第12卷 第1期 自动化学报 Vol. 12, No. 1
1986年1月 ACTA AUTOMATICA SINICA Jan., 1986

模式识别的统计模糊方法和模糊统计方法

陶笃纯
(中国科学院声学研究所)

摘　　要

本文在论述模式识别的统计方法和模糊方法的共同性、差异以及各自适用范围的基础上，研究了模式识别的统计模糊方法和模糊统计方法. 统计模糊方法是在模糊分类器中充分利用模式分量统计信息的隶属函数，使分类性能优于普通的模糊分类器. 模糊统计方法是在以统计方法为基础的分类器中，用模式分量的模糊隶属函数代替模式分量作为分类器输入. 从对本文中几个数据集所作的分类试验结果看，这种方法只需要不大的训练样本集便可使分类性能接近于 Bayes 分类器的最佳水平.

一、引　　言

模式识别的统计方法已经得到充分的发展，模糊方法也开始在模式识别的许多应用中取得了成功. 事实上，模式识别的统计方法和模糊方法在本质上是一致的，它们有相同的模式分类器模型. 假定输入模式 $X=[x_1\cdots x_l\cdots x_d]^T$ 由 d 个分量组成，来自 K 个类别 $C_i(i=1,2,\cdots,K)$，则模式分类器的基本模型可用图1表示. 所不同的只是鉴别函数 $g_i(X)$. 当使用 Bayes 假设检验的统计方法时，鉴别函数是各个类别的后验概率. 当采用模糊方法时，鉴别函数则是对于各个类别的隶属函数.

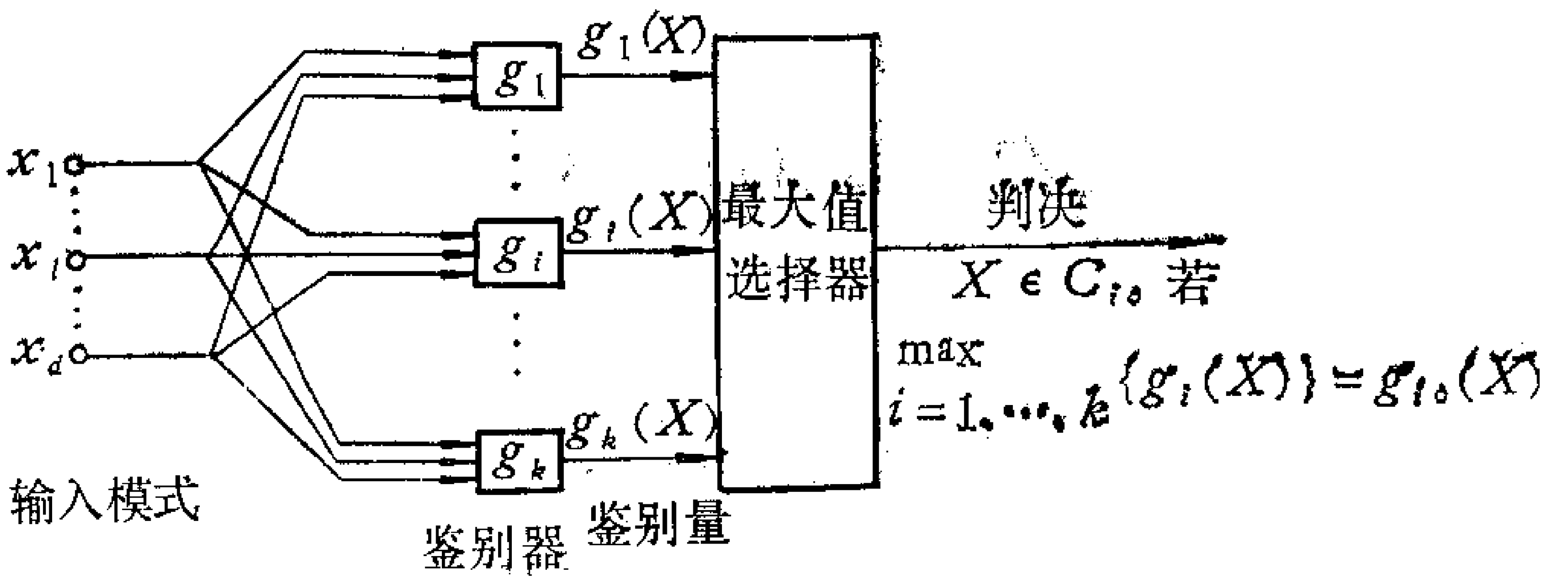

图1　模式分类器的基本模型

计算后验概率需要知道每一类各个分量的概率分布以及每一类的先验概率. 因此使用 Bayes 统计方法的先决条件，是事先掌握这些统计信息，或者有大的训练样本集可用，能够通过学习得到所需的统计知识. 如果不具备上述条件，在对各类模式只有大致了解，

本文于1983年12月22日收到.

或者训练集很小的情况下，用模糊方法往往能取得不坏的效果. 当然这还要取决于隶属函数形式的选择和其中参数的估计.

