix}$$

则由定义 2 有：$S\subseteq T$，并且

$$R\cap S=\begin{pmatrix}0.4\wedge 0.9 & 0.7\wedge 0.4\\ 0.5\wedge 0.2 & 0.1\wedge 0.7\end{pmatrix}=\begin{pmatrix}0.4 & 0.4\\ 0.2 & 0.1\end{pmatrix}，R\cup S=\begin{pmatrix}0.4\vee 0.9 & 0.7\vee 0.4\\ 0.5\vee 0.2 & 0.1\vee 0.7\end{pmatrix}=\begin{pmatrix}0.9 & 0.7\\ 0.5 & 0.7\end{pmatrix}$$

$$R^C=\begin{pmatrix}1-0.4 & 1-0.7\\ 1-0.5 & 1-0.1\end{pmatrix}=\begin{pmatrix}0.6 & 0.3\\ 0.5 & 0.9\end{pmatrix}$$

模糊矩阵的交、并、余运算满足如下算律。

定理 1　设 $R, S, T, W\in M_{m\times n}$，则有

(1) 幂等律：$R\cap R=R$，$R\cup R=R$；

(2) 交换律：$R\cap S=S\cap R$，$R\cup S=S\cup R$；

(3) 结合律：$(R\cap S)\cap T=R\cap(S\cap T)$，$(R\cup S)\cup T=R\cup(S\cup T)$；

(4) 吸收律：$(R\cap S)\cup R=R$，$(R\cup S)\cap R=R$；

(5) 分配律：$R\cap(S\cup T)=(R\cap S)\cup(R\cap T)$，$R\cup(S\cap T)=(R\cup S)\cap(R\cup T)$；

(6) 复原律(对合律)：$(R^C)^C=R$；

(7) 对偶律(De Morgan 律)：$(R\cap S)^C=R^C\cup S^C$，$(R\cup S)^C=R^C\cap S^C$；

(8) 零壹律：$R\cap O=O$，$R\cap E=R$，$R\cup O=R$，$R\cup E=E$，其中

$$O=\begin{pmatrix}0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \cdots & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix},\quad E=\begin{pmatrix}1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & \cdots & \vdots\\ 1 & 1 & \cdots & 1\end{pmatrix}$$

(9) 单调性：$R\subseteq S, T\subseteq W\Rightarrow R\cap T\subseteq S\cap W$，$R\cup T\subseteq S\cup W$；

(10) 逆序性：$R\subseteq S\Leftrightarrow S^C\subseteq R^C$；

(11) $R\subseteq S\Leftrightarrow R\cap S=R$，$R\cup S=S$。

必须指出，一般 $R\cap R^C\neq O$，$R\cup R^C\neq E$，即模糊矩阵的交、并、余运算不满足互补律。

模糊矩阵的交、并运算还可以推广到任意多个模糊矩阵上去。

定义 3　设 T 是指标集，$R^{(t)}=(r_{ij}{}^{(t)})_{m\times n}\in M_{m\times n}(t\in T)$，则分别称

$$\bigcap_{t\in T}R^{(t)}=\left(\bigwedge_{t\in T}r_{ij}{}^{(t)}\right)_{m\times n}\ \text{和}\ \bigcup_{t\in T}R^{(t)}=\left(\bigvee_{t\in T}r_{ij}{}^{(t)}\right)_{m\times n}$$

为模糊矩阵 $\{R^{(t)}\mid t\in T\}$ 的**交**和**并**。

定理 2　设 T 是指标集，$S, R^{(t)}\in M_{m\times n}(t\in T)$，则有

(1) $S\bigcap\left(\bigcup_{t\in T}R^{(t)}\right)=\bigcup_{t\in T}(S\bigcap R^{(t)})$，$S\bigcup\left(\bigcap_{t\in T}R^{(t)}\right)=\bigcap_{t\in T}(S\bigcup R^{(t)})$；

(2) $\left(\bigcap_{t\in T}R^{(t)}\right)^C=\bigcup_{t\in T}(R^{(t)})^C$，$\left(\bigcup_{t\in T}R^{(t)}\right)^C=\bigcap_{t\in T}(R^{(t)})^C$。

6.1.2　模糊矩阵的截矩阵

定义 4　设 $R\in M_{m\times n}$，且 $R=(r_{ij})_{m\times n}$。如果 $R_\lambda=(r_{ij}(\lambda))_{m\times n}$ 满足

$$r_{ij}(\lambda)=\begin{cases}1, & r_{ij}\ge\lambda\\ 0, & r_{ij}<\lambda\end{cases}$$

则称 R_λ 为模糊矩阵 R 的 **λ 截矩阵**。

显然，模糊矩阵的 λ 截矩阵是布尔矩阵。并且容易证明：对 $\forall R, S\in M_{m\times n}$，

(1) $(R\cap S)_\lambda = R_\lambda\cap S_\lambda$，$(R\cup S)_\lambda = R_\lambda\cup S_\lambda$；

(2) $\lambda\le\alpha\Rightarrow R_\alpha\subseteq R_\lambda$；

(3) $R\subseteq S\Leftrightarrow R_\lambda\subseteq S_\lambda$，$\forall\lambda\in[0, 1]$。

定理 3 （**模糊矩阵的分解定理**）设 $R\in M_{m\times n}$，且 $R = (r_{ij})_{m\times n}$。则

$$R=\bigcup_{\lambda\in[0,1]}(\lambda R_\lambda) \tag{6.2}$$

其中 $\lambda R_\lambda = (r_{\lambda ij})_{m\times n}$，且

$$r_{\lambda ij}=\lambda\wedge r_{ij}(\lambda)=\begin{cases}\lambda, & r_{ij}\ge\lambda\\ 0, & r_{ij}<\lambda\end{cases}$$

6.1.3　模糊矩阵的合成与转置

定义 5　设 $R\in M_{m\times t}$，$S\in M_{s\times n}$ 且 $R = (r_{ij})_{m\times t}$，$S = (s_{ij})_{t\times n}$。令 $T\in M_{m\times n}$ 且 $T = (t_{ij})_{m\times n}$，其中

$$t_{ij}=\bigvee_{k=1}^{t}(r_{ik}\wedge s_{kj}) \tag{6.3}$$

则称 T 为 R 与 S 的**合成**(**模糊乘积**)，记为 $T = R\circ S$。

例 2　设 $R\in M_{2\times 3}$，$S\in M_{3\times 2}$，其中

$$R=\begin{pmatrix}0.2 & 0.5 & 1\\ 0.7 & 0.1 & 0.8\end{pmatrix},\quad S=\begin{pmatrix}0.6 & 0.5\\ 0.4 & 1\\ 0.1 & 0.9\end{pmatrix}$$

则

$$R\circ S=\begin{pmatrix}(0.2\wedge0.6)\vee(0.5\wedge0.4)\vee(1\wedge0.1) & (0.2\wedge0.5)\vee(0.5\wedge1)\vee(1\wedge0.9)\\ (0.7\wedge0.6)\vee(0.1\wedge0.4)\vee(0.8\wedge0.1) & (0.7\wedge0.5)\vee(0.1\wedge1)\vee(0.8\wedge0.9)\end{pmatrix}=\begin{pmatrix}0.4 & 0.9\\ 0.6 & 0.8\end{pmatrix}$$

模糊矩阵的合成运算满足如下算律。

定理 4　设 R, S, T 为满足相应运算的模糊矩阵，则有

(1) 结合律：$(R\circ S)\circ T = R\circ(S\circ T)$；特别地，$R^m\circ R^n = R^{m+n}$，$(R^m)^n = R^{mn}$；

(2) 分配律：$(R\cup S)\circ T = (R\circ T)\cup(S\circ T)$，$R\circ(S\cup T) = (R\circ S)\cup(R\circ T)$；

(3) 弱分配律：$(R\cap S)\circ T\subseteq (R\circ T)\cap(S\circ T)$，$R\circ(S\cap T)\subseteq(R\circ S)\cap(R\circ T)$；

(4) 单调性：$R\subseteq S\Rightarrow R\circ T\subseteq S\circ T$，$T\circ R\subseteq T\circ S$。

(5) $(R\circ S)_\lambda = R_\lambda\circ S_\lambda$。

一般来讲，模糊矩阵的合成不满足交换律。例如，考虑例 1 中的模糊矩阵 R 和 S，显然有

$$R\circ S=\begin{pmatrix}0.7 & 0.7\\ 0.5 & 0.4\end{pmatrix},\quad S\circ R=\begin{pmatrix}0.4 & 0.7\\ 0.5 & 0.2\end{pmatrix}$$

即 $R\circ S\neq S\circ R$。

定义 6　设 $R\in M_{m\times n}$ 且 $R=(r_{ij})_{m\times n}$，则称 $R^{\mathrm{T}}=(r_{ji})_{n\times m}$ 为 R 的**转置**。

模糊矩阵的转置矩阵与普通数值矩阵的转置矩阵概念是相同的，即把相应的行变为列，列变为行，并且满足如下的性质。

定理 5　设 R 和 S 为满足相应运算的模糊矩阵，则有

(1) $(R^{\mathrm{T}})^{\mathrm{T}}=R$；

(2) $(R\cap S)^{\mathrm{T}}=R^{\mathrm{T}}\cap S^{\mathrm{T}}$，$(R\cup S)^{\mathrm{T}}=R^{\mathrm{T}}\cup S^{\mathrm{T}}$；

(3) $R\subseteq S\Leftrightarrow R^{\mathrm{T}}\subseteq S^{\mathrm{T}}$；

(4) $(R^{\mathrm{T}})_\lambda=(R_\lambda)^{\mathrm{T}}$；

(5) $(R\circ S)^{\mathrm{T}}=R^{\mathrm{T}}\circ S^{\mathrm{T}}$，$(R^n)^{\mathrm{T}}=(R^{\mathrm{T}})^n$；

(6) $(R^C)^{\mathrm{T}}=(R^{\mathrm{T}})^C$。

6.2　模糊关系

模糊关系是经典关系的推广。经典关系描述元素之间是否适合某种关系，而模糊关系则是描述元素之间适合某种关系的程度大小。在模糊集合理论中，模糊关系占有相当重要的地位。

6.2.1　模糊关系的定义及其运算

由第一章知道，经典关系是论域直积的经典子集。事实上，模糊关系是论域直积的模糊子集，因而模糊关系可由模糊集合及其隶属函数来刻画。

定义 7　给定非空集合 X 和 Y。如果 $R\in\mathcal{F}(X\times Y)$，则称 R 为从 X 到 Y 的一个**模糊(二元)关系**，而 R 的隶属函数 $R(x, y)$ 表示 X 中的元素 x 与 Y 中的元素 y 适合这种关系的程度。特别地，当 $X=Y$ 时，从 X 到 Y 的模糊(二元)关系 R 称为 X 上的模糊(二元)关系。

