模糊聚类分析在企业分类上的应用

唐 莉 唐 军
（长沙交通学院） （无锡轻工业学院）

摘 要 运用模糊聚类分析法对企业按经济效益划分等级并归类。模糊集合论的应用涉及社会经济领域的许多方面，文章结合企业实际进行了实证分析。

关键词 模糊聚类分析 评价指标 企业分类

一、引言

聚类分析是研究“物以类聚”的一种方法，是研究分类问题的一种多元统计方法。将模糊集概念用到聚类分析中便产生了模糊聚类分析。由于企业的经济效益要用多个指标构成的指标体系来加以反映，这使得对同行业内的企业进行分类变得困难。企业经济效益的好与差，并没有明确的数量界限，即不存在明确的外延。企业经济效益的表现具有一定的模糊不确定性。解决上述问题可运用模糊聚类分析方法。

二、模型

模糊聚类分析的实质就是根据研究对象本身的属性而构造模糊矩阵，在此基础上根据一定的隶属度来确定其分类关系。

$$A(x)=\begin{cases}1 & x\in A\\ 0 & x\overline{\in} A\end{cases}$$

称为集合 A 的特征函数。如直接根据特征函数来描述企业的经济效益或计划完成程度等是不行的。模糊数学把它推广到〔0,1〕闭区间，即用0和1之间的一个数去度量它，此数称隶属度，当用函数来表示隶属度的变化规律时，就叫隶属函数。即 $0\leqslant A(x)\leqslant 1$

显然隶属度概念是特征函数概念的拓广。特征函数描述空间的元素之间是否有关联，而隶属度描述了元素之间的关联是多少。

模糊聚类分析方法就是用分类集合 X 上各元素之间的模糊关系，将 X 中的元素分成若干类。计算步骤如下：

1. 对原始数据进行变换。通常有标准化变换、极差变换、对数变换等。

2. 计算模糊相似矩阵 $\underset{\sim}{R}$

收稿日期：2001－04－28

$$\widetilde{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & & & \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

其中 r_{ij} 是表示元素 x_i 与 x_j 之间相似程度的量，称为相似系数。r_{ij} 是将经过标准化后的 x_i 与 x_j 的各项统计指标按一定方法进行计算后得到的。计算方法有欧氏距离法、数量积法、夹角余弦法、相关系数法、指数相似系数法等等。

3. 建立模糊等价矩阵。对模糊矩阵进行褶积计算：$R \to R^2 \to R^3 \to \cdots \to R^k$。经有限次褶积后使得 $R^k \cdot R = R^k$。

4. 进行聚类。给定不同的置信水平 λ，求 $\widetilde{R}_\lambda$ 截阵，得普通的分类关系 $\widetilde{R}_\lambda$。

三、实例分析

现以某市六个工业企业 ×× 年的经济效益指标为依据，对六个企业按经济效益的相关程度分类。

设六个企业组成一个分类集合：$X = (X_1, X_2, \cdots, X_6)$，每一个企业经济效益均采用八项统计指标表明，即有 $X_{ij} = (X_{i1}, X_{i2}, \cdots X_{i8})$。这里 X_{ij} 为第 i 个企业的第 j 项指标（$i = 1,2,\cdots,6; j = 1,2,\cdots 8$），这八项经济效益指标分别是：全员劳动生产率（$X_{i1}$）、资金利税率（$X_{i2}$）、产值利税率（$X_{i3}$）、销售收入利润率（$X_{i4}$）、万元产值占用定额流动资金（$X_{i5}$）、万元固定资产提供利税额（$X_{i6}$）、流动资金周转天数（$X_{i7}$）、增加值率（$X_{i8}$）。各企业经济效益指标值见表 1：

表 1　六个企业经济效益指标

指标 \ X_{ij} \ 企业	一厂	二厂	三厂	四厂	五厂	六厂
X_{ij}(万元/人)	7.46	5.29	6.54	6.87	5.51	5.49
X_{i2}(%)	26	29	28	27	21	34
X_{i3}(%)	20	24	23	27	27	29
X_{i4}(%)	11.20	10.84	11.35	12.04	10.19	10.79
X_{i5}(百元)	19.86	26.57	30	29	48	21.68
X_{i6}(百元)	45	50	53	44	46	57
X_{i7}(天)	118	109	144	124	204	84
X_{i8}(%)	38.28	34.12	37.38	36.72	39.83	31.53

