

# 统计图象识别的一种数值方法

徐宏 李洪勳

(空军气象研究所)

## 提 要

本文提出了统计图象识别的一种数值方法,它不需要假定预报对象服从某种分布,而是根据其实际分布利用微机直接建立预报量和预报因子之间的统计关系,识别图象类别。与常用的解析判别方法相比,它提高了预报效果。在这种方法中,由于能较为清楚地看到预报因子对预报量的影响,透明性好,因此对因果关系较为明显的短时预报更适宜。

## 引 言

统计图象识别也称模式识别,是指把几类研究对象构成的总体,根据统计特征进行识别分类。在天气预报中, Fisher 判别, Bayes 判别分析, 以及逐步判别, 即属统计图象识别方法。这类统计方法或假定随机变量呈正态分布, 或使用线性的数学模型, 得出判别函数的解析表达式, 根据函数值的大小判定图象的类别。然而实际上随机变量的分布有多种形式, 预报量与预报因子之间的关系也多为非线性, 上面得到的直线或超平面形式的判别函数难免歪曲, 甚至严重歪曲实际存在的量之间的统计关系。这样, 虽然在建立判别函数时是严格的, 但有时实际使用效果不佳。针对这一问题, 一些非线性统计模式被提出来了<sup>[1]</sup>。在这类模式中由于解系数方程组的困难, 若增加判别函数中的多项式的次数, 则它所能包含的因子数就受到限制。可见如同许多微分方程很难求得解析解一样, 预报量与多种预报因子之间实际存在的统计关系, 用解析的判别函数表示, 常常是困难的。

本文提出一种新的方法, 考虑到与使用判别函数判别的解析方法的区别, 我们称它为统计图象识别的数值方法。它不需要假定预报对象服从某种分布, 而是根据随机变量的实际分布, 研究建立预报量与预报因子间的实际统计关系, 进行判别分类。这里以 Bayes 方法为例, 阐述统计图象识别的数值方法。

## 二、基本方法

设对某一图象 $\xi$ 可以用 $n$ 个有序的数 $x_1, x_2, \dots, x_n$ 来表示, 每一个 $\xi$ 在 $n$ 维空间 $R_n$ 中都对应着一个向量 $\bar{x} = (x_1, x_2, \dots, x_n)^T$ 或一个点。如果已知图象的集合 $B$ 由 $m$ 类互不相容的图象所组成,  $\bigcup_{i=1}^m B_i = B, B_i \cap B_k = \phi$

$$(i \neq k, i, k = 1, 2, \dots, m)$$

$$B_i: \bar{x}_j^{(i)} = (x_{j1}^{(i)}, x_{j2}^{(i)}, \dots, x_{jn}^{(i)})^T \quad (j = 1, 2, \dots, N_i) \quad (1)$$

这里 $N = \sum_{i=1}^m N_i$ , 为 $B$ 中的总样本数。当样本数足够大时, 相应 $B_i$ 类图象出现的频率可近似看作气候概率,

$$P(B_i) \approx \frac{N_i}{N}, \quad \sum_{i=1}^m P(B_i) = 1. \quad (2)$$

如果将每一个图象 $\xi \in B_i$ 看作随机过程, 其条件概率密度分布为 $p(\xi/B_i)$ , 对应 $R_n$ 空间有条件概率密度分布 $p(\bar{x}/B_i)$ , 根据各类图象的样本集和气候概率就可以建立统计识别方法。

### 1. 损失函数

统计图象识别一般总会产生误判, 考虑到识别的重点不同, 各类识别图象的气候概率不同, 误判的后果是不同的。比如在冰雹这类小概率事件预报中, 把冰雹误判为无冰雹和把无冰雹判为有冰雹其损失程度是不同的。为此需要根据识别对象设定损失函数 $H(c(\bar{x}), j)$ 。对于被识别的 $m$ 类图象,  $H$ 为 $m^2$ 个常数。 $h_{ij}$  ( $i, j = 1, 2, \dots, m$ ) 为当 $\xi$ 是

