

# 任意分布的模糊多级判别法

陈 治 典

(管理信息科学系)

**摘要** 本方法适用于任意分布的样本数据,能利用更多的因子信息且具有处理模糊信息的能力,可作趋势及定值预报。比通常的逐步判别法适用范围更广,效果更好。

**关键词** 判别分析, 分布函数, 模糊信息

## 0 引言

常用的Bayes意义下的逐步判别分析,要求样本数据满足正态分布及各类总体的协方差阵相等的条件<sup>[1]</sup>,但这对于许多实际问题是难以达到或判定的。另一方面,使用电子计算机对一个预报对象往往可以挑选出一大批因子,可是用通常的逐步判别法只能选用几个线性相关的因子进入判别函数,因而浪费了大量因子的信息。本文针对上述问题,首先对预报对象及初选因子的数据作秩变换,用聚类分析或试探组合的方法将若干个因子组成一组,取其秩数之和作为一新因子;然后利用模糊模式识别和逐步判别分析相结合,建立多级判别函数(即各类从属函数),据此判定预报对象的级别(趋势);最后再采用模糊模式识别法作定值预报。根据文献<sup>[2]</sup>,经过秩变换后,不论原样本数据服从何种概率分布,在引进和剔除因子中用作显著性检验的统计量均近似地服从 $F$ 分布,因而本方法适用于任意分布的样本数据,扩大了多级判别与预测的应用范围,并且由于采用模糊数学方法,又使结果更为合理和有效。通过实例证实了本文的方法比常用的逐步判别法效果好得多。

## 1 具体方法步骤

### 1.1 对样本数据作秩变换并构造秩和因子

该预报对象为 $y$ ,利用计算机由秩相关系数检验法,选出一批跟 $y$ 关系较密切的因子 $x_1, x_2, \dots, x_l$ , $y$ 及各因子都有 $n$ 个历史观测值:

$$y_1, y_2, \dots, y_n; \quad x_{i1}, x_{i2}, \dots, x_{in}, \quad i = 1, 2, \dots, l.$$

本文1989—04—12收到。

将  $y$  的观测值按从小到大的顺序排列定其秩。若因子与  $y$  正相关, 则将其观测值按从小到大的顺序定其秩; 若因子与  $y$  负相关, 则将其观测值按从大到小的顺序定其秩。至于因子的当前值, 就以它跟历史值中最接近的那个数据之秩为其秩 (严格说来应绘出各因子的观测值与其秩的关系图, 由此得出各因子的当前值之秩)。用试探组合法将  $k$  个因子组成一组使其秩数之和跟  $y$  的秩的相关系数尽可能地大 (或用  $R$  型聚类法将彼此间相关较密切的  $k$  个因子归为一组), 取它们的秩数之和作为一个新因子。这样的“秩和”因子共有  $m = l/k$  个, 分别记作  $x_1', x_2', \dots, x_m'$ , 其历史观测值及当前值分别记作  $x'_{i1}, x'_{i2}, \dots, x'_{in}$  和  $x'_{i0}$ ,  $i = 1, 2, \dots, m$ 。每个“秩和”因子都集中了  $k$  个初选因子的信息, 并且由于经过了秩变换, 原初选因子跟  $y$  的非线性关系亦线性化了, 故“秩和”因子跟  $y$  的相关系数一般地都比原选因子跟  $y$  的相关系数有显著提高 (见预报实例)。

## 1.2 对 $y$ 及各“秩和”因子分级并求各级中心

根据实际需要, 将  $y$  的数据适当地分成  $s$  级 (类), 使各级之间有明显差异, 分别记作:

$$Y^{(k)} = \{y_1^{(k)}, y_2^{(k)}, \dots, y_{n_k}^{(k)}\}, \quad k = 1, 2, \dots, s \left( \sum_{k=1}^s n_k = n \right)$$