对各项统计指标进行标准化处理：

$$X'_{ij} = \frac{X_{ij} - X_{min}}{X_{max} - X_{min}}$$

式中 X_{ij} 是第 i 厂第 j 项经济效益指标的原始数据，X_{min} 和 X_{max} 分别为不同厂的同一项经济效益指标的最小值与最大值。X'_{ij} 为第 i 个企业第 j 项经济效益指标的标准化数值。当 $X_{ij}=X_{min}$ 时，$X'=0$，当 $X_{ij}=X_{max}$ 时，$X'=1$。

按上述公式计算得到六个企业八项经济效益指标的标准化数值如表2:

表2　经济效益指标值的标准化数值表

企业 / X'_{ij} / 指标	一厂	二厂	三厂	四厂	五厂	六厂
X_{i1}	1	0	0.5760	0.7084	0.1014	0.0922
X_{i2}	0.3846	0.6154	0.5383	0.4615	0	1
X_{i3}	0	0.4444	0.3333	0.7778	0.7778	1
X_{i4}	0.5459	0.3514	0.9270	1	0	0.3243
X_{i5}	0	0.2385	0.3603	0.6802	1	0.0647
X_{i6}	0.0769	0.4615	0.6923	0.1540	0.1538	1
X_{i7}	0.2833	0.2083	0.5000	0.3333	1	0
X_{i8}	1	0.5077	0.8935	0.8154	0	0.2012

再根据标准化数值建立各企业之间八项经济效益指标的相似关系矩阵。相似关系矩阵的基本形式为:

$$\tilde{R}=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{16} \\ r_{21} & r_{22} & \cdots & r_{26} \\ \vdots & \vdots & \vdots & \vdots \\ r_{61} & r_{62} & \cdots & r_{66} \end{bmatrix}$$

其中，$r_{ij}\in[0,1]$，$(i=1,2,\cdots,6;j=1,2,\cdots,6)$ 是表示第 i 个企业与第 j 个企业之间在八项经济效益指标上相似程度的量。本例采用夹角余弦来计算 r_{ij}，其计算公式为:

$$r_{ij}=\frac{\sum_{k=1}^{n}X_{ki}X_{kj}}{\sqrt{\sum_{k=1}^{n}X_{ki}^{2}\sum_{k=1}^{n}X_{kj}^{2}}}$$

将表2中的标准化数值代入上述公式，可计算得到六个企业八项经济效益指标的相似关系矩阵 $\tilde{R}$:

$$\tilde{R}=\begin{bmatrix} 1 & 0.58 & 0.84 & 0.78 & 0.15 & 0.33 \\ 0.58 & 1 & 0.88 & 0.76 & 0.47 & 0.87 \\ 0.84 & 0.88 & 1 & 0.52 & 0.47 & 0.68 \\ 0.78 & 0.76 & 0.52 & 1 & 0.55 & 0.54 \\ 0.15 & 0.47 & 0.47 & 0.55 & 1 & 0.35 \\ 0.33 & 0.87 & 0.68 & 0.54 & 0.35 & 1 \end{bmatrix}$$

这是一个模糊相似矩阵，它是进行聚类分析的基础。但它必须满足下列三个条件：

(1) 自反性：即任何一个对象必需和自己同在一类；

(2) 对称性：若对象 a 与对象 b 同类，则 b 与 a 也应同类；

(3) 传递性：如果对象 a 与对象 b 同类，而 b 又与对象 c 同类，则 a 与 c 同类。

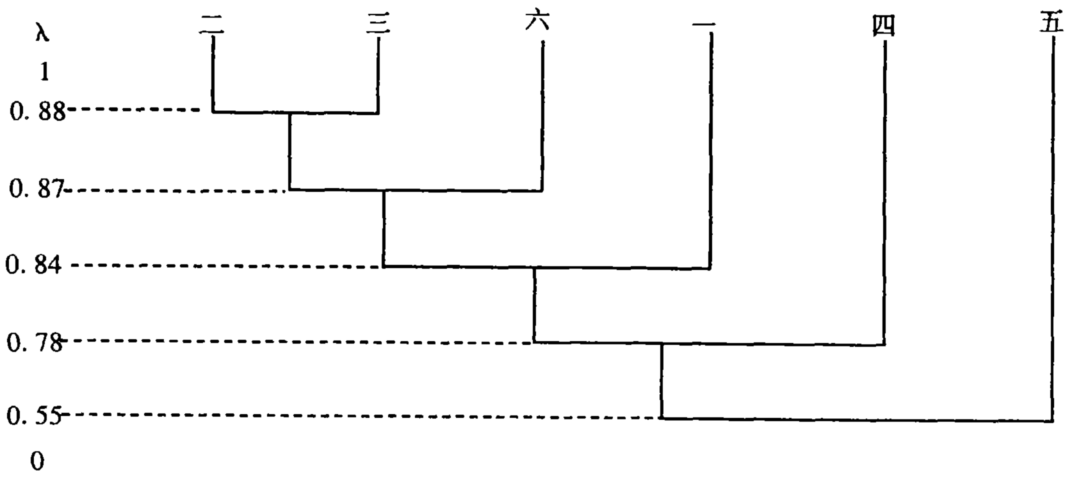
模糊相似矩阵如果同时满足这三个性质，就成为模糊等价关系矩阵，它是进行模糊聚类分析的直接依据。