第j类图象, 根据判别函数 $c(\bar{x}) = i$ 被错判为第i类时所带来的损失。如上所述, 若冰雹属 $B_1$ 类, 无冰雹属 $B_2$ 类, 那么应当选择损失函数时取 $h_{12} < h_{21}$ , 也就是说需要着重报出的对象被错判, 损失函数要取大些。当 $i = j$ 即识别正确, 此时 $h_{ij} = 0$ 。

### 2. 判别函数

设在n维特征空间上存在一个以 $\{1, 2, \dots, m\}$ 为值域的函数 $c(\bar{x})$ , 当 $c(\bar{x}) = i$ 时就判决 $\bar{x}$ 所对应的 $\xi \in B_i$ , 则称函数 $c(\bar{x})$ 为判决函数。为此下面的问题是需要把 $R_n$ 空间划分成m个互不相交的子集合 $r_i$  ( $1 \leq r \leq m$ ), 当 $\bar{x} \in r_i$ 时令 $c(\bar{x}) = i$ , 判决 $\xi \in B_i$ , 使平均损失<sup>[2]</sup>为最小。

$$L = \sum_{i=1}^m \int \sum_{j=1}^m h_{ij} P_j p(\bar{x}/B_j) d\bar{x} \quad (3)$$

式中 $P_j$ 为第j类图象的气候概率, 可近似用频率代替。欲使(3)式满足, 被积函数 $\sum_{j=1}^m h_{ij} P_j p(\bar{x}/B_j)$ 的值应比其他(m-

1)个 $\sum_{j=1}^m h_{ij} P_j p(\bar{x}/B_j)$ 都小。于是Bayes方法的判别准则:

$$\sum_{j=1}^m h_{il} P_j p(\bar{x}/B_j) \leq \sum_{j=1}^m h_{ij} P_j p(\bar{x}/B_j) \quad (i, l = 1, 2, \dots, m, l \neq i) \quad (4)$$

判决 $c(\bar{x}) = i$ , 即 $\xi$ 属第i类图象。

### 3. 条件概率密度函数

从公式(4)可知, 进行图象识别需要知道条件概率密度函数 $p(\bar{x}/B_i)$ 。这里从样本的实际分布着手, 求出其所对应的条件概率密度函数。

由样本集(1), 可利用位势函数的性质构造条件概率密度函数

$$p(\bar{x}/B_i, b^2) = \frac{1}{N_i} \sum_{j=1}^{N_i} k(|\bar{x} - \bar{x}_{ij}|, b^2) \quad (5)$$

式中 $k(|\bar{x} - \bar{x}_{ij}|, b^2)$ 是位势函数, 它具有如下性质:  $k(|\bar{x} - \bar{x}_{ij}|, b^2) = k(|\bar{x}_{ij} - \bar{x}|, b^2)$ ; 当 $\bar{x} = \bar{x}_{ij}$ 时 $k(|\bar{x} - \bar{x}_{ij}|, b^2)$ 达极大值; 随着 $|\bar{x} - \bar{x}_{ij}|$ 增大,  $k$ 值单调减小, 直至趋于或等于0。 $b^2$ 为尺度参数。在分析中位势函数选择如下形式,

$$k(|\bar{x} - \bar{x}_{ij}|, b^2) = \frac{1}{(b^2\pi)^{n/2}} \exp\left(-\frac{(X_1 - X_{1j})^2 + (X_2 - X_{2j})^2 + \dots + (X_n - X_{nj})^2}{b^2}\right)$$

这里的 $x_1, x_2, \dots, x_n$ 为经过标准化后的变量, 将上式代入(5)式, 则

$$p(\bar{x}/B_i, b^2) = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{(b^2\pi)^{n/2}} \exp\left(-\frac{(X_1 - X_{1j})^2 + (X_2 - X_{2j})^2 + \dots + (X_n - X_{nj})^2}{b^2}\right) \quad (6)$$

于是 $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(\bar{x}/B_i, b^2) d\bar{x} = 1$ 。

$p(\bar{x}/B_i, b^2)$ 为条件概率密度函数 $p(\bar{x}/B_i)$ 的渐近无偏估计。

由(6)式可知,  $b^2$ 表征的是样本点 $\bar{x}_{ij}$ 对空间上点 $\bar{x}$ 的影响程度。当 $b^2$ 值较小时, 将突出 $\bar{x}$ 附近样本点的作用; 而 $b^2$ 较大时, 远处样本点对 $\bar{x}$ 的作用也不可忽视。它是样本数 $N_i$ 的函数, 当样本图象很多时,  $b^2$ 可以取得较小; 反之当样本较少时, 则 $b^2$ 要取得大些。