假定各“秩和”因子的数据均含有对  $y$  分级的信息, 亦即认为各“秩和”的因子数据均来自  $s$  个总体且与  $y$  的各级有一定程度的对应关系 (实际上只是一种模糊关系)。按  $y$  的分级, 将同一级的  $y$  值及对应的各“秩和”因子之值归在一起列成表 1。将  $(x'_{1j}, x'_{2j}, \dots, x'_{mj})$  当作一个样本, 记作  $X_j$ , 其中  $x'_{ij}$  ( $i = 1, 2, \dots, m$ ) 称为该样本的特性指标。表 1 最末一行就是各“秩和”因子的当前值, 即待分类样本, 记作  $X_0$ 。“秩和”因子  $x_i'$  的第  $k$  级的数据用  $x_{ij}^{(k)}$  ( $j = 1, 2, \dots, n_k$ ) 表示。

由表 1 求出各级 (类) 样本的聚类中心作为该类的模式。现以各级的重心为其聚类中心, 记第  $g$  类的重心为

$$\bar{X}^{(g)} = (\bar{x}_1^{(g)}, \bar{x}_2^{(g)}, \dots, \bar{x}_m^{(g)}), \quad (1)$$

其中

$$\bar{x}_i^{(g)} = \frac{1}{n_g} \sum_{k=1}^{n_g} x_{ik}^{(g)}, \quad i = 1, 2, \dots, m \quad (g = 1, 2, \dots, s; \sum_{g=1}^s n_g = n) \quad (2)$$

(见表 1)。又记

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x'_{ij} = \sum_{g=1}^s n_g \bar{x}_i^{(g)}, \quad (3)$$

此即“秩和”因子  $x_i'$  的历史数据之总平均 ( $i = 1, 2, \dots, m$ )。

## 1.3 选择作用显著的“秩和”因子并确定其权重

利用逐步判别分析法<sup>[1]</sup>来引进和剔除因子。记

$$w_{ij} = \sum_{g=1}^s \sum_{k=1}^{n_g} (x_{ik}^{(g)} - \bar{x}_i^{(g)})(x_{jk}^{(g)} - \bar{x}_j^{(g)}), \quad i, j = 1, 2, \dots, m,$$

$$b_{ij} = \sum_{g=1}^s n_g (\bar{x}_i^{(g)} - \bar{x}_i)(\bar{x}_j^{(g)} - \bar{x}_j), \quad i, j = 1, 2, \dots, m,$$

$$t_{ij} = \sum_{g=1}^s \sum_{k=1}^{n_g} (x_{ik}^{(g)} - \bar{x}_i)(x_{jk}^{(g)} - \bar{x}_j), \quad i, j = 1, 2, \dots, m.$$