给定非空集合 $X_1, X_2, ..., X_n$。如果 $R\in\mathcal{F}(X_1\times X_2\times...\times X_n)$，则称 R 为 $X_1, X_2, ..., X_n$ 的一个**模糊 n 元关系**。特别地，X^n 的模糊关系 R 称为 X 上的模糊 n 元关系。

在此我们仅讨论模糊二元关系，也常常简称为模糊关系。

由定义 7 可见，模糊关系本质上就是模糊集合，完全由其隶属函数来刻画。因此，当 $R(x, y)$ 仅取 1 或者 0 时，R 退化为经典二元关系，即经典关系是模糊关系的特例。

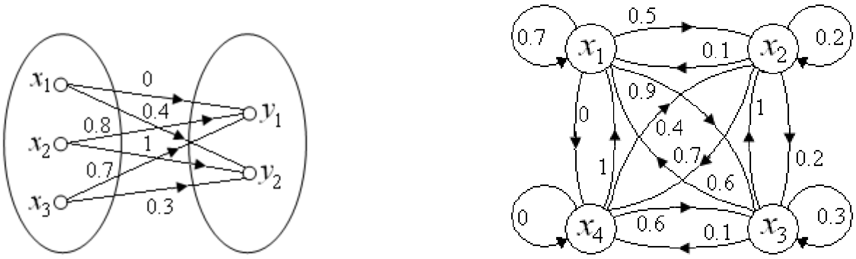
与经典二元关系类似，在有限论域的情况下，模糊二元关系 R 可以直观地用模糊矩阵或者赋权图来表示。设 $X = \{x_1, x_2, \ldots, x_m\}$，$Y = \{y_1, y_2, \ldots, y_n\}$ 为两个非空的有限集合，若 $R\in\mathcal{F}(X\times Y)$，则模糊二元关系 R 可由一个 $m\times n$ 的模糊矩阵 $R = (r_{ij})_{m\times n}$ 表示，其中

$$r_{ij} = R(x_i, y_j),\quad 1 \le i \le m,\ 1 \le j \le n$$

并且模糊二元关系 R 也可用关系图来描述，如图 1(a) 所示。若 $S\in\mathcal{F}(X^2)$，则模糊二元关系 S 可由一个 m 阶的模糊方阵 $S = (s_{ij})_{m\times m}$ 表示，其中

$$s_{ij} = S(x_i, x_j),\quad 1 \le i, j \le m$$

并且模糊二元关系 S 也可用赋权图来描述，如图 1(b) 所示。



(a) 有限集到有限集的模糊二元关系　　　　(b) 有限集上的模糊二元关系

图 1　有限集到有限集的模糊二元关系图

例 3　科学家们相信，大多数人的性格特征是先天和后天两种因素共同影响下形成的。因此，家庭成员之间在性格上总有一定的相似程度。在某家庭的三代成员中，设 $X = \{$长子 (x_1), 长女 (x_2), 次子 $(x_3)\}$，$Y = \{$父亲 (y_1), 母亲 $(y_2)\}$，$Z = \{$祖父 (z_1), 祖母 $(z_2)\}$。如果用 R 表示子女与父母之间的性格相似关系，则它是一个从 X 到 Y 的模糊二元关系；如果用 S 来表示父母与祖父母之间的性格相似关系，则它是一个从 Y 到 Z 的模糊二元关系。经过评估，这两个模糊二元关系可用如下的模糊矩阵来表达

$$R = \begin{pmatrix} 0.2 & 0.4 \\ 0.8 & 0.5 \\ 0.7 & 0.8 \end{pmatrix},\quad S = \begin{pmatrix} 0.5 & 0.7 \\ 0.9 & 0.8 \end{pmatrix}$$

也可用图 2 和图 3 所示的关系图描述。

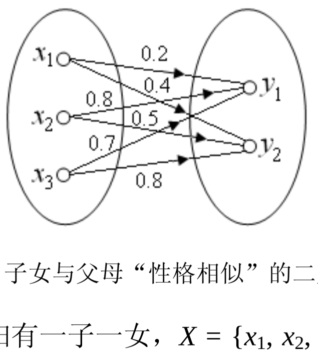


图 2　子女与父母“性格相似”的二元关系

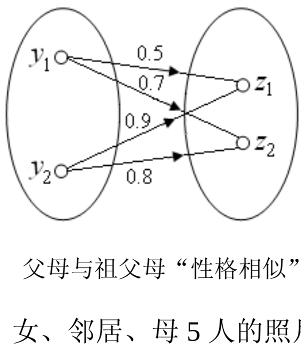


图 3　父母与祖父母“性格相似”的模糊关系

例 4　某夫妇有一子一女，$X = \{x_1, x_2, x_3, x_4, x_5\}$ 为父、子、女、邻居、母 5 人的照片集合。请陌生人根据照片对 5 人相貌的“相像”程度两两打分，得到 X 上的一个模糊二元关系 R，可用模糊矩阵表示如下

$$R=\begin{bmatrix}1 & 0.8 & 0.6 & 0.1 & 0.2\\ 0.8 & 1 & 0.8 & 0.2 & 0.85\\ 0.6 & 0.8 & 1 & 0 & 0.9\\ 0.1 & 0.2 & 0 & 1 & 0.1\\ 0.2 & 0.85 & 0.9 & 0.1 & 1\end{bmatrix}$$

例 5　设 $X\times Y$ 为实数集 **R** 的直积，如果用 **R** 上的一个模糊二元关系 R 描述“x 接近 y”，则 R 的隶属函数可规定为

$$R(x,y)=e^{-|x-y|},\ \forall x,y\in U$$

例 6　设 $X\times Y$ 为实数集 **R** 的直积，如果用 **R** 上的一个模糊二元关系 R 描述“x 远大于 y”的关系，则 R 的隶属函数可规定为

$$R(x,y)=\begin{cases}0, & x\le y\\ [1+100(x-y)^{-2}]^{-1}, & x>y\end{cases}$$

当然，也可用 **R** 上的另一个模糊二元关系 S 来描述“x 远大于 y”的关系，其隶属函数为

$$S(x,y)=\begin{cases}0, & x\le y\\ \dfrac{x-y}{9y}, & y<x\le 10y\\ 1, & x>10y\end{cases}\quad \text{或}\quad S(x,y)=\begin{cases}0, & x\le y\\ \min\left\{1,\ \dfrac{x-y}{9y}\right\}, & x>y\end{cases}$$

由于从 X 到 Y 的模糊二元关系是 $X\times Y$ 的模糊子集，因此可按模糊集合的运算及性质来定义和讨论模糊二元关系的运算和性质。

定义 8　设 R 与 S 为从论域 U 到论域 V 的两个模糊二元关系，即 $R, S\in\mathcal{F}(U\times V)$，则

(1) R 与 S **相等**定义为：$R=S\Leftrightarrow R(x,y)=S(x,y),\ \forall(x,y)\in U\times V$；

(2) R **包含于** S 定义为：$R\subseteq S\Leftrightarrow R(x,y)\le S(x,y),\ \forall(x,y)\in U\times V$；

(3) R 与 S 的**交**仍为从 U 到 V 的模糊二元关系，隶属函数定义为：

$$(R\cap S)(x,y)=\min\{R(x,y),S(x,y)\},\ \forall(x,y)\in U\times V;$$

(4) R 与 S 的**并**仍为从 U 到 V 的模糊二元关系，隶属函数定义为：

$$(R\cup S)(x,y)=\max\{R(x,y),S(x,y)\},\ \forall(x,y)\in U\times V;$$

(5) R 的**余**仍为从 U 到 V 的模糊二元关系，隶属函数定义为：

$$R^C(x,y)=1-R(x,y),\ \forall(x,y)\in U\times V;$$

(6) R 的**逆关系**为从 V 到 U 的模糊二元关系，隶属函数定义为：

$$R^{-1}(y,x)=R(x,y),\ \forall(y,x)\in V\times U。$$

定义 9　设 R 为从论域 U 到论域 V 的模糊二元关系，即 $R\in\mathcal{F}(U\times V)$。令

$$R_\lambda=\{(x,y)\,|\,R(x,y)\ge\lambda,(x,y)\in U\times V\}$$

则称 R_λ 为模糊关系 R 的 $\boldsymbol{\lambda}$ **截关系**。

显然，R_λ 是从 U 到 V 的经典二元关系，并且 $\forall(x, y)\in U\times V$，有

$$R_\lambda(x, y) = 1 \Leftrightarrow R(x, y) \geq \lambda,\quad R_\lambda(x, y) = 0 \Leftrightarrow R(x, y) < \lambda$$

即 x、y 在 λ 水平上才有关系，否则无关系。

模糊二元关系及其截关系具有模糊集合及其截集所具有的一切运算性质，不再赘述，参见第二章。

6.2.2　模糊关系的合成

模糊关系的合成是常用的运算之一，许多模糊关系的性质都与合成有关。自然，模糊二元关系的合成是经典二元关系合成的推广。

由定义 1 知道，模糊关系是经典关系的推广且完全由它的隶属函数来描述，因而模糊二元关系的合成也应是经典二元关系合成的推广且通过隶属函数来实施。于是，将经典二元关系合成运算的特征函数表示形式(参见 1.2.3 节)自然地推广为隶属函数形式，即可定义模糊二元关系的合成运算。

定义 10　给定论域 X、Y、Z，且 $R\in\mathcal{F}(X\times Y)$，$S\in\mathcal{F}(Y\times Z)$。所谓 R 与 S 的合成，就是从 X 到 Z 的一个模糊二元关系，记作 $R\circ S$，其隶属函数为

$$(R\circ S)(x,z) = \bigvee_{y\in Y}[R(x,y)\wedge S(y,z)],\quad \forall (x,z)\in X\times Z \tag{6.4}$$

当 $R\in\mathcal{F}(X\times X)$ 时，记 $R^2 = R\circ R$，$R^n = R^{n-1}\circ R$。

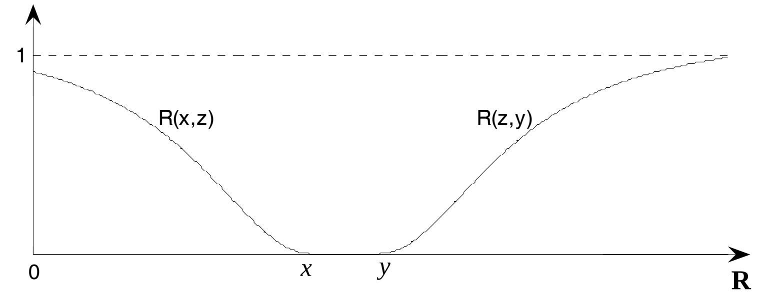
例 7　考虑例 6 中描述“x 远大于 y”的模糊二元关系 R，则可以用 R^2 来描述“x 远远大于 y”的模糊二元关系。根据定义 10，我们有

$$R^2(x,y) = (R\circ R)(x,y) = \bigvee_{z\in \mathrm{R}}[R(x,z)\wedge R(z,y)],\quad \forall x, y\in\mathbf{R}$$