对于确定的识别对象, 各类图象的气候概率是已知的, 当给定损失函数 $\{h_{ij}\}$ 时, 就可以根据(4)、(6)式对识别对象分类。

一般说来, 根据(4)、(6)式可以对n维空间上的m类图象进行分类。但是当n增大时, 计算量大大增加。所以, 使用这种方法作预报时, 应注意选择与预报对象密切相关且相互独立的因子作为坐标, 以便在维数较低时, 完成对m类图象识别。具体的步骤可以从n个因子两两组合中根据评分标准选择判别质量最好的一组作为第1、2坐标, 区分出m类图象中肯定区和不定区。然后从(n-2)因子选出与前两个坐标配合较好的第3个坐标, 对不定区进行分类, 类推。这种方法称为序贯判别方法。使用序贯判别方法可减少计算量和所占计算机的内存。

### 三、方法的使用说明

二维问题图示方便直观, 又是序贯判别等常用方法的基础, 这里以两类二维识别问题为例。对于二维问题的图象数值识别问题, 也可以看作是点聚图制作和修改使用的客观

化、自动化问题,是传统点聚图方法的提高。

我们引用汉口雷暴预报方法中的无切变型资料作样本,其中雷暴26天,无雷暴148天,雷暴日气候概率为15%。用凝结高度 $H_c$ 与释放不稳定能量气层顶的气压 $P_h$ 为因子区分未来12小时有无雷暴<sup>[3]</sup>。

在二维问题中,(6)式简化为

$$p(\bar{x}/B_i, b^2) = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{b^2 \pi} \exp \left( -\frac{(x_1 - x_{1j})^2 + (x_2 - x_{2j})^2}{b^2} \right) \quad (7)$$

在工作中,我们对标准化后的资料取网格 $15^\circ \times 15^\circ$ ,试验了 $b^2 = 1-10$ 和各种损失函数的选取。根据(7)式分别计算了识别空格点上有雷暴 $B_1$ 类和无雷暴 $B_2$ 类的条件概率密度分布 $p(\bar{x}/B_1, b^2)$ 、 $p(\bar{x}/B_2, b^2)$ 。图1、2分别给出了两类对象条件概率密度分布形态。其中虚线是 $b^2 = 1$ ,实线是 $b^2 = 5$ 的情况。从这两张图中可以看到,当样本一定时, $b^2$ 取值不同,条件概率密度等值线亦随之变化。当 $b^2$ 小时,可较细微地反映分布的局部特征; $b^2$ 较大时,等值线分布则比较平滑,有利于反映分布的总体特征。

对两类问题, $m = 2$ , (4)式简化为划分两类区间 $r_1$ 、 $r_2$ ,使得满足

$$r_1: h_{12}P_2p(\bar{x}/B_2, b^2) \leq h_{21}P_1p(\bar{x}/B_1, b^2)$$

$$r_2: h_{21}P_1p(\bar{x}/B_1, b^2) < h_{12}P_2p(\bar{x}/B_2, b^2)$$

令

$$d(\bar{x}) = -\ln \frac{h_{21}P_1p(\bar{x}/B_1, b^2)}{h_{12}P_2p(\bar{x}/B_2, b^2)} \quad (8)$$

比较上两式,可知

$$\text{当 } d(\bar{x}) \leq 0, \text{ 则判决 } \bar{x} \in r_1$$

$$\text{当 } d(\bar{x}) > 0, \text{ 则判决 } \bar{x} \in r_2$$

函数 $d(\bar{x})$ 也可称为判决函数,方程 $d(\bar{x}) = 0$ 称为判决边界。它所确定的曲线即为两类识别对象的分界线。但由(7)式可知,一般我们难以得到 $d(\bar{x}) = 0$ 的解析表达式。