表 1 求各级样本聚类中心

级别	样本序号	$y$	$x_1'$	$x_2'$	$\dots$	$x_m'$
第 1 级	1(1)	$y_1^{(1)}$	$x_{11}^{(1)}$	$x_{21}^{(1)}$	$\dots$	$x_{m1}^{(1)}$
	1(2)	$y_2^{(1)}$	$x_{12}^{(1)}$	$x_{22}^{(1)}$	$\dots$	$x_{m2}^{(1)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	1( $n_1$ )	$y_{n_1}^{(1)}$	$x_{1n_1}^{(1)}$	$x_{2n_1}^{(1)}$	$\dots$	$x_{mn_1}^{(1)}$
级	平均		$\bar{x}_1^{(1)} = \frac{1}{n_1} \sum_{k=1}^{n_1} x_{1k}^{(1)}$	$\bar{x}_2^{(1)} = \frac{1}{n_1} \sum_{k=1}^{n_1} x_{2k}^{(1)}$	$\dots$	$\bar{x}_m^{(1)} = \frac{1}{n_1} \sum_{k=1}^{n_1} x_{mk}^{(1)}$
第 2 级	2(1)	$y_1^{(2)}$	$x_{11}^{(2)}$	$x_{21}^{(2)}$	$\dots$	$x_{m1}^{(2)}$
	2(2)	$y_2^{(2)}$	$x_{12}^{(2)}$	$x_{22}^{(2)}$	$\dots$	$x_{m2}^{(2)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	2( $n_2$ )	$y_{n_2}^{(2)}$	$x_{1n_2}^{(2)}$	$x_{2n_2}^{(2)}$	$\dots$	$x_{mn_2}^{(2)}$
级	平均		$\bar{x}_1^{(2)} = \frac{1}{n_2} \sum_{k=1}^{n_2} x_{1k}^{(2)}$	$\bar{x}_2^{(2)} = \frac{1}{n_2} \sum_{k=1}^{n_2} x_{2k}^{(2)}$	$\dots$	$\bar{x}_m^{(2)} = \frac{1}{n_2} \sum_{k=1}^{n_2} x_{mk}^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
第 S 级	S(1)	$y_1^{(s)}$	$x_{11}^{(s)}$	$x_{21}^{(s)}$	$\dots$	$x_{m1}^{(s)}$
	S(2)	$y_2^{(s)}$	$x_{12}^{(s)}$	$x_{22}^{(s)}$	$\dots$	$x_{m2}^{(s)}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	S( $n_s$ )	$y_{n_s}^{(s)}$	$x_{1n_s}^{(s)}$	$x_{2n_s}^{(s)}$	$\dots$	$x_{mn_s}^{(s)}$
级	平均		$\bar{x}_1^{(s)} = \frac{1}{n_s} \sum_{k=1}^{n_s} x_{1k}^{(s)}$	$\bar{x}_2^{(s)} = \frac{1}{n_s} \sum_{k=1}^{n_s} x_{2k}^{(s)}$	$\dots$	$\bar{x}_m^{(s)} = \frac{1}{n_s} \sum_{k=1}^{n_s} x_{mk}^{(s)}$
当前值	待报		$x'_{10}$	$x'_{20}$	$\dots$	$x'_{m0}$

易知  $t_{ij} = w_{ij} + b_{ij}$ 。若记  $T = [t_{ij}]_{m \times m}$ ,  $W = [w_{ij}]_{m \times m}$ ,  $B = [b_{ij}]_{m \times m}$ , 则  $T = W + B$ 。T 反映各“秩和”因子数据的总的离差, W 反映类内离差, B 反映类间离差。由 Wilks U 准则, 判别效果希望统计量

$$U = |W|/|T|$$

越小越好。对矩阵 W 及 T 按一般常用的高斯消去法作求解求逆变换<sup>[1]</sup>, 以  $w_{ij}^{(p)}$  及  $t_{ij}^{(p)}$  分别表示对 W 及 T 经 p 次变换后的元素 ( $i, j = 1, 2, \dots, m$ ), 当消去顺序与自然顺序相同时, 即有

$$|W| = w_{11} \cdot w_{22}^{(1)} \cdot w_{33}^{(2)} \cdot \dots \cdot w_{mm}^{(m-1)}, \quad |T| = t_{11} \cdot t_{22}^{(1)} \cdot t_{33}^{(2)} \cdot \dots \cdot t_{mm}^{(m-1)}$$

$$|T| = t_{11} \cdot t_{22}^{(1)} \cdot t_{33}^{(2)} \cdots t_{mm}^{(m-1)}.$$

若经过  $p$  步已引进  $x'_{k1}, x'_{k2}, \dots, x'_{kp}$  入判别函数, 考虑其中是否有因子要剔除, 则计算

$$U_- = \max_{1 \leq l \leq p} \left\{ \frac{t_{kl}^{(p)}}{w_{kl}^{(p)}} \right\} = \frac{t_{k0}^{(p)}}{w_{k0}^{(p)}},$$