矩阵 $\tilde{R}$ 满足自反性，即 $r_{ij}(X_i, X_j) = 1(i = j)$，同时具有对称性，即满足 $(r_{ij}) = (r_{ji})$。但是，矩阵 $\tilde{R}$ 不具有传递性，因为 $\tilde{R}^2 \neq \tilde{R}$，这说明要对 $\tilde{R}$ 进行改造。对模糊矩阵进行褶积计算，经计算最后可得到 $\tilde{R}^{16} = \tilde{R}^8 \cdot \tilde{R}^8 = \tilde{R}^8$ 的结果，故 $\tilde{R}^8$ 已具有传递性。它就是模糊等价关系矩阵。

$$\tilde{R}^8 = \begin{bmatrix} 1 & 0.84 & 0.84 & 0.78 & 0.55 & 0.84 \\ 0.84 & 1 & 0.88 & 0.78 & 0.55 & 0.87 \\ 0.84 & 0.88 & 1 & 0.78 & 0.55 & 0.87 \\ 0.78 & 0.78 & 0.78 & 1 & 0.55 & 0.78 \\ 0.55 & 0.55 & 0.55 & 0.55 & 1 & 0.55 \\ 0.84 & 0.87 & 0.87 & 0.78 & 0.55 & 1 \end{bmatrix}$$

依据 $\tilde{R}^8$ 可对六个企业进行分类。对满足传递性的模糊分类关系的 $\tilde{R}^8$ 进行聚类处理，给定不同置信水平的 λ，求 $\tilde{R}_\lambda$ 阵，找出 $\tilde{R}$ 的 λ 显示，得到普通的分类关系。据实例中的模糊等价矩阵的各个元素及分类的要求，按不同的置信水平 λ 对六个企业进行模糊分类，将会得到不同的分类结果。

模糊分类关系经过 λ 截阵的显示，得到聚类谱系图如下：



注：图中数序可为厂也可为企业

上述分类结果中，按 $0.88 < \lambda \leqslant 1$ 进行分类，由于过分强调六个企业在八项经济效益

指标上的差异,而没有注意各指标的相互影响关系,没有真正起到分类的作用,因而不可取。按 $0.55<\lambda\leqslant0.78$ 和 $0<\lambda\leqslant0.55$ 分类又完全忽视了六个企业在经济效益上所表现出的各种差异,分类太粗。按 $0.87<\lambda\leqslant0.88$ 分类,将六个企业分为五类,用集合表示为:{一企},{二企,三企},{四企},{五企},{六企}。按 $0.84<\lambda\leqslant0.87$ 分类,六个企业分为四类,{一企},{二企,三企,六企},{四企},{五企},这说明二企业、三企业、六企业的经济效益状况基本接近,其关联程度为 0.84 ~ 0.87。按 $0.78<\lambda\leqslant0.84$ 分类,六个企业分为三类等等。本例的模糊聚类按 0.87 ~ 0.88,0.84 ~ 0.87,0.78 ~ 0.84 分类较好,不仅将具有相同特征的企业归并到了一块,而且还将有不同特征的企业区分开来。企业的分类利于企业经济效益的比较、排队,为企业经济管理、生产经营活动提供咨询服务,为企业投资决策、产业结构调整提供信息服务。运用模糊聚类分析法可使分类更科学合理,符合实际。

参 考 文 献

1 何晓群.现代统计分析方法与应用.北京:中国人民大学出版社,1998
2 张海波.模糊聚类分析在统计分组中的应用研究,统计与决策,1991(5)
3 何晓亚.企业经济效益综合评价方法及应用.数理统计与管理,1997(7)
4 Zadeh L A. Fuzzy Sets as A Basis for A Theory of Possibility. Fuzzy Sets and Systems,1978