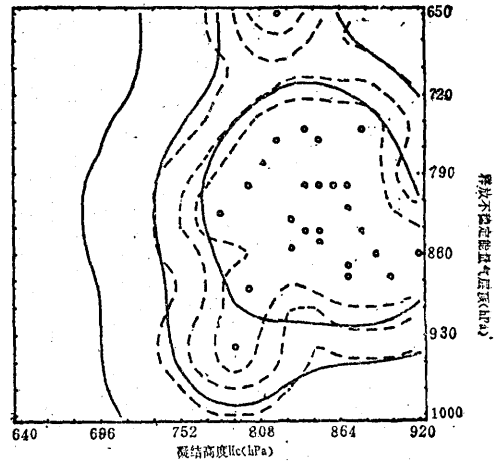


图1 条件概率密度函数 $p(\bar{x}/B_1, b^2)$ 分布  
虚线 $b^2 = 1$ ;实线 $b^2 = 5$ ;  
横坐标是凝结高度 $H_c$ (hPa),纵坐标是释放不稳定能量  
气层顶(hPa)

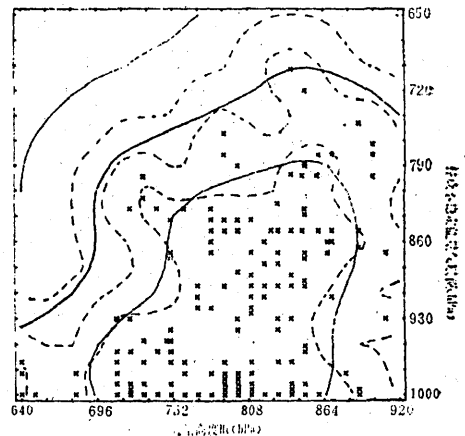


图2 条件概率密度函数 $p(\bar{x}/B_2, b^2)$ 分布  
虚线 $b^2 = 1$ ;实线 $b^2 = 5$

在实际工作中,我们是通过求取各网格点的 $d(\bar{x})$ 值,进而插值得出 $d(\bar{x}) = 0$ 的点,通过样条函数连接而成。图3则是图1和图2两种条件概率密度分布情况,当 $h_{12} = 0.2$ , $h_{21} = 0.8$ 时,由计算机输出的判别分界线,同样图中虚线对应 $b^2 = 1$ ,实线对应 $b^2 = 5$ 。从图中也能看到, $b^2$ 取值不同,对判别边界的形状影响很大。 $b^2$ 值越大,判别边界越平缓;反之 $b^2$ 值较小,则判别边界较复杂。

由(8)式可知,损失函数取法不同,判

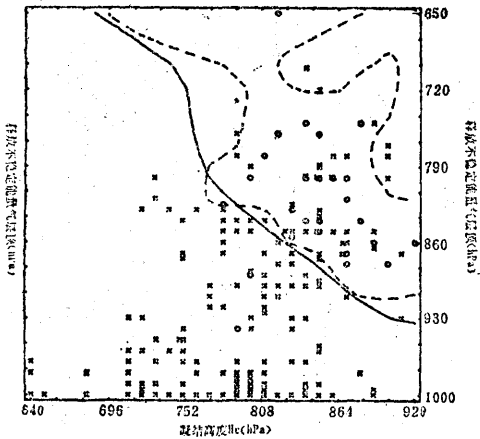


图3  $h_{12} = 0.2, h_{21} = 0.8$ 时的数值判别图  
虚线 $b^2 = 1$ , 实线 $b^2 = 5$

别分界线也不同。实际上, 判别的标准是在给定评分标准情况下, 使错判的平均损失最小, 使评分达最高。当前评分标准尚不统一, 有用准确率, 漏报率, 空报率, CSI等等。表1给出了当 $b^2 = 1$ 时, 损失函数 $h_{12} = h_{21} = 0.5$  (即漏报与空报同等价值, 对应于准确率的评分方法);  $h_{12} = 0.3, h_{21} = 0.7$ ; 和 $h_{12} = 0.15, h_{21} = 0.85$ 时三种数值判别结果。总样本数是174, 天气出现次数为26, 表中的数字为所对应预报结果的次数, 其中括号中的数字分别是所对应的概率。从表中第一行可以明显看到虽然其准确率较高, 但漏报过多。也就是说, 对于雷暴这类小概率事件仅以准确率作为判别标准, 效果并不好。相比之下, 后两行的结果虽然准确率要低一些, 但漏报大为减少。CSI得分明显高于第一行, 显得合理一些。这也说明通过变换损失函数, 可以弥补两类样品量相差较大造成的判别困难。