检验统计量取

$$F_- = \frac{(1 - U_-)(n - s - p + 1)}{U_-(s - 1)}.$$

若无可剔除的“秩和”因子, 则进一步考虑引进新的“秩和”因子, 计算

$$U_+ = \min_{\substack{k \approx k_l \\ 1 \leq l \leq p}} \left\{ \frac{w_{r0}^{(p)}}{t_{r0}^{(p)}} \right\} = \frac{w_{r0}^{(p)}}{t_{r0}^{(p)}},$$

检验统计量取

$$F_+ = \frac{(1 - U_+)(n - s - p)}{U_+(s - 1)}.$$

文献[2]证明了: 不论原样本数据服从何种概率分布, 当样本数据经过秩变换后, 上述统计量  $F_+$  及  $F_-$  均近似地服从  $F$  分布 (其第一自由度都是  $s-1$ , 第二自由度分别为  $n-s-p$  及  $n-s-p+1$ ), 可用  $F$  检验法决定是否引进或剔除“秩和”因子. 由此可见 (以及采用模糊模式识别法作预报), 本方法适用于任意分布的样本数据.

引进和剔除因子的过程不断进行, 直至既无因子可引进, 又无因子可剔除为止. 设最后引进判别函数的“秩和”因子为  $x'_{k1}, x'_{k2}, \dots, x'_{kr}$ , (假定经过  $P$  步变换), 取

$$C_{ig} = (n-s) \sum_{i=1}^r w_{ik_i}^{(P)} \bar{x}_{k_i}^{(g)}, \quad i = k_1, k_2, \dots, k_r, \quad g = 1, 2, \dots, s \quad (5)$$

作为选出的作用显著的“秩和”因子  $x_{i'}$  ( $i = k_1, k_2, \dots, k_r$ ) 在第  $g$  类判别函数中的权重.

#### 1.4 建立各类的隶属函数和判别法则

记  $r$  维向量空间为  $\mathcal{R}_r$ , 称之为样本空间. 任一样本  $X_i = (x'_{k1j}, x'_{k2j}, \dots, x'_{krj})$  均为  $\mathcal{R}_r$  中的元素, 对于  $y$  的任一级  $Y^{(g)}$ , 能判别它的样本全体应视为  $\mathcal{R}_r$  上的一个模糊子集, 记作  $X^{(g)}$ , 其隶属函数规定如下:

$$\mu_{X^{(g)}}(X_j) = 1 - \frac{\sum_{i=1}^r C_{ig} |x'_{k_ij} - \bar{x}_{k_i}^{(g)}|}{\max \left\{ \sum_{i=1}^r C_{ih} |x'_{k_ij} - \bar{x}_{k_i}^{(h)}| \right\}},$$

$$X_j = (x'_{k1j}, x'_{k2j}, \dots, x'_{krj}) \in \mathcal{R}_r, \quad (g = 1, 2, \dots, s), \quad (6)$$

取  $Y = \bigcup_{g=1}^s X^{(g)}$ , 其隶属函数为

$$\mu_Y(X_j) = \bigvee_{g=1}^s \mu_{X^{(g)}}(X_j), \quad X_j \in \mathcal{R}_r. \quad (7)$$

对于任意给定的样本  $X \in \mathcal{R}_r$ , 若有某个  $k$  使得

$$\mu_{X^{(k)}}(X) = \mu_Y(X), \quad (8)$$

则判断  $X$  所对应的  $y$  属于第  $k$  级 ( $1 \leq k \leq s$ )<sup>[3]</sup>.

### 1.5 预报 $y$ 的趋势及数值

将隶属函数 (6) 中各“秩和”因子的当前值组成的样本 (即待分类样本)  $X_0 = (x'_{k_1 0}, x'_{k_2 0}, \dots, x'_{k_r 0})$  代入式 (6), 求出  $X_0$  隶属于各类 (级) 的隶属度  $\mu_{X^{(g)}}(X_0), g = 1, 2, \dots,$

$s$ . 若有某个  $h (1 \leq h \leq s)$  使得

$$\mu_{X^{(h)}}(X_0) = \mu_Y(X_0) = \max_{1 \leq g \leq s} \mu_{X^{(g)}}(X_0),$$

则判定  $X_0$  属于第  $h$  类, 从而相应地预报  $y_{n+1}$  将属于第  $h$  级, 这就预报了  $y$  的未来趋势.