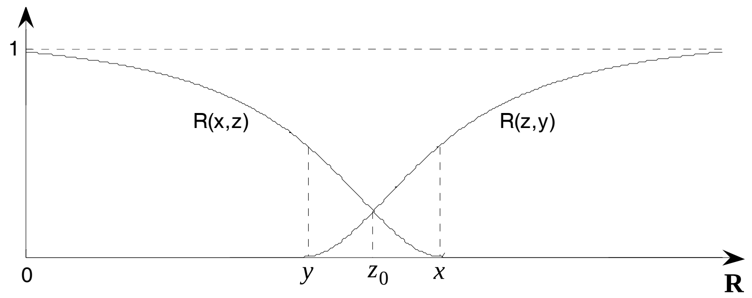
(1) 如果 $x\leq y$，则对 $\forall z\in\mathbf{R}$，有

$$R(x,z) = \begin{cases} 0, & x\leq z \\ [1+100(x-z)^{-2}]^{-1}, & x>z \end{cases},\quad R(z,y) = \begin{cases} 0, & z\leq y \\ [1+100(z-y)^{-2}]^{-1}, & z>y \end{cases}$$

于是当 $x\leq y$ 时有 $R^2(x,y) = \bigvee_{z\in\mathrm{R}}[R(x,z)\wedge R(z,y)] = 0$，如图 4(a) 所示。



(a) 当 $x\leq y$ 时的情形



(b) 当 $x>y$ 时的情形

图 4　模糊关系“x 远远大于 y”的构造

(2) 如果 $x>y$，则对 $\forall z\in\mathbf{R}$，有

$$R(x,z)=\begin{cases}0, & x\le z\\ [1+100(x-z)^{-2}]^{-1}, & x>z\end{cases},\quad R(z,y)=\begin{cases}0, & z\le y\\ [1+100(z-y)^{-2}]^{-1}, & z>y\end{cases}$$

令 z_0 为 $R(x,z)$ 与 $R(z,y)$ 的交点坐标，如图 4(b) 所示，则

$$\begin{aligned}R^2(x,y)&=\bigvee_{z\in\mathrm{R}}[R(x,z)\wedge R(z,y)]\\&=\left[\bigvee_{z\le y<x}[R(x,z)\wedge R(z,y)]\right]\bigvee\left[\bigvee_{y<z\le x}[R(x,z)\wedge R(z,y)]\right]\bigvee\left[\bigvee_{y<x<z}[R(x,z)\wedge R(z,y)]\right]\\&=0\bigvee\left[\bigvee_{y<z\le x}[R(x,z)\wedge R(z,y)]\right]\bigvee 0=R(x,z_0)=R(z_0,y)\end{aligned}$$

令 $R(x,z)=R(z,y)$，则有

$$\left[1+\frac{100}{(x-z)^2}\right]^{-1}=\left[1+\frac{100}{(z-y)^2}\right]^{-1}\Rightarrow|x-z|=|z-y|\Rightarrow z-x=y-z\text{（考虑到 }y\le z\le x\text{）}$$

解得 $z_0=(x+y)/2$，于是当 $x>y$ 时有

$$R^2(x,y)=\left[1+100\bigg/\left(\frac{x+y}{2}\right)^2\right]^{-1}$$

综上，我们有

$$R^2(x,y)=\begin{cases}0, & x\le y\\ \left[1+100\bigg/\left(\dfrac{x+y}{2}\right)^2\right]^{-1} & x>y\end{cases}$$

在有限论域的情形下，由于模糊二元关系可以用模糊矩阵来表示，因此模糊二元关系的合成也可转化为模糊矩阵的合成运算。

例 8　考虑例 3 中的两个模糊关系 R 和 S：

$$R=\begin{pmatrix}0.2&0.4\\0.8&0.5\\0.7&0.8\end{pmatrix},\quad S=\begin{pmatrix}0.5&0.7\\0.9&0.8\end{pmatrix}$$

则子女与祖父母之间的性格相似关系可由 R 与 S 的合成来描述，并且通过模糊矩阵的合成运算获得：

$$R\circ S=\begin{pmatrix}0.2 & 0.4\\ 0.8 & 0.5\\ 0.7 & 0.8\end{pmatrix}\circ\begin{pmatrix}0.5 & 0.7\\ 0.9 & 0.8\end{pmatrix}=\begin{pmatrix}0.4 & 0.4\\ 0.5 & 0.7\\ 0.8 & 0.8\end{pmatrix}$$

模糊关系的合成运算满足如下算律。

定理 6　设 R, S, T 为满足相应运算的模糊二元关系，则有

(1) 结合律：$(R\circ S)\circ T = R\circ(S\circ T)$；特别地，$R^m\circ R^n = R^{m+n}$，$(R^m)^n = R^{mn}$；

(2) 分配律：$(R\cup S)\circ T = (R\circ T)\cup(S\circ T)$，$R\circ(S\cup T) = (R\circ S)\cup(R\circ T)$；

(3) 弱分配律：$(R\cap S)\circ T\subseteq(R\circ T)\cap(S\circ T)$，$R\circ(S\cap T)\subseteq(R\circ S)\cap(R\circ T)$；

(4) 单调性：$R\subseteq S \Rightarrow R\circ T\subseteq S\circ T$，$T\circ R\subseteq T\circ S$。

(5) $(R\circ S)_\lambda = R_\lambda\circ S_\lambda$。

6.2.3　模糊等价关系

经典的等价关系是具有自反性、对称性、传递性的二元关系，利用等价关系可以对一个集合进行划分（分类）。同样地，当模糊二元关系具有自反性、对称性和传递性时，就成为一个模糊等价关系。模糊等价关系也是某些模糊聚类方法的理论基础。

定义 11　设 R 是论域 U 上的一个模糊二元关系，即 $R\in\mathcal{F}(U\times U)$。

(1) 如果 $\forall x\in U$，都有 $R(x, x) = 1$，则称 R 具有**自反性**；

(2) 如果 $\forall x, y\in U$，都有 $R(x, y) = R(y, x)$，则称 R 具有**对称性**；

(3) 如果 $R\supseteq R^2$，则称 R 具有**传递性**。

定理 7　设论域 U 上的模糊二元关系 $R\in\mathcal{F}(U\times U)$ 具有自反性，则

(1) $R^{n+1}\supseteq R^n\,(n\geq 1)$。

(2) R^n 也具有自反性；

证明　(1) 由模糊关系合成运算的定义，及 R 具有自反性，则 $\forall x, y\in U$

$$R^2(x,y)=\bigvee_{z\in U}[R(x,z)\wedge R(z,y)]\geq R(x,y)\wedge R(y,y)=R(x,y)$$

即 $R^2\supseteq R$。若对于某个给定的自然数 k，有 $R^{k+1}\supseteq R^k$，则由定理 6(4) 有 $R^{k+2}\supseteq R^{k+1}$，从而 $R^{n+1}\supseteq R^n\,(n\geq 1)$。

(2) 由 (1) 有：$R^{n+1}\supseteq R^n\supseteq \dots \supseteq R^2\supseteq R$。于是，对于 $\forall x\in U$ 均有 $R^n(x, x) = R(x, x) = 1$，即 R^n 也具有自反性。

定理 8　设 $R\in\mathcal{F}(U\times U)$ 是论域 U 上的模糊二元关系，则

(1) R 具有对称性当且仅当 $R = R^{-1}$；

(2) 若 R 具有对称性，则 R^n 也具有对称性 $(n\geq 1)$；

(3) $R\circ R^{-1}$ 是 U 上的对称模糊关系。

证明　(1) 由对称关系和逆关系的定义可得。

(2) 根据模糊关系合成运算的定义：$\forall x, y\in U$

$$\begin{aligned}(R^n)^{-1}(x,y) &= R^n(y,x)=(R^{n-1}\circ R)(y,x)=\bigvee_{z\in U}[R^{n-1}(y,z)\wedge R(z,x)]\\ &=\bigvee_{z\in U}[(R^{n-1})^{-1}(z,y)\wedge R^{-1}(x,z)]\\ &=\bigvee_{z\in U}[R^{-1}(x,z)\wedge(R^{n-1})^{-1}(z,y)]\\ &=[R^{-1}\circ(R^{n-1})^{-1}](x,y)\end{aligned}$$

从而

$$(R^n)^{-1}=R^{-1}\circ(R^{n-1})^{-1}=R^{-1}\circ R^{-1}\circ(R^{n-2})^{-1}=\ldots=(R^{-1})^n$$

因为 R 具有对称性，于是由 (1) 式有：$(R^n)^{-1}=(R^{-1})^n=R^n$，即 R^n 也具有对称性（$n\geq 1$）。

(3) 根据模糊关系合成运算的定义：$\forall x, y\in U$

$$\begin{aligned}(R\circ R^{-1})^{-1}(x,y) &= (R\circ R^{-1})(y,x)=\bigvee_{z\in U}[R(y,z)\wedge R^{-1}(z,x)]\\ &=\bigvee_{z\in U}[R^{-1}(z,y)\wedge R(x,z)]=\bigvee_{z\in U}[R(x,z)\wedge R^{-1}(z,y)]=(R\circ R^{-1})(x,y)\end{aligned}$$

即 $(R\circ R^{-1})^{-1}=R\circ R^{-1}$，于是由 (1) 式知：$R\circ R^{-1}$ 在 U 上是对称的。

定理 9　设 $R\in\mathcal{F}(U\times U)$ 是论域 U 上的模糊二元关系，则

(1) 若 R 具有自反性，则 R 具有传递性当且仅当 $R=R^2$；

(2) 若 R 具有传递性，则 R^n 也具有传递性（$n\geq 1$）；

证明　(1) 由于 $R=R^2$ 是 $R\supseteq R^2$ 的特例，故充分性成立。

下面证必要性。若 R 具有传递性，因为 R 具有自反性，故 $\forall x, y\in U$

$$R(x,y)\geq R^2(x,y)=\bigvee_{z\in U}[R(x,z)\wedge R(z,y)]\geq R(x,y)\wedge R(y,y)=R(x,y)$$

即 $R=R^2$ 成立。

(2) 因为 R 具有传递性，故 $R\supseteq R^2$，于是由定理 6(4)：$R^n\supseteq(R^2)^n$。而 $(R^2)^n=(R^n)^2$，故 $R^n\supseteq(R^n)^2$，即 R^n 也具有传递性。

定义 12　设 R 是论域 U 上的一个模糊二元关系，即 $R\in\mathcal{F}(U\times U)$。如果 R 具有自反性和对称性，则称 R 为**模糊相似关系**，且称隶属度 $R(x, y)$ 为 x 与 y 关于 R 的**相似程度**；如果 R 具有自反性、对称性和传递性，则称 R 为**模糊等价关系**。