表1  $b^2 = 1$ 时数值判别结果

$h_{12}$	$h_{21}$	准确(率)	漏报(率)	空报(率)	CSI
0.5	0.5	156 (0.90)	17 (0.65)	1 (0.10)	0.33
0.3	0.7	150 (0.86)	4 (0.15)	20 (0.48)	0.48
0.15	0.85	144 (0.83)	1 (0.04)	29 (0.54)	0.45

分界线的形状可能是多种多样, 我们根据分界线的走向建立了判别各点属于那类

的程序。建立预报方法过程中, 在显示分界线的同时, 还可得出相应评分结果。在损失函数给定后, 调整 $b^2, p(\bar{x}/B_i, b^2)$ 分布变化, 分界线也随之变化。在确定分界线时, 一方面要考虑评分的高低, 另一方面还要考虑使用效果的稳定。一般来说, 当 $b^2$ 值小时, 评分结果比 $b^2$ 值大时好, 但除了在样本点密集的区域分界线可有较大变化外, 一般以较光滑为好。在图3中两条分界线有较大的不同。 $b^2 = 1$ 时的准确率为84%, 漏报率为8%,  $b^2 = 5$ 时分别为78%、12%。然而从分界线的代表性、使用的稳定性来看, 后者可能好些。

对二维问题分界线确定后, 或多维问题 $r_1$ 划定后, 就可进行识别了。在使用时, 根据预报对象的坐标 $\bar{x}$ , 可以得到其所属的类别, 并同时显示预报点附近各类对象的条件概率密度, 供预报员参考。当实况出现后, 预报点即变成历史样本点。在使用过程中, 方法应自我调整, 每用一次就相当于第 $i$ 类图象样本数由 $N_i$ 变为 $(N_i + 1)$ 。由(6)式在 $N_i$ 的基础上算出 $(N_i + 1)$ 个样本的 $p(\bar{x}/B_i, b^2)$ 的分布。进而修正分界线或区域 $r_1$ 的划分。这种调整方法称自适应方法。

#### 四、与解析的判别分析方法的比较

天气现象预报大多可归结为对预报对象分类问题。为了得到判别函数的数学表达式, 在判别分析中总要做一些假设, 使问题简化。这里以Bayes判别为例对解析判别方法与数值判别方法作一分析。

在Bayes判别中, 为了能够得到判别函数解析表达式, 通常假定各类识别对象的条件概率密度函数都是多元正态分布。

$$p(\bar{x}/B_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|} \exp$$

$$\left[ -\frac{1}{2} (\bar{x} - \mu^{(i)})^T \Sigma_i^{-1} (\bar{x} - \mu^{(i)}) \right] \quad (9)$$

其中 $\mu^{(i)}$ 是随机向量 $\bar{x}$ 在 $B_i$ 类中的数学期望,  $\Sigma_i$ 是 $\bar{x}$ 的协方差矩阵;  $|\Sigma_i|$ 为 $\Sigma_i$ 行列式。在实际工作中, 我们以样本资料直接估计出

正态分布的数学期望、协方差矩阵等参数后，就可直接应用判别公式进行判别。但是当各类总体协方差矩阵不同时，计算是复杂的。业务工作中使用的线性判别函数是在各类协方差矩阵相同的假定下得到的<sup>[4]</sup>。实际上将(9)式代入(8)式，即有：

$$d(\bar{x}) = -\ln \frac{p(\bar{x}/B_1)}{p(\bar{x}/B_2)} + C$$

$$= \frac{1}{2}(\bar{x} - \mu^{(1)})^T \Sigma^{-1}(\bar{x} - \mu^{(1)}) - \frac{1}{2}(\bar{x} - \mu^{(2)})^T \Sigma^{-1}(\bar{x} - \mu^{(2)}) + C$$

$$= \bar{x} \Sigma^{-1}(\mu^{(2)} - \mu^{(1)}) - (\mu^{(1)} + \mu^{(2)})^T \Sigma^{-1}(\mu^{(2)} - \mu^{(1)}) + C \quad (10)$$