若还要进一步预报  $y$  的数值, 则将  $y$  的历史资料中凡属第  $h$  级的数据及相应的各选中的“秩和”因子的数据列成表 2. 表 2 的末行填上各选中的“秩和”因子在第  $h$  类隶属函数中的权重.

表 2 第  $h$  级数据及当前样本的隶属函数值

样本序号	$Y^{(h)}$	$x'_{k_1}$	$x'_{k_2}$	...	$x'_{k_r}$	$X_0$ 的隶属函数值
$h(1)$	$y_1^{(h)}$	$x'_{k_1 h(1)}$	$x'_{k_2 h(1)}$	...	$x'_{k_r h(1)}$	$\mu_{X_1^{(h)}}(X_0)$
$h(2)$	$y_2^{(h)}$	$x'_{k_1 h(2)}$	$x'_{k_2 h(2)}$	...	$x'_{k_r h(2)}$	$\mu_{X_2^{(h)}}(X_0)$
...	...	...	...	...	...	...
$h(n_h)$	$y_{n_h}^{(h)}$	$x'_{k_1 h(n_h)}$	$x'_{k_2 h(n_h)}$	...	$x'_{k_r h(n_h)}$	$\mu_{X_{n_h}^{(h)}}(X_0)$
当前值		$x'_{k_1 0}$	$x'_{k_2 0}$	...	$x'_{k_r 0}$	
权系数		$C_{1h}$	$C_{2h}$	...	$C_{nh}$	

对于表 2 中的任一  $y_j^{(h)} (1 \leq j \leq n_h)$ , 能预报它的样本全体应视为样本空间  $\mathcal{H}_r$  上的模糊子集, 记作  $X_j^{(h)}$ , 其隶属函数规定为

$$\mu_{X_j^{(h)}}(X) = 1 - \frac{\sum_{i=1}^r C_{ih} |x'_{k_i} - x'_{k_i h(j)}|}{\max_{1 \leq l \leq n} \left\{ \sum_{i=1}^r C_{ih} |x'_{k_i} - x'_{k_i h(l)}| \right\}},$$

$$X = (x'_{k_1}, x'_{k_2}, \dots, x'_{k_r}) \in \mathcal{H}_r, \quad j = 1, 2, \dots, n_h. \quad (9)$$

取  $Y = \bigcup_{j=1}^{n_h} X_j^{(h)}$ , 其隶属函数为

$$\mu_Y(X) = \bigvee_{j=1}^{n_h} \mu_{X_j^{(h)}}(X), X = (x'_{k_1}, x'_{k_2}, \dots, x'_{k_r}) \in \mathcal{H}_r. \quad (10)$$

若以各选中的“秩和”因子的当前值  $X_0 = (x'_{k_1 0}, x'_{k_2 0}, \dots, x'_{k_r 0})$  代入式 (9) 计算得

$$\mu_{X_{j_0}^{(1)}}(X_0) = \mu_Y(X_0), \quad (1 \leq j_0 \leq n_h),$$

则可预报  $\hat{y}_{n+1} \approx y_{j_0}^{(1)}$ .

## 2 预报实例及效果比较

根据泉州市气象台1959—1985年实测资料,预报惠安县1986年8月中旬降雨量(对象用  $y$  表示)。

利用秩相关系数检验法<sup>[4]</sup>从631个气象因子中选出35个跟  $y$  关系较为密切的因子(为节省篇幅,发表时不列出实测数据),对它们作秩变换,并利用试探最好组合的办法将5个因子组成一组,取其秩数之和构成一个新因子,分别记作  $x_1', x_2', \dots, x_7'$ . 各“秩和”因子及  $y$  的数据见表3,“秩和”因子的组成及其与  $y$  的相关系数见表4. 从表4可见“秩和”因子跟  $y$  的相关系数显著地高于原初选因子跟  $y$  的相关系数。

将  $y$  的观测值分成三级,分级标准是:

第一级  $y_j \leq 1.0 \text{ mm}$ ;

第二级  $1.0 \text{ mm} < y_j \leq 44.0 \text{ mm}$ ;

第三级  $y_j > 44.0 \text{ mm}$ .