如果论域 U 有限，则描述模糊相似关系和模糊等价关系的模糊矩阵分别称为**模糊相似矩阵**和**模糊等价矩阵**。显然，如果一个模糊矩阵 R 满足

(1) $R\supseteq I$（I 为单位矩阵）

(2) $R=R^{\mathrm{T}}$

则 R 是模糊相似矩阵；如果还满足

(3) $R \supseteq R^2$

则 R 是模糊等价矩阵。

满足 $R = R^2$ 的模糊相似矩阵一定是模糊等价矩阵。

定义 13　设 R 是一个 n 阶模糊矩阵，即 $R \in M_{n\times n}$。

(1) 如果 R 满足 $R \supseteq R^2$，则称 R 为**模糊传递矩阵**；

(2) 在 $M_{n\times n}$ 中包含 R 的最小的模糊传递矩阵称为 R 的传递闭包，记为 $t(R)$。

一个模糊矩阵 R 的传递闭包 $t(R)$ 满足

(1) 传递性：$t(R)\circ t(R) \subseteq t(R)$；

(2) 包含性：$R \subseteq t(R)$；

(3) 最小性：$R \subseteq S$ 且 $S\circ S \subseteq S \Rightarrow t(R) \subseteq S$。

定理 10　设 $R \in M_{n\times n}$ 是 n 阶模糊相似矩阵。则存在一个最小自然数 $k\,(k \le n)$，使得传递闭包 $t(R) = R^k$，且对一切大于 k 的自然数 l，恒有 $R^l = R^k$。

证明　(1) 首先证明：

$$t(R) = \bigcup_{s=1}^{\infty} R^s$$

事实上，1) 我们有

$$\left(\bigcup_{s=1}^{\infty} R^s\right)\circ\left(\bigcup_{k=1}^{\infty} R^k\right) = \bigcup_{s=1}^{\infty}\left(R^s \circ \bigcup_{k=1}^{\infty} R^k\right) = \bigcup_{s=1}^{\infty}\left(\bigcup_{k=1}^{\infty}(R^s \circ R^k)\right) = \bigcup_{m=2}^{\infty}\left(\bigcap_{s+k=m} R^{s+k}\right) = \bigcup_{m=2}^{\infty} R^m \subseteq \bigcup_{m=1}^{\infty} R^m$$

所以 $\bigcup\limits_{s=1}^{\infty} R^s$ 是模糊传递矩阵。

2) 显然，$\bigcup\limits_{s=1}^{\infty} R^s \supseteq R$。

3) 设 $R \subseteq S$ 且 S 为模糊传递矩阵，则 $S^2 \subseteq S$，进而 $S^s \subseteq S$。于是根据模糊关系合成的性质

$$R^s \subseteq S^s \subseteq S \;\Rightarrow\; \bigcup_{s=1}^{\infty} R^s \subseteq S$$

即 $\bigcup\limits_{s=1}^{\infty} R^s$ 是包含 R 的最小的模糊传递矩阵。所以 $t(R) = \bigcup\limits_{s=1}^{\infty} R^s$。

(2) 其次证明：

$$t(R) = \bigcup_{m=1}^{n} R^m$$

记 $R^s = (r_{ij}^{(s)})_{n\times n}$，则有

$$r_{ij}^{(2)} = \bigvee_{k_1=1}^{n}(r_{ik_1} \wedge r_{k_1 j})$$

$$r_{ij}^{(3)} = \bigvee_{k_1=1}^{n}(r_{ik_1} \wedge r_{k_1 j}^{(2)}) = \bigvee_{k_1=1}^{n}[r_{ik_1} \wedge (\bigvee_{k_2=1}^{n}(r_{ik_2} \wedge r_{k_2 j}))] = \bigvee_{k_1=1}^{n}\bigvee_{k_2=1}^{n}(r_{ik_1} \wedge r_{k_1 k_2} \wedge r_{k_2 j})$$

一般地，利用数学归纳法可以证明

$$r_{ij}^{(s)}=\bigvee_{k_1=1}^{n}\cdots\bigvee_{k_{s-1}=1}^{n}(\underbrace{r_{ik_1}\wedge r_{k_1k_2}\wedge\cdots\wedge r_{k_{s-1}j}}_{s\text{个}})$$

当 $s>n$ 时，上式右端每项脚标 $k_1, k_1, \ldots, k_{s-1}, k$ 中必有重复出现者。若 $k_l=k_t$，则略去

$$r_{k_lk_{l+1}}\wedge r_{k_{l+1}k_{l+2}}\wedge\cdots\wedge r_{k_{t-1}j_t}$$

我们有

$$r_{ik_1}\wedge r_{k_1k_2}\wedge\cdots\wedge r_{k_{s-1}j}\le r_{ik_1}\wedge r_{k_1k_2}\wedge\cdots\wedge r_{k_{l-1}j_l}\wedge r_{k_tk_{t+1}}\wedge\cdots\wedge r_{k_{s-1}j}\quad(k_l=k_t)$$

（若 $k_1=k_t$，则略去 $r_{k_1k_2}\wedge\cdots\wedge r_{k_{t-1}j_t}$；若 $k_l=j$，则略去 $r_{k_lk_{l+1}}\wedge\cdots\wedge r_{k_{s-1}j}$）

也就是说，我们可以减少 $t-l$ 个因子(对模糊算子 $\wedge$ 来说)。以此类推，对 $r_{ij}^{(s)}$ 表达式中的每一项

$$r_{ik_1}\wedge r_{k_1k_2}\wedge\cdots\wedge r_{k_{s-1}j}$$

总存在 $m\le n$，使得

$$\underbrace{r_{ik_1}\wedge r_{k_1k_2}\wedge\cdots\wedge r_{k_{s-1}j}}_{s\text{个}}\le\underbrace{r_{il_1}\wedge r_{l_1l_2}\wedge\cdots\wedge r_{l_{m-1}j}}_{m\text{个}}$$

显然，右端乃是

$$r_{ij}^{(m)}=\bigvee_{l_1=1}^{n}\cdots\bigvee_{l_{m-1}=1}^{n}(\underbrace{r_{il_1}\wedge r_{l_1l_2}\wedge\cdots\wedge r_{l_{m-1}j}}_{m\text{个}})$$

的一项。因此

$$r_{ik_1}\wedge r_{k_1k_2}\wedge\cdots\wedge r_{k_{s-1}j}\le r_{ij}^{(m)}$$

上述对应不同项 $r_{ik_1}\wedge r_{k_1k_2}\wedge\cdots\wedge r_{k_{s-1}j}$ 的 m 可以不同，但必须满足 $m\le n$。因此

$$r_{ij}^{(s)}=\bigvee_{k_1=1}^{n}\cdots\bigvee_{k_{s-1}=1}^{n}(\underbrace{r_{ik_1}\wedge r_{k_1k_2}\wedge\cdots\wedge r_{k_{s-1}j}}_{s\text{个}})\le\bigvee_{m=1}^{n}r_{ij}^{(m)}$$

即对 $\forall s>n$，恒有 $R^s\subseteq\bigcup\limits_{m=1}^{n}R^m$。于是

$$t(R)=\bigcup_{s=1}^{\infty}R^s=\left(\bigcup_{s=1}^{n}R^s\right)\cup\left(\bigcup_{s=n+1}^{\infty}R^s\right)=\bigcup_{m=1}^{n}R^m$$

(3) 最后证明：$\exists k\le n$ 使得 $t(R)=R^k$。

根据定理 7 及定理 8，对于模糊相似矩阵 R 有：$R\subseteq R^2\subseteq\ldots\subseteq R^k\subseteq\ldots$，且 R^k 仍然是模糊相似矩阵。于是

$$t(R)=\bigcup_{m=1}^{n}R^m=R^n$$

由于 n 为有限数，因此存在一个最小自然数 $k(k\le n)$，使得传递闭包 $t(R)=R^k$。

此外，对于 $\forall\, l \geq k$，我们有

$$t(R) = R^k \subseteq R^k \supseteq \bigcup_{m=1}^{n} R^m = t(R)$$

因此，对一切大于 k 的自然数 l，恒有 $R^l = R^k$。

结合定理 7~10，我们可以得到将模糊相似矩阵 R 改造为模糊等价矩阵的方法：通过“平方法”依次计算 $R, R^2, R^4, R^8, \ldots$ 当第一次出现$(R^k)^2 = R^k$时，R^k就是 R 的传递闭包 $t(R)$，而 $t(R)$就是由 R 经过改造后获得的模糊等价矩阵。

这也表明，对于任何模糊相似关系，通过至多 $n-1$ 次复合就可以重新组合成一个模糊等价关系。

在实际应用中，往往需要根据具体情况建立一个模糊等价矩阵，它需要同时满足自反性、对称性和传递性，这在技术上常常会遇到困难。因而，我们可以暂不考虑传递性，建立一个具有自反性和对称性的模糊相似矩阵，然后再将它进行改造，使它满足传递性而且保留原有的自反性和对称性。

例 9 考虑例 4 中描述相貌“相像”关系的模糊矩阵

$$R = \begin{bmatrix} 1 & 0.8 & 0.6 & 0.1 & 0.2 \\ 0.8 & 1 & 0.8 & 0.2 & 0.85 \\ 0.6 & 0.8 & 1 & 0 & 0.9 \\ 0.1 & 0.2 & 0 & 1 & 0.1 \\ 0.2 & 0.85 & 0.9 & 0.1 & 1 \end{bmatrix}$$

由于 $R \supseteq I$ 且 $R = R^{\mathrm{T}}$，故 R 是模糊相似矩阵。因为

$$R^2 = \begin{bmatrix} 1 & 0.8 & 0.8 & 0.2 & 0.8 \\ 0.8 & 1 & 0.85 & 0.2 & 0.85 \\ 0.8 & 0.85 & 1 & 0.2 & 0.9 \\ 0.2 & 0.2 & 0.2 & 1 & 0.2 \\ 0.8 & 0.85 & 0.9 & 0.2 & 1 \end{bmatrix} \neq R\text{，}\quad R^4 = \begin{bmatrix} 1 & 0.8 & 0.8 & 0.2 & 0.8 \\ 0.8 & 1 & 0.85 & 0.2 & 0.85 \\ 0.8 & 0.85 & 1 & 0.2 & 0.9 \\ 0.2 & 0.2 & 0.2 & 1 & 0.2 \\ 0.8 & 0.85 & 0.9 & 0.2 & 1 \end{bmatrix} = R^2$$

于是 R 的传递闭包为 $t(R) = R^4 = R^2$。从而 $t(R)$ 是由 R 经过改造后得到的模糊等价矩阵。

6.2.4 模糊关系的广义合成运算

在 6.2.2 中给出的模糊关系的合成运算，是通过对关系程度逐点取小再取大来实现的(称之为最大–最小合成或 max–min 合成)，它是经典二元关系合成运算的特征函数表示形式的自然推广。