显然它是  $\bar{x}$  的分量  $x_1, x_2, \dots, x_n$  的线性函数，同样当  $n > 2$  时，判决边界就是  $n$  维空间的超平面。仍以上面雷暴预报的样本为例，在两类对象为正态分布，且协方差矩阵相同的假定下，计算其条件概率密度。我们取两类对象相应的协方差算术平均作为公共的协方差。图4中给出两类对象的等概率密度分布。由图可见，两类曲线在平面上的形状完

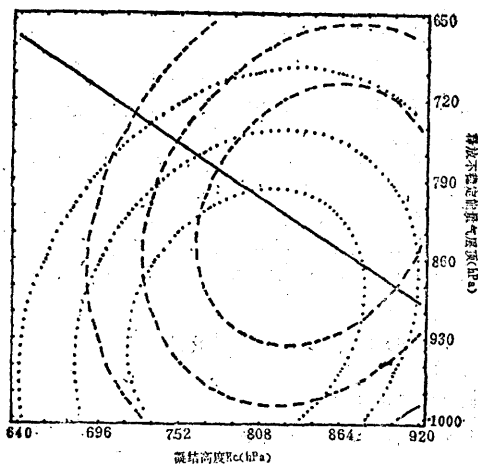


图4 协方差相同时，正态分布的条件概率密度和判别分界线

点线为等  $p(\bar{x}/B_1)$  线；虚线为等  $p(\bar{x}/B_2)$  线；实线为判别边界（下同）

全相同，由于它们的数学期望不同，其大值中心不同，一类分布曲线亦可看作是另一类平移得到。图中的判别分界线是取损失函数  $h_{12} = 0.3, h_{21} = 0.7$  得到的直线。与图1、2、

3 比较，其分布可看作是实际分布的一种简化，值得注意的是对 1 类对象歪曲较大。

我们若只保留正态分布的假定，计算了协方差矩阵不同的情况。这时，两类对象的条件概率密度等值线及分界线结果如图5所示。对比图4，两组闭合线的长轴走向已不一致，相应的分界线也由直线变为二次曲线。这条分界线的解析表达式为

$$d(\bar{x}) = -\frac{1}{2} \ln \frac{|\Sigma_2|}{|\Sigma_1|} + \frac{1}{2}(\bar{x} - \mu^{(1)})^T \Sigma_1^{-1}(\bar{x} - \mu^{(1)}) - \frac{1}{2}(\bar{x} - \mu^{(2)})^T \Sigma_2^{-1}(\bar{x} - \mu^{(2)}) + C \quad (11)$$

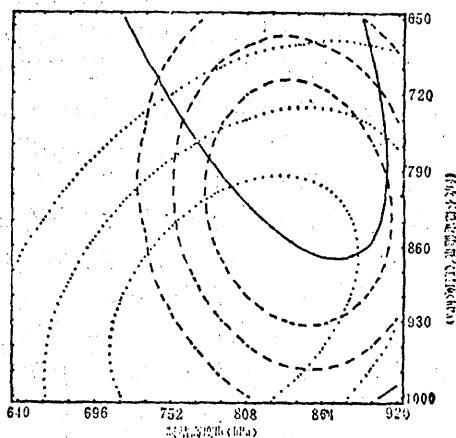


图5 协方差不同时，正态分布的条件概率密度和判别分界线

它已是一个二次函数了，一般地是一个二次超曲面。可见协方差矩阵相同与否对判别边界影响很大。对比(8)式可知，(10)式、(11)式中

$$C = \ln \frac{h_{12} p_2}{h_{21} p_1}$$

这里  $p_1, p_2$  是已知的，改变  $h_{12}, h_{21}$  值能够得到一族判别边界。这簇曲线相互平行。因此，通常在解析判别中，不引用损失函数的概念，通过改变常数项，平移边界线，选取适当的分界线。对比图2、3，图5中相应的等密度线和分界线与其比较接近，是实际分布的较好近似。这是由于本例两类对象实际分布近于正态分布。表2是损失函数取  $h_{12} = 0.3, h_{21} = 0.7$  时，两种正态分布和