据此确定  $y$  的各年数据的级别,列入表3. 将同一级的  $y$  及相应各“秩和”因子的数据归在

表4 “秩和”因子的组成及其与  $y$  的相关系数

“秩和”因子构成情况	“秩和”因子跟 $y$ 的相关系数
$x_1'$ ( $x_{478}, x_{456}, x_{69}, x_{452}, x_{164}$ )	0.9153 (-0.6867, -0.5725, 0.4753, -0.6259, 0.6018)
$x_2'$ ( $x_{390}, x_{418}, x_{427}, x_{248}, x_{402}$ )	0.8282 (-0.6490, -0.4585, 0.4992, -0.4417, -0.6413)
$x_3'$ ( $x_{488}, x_{263}, x_{602}, x_{255}, x_{85}$ )	0.8410 (-0.4303, 0.4270, -0.4110, -0.4698, 0.4414)
$x_4'$ ( $x_{19}, x_{430}, x_{119}, x_{15}, x_{57}$ )	0.8286 (-0.4014, -0.4008, -0.4098, 0.3976, -0.4689)
$x_5'$ ( $x_{115}, x_{86}, x_{167}, x_{91}, x_{63}$ )	0.8525 (-0.3944, 0.3869, -0.3849, -0.3843, -0.5644)
$x_6'$ ( $x_{13}, x_{389}, x_{615}, x_{211}, x_{464}$ )	0.8504 (0.3781, -0.3739, 0.3729, -0.3701, -0.4800)
$x_7'$ ( $x_{81}, x_{161}, x_{621}, x_{139}, x_{88}$ )	0.7918 (-0.3678, 0.3655, 0.3932, 0.3608, 0.4096)

一起,求出各级中各“秩和”因子的平均值  $\bar{x}_i^{(g)}$  ( $i=1, 2, \dots, 7$ ;  $g=1, 2, 3$ ). 利用逐步判别法在信度  $\alpha=0.01$  下选出了两个作用显著的因子  $x_1'$  及  $x_4'$ , 其权重及均值如表5所示. 以这些权重及均值代入式(6)便建立起判别  $y$  的各级的模糊集的隶属函数. 以  $x_1', x_4'$  的历史数据代入各隶属函数再按最大隶属原则(7), (8)便可得  $y$  的相应数据<sup>[5]</sup>预报级别,