然而，当用特征函数来表达经典二元关系的合成运算时，除了采用对关系程度逐点取小再取大的形式之外，也可以采用对关系程度逐点做 T 模运算再取大。

于是，类似于模糊集合的广义运算，我们也可以引入模糊关系的广义合成运算。

定义 14 给定论域 X、Y、Z，且 $R \in \mathcal{F}(X \times Y)$，$S \in \mathcal{F}(Y \times Z)$。所谓 R 与 S 的广义合成，就是从 X 到 Z 的一个模糊二元关系，记作 $R \circ S$，其隶属函数为

$$(R \circ S)(x,z) = \bigvee_{y \in Y} T(R(x,y), S(y,z)), \quad \forall (x,z) \in X \times Z \tag{6.5}$$

其中 T 为三角模。

特别地，当取 $T = \min$ 时，就得到最大–最小合成运算；当取 T 为代数积时，就得到最大–代数积合成运算；等等。

6.3　模糊聚类分析

聚类分析就是将一个没有类别标记的样本集按照某种准则划分成若干个子集(类)，使相似的样本尽可能归为一类，而不相似的样本尽可能划分到不同的类中。由于在对样本集进行聚类的过程中，没有任何关于类别的先验知识，所以聚类分析属于无监督分类的范畴。

传统的聚类分析是一种硬划分，它将每个待识别的对象严格地划分到某个类中，类别划分的界限是分明的，具有“非此即彼”的性质。而现实世界中，一组对象根据其亲疏程度和相似性是否形成一个类群，或一个对象是否属于一个类别，其界限往往是不分明的，具有“亦此亦彼”的性质。对于这种带有不确定性的聚类问题，模糊聚类分析提供了有力的分析工具。

模糊聚类分析能够建立样本对于类别的不确定性描述，表达样本类属的中介性，已经成为聚类分析研究的主流。粗略来讲，模糊聚类分析方法可分为两类：基于模糊等价关系的聚类方法和基于目标函数的聚类方法。有时，这两类方法也结合起来使用。

6.3.1　数据预处理

在模糊聚类分析中，我们称待分类的对象为样本。要对样本进行合理的分类，首先应考虑样本的各种特性指标(观测数据)。设有 n 个被分类对象，即样本集为

$$X = \{x_1, x_2, \dots, x_n\}$$

每一个 x_i 有 p 个特性指标，即 x_i 可表示为特性指标向量

$$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$$

其中 x_{ij} 表示第 i 个样本的第 j 个特性指标。于是，n 个样本的特性指标矩阵为

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

通常，我们也将样本集记为特性指标矩阵的形式，即 $X = (x_{ij})_{n \times p}$。

如果 p 个特性指标的量纲和数量级都不相同，在运算过程中就可能会因为突出某些数量级特别大的特性指标对分类的作用，而降低甚至排除某些数量级很小的特性指标的作用，致使对各特性指标的分类缺乏

一个统一的尺度。所以，为了消除特性指标单位的差别和数量级不同的影响，当特性指标的量纲和数量级不相同时，通常事先对各种指标值实施数据标准化(规格化)，从而使得各个指标值都统一于某种共同的数值特性范围。我们称之为数据预处理。

常用的数据标准化方法有两种：均值方差标准化和极大极小标准化。

(1) 均值方差标准化

设给定的样本集为 $X = (x_{ij})_{n\times p}$，标准化之后的样本集为 $X = (x'_{ij})_{n\times p}$，则

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}\text{，}\ i = 1, 2, \ldots, n\text{，}\ j = 1, 2, \ldots, p$$

式中

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}\text{，}\quad \sigma_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}\text{，}\ j = 1, 2, \ldots, p$$

(2) 极大极小标准化

设给定的样本集为 $X = (x_{ij})_{n\times p}$，标准化之后的样本集为 $X = (x'_{ij})_{n\times p}$，则

$$x'_{ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}}\text{，}\ i = 1, 2, \ldots, n\text{，}\ j = 1, 2, \ldots, p$$

式中

$$x_{j\min} = \min_{1\le i\le n}\{x_{ij}\}\text{，}\quad x_{j\max} = \max_{1\le i\le n}\{x_{ij}\}\text{，}\ j = 1, 2, \ldots, p$$

显然，实施数据标准化之后，每个指标值均在区间 [0, 1] 中。

6.3.2 基于模糊等价关系的聚类方法

（一）模糊等价矩阵聚类方法

模糊等价矩阵聚类方法是典型的基于模糊等价关系的聚类方法之一。其主要思想就是从计算各个样本之间的相似性统计量出发，建立样本集 X 上的模糊相似矩阵(关系)；通过改造模糊相似矩阵为模糊等价矩阵，达到对样本集 X 进行模糊聚类的目的。

模糊等价矩阵聚类法

1° 选择适当的相似性统计量；

2° 构造样本集上的模糊相似矩阵；

3° 将模糊相似矩阵改造为模糊等价矩阵；

4° 聚类；画出聚类的谱系图。

1. 建立模糊相似矩阵

设待分类的样本集为 $X = \{x_1, x_2, \dots, x_n\}$ 或 $X = (x_{ij})_{n\times p}$，并已经标准化。如果能够计算出衡量样本 x_i 与 x_j 之间相似程度的相似性统计量 r_{ij}，使得 $0 \le r_{ij} \le 1$（$i, j = 1, 2, \dots, n$，其中，$r_{ij} = 0$ 表示样本 x_i 与 x_j 之间毫不相似，$r_{ij} = 1$ 表示样本 x_i 与 x_j 之间完全相似或者等同，r_{ii} 表示样本 x_i 自己与自己的相似程度，恒取为 1，即 $r_{ii} = 1$，$i = 1, 2, \dots, n$），那么，描述样本之间的模糊相似关系、建立在样本集 X 上的模糊相似矩阵为

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{pmatrix},\ \text{其中} r_{ij} = r_{ji} \in [0,1]$$

常用的计算样本的相似性统计量的方法有如下几种：

(1) 相关系数法

$$r_{ij} = \frac{\sum_{k=1}^{p} |x_{ik} - \bar{x}_i| \cdot |x_{jk} - \bar{x}_j|}{\sqrt{\sum_{k=1}^{p}(x_{ik} - \bar{x}_i)^2} \cdot \sqrt{\sum_{k=1}^{p}(x_{ik} - \bar{x}_j)^2}}$$

其中

$$\bar{x}_i = \frac{1}{p}\sum_{k=1}^{p} x_{ik}\ ,\quad \bar{x}_j = \frac{1}{p}\sum_{k=1}^{p} x_{jk}$$

(2) 指数相似系数法

$$r_{ij} = \frac{1}{p}\sum_{k=1}^{p} e^{-\frac{4}{3}\cdot\frac{(x_{ik}-x_{jk})^2}{S_k^2}}$$

其中 S_k 是第 k 个特征的标准差：

$$S_k^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_{ik} - \bar{x}_k\right)^2,\quad \bar{x}_k = \frac{1}{n}\sum_{i=1}^{n} x_{ik}$$

(3) 夹角余弦法

$$r_{ij} = \frac{\sum_{k=1}^{p} x_{ik} \cdot x_{jk}}{\sum_{k=1}^{p} \sqrt{x^2{}_{ik}} \cdot \sqrt{x^2{}_{jk}}}$$

(4) 数量积法

$$r_{ij} = \begin{cases} 1, & i = j \\ \dfrac{1}{M}\sum_{k=1}^{p} x_{ik} \cdot x_{jk}, & i \ne j \end{cases}$$

其中 M 为一适当选取的正数，满足

$$M \ge \max_{i,\,j}\left\{\sum_{k=1}^{p} x_{ik} \cdot x_{jk}\right\}$$

(5) 最大最小法

$$r_{ij}=\frac{\sum_{k=1}^{p}\min(x_{ik},x_{jk})}{\sum_{k=1}^{p}\max(x_{ik},x_{jk})}$$

(6) 算术平均最小法

$$r_{ij}=\frac{\sum_{k=1}^{p}\min(x_{ik},x_{jk})}{\frac{1}{2}\sum_{k=1}^{p}(x_{ik}+x_{jk})}$$

(7) 几何平均最小法

$$r_{ij}=\frac{\sum_{k=1}^{p}\min(x_{ik},x_{jk})}{\sum_{k=1}^{p}\sqrt{x_{ik}\cdot x_{jk}}}$$

(8) 绝对值指数法

$$r_{ij}=e^{-\sum_{k=1}^{p}|x_{ik}-x_{jk}|}$$

(9) 绝对值减数法

$$r_{ij}=\begin{cases}1, & i=j\\ 1-c\sum_{k=1}^{p}|x_{ik}-x_{jk}|, & i\neq j\end{cases}$$

其中，c 是一个适当选取的数，使得 $0\le r_{ij}\le 1$。

(10) 绝对值倒数法

$$r_{ij}=\begin{cases}1, & i=j\\ \dfrac{M}{\sum_{k=1}^{p}|x_{ik}-x_{jk}|}, & i\neq j\end{cases}$$

其中 M 为适当选取的正数，使得 $0\le r_{ij}\le 1$。

在上述计算样本相似性统计量的方法中：(1)~(7) 是利用两个样本的相似系数来描述它们之间的相似性，其中 (3)~(7) 还要求 $x_{ij}\ge 0$（否则需要进行适当的变换）；而 (8)~(10) 则是利用两个样本的距离来描述它们之间的相似性，所以其中的 $\sum|x_{ik}-x_{jk}|$ 也可以用一般的向量范数 $\|x_i-x_j\|_A$ 来替代。

2．改造模糊相似矩阵为模糊等价矩阵

根据计算相似性统计量得到的模糊矩阵 R，一般只满足自反性和对称性，即 R 是相似矩阵。为了进行模糊聚类，需将 R 改造成模糊等价矩阵。为此采用平方法求出 R 的传递闭包 $t(R)$，$t(R)$ 便是所求的模糊等价矩阵。

3．聚类

根据得到的模糊等价矩阵 $t(R)$，我们就可以利用不同水平下的截矩阵得到该水平下的聚类结果。所有不同水平的聚类结果形成聚类的谱系图。

例 10　环境单元分类。每个环境单元包括空气、水分、土壤、作物四个要素。环境单元的污染状况由污染物在四要素中含量的超限度来描述。现有五个环境单元，它们的污染数据如下：

$$\mathrm{I} = (5, 5, 3, 2),\ \mathrm{II} = (2, 3, 4, 5),\ \mathrm{III} = (5, 5, 2, 3),\ \mathrm{IV} = (1, 5, 3, 1),\ \mathrm{V} = (2, 4, 5, 1)$$

设 $U = \{\mathrm{I, II, III, IV, V}\}$，试对 U 进行分类。

解　样本集的特性指标矩阵为

$$\begin{pmatrix} 5 & 5 & 3 & 2 \\ 2 & 3 & 4 & 5 \\ 5 & 5 & 2 & 3 \\ 1 & 5 & 3 & 1 \\ 2 & 4 & 5 & 1 \end{pmatrix}$$