$b^2 = 1$ ,  $b^2 = 5$  的数值判别结果。由表可知, 协方差不同时 ( $\Sigma_1 \neq \Sigma_2$ ) 的效果优于协方差相同 ( $\Sigma_1 = \Sigma_2$ ), 而与  $b^2 = 5$  的数值判别结果接近。

表 2  $h_{12} = 0.3, h_{21} = 0.7$  时, 不同假定下的判别结果

	准确 (率)	漏报 (率)	空报 (率)	CSI
$\Sigma_1 = \Sigma_2$	141 (0.81)	11 (0.42)	22 (0.59)	0.31
$\Sigma_1 \neq \Sigma_2$	139 (0.80)	7 (0.27)	28 (0.60)	0.35
$b^2 = 5$	141 (0.81)	7 (0.27)	26 (0.58)	0.37
$b^2 = 1$	150 (0.86)	4 (0.15)	20 (0.48)	0.48

当各类对象的条件概率密度与正态分布差别较大时, 正态分布的假设便不成立。我们利用文献(1)表 1 中应用 logistic 判别分析所用的资料作了计算。其中  $x_1$  是 2 月下旬到 6 月上旬内最高气温  $\geq 35^\circ\text{C}$  的日数;  $x_2$  是 6 月下旬内最高气温  $\geq 35^\circ\text{C}$  的日数, 判别 7 月下旬最高气温  $\geq 35^\circ\text{C}$  的日数大于等于 2 天与小于 2 天问题。我们分别计算了按正态分布协方差矩阵相同, 正态协方差矩阵不同, 和数值方法三种情况下的等概率密度曲线以及损失函数相同时的判别分界线 (图略), 三条线的预报准确率分别为 80%, 80% 和 100%。两类对象的实际分布与正态分布差别较大。这时利用解析方法判别, 不论协方差矩阵相同与否, 其判别结果都明显不如数值方法识别结果。

由此可见, 当样本实际分布与解析判别分析的假设接近时, 解析判别方法与数值方

法求得的结果较接近。反之, 当实际分布与假设差异较大时, 其结果必然有显著差异。还需指出不求解系数的方程组, 应用本文的计算方法, 也可得到与解析判别方法相同的判别边界。即图 4、5 中的分界可根据解析表达算出, 也可由数值方法准确地绘出。因此, 就解决实际问题而言, 应用微机作统计图象的数值识别是较为有效的判别方法。

## 五、结 语

根据预报业务的需要, 考虑到计算机的应用, 本文提出了统计图象识别的一种数值方法。这种方法没有解析判别方法中的那样一些先决条件的限定, 扩大了判别分析方法的应用范围。它从样本的实际分布出发, 判别界限的划分比较切合实际。与解析方法相比, 判别效果较好。在解析判别方法中, 一些计算过程常常掩盖各因子以及样本点对判别界限的具体影响, 而数值方法对这种影响常常可直观分析, 透明性好, 对失误的检查, 方法的修改比较方便。因此, 这一方法对天气现象预报特别是因果关系明显的短时预报有实际意义。当然, 解析判别方法也在发展, 现存的问题也是可望改进的。

## 参 考 文 献

- (1) 吕纯谦, 陈舜华, 二次 logistic 判别分析及其气象应用, 气象学报, Vol. 44, No. 3, 1986。
- (2) 程民德等, 图象识别导论, 上海科学技术出版社, 1983。
- (3) 肖桂珍, 李洪勤, 汉口王家墩七、八月雷暴预报方法研究, 航空气象, 1978, 5, No. 1。
- (4) 屠其璞等, 气象应用概率统计学, 气象出版社, 1984。

# A numerical method of statistical pattern recognition

Xu Hong                      LiHongji

(Air Force Meteorological Research Institute)

## Abstract

A numerical method of statistical pattern recognition is proposed. it needn't assume that the predictand meets the requirement of certain distribution. A statistical relationship between the predictand and the predictor was obtained directly by computer from its practical distribution, and then the type of pattern was recognized. Compared with the usual analytic discriminatory method, the Predictive effectiveness is improved. In this method, the influence of transparency is better. Therefore, it is better to be used in the very short-range forecast, in which the relationship between cause and effect is more obvious.