表 3 各“秩和”因子及预报对象的数据

年 份	$y$		“秩和因子数据”						
	数据	级别	$x_1'$	$x_2'$	$x_3'$	$x_4'$	$x_5'$	$x_6'$	$x_7'$
1960	54.1	3	98.0	87.5	70.5	74.5	81.5	87.5	79.5
1961	0.0	1	35.5	41.0	59.0	53.0	60.0	57.0	54.5
1962	0.0	1	34.0	23.0	39.5	38.0	35.0	41.0	36.5
1963	0.6	1	33.0	44.0	59.0	53.0	60.5	52.0	69.5
1964	1.3	2	46.5	53.0	60.0	79.5	47.0	55.5	64.0
1965	150.4	3	108.0	114.5	99.0	92.0	93.5	86.5	106.0
1966	118.3	3	95.5	97.0	99.0	77.5	73.0	88.5	80.5
1967	2.9	2	63.0	52.0	79.0	65.5	58.5	59.0	76.5
1968	0.3	1	50.5	71.5	39.5	36.0	48.5	59.5	38.5
1969	8.2	2	88.0	71.5	87.0	65.0	91.0	79.0	86.5
1970	0.1	1	36.0	47.0	48.0	39.0	56.0	46.0	59.0
1971	4.5	2	54.5	42.5	56.5	40.5	63.5	52.0	51.5
1972	266.1	3	113.0	118.0	96.5	87.0	89.0	91.0	107.5
1973	54.8	3	89.5	93.0	100.5	83.5	79.5	81.5	72.0
1974	98.5	3	79.0	63.0	78.5	97.5	98.0	63.5	69.0
1975	6.6	2	84.0	91.0	70.0	81.0	63.0	79.0	64.5
1976	46.7	3	101.5	73.5	80.5	84.5	85.5	92.0	78.5
1977	0.5	1	59.0	55.5	54.0	45.5	50.5	46.5	51.5
1978	43.2	2	72.5	69.0	63.5	75.0	71.5	99.5	73.0
1979	3.1	2	47.0	65.0	60.0	75.5	67.5	78.5	55.0
1980	7.1	2	85.5	67.0	61.0	77.0	67.0	67.5	73.0
1981	0.0	1	36.5	33.5	35.0	50.0	34.0	32.5	46.0
1982	22.5	2	66.0	83.5	72.0	81.0	98.5	82.5	89.0
1983	2.0	2	54.0	69.0	74.5	66.0	57.5	54.0	64.5
1984	0.8	1	44.0	30.5	52.5	63.5	41.0	47.0	51.0
1985	32.8	2	81.0	99.0	55.5	74.0	84.5	76.5	58.0
当前	待 报		51.0	52.5	35.5	25.0	52.0	60.0	55.5

表 6 按本文方法的回报及预报

表 7 按通常逐年判别法的回报及预报

年 份	第一类 从属函 数 值	第二类 从属函 数 值	第三类 从属函 数 值	回报 $y_j$ 级别	$y_j$ 实际 级别	年 份	第一类 函数值 $f_1(X)$	第二类 函数值 $f_2(X)$	第三类 函数值 $f_3(X)$	回报 $y_j$ 级别	$y_j$ 实际 级别	报 对 与 否
1960	0.0000	0.4469	0.6558	3	3	1960	214.8	225.1	237.5	3	3	✓
1961	0.9324	0.6042	0.0000	1	1	1961	310.5	323.0	250.4	2	1	×
1962	0.9207	0.5255	0.0000	1	1	1962	203.5	286.2	250.7	2	1	×
1963	0.9289	0.5949	0.0000	1	1	1963	205.9	205.0	273.1	2	1	×
1964	0.5245	0.5872	0.0000	2	2	1964	270.7	233.4	267.4	2	2	✓
1965	0.0000	0.1654	0.6359	3	3	1965	224.9	240.6	241.8	3	3	✓
1966	0.0000	0.4223	0.7139	3	3	1966	238.5	252.7	247.8	2	3	×
1967	0.6245	0.8466	0.0000	2	2	1967	323.9	342.4	312.2	2	2	✓
1968	0.8853	0.5443	0.0000	1	1	1968	276.0	286.8	264.3	2	1	×
1969	0.1367	0.4570	0.0000	2	2	1969	270.7	267.7	267.7	2	2	✓
1970	0.5313	0.5327	0.0000	1	1	1970	265.2	276.2	258.3	2	1	×
1971	0.8976	0.5764	0.0000	1	2	1971	330.8	347.5	213.9	2	2	✓
1972	0.0000	0.1823	0.7108	3	3	1972	221.7	237.9	243.5	3	3	✓
1973	0.0000	0.3858	0.7810	3	3	1973	211.6	226.9	225.6	2	3	×
1974	0.0000	0.2885	0.3513	3	3	1974	254.2	265.4	262.2	2	3	×
1975	0.0000	0.4779	0.5573	3	2	1975	315.5	334.0	309.7	2	2	✓
1976	0.0000	0.2794	0.9134	3	3	1976	210.5	226.5	231.9	3	3	✓
1977	0.9083	0.6196	0.0000	1	1	1977	300.7	304.1	279.2	2	1	×
1978	0.1175	0.7908	0.0000	2	2	1978	305.8	322.6	298.0	2	2	✓
1979	0.6099	0.7004	0.0000	2	2	1979	299.8	313.2	291.6	2	2	✓
1980	0.0000	0.5358	0.4293	2	2	1980	297.2	315.3	295.9	2	2	✓
1981	0.9639	0.5866	0.0000	1	1	1981	304.8	316.8	297.6	2	1	×
1982	0.0102	0.6519	0.0000	2	2	1982	330.5	353.2	323.3	2	2	✓
1983	0.7139	0.7392	0.0000	2	2	1983	302.4	322.7	300.3	2	2	✓
1984	0.8315	0.7077	0.0000	1	1	1984	315.6	324.0	301.1	2	1	×
1985	0.0000	0.6595	0.1328	2	2	1985	274.9	293.3	277.8	2	2	✓
预报1986	0.8338	0.4998	0.0000	1	1	预测1986	300.3	315.2	292.5	2	1	×