由于数据不存在量纲和数量级的差异，故不需进行数据标准化，直接进入构造模糊相似矩阵步骤。按照绝对值减数法建立模糊相似关系，取 $c = 0.1$，得模糊相似矩阵

$$R = \begin{pmatrix} 1 & 0.1 & 0.8 & 0.5 & 0.3 \\ 0.1 & 1 & 0.1 & 0.2 & 0.4 \\ 0.8 & 0.1 & 1 & 0.5 & 0.3 \\ 0.5 & 0.2 & 0.5 & 1 & 0.6 \\ 0.3 & 0.4 & 0.3 & 0.6 & 1 \end{pmatrix}$$

用平方法求传递闭包，以便将模糊相似矩阵改造成模糊等价矩阵，我们有：

$$R^2 = \begin{pmatrix} 1 & 0.3 & 0.8 & 0.5 & 0.5 \\ 0.3 & 1 & 0.3 & 0.4 & 0.4 \\ 0.8 & 0.3 & 1 & 0.5 & 0.5 \\ 0.5 & 0.4 & 0.5 & 1 & 0.6 \\ 0.5 & 0.4 & 0.5 & 0.6 & 1 \end{pmatrix},\quad R^4 = \begin{pmatrix} 1 & 0.4 & 0.8 & 0.5 & 0.5 \\ 0.4 & 1 & 0.4 & 0.4 & 0.4 \\ 0.8 & 0.4 & 1 & 0.5 & 0.5 \\ 0.5 & 0.4 & 0.5 & 1 & 0.6 \\ 0.5 & 0.4 & 0.5 & 0.6 & 1 \end{pmatrix},$$

$$R^8 = \begin{pmatrix} 1 & 0.4 & 0.8 & 0.5 & 0.5 \\ 0.4 & 1 & 0.4 & 0.4 & 0.4 \\ 0.8 & 0.4 & 1 & 0.5 & 0.5 \\ 0.5 & 0.4 & 0.5 & 1 & 0.6 \\ 0.5 & 0.4 & 0.5 & 0.6 & 1 \end{pmatrix} = R^4$$

于是，传递闭包 $t(R) = R^4$ 就是所求的模糊等价矩阵。根据得到的模糊等价矩阵 $t(R)$，利用不同水平下的截矩阵得到各个水平下的聚类结果如下：

当 $0.0 \le \lambda \le 0.4$ 时，U 分为一类：$\{\mathrm{I, II, III, IV, V}\}$；

当 $0.4 < \lambda \le 0.5$ 时，U 分为二类：$\{\mathrm{I, III, IV, V}\}$，$\{\mathrm{II}\}$；

当 $0.5 < \lambda \le 0.6$ 时，U 分为三类：{I, III}，{IV, V}，{II}；

当 $0.6 < \lambda \le 0.8$ 时，U 分为四类：{I, III}，{II}，{IV}，{V}；

当 $0.8 < \lambda \le 1.0$ 时，U 分为五类：{I}，{II}，{III}，{IV}，{V}。

最后，将所有不同水平下的聚类结果形成聚类的谱系图如下：

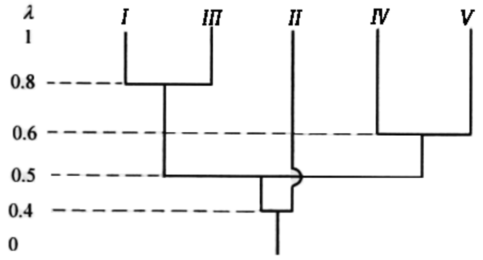


图 5　不同水平下聚类结果的谱系图

（二）模糊最大支撑树聚类方法

模糊最大支撑树聚类方法是另一个典型的基于模糊等价关系的聚类方法。它首先构造一个完全赋权图 $K_{|X|}$，$K_{|X|}$ 中的顶点为待分类的样本点，边权为相应的两个样本之间的相似性统计量值；然后通过寻找完全赋权图 $K_{|X|}$ 的最大支撑树，来进行聚类。

模糊最大支撑树聚类法

1° 选择适当的相似性统计量；

2° 构造样本集上的模糊相似矩阵；

3° 根据模糊相似关系矩阵构造完全赋权图；

4° 寻找完全赋权图的最大支撑树；

5° 由最大支撑树进行聚类分析。

1．建立模糊相似矩阵

设待分类的样本集为 $X = \{x_1, x_2, \dots, x_n\}$ 或 $X = (x_{ij})_{n\times p}$，并已经标准化。如果能够计算出衡量样本 x_i 与 x_j 之间相似程度的相似性统计量 r_{ij}，使得

$$0 \le r_{ij} \le 1,,\quad i, j = 1, 2, \dots, n$$

其中，$r_{ij} = 0$ 表示样本 x_i 与 x_j 之间毫不相似，$r_{ij} = 1$ 表示样本 x_i 与 x_j 之间完全相似或者等同，r_{ii} 表示样本 x_i 自己与自己的相似程度，恒取为 1，即 $r_{ii} = 1$，$i = 1, 2, \dots, n$，那么，描述样本之间的模糊相似关系、建立在样本集 X 上的模糊相似矩阵为

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{pmatrix},\ \text{其中} r_{ij} = r_{ji} \in [0,1]$$

建立模糊相似矩阵的方法与模糊等价矩阵聚类法完全相同。

2．构造完全赋权图

构造一个完全赋权图 $K_{|X|}$，$K_{|X|}$ 中顶点集为 $X = \{x_1, x_2, \ldots, x_n\}$，每条边 x_ix_j 的权值为 x_i 与 x_j 之间的相关系数(模糊相似关系矩阵中的元素) r_{ij}。

3．寻找 $K_{|X|}$ 的最大支撑树

用 Kruskal 算法或者 Prim 算法，求完全赋权图 $K_{|X|}$ 的最大支撑树(生成树) T。

Kruskal 算法和 Prim 算法可参见相关的图论著作。

4．聚类

适当选取阈值 λ，砍去最大支撑树 T 中权值小于 λ 的边，相互连通的顶点(样本点)归为同一类。

例如，设 $X = \{x_1, \ldots, x_6\}$，已知求得的模糊相似关系矩阵为：

$$R = \begin{pmatrix} 1 & & & & & \\ 0.53 & 1 & & & & \\ 0.38 & 0.38 & 1 & & & \\ 0.45 & 0.45 & 0.38 & 1 & & \\ 0.64 & 0.53 & 0.38 & 0.45 & 1 & \\ 0.60 & 0.38 & 0.53 & 0.55 & 0.59 & 1 \end{pmatrix}$$

利用图论中的 Prim 算法，可求得矩阵

$$T = \begin{pmatrix} 1 & 1 & 6 & 1 & 6 \\ 5 & 6 & 4 & 2 & 3 \\ 0.64 & 0.60 & 0.55 & 0.53 & 0.53 \end{pmatrix}$$

其中，矩阵 T 中的第一、第二行表示构成最大支撑树 T 的边的端点标号，第三行表示对应的边权值，如图。

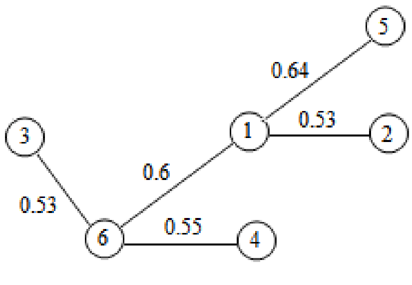


图 6　最大支撑树 T

根据最大支撑树 T，可以得到不同水平下的聚类结果如下：

当 $0.00 \le \lambda \le 0.53$ 时，X 分为一类：{1, 2, 3, 4, 5, 6}；

当 $0.53 < \lambda \le 0.55$ 时，X 分为三类：{1, 4, 5, 6}，{2}，{3}；

当 $0.55 < \lambda \le 0.60$ 时，X 分为四类：{1, 5, 6}，{2}，{3}，{4}；

当 $0.60 < \lambda \le 0.64$ 时，X 分为五类：{1, 5}，{2}，{3}，{4}，{6}；

当 $0.64 < \lambda \le 1.00$ 时，X 分为六类：{1}，{2}，{3}，{4}，{5}，{6}。

最后，将所有不同水平下的聚类结果形成聚类的谱系图如下：

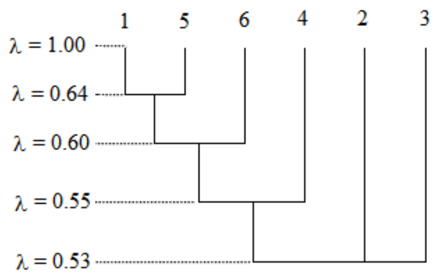


图 7　不同水平下聚类结果的谱系图

（三）最佳阈值的确定

基于模糊等价关系的模糊聚类算法，如上述的模糊等价矩阵法和模糊最大支撑树法，本质上是一种动态的聚类方法，最终给出的只是在不同水平下的聚类结果。如何选择阈值 λ，使得在 λ 水平下样本集的聚类结果更为合理，我们称之为最佳阈值的确定问题。最佳阈值通常根据问题的背景和经验知识来确定，也可以运用 F 统计量来选择理论最佳阈值。

如果样本集 $X = \{x_1, x_2, \ldots, x_n\}$ 第 k 个样本的特性指标向量为 $x_k = \{x_{k1}, x_{k2}, \ldots, x_{kp}\}$，$k = 1, \ldots, n$，$n_i$ 为第 i 个类的样本数，第 i 个类的样本为 $\{x_1^i, x_2^i, \ldots, x_{ni}^i\}$，$x_k^i$ 的特性指标向量为 $x_k^i = (x_{k1}^i, x_{k2}^i, \ldots, x^i_{kpx_k^i=(x_{k1}^i, x_{k2}^i, \ldots, x_{km}^i)})$，$k = 1, \ldots, n_i$，则第 i 个类的聚类中心向量为

$$v_i = (v_1^i, v_2^i, \ldots, v_p^i),\quad v_j^i = \frac{1}{n_i}\sum_{k=1}^{n_i} x_{kj}^i,\quad j = 1, \ldots, p$$

样本集的中心向量为

$$\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p),\quad \bar{x}_j = \frac{1}{n}\sum_{k=1}^{n} x_j,\quad j = 1, \ldots, p$$

设 C_λ 为对应 λ 值的聚类数，构造 F 统计量

$$F_\lambda = \frac{\dfrac{1}{C_\lambda - 1}\sum_{i=1}^{C_\lambda} n_i \| v_i - \bar{x} \|^2}{\dfrac{1}{n - C_\lambda}\sum_{i=1}^{C_\lambda}\sum_{j=1}^{n_i} \| x_j^i - v_i \|^2}$$

式中分子表示类间距离，分母表示类内距离，F_λ 越大说明分类越合理，对应 F 统计量最大的阈值 λ^* 为最佳阈值。

6.3.3　模糊 *C* 均值聚类方法

模糊 C 均值(FCM，Fuzzy C Means)聚类算法是一个通用的数据聚类算法，其中每个样本点属于一个类的程度，以一个隶属度来指定。在基于目标函数的模糊聚类方法中，FCM 算法的理论最为完善，应用最为广泛。

（一）模糊 C 均值(FCM)聚类算法

设待分类的样本集为 $X=\{x_1, x_2, \ldots, x_n\}$ 或 $X=(x_{ij})_{n\times p}$，FCM 算法就是将样本集 X 划分成 c 个模糊群组，并且在每个模糊群组中寻找一个聚类中心，使得一个基于距离测度的目标函数最小化。它兼顾了类之间的交迭，允许对象对所有的类有部分归属。每一个样本点都以 0 和 1 之间的一个隶属程度属于任何群组，聚类的结果是用下面的 $c\times n$ 矩阵

$$U=(\mu_{ik})_{c\times n}=\begin{pmatrix}\mu_{11} & \mu_{12} & \cdots & \mu_{1n}\\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n}\\ \vdots & \vdots & \cdots & \vdots\\ \mu_{c1} & \mu_{c2} & \cdots & \mu_{cn}\end{pmatrix}$$

表示的一个模糊 c 划分：

$$M_{fc}=\{U\in R^{c\times n}\mid \mu_{ik}\in[0,1];\ \ \textstyle\sum_{i=1}^{c}\mu_{ik}=1，\ \forall k;\ \ n>\sum_{k=1}^{n}\mu_{ik}>0，\ \forall i\}$$

矩阵 U 的第 i 行给出了描述样本集 X 中第 i 个类的模糊子集 μ_i 的隶属函数，$i=1, \ldots, c$；U 的第 k 列表示样本点 x_k 在 X 的 c 个模糊子集中的隶属度值；$\mu_{ik}=\mu_i(x_k)$ 表示 x_k 隶属于 X 的第 i 个类(模糊子集)的隶属度值。

设 J_m：$M_{fc}\times R^{c\times p}\to R^+$，

$$J_m(U, V)=\sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}^{m}d_{ik}^{2} \tag{6.6}$$

这里，$U\in M_{fc}$ 是 X 的模糊 c 划分；$V=\{v_1, \ldots, v_c\}$，$v_i\in R^p$ 是第 i 个模糊群组 μ_i 的聚类中心，$i=1, \ldots, c$；d_{ik}^2 是样本点 x_k 与第 i 类的中心向量 v_i 之间的距离测度，

$$d_{ik}^2=\|x_k-v_i\|_A=(x_k-v_i)^{\mathrm{T}}A(x_k-v_i)，A 是 p 阶对称正定矩阵 \tag{6.7}$$

$m\in[1, \infty]$ 是控制最后划分的模糊性的加权指数，增大 m 将增加函数的模糊性。显然， $J_m(U, V)$ 是一个平方误差聚类准则，算法的目的是寻找合适的 U 和 V，使得 $J_m(U, V)$ 达到最小值，因此，聚类问题即转化为求解如下的非线性优化问题

$$\begin{cases}\min J_m(U,V)=\displaystyle\sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}^{m}d_{ik}^{2}\\ \text{s.t.}\quad U\in M_{fc}\end{cases} \tag{6.8}$$

由于 U 中各列都是独立的，因此

$$\min\{J_m(U,V)\}=\min\left\{\sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}^{m}d_{ik}^{2}\right\}=\sum_{k=1}^{n}\min\left\{\sum_{i=1}^{c}\mu_{ik}^{m}d_{ik}^{2}\right\} \tag{6.9}$$

考虑到约束条件 $\sum_{i=1}^{c}\mu_{ik}=1$，使用 Lagrange 乘数法，令

$$F=\sum_{i=1}^{c}\mu_{ik}^{m}d_{ik}^{2}+\lambda\left(\sum_{i=1}^{c}\mu_{ik}-1\right) \tag{6.10}$$

则最优化的一阶必要条件为

$$\frac{\partial F}{\partial \lambda}=\sum_{i=1}^{c}\mu_{ik}-1=0 \tag{6.11a}$$

且

$$\frac{\partial F}{\partial \mu_{jt}}=m\mu_{jt}^{m-1}d_{jt}^{2}-\lambda=0 \tag{6.11b}$$

由(6.11b)得

$$\mu_{jt}=\left[\frac{\lambda}{md_{jt}^{2}}\right]^{\frac{1}{m-1}} \tag{6.12}$$

将上式代入(6.11a)得

$$\sum_{l=1}^{c}\mu_{lt}=\sum_{l=1}^{c}\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\left[\frac{1}{d_{lt}^{2}}\right]^{\frac{1}{m-1}}=\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\left\{\sum_{l=1}^{c}\left[\frac{1}{d_{lt}^{2}}\right]^{\frac{1}{m-1}}\right\}=1$$

因而

$$\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}=\frac{1}{\sum_{l=1}^{c}\left[\frac{1}{d_{lt}^{2}}\right]^{\frac{1}{m-1}}}$$

将上式代入(6.12)得

$$\mu_{jt}=\frac{1}{\sum_{l=1}^{c}\left[\frac{d_{jt}}{d_{lt}}\right]^{\frac{2}{m-1}}}$$

考虑到 d_{ik} 可能为 0，应分两种情况加以讨论。对于 $\forall k$，定义集合

$$I_k=\{i \mid 1\le i\le c,\ d_{ik}=0\}，\ \tilde{I}_k=\{1,\ \ldots,\ c\}-I_k,$$

则使得 $J_m(U, V)$ 达到最小的 μ_{ik} 值为：

$$\begin{cases}\mu_{ik}=\dfrac{1}{\sum_{j=1}^{c}\left[\dfrac{d_{ik}}{d_{jk}}\right]^{\frac{2}{m-1}}}, & I_k=\varnothing\\ \mu_{ik}=0,\ \forall i\in\tilde{I}_k,\ \sum_{i\in I_k}\mu_{ik}=1, & I_k\ne\varnothing\end{cases} \tag{6.13}$$

用类似的方法可以获得 $J_m(U, V)$ 达到最小的 v_i 值。令

$$\frac{\partial}{\partial v_i}J_m(U,V)=0$$

得到

$$\sum_{k=1}^{n}\mu_{ik}^{m}\frac{\partial}{\partial v_i}[(x_k-v_i)^{\mathrm{T}}A(x_k-v_i)]=0$$

于是

$$\sum_{k=1}^{n}\mu_{ik}^{m}[-2A(x_k-v_i)]=0$$

即

$$\sum_{k=1}^{n}\mu_{ik}^{m}(x_k-v_i)=0$$

由此得

$$v_i=\frac{1}{\sum_{k=1}^{n}\mu_{ik}^{m}}\sum_{k=1}^{n}\mu_{ik}^{m}x_k \tag{6.14}$$

若样本集 X、聚类数 c 和权重 m 的值已知，就能由式(6.13)和(6.14)确定最佳模糊分类矩阵和聚类中心。算法的迭代步骤如下：

FCM 聚类算法

1° 初始化：给定聚类数 c，$0\le c\le n$，设定迭代停止阈值 ε，初始化中心向量矩阵 $V^{(0)}$，设置迭代计数器 $l=0$；

2° 更新划分矩阵 $U^{(l)}$：对于 $\forall i$，k，如果 $\exists\, d_{ik}^{(l)}>0$，则有

$$\mu_{ik}^{(l)}=\left\{\sum_{j=1}^{c}\left[\frac{d_{ik}^{(l)}}{d_{jk}^{(l)}}\right]^{\frac{2}{m-1}}\right\}^{-1} \tag{6.15a}$$

如果 $\exists i$，r 使得 $d_{ir}^{(l)}=0$，则有

$$\mu_{ir}^{(l)}=1，\text{且对 } j\neq r,\quad \mu_{ij}^{(l)}=0 \tag{6.15b}$$

3° 更新聚类中心向量矩阵 $V^{(l+1)}$：对于 $\forall i$，k，如果 $\exists\, d_{ik}^{(l)}>0$，则有

$$v_i^{(l+1)}=\frac{\sum_{k=1}^{n}\left(\mu_{ik}^{(l+1)}\right)^m\cdot x_k}{\sum_{k=1}^{n}\left(\mu_{ik}^{(l+1)}\right)^m} \tag{6.16}$$

4° 判断与终止：如果 $\|V^{(l)}-V^{(l+1)}\|<\varepsilon$，则算法停止，并输出划分矩阵 U 和聚类中心矩阵 V；否则，令 $l=l+1$，转向 2°。

FCM 是一个实现模糊聚类的典型算法，它以最小化类内距的方式来更新隶属度。一旦样本点的隶属度确定下来，按照最大隶属原则，样本点将被指派给具有最高隶属度的类别。

（二）聚类有效性分析

与其它基于目标函数的模糊聚类方法一样，FCM 要求在实施算法之前，必须事先指定分类数目。如果我们指定的聚类数不正确，即使使用很好的聚类算法也不会得到最优的聚类结果。例如，图 8 给出一个数据集，我们从视觉上很容易得知这个数据集分三类。然而，当我们使用 FCM 聚类算法对其进行聚类，并把类别数设为 4 时，我们会得到如图 9 所示的结果。FCM 算法可能得到了类别数为 4 时的最好的聚类效果，然而这个结果对于这个数据集来说却不是最优的，因为它没有反应出数据集的真实结构。

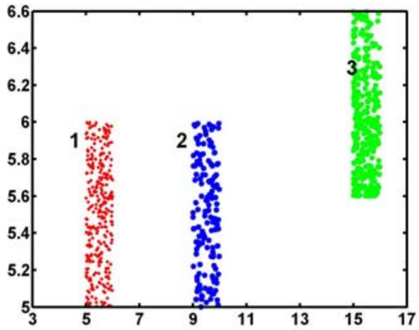


图 8　由三类数据组成的数据集

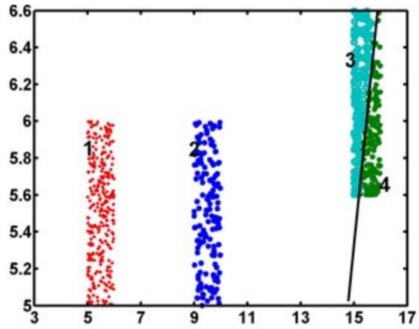


图 9 数据集被 FCM 聚成 4 类的结果

然而，在实际应用中通常事先并不知道或难以确切知道样本集的聚类数。因此，根据给定样本集中数据的内在结构来自动判断样本集最应该具有的聚类数(称之为最佳聚类数)，是一个非常有意义的课题。这种自动寻求样本集的最佳聚类数问题，我们称之为聚类有效性分析，可以借助有效性函数来实现。聚类的有效性函数，均以聚类算法的相关参数为自变量，目的是评价在选定的参数值下，算法所得的分类结果是否或以多大程度有效地反映了数据本身固有的分类结构。经过深入研究，人们基本取得一致的认识：反映聚类有效性的主要特征是类内的紧致性和类间的分离性，因而人们设计的聚类有效性函数，都始于这一基本出发点。