表5 作用显著的因子之权重及均值

选中因子	第一级		第二级		第三级	
	权	重 平均值	权	重 平均值	权	重 平均值
$x_1'$	$C_{11}=0.219$	$\bar{x}_1^{(1)}=41.0625$	$C_{12}=0.364$	$\bar{x}_1^{(2)}=67.4545$	$C_{13}=0.539$	$\bar{x}_1^{(3)}=97.7857$
$x_4'$	$C_{41}=0.436$	$\bar{x}_4^{(1)}=47.3125$	$C_{42}=0.650$	$\bar{x}_4^{(2)}=70.9091$	$C_{43}=0.772$	$\bar{x}_4^{(3)}=85.2143$

见表6。回报准确率达92.31%。以 $x_1', x_4'$ 的当前值(见表3末行)代入各类隶属函数则可判定 $\hat{y}_{1975}$ 将属于第一级(表6末行),即预报惠安县1986年8月中旬降雨量小于1.0mm,跟实际相符。再进一步按式(9), (10)预报 $y$ 之值,可得 $\hat{y}_{1986} \approx y_{1986} = 0.3\text{mm}$ ,跟实况(0.0mm)很接近。

为了对比,用原初选的35个因子按通常的逐步判别法在信度 $\alpha=0.01$ 下建立起各类判别式:

$$f_1(X) = -89.447 + 45.398x_{161} - 15.702x_{164} - 1.262x_{402} + 13.501x_{452},$$

$$f_2(X) = -90.576 + 46.618x_{161} - 14.860x_{164} - 1.663x_{402} + 14.652x_{452},$$

$$f_3(X) = -34.302 + 38.372x_{161} - 10.704x_{164} - 1.167x_{402} + 10.218x_{452}.$$

据此作逐年回报及预报(见表7),回报准确率仅57.69%,预报 $\hat{y}_{1986}$ 属于第二级,跟实况不符。可见本文方法优越得多。

### 参 考 文 献

- [1] 丁士晟,多元分析方法及其应用,吉林人民出版社,(1981).
- [2] Conover, W. J., *Practical Nonparametric Statistics*, Second Edition, (1980).
- [3] 冯德益、楼世博等,模糊数学方法与应用,地震出版社,(1983).
- [4] 朱伯承,统计天气预报,上海科学技术出版社,(1981).

## A Fuzzy Multiple Order Discriminate Method Applicable to the Same Data of Arbitrary Distribution

Chen Zhiqian

(Department of Management Information Science)

**Abstract** This paper presents a method applicable to the sample data of arbitrary distribution. Much more factor information can be utilized and fuzzy information can be treated by this method. Trend and definite value of an object can be forecasted by this method. As compared with the ordinary stepwise discriminate analysis, this is a method of widely use with better results.

**Key words** discriminate analysis, distribution function, fuzzy information