已经提出的聚类有效性函数有许多，并且人们还在逐步改进。比较有代表性的有如下几个。

(1) 划分系数

如果将给定样本集分成 c 类，对应的模糊划分矩阵为 $U = (\mu_{ij})_{c\times n}$，则相应的划分系数为

$$F_c = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2 = \frac{1}{n}\mathrm{tr}(UU^{\mathrm{T}})$$

F_c 的最大值对应于最佳的聚类数 c^*。

事实上，若令 $UU^{\mathrm{T}} = (s_{ij})_{c\times c}$，则 s_{ij} 是 U 的第 i 行与第 j 行的内积(即相关性)，反映的是在模糊划分矩阵为 U 的模糊聚类中，第 i 类与第 j 类两类相互覆盖的程度。

考虑两种特殊的情况：

情形 1：分类是清晰的。此时，模糊划分矩阵 U 为如下形式(以 $c = 3$，$n = 7$ 为例)：

$$U=\begin{bmatrix}1&1&0&0&0&0&0\\0&0&1&0&1&0&0\\0&0&0&1&0&1&1\end{bmatrix}$$

相应的 UU^{T} 为：

$$UU^{\mathrm{T}}=\begin{bmatrix}2&0&0\\0&2&0\\0&0&3\end{bmatrix}$$

划分系数为 $F_c = 1$。这说明任意两类之间都不相互覆盖。

情形 2：分类是最模糊的。此时，分类矩阵 R 为如下形式：

$$U=\begin{bmatrix}\frac{1}{c}&\cdots&\frac{1}{c}\\\vdots&\cdots&\vdots\\\frac{1}{c}&\cdots&\frac{1}{c}\end{bmatrix}_{c\times n}$$

相应的 UU^{T} 为：

$$UU^{\mathrm{T}}=\begin{bmatrix}\frac{n}{c^2}&\cdots&\frac{n}{c^2}\\\vdots&\cdots&\vdots\\\frac{n}{c^2}&\cdots&\frac{n}{c^2}\end{bmatrix}_{c\times c}$$

划分系数为 $F_c = 1/c$。这说明任意两类之间都最大程度地相互覆盖。

可以证明：$1/c \le F_c \le 1$。

(2) 平均模糊熵

如果将给定样本集分成 c 类，对应的模糊划分矩阵为 $U = (\mu_{ij})_{c\times n}$，则相应的平均模糊熵为

$$H_c=-\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}\ln(\mu_{ij})$$

H_c 的最小值对应于最佳的聚类数 c^*。

平均模糊熵是借用熵函数的形式来度量对应于某一个模糊划分的模糊性。显然，如果分类是清晰的，则模糊划分矩阵退化为一个硬划分矩阵：$\mu_{ij}\in\{0, 1\}$，此时 $H_c = 0$。

上面提到的划分系数和平均模糊熵这两个有效性函数，只用隶属度来衡量聚类的有效性，而没有考虑到数据集的结构，通常它们有如下几个缺陷：

- 随着聚类数的变化而具有单调性
- 对模糊聚类算法中的加权指数(m)敏感
- 和样本集的数据结构缺少直接联系

而下面有效性函数，则同时兼顾了模糊划分的隶属度和样本集的数据结构。

(3) 变差––分离度

如果将给定样本集分成 c 类，对应的模糊划分矩阵为 $U=(\mu_{ij})_{c\times n}$，聚类中心矩阵为 $V=(v_1, v_2, \ldots, v_c)^{\mathrm{T}}$，$v_i$ 为第 i 类的聚类中心向量，则相应的变差–分离度为

$$V_{\mathrm{W}}=\frac{\mathrm{Var}^N(U,V)}{\mathrm{Sep}^N(c,U)}$$

V_w 的最小值对应于最佳的聚类数 c^*，其中

$$\mathrm{Var}^N(U,V)=\frac{\mathrm{Var}(U,V)}{\mathrm{Var}_{\max}}，\ \mathrm{Var}_{\max}=\max_{c\in\{2,3,\cdots,c_{\max}\}}\mathrm{Var}(U,V)，$$

$$\mathrm{Sep}^N(c,U)=\frac{\mathrm{Sep}(c,U)}{\mathrm{Sep}_{\max}}，\ \mathrm{Sep}_{\max}=\max_{c\in\{2,3,\cdots,c_{\max}\}}\mathrm{Sep}(c,U)$$

我们称 $\mathrm{Var}(U, V)$ 为变差度，其定义如下：

$$\mathrm{Var}(U,V)=\left[\sum_{i=1}^{c}\sum_{j=1}^{n}\frac{\mu_{ij}d^2(x_j,v_i)}{n(i)}\right]\times\left(\frac{c+1}{c-1}\right)^{1/2}$$

这里，$n(i)$ 是第 i 类的样本数，$d(x, y)$ 是一个度量，定义为：

$$d(x, y) = [1-\exp(-\beta\|x-y\|^2)]^{1/2}$$

式中

$$\beta=\frac{n}{\sum_{j=1}^{n}\|x_j-\bar{x}\|^2}，\quad \bar{x}=\frac{1}{n}\sum_{j=1}^{c}x_j$$

变差度所反映的是样本集整体的平均类内分散程度，分类的效果越好，类内的数据越紧密，变差度就越小。

称 $\mathrm{Sep}(c, U)$ 为分离度，其定义如下：

$$\mathrm{Sep}(c,U)=1-\max_{i\neq j}\left\{S(F_i,F_j)\right\}$$

式中

$$S(F_i,F_j)=\max_{x_k\in X}\left\{\min(\mu_{ik},\mu_{jk})\right\}$$

分离度通过模糊集之间的不相似性来描述各个类之间的隔离程度，分类效果越好，分离度越大。

注：最佳聚类数 c^* 的求取，通常是对 $c = 2, 3, \ldots, c_{\max}$ 逐次实施模糊聚类算法，然后根据相应的模糊划分矩阵 U 和聚类中心矩阵 V 计算有效性函数值，最后比较得出最佳的聚类数 c^*。在实际应用中，需要处理的数据集其样本数可能非常大，但真实的分类数并不很大。因而通常取 $c_{\max} = n^{1/2}$，或者根据实际情况将 $c_{\max}$ 限定的更小，以减少算法运行的费用。

例 11　设给定的数据集如表 1 所示，样本点分布的散点图如图 10 所示。

分别使用划分系数、平均模糊熵和变差–分离度对给定数据进行 15 次有效性检验，得到的检验结果如表 2~4。从有效性函数的平均检验值可以看出，给定数据集的最佳聚类数为 2，即从数据的分布结构来

看，分成两类是最合理的。

如果将数据集分成两类，根据 FCM 聚类算法，得到的聚类结果如表 5。

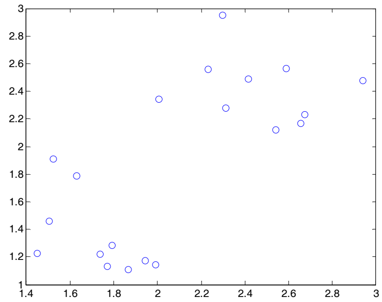


图 10　数据的散点图

表 1　给定的数据集

样本序号	指标 1	指标 2
1	1.9425	1.1715
2	1.7701	1.1307
3	1.7374	1.2188
4	1.8663	1.1055
5	1.9909	1.1414
6	1.5039	1.4570
7	1.6291	1.7881
8	1.7926	1.2811
9	1.4486	1.2248
10	1.5244	1.9089
11	2.0073	2.3431
12	2.5887	2.5630
13	2.5421	2.1189
14	2.6535	2.1690
15	2.3134	2.2789
16	2.2312	2.5568
17	2.4161	2.4856
18	2.2988	2.9522
19	2.6724	2.2319
20	2.9383	2.4787

表 2　划分系数

运行次数	$c = 2$	$c = 3$	$c = 4$	$c = 5$
1	0.8923	0.7636	0.7565	0.7311
2	0.8927	0.8164	0.7565	0.7361
3	0.8927	0.7631	0.7565	0.7359
4	0.8927	0.7631	0.7564	0.7309
5	0.8927	0.8164	0.7565	0.7309
6	0.8927	0.7636	0.7564	0.7309
7	0.8927	0.8165	0.7566	0.7309
8	0.8927	0.8165	0.7627	0.7359
9	0.8927	0.8164	0.7565	0.7309
10	0.8927	0.7631	0.7564	0.7426
11	0.8927	0.8164	0.7565	0.7309
12	0.8927	0.7631	0.7566	0.7316
13	0.8927	0.8164	0.7565	0.7432
14	0.8927	0.8165	0.7564	0.7361
15	0.8927	0.8165	0.7564	0.7361
平均值	0.8927	0.7952	0.7569	0.7343

表 3　平均熵

运行次数	$c = 2$	$c = 3$	$c = 4$	$c = 5$
1	0.2013	0.4228	0.4648	0.5417
2	0.2013	0.3542	0.4648	0.5367
3	0.2013	0.4238	0.4648	0.5369
4	0.2013	0.4236	0.4649	0.5424
5	0.2013	0.3542	0.4648	0.5424
6	0.2013	0.4228	0.4649	0.5422
7	0.2013	0.3540	0.4645	0.5423
8	0.2013	0.3540	0.4939	0.5369
9	0.2013	0.3542	0.4647	0.5423
10	0.2013	0.4237	0.4649	0.5255
11	0.2013	0.3541	0.4648	0.5422
12	0.2013	0.4238	0.4645	0.5400
13	0.2013	0.3542	0.4648	0.5243
14	0.2013	0.3540	0.4649	0.5367
15	0.2013	0.3540	0.4649	0.5367
平均值	0.2013	0.3818	0.4667	0.5379

表 4　变差––分离度

运行次数	$c = 2$	$c = 3$	$c = 4$	$c = 5$
1	0.8652	1.4935	1.1813	1.2612
2	0.9992	1.7244	1.3642	1.4556
3	0.8615	1.2769	1.1768	1.2556
4	0.5836	0.9993	0.7966	0.8512
5	0.9994	1.7128	1.3652	1.4340
6	0.9699	1.6738	1.3132	1.3924
7	0.8647	1.4914	1.1698	1.2408
8	0.9994	1.7254	1.3652	1.4333
9	0.9994	1.4787	1.3520	1.4467
10	0.9995	1.4803	1.3614	1.4292
11	0.9995	1.7141	1.35213	1.4579
12	0.8598	1.2738	1.17469	1.2526
13	0.8614	1.4861	1.17330	1.2552
14	0.9992	1.7134	1.35106	1.4570
15	0.9992	1.7244	1.36540	1.4200
平均值	0.9241	1.5312	1.2575	1.3362

表 5　分成两类的聚类结果

样本序号	1	2	3	4	5	6	7
所属类标号	1	1	1	1	1	1	1
样本序号	8	9	10	11	12	13	14
所属类标号	1	1	1	2	2	2	2
样本序号	15	16	17	18	19	20	
所属类标号	2	2	2	2	2	2	