模糊方法的分类性能比不上 Bayes 统计方法[1,2,3]. 但是，模糊方法可以在对各类输入模式的统计知识了解得很少的场合使用，这是它的一个重要优点.

统计方法和模糊方法各有所长，且有共同的基础，所以把这两种方法结合起来，发展模式识别的统计模糊方法和模糊统计方法，在适当场合分别发挥统计和模糊这两种方法的长处，以达到提高分类性能的目的.

本文将结合具体的识别问题来研究新的识别方法. 问题是要根据机器振动谱，选定若干特征来判断机器的工作状态是否正常. 在基本数据集 I 中，待识别模式来自四个类别 C_1, C_2, C_3 和 C_4，其中 C_4 为正常状态；C_1, C_2 和 C_3 分别为三种不正常状态. 每个模式矢量由与五个特征相对应的五个相互独立的分量组成. 每个分量都有正态分布，它们的参数列于表1. 在每一格中，前一个数代表相应分量的均值 $\bar{x}$，后一个数则代表其标准偏差 σ，单位均为分贝. 由表可见，三种异常状态都是某个或某几个特征数值超过了正常值.

表 1 基本数据集 I 的参数

$\bar{x},\sigma(dB)$ 分量 / 类别	No. 1	No. 2	No. 3	No. 4	No. 5
C_1	10,6	16,10	8,5	22,16	4,1
C_2	10,5	12,7	14,10	17,13	6,2
C_3	16,12	12,7	8,4	17,13	8,4
C_4	10,5	12,7	8,4	17,13	4,0

本文还使用了两个辅助数据集 (II、III). 数据集 II 与数据集 I 一样，也是四类五特征问题的数据集，只是参数不同. 数据集 III 则是一个三类三特征问题的数据集.

二、以 Bayes 统计方法为基础的隶属函数及统计模糊分类方法

1. 模糊分类方法

如果只知道正常状态下诸模式分量的均值以及各个异常状态分别有哪几个特征不正常(大于正常值)，而对异常值比正常均值大多少以及各分量的概率分布均不了解，这时显然无法利用 Bayes 统计方法进行分类，但可以采用模糊方法.

假定每一个模式分量对其异常状态模糊子集的隶属函数为

$$Y_l(x_l) = \frac{1}{1+A_l/(x_l/x_{0l})^2}. \tag{1}$$

式中 x_{0l} 为第 l 个模式分量的正常均值；x_l 为输入模式的第 l 个分量；A_l 为参数. 相应的正常状态模糊子集的隶属函数为

$$Z_l(x_l) = 1 - Y_l(x_l). \tag{2}$$

输入模式矢量 X 对于各类的隶属函数取下列形式：

$$\mu_i(X) = \sum_{l=1}^{d} w_{il} U_l(x_l), \quad i = 1, \cdots, K. \tag{3}$$

而

$$\sum_{l=1}^{d} w_{il} = 1. \tag{4}$$

其中 $U_l(x_l)$ 为输入模式第 l 个分量的隶属函数，第 i 类的第 l 个分量正常时，取 $U_l(x_l) = Z_l(x_l)$；否则取 $U_l(x_l) = Y_l(x_l)$。w_{il} 为权重系数。

模糊分类器计算输入模式矢量 X 对于各类的隶属函数 $\mu_i(X), i = 1, \cdots, K$，判决规则为

如果

$$\max_{i=1,\cdots,K} \mu_i(X) = \mu_{i0}(X), \text{ 则判决 } X \in C_{i0} \tag{5}$$

2. 以 Bayes 统计方法为基础的模糊隶属函数

在某些场合，对于几个模式分量的统计性质有充分的先验知识，而对另一些分量则了解得很少。这时，因不满足设计 Bayes 分类器的要求，就整体而言，只能采用模糊方法，但在构造分量隶属函数时，不能象(1)式那样很少利用已有信息，而应设法充分利用已知信息。现在的关键问题，是要寻找一种既满足模糊隶属函数的基本要求，又充分利用关于模式分量先验知识的分量隶属函数形式。

用 $Z(x)$ 表示模式分量 x 对于其正常状态模糊子集的隶属度，用 $Y(x)$ 表示模式分量 x 对其异常状态模糊子集的隶属度，恒有 $Y(x) + Z(x) = 1$。假定能得到模式分量 x 对于正常和异常状态的后验概率 $P_1(x)$ 和 $P_2(x)$，则可以采用下列形式的隶属函数：

$$Z(x) = \frac{2^{-P_2(x)/P_1(x)}}{2^{-P_2(x)/P_1(x)} + 2^{-P_1(x)/P_2(x)}}, \tag{6-1}$$

$$Y(x) = \frac{2^{-P_1(x)/P_2(x)}}{2^{-P_1(x)/P_2(x)} + 2^{-P_2(x)/P_1(x)}}. \tag{6-2}$$

这种隶属函数利用了比(1)式多得多的有关模式分量的统计信息，且符合对模糊隶属函数

表 2　数据集 I 各种分类程序的性能

程　　序	所用隶属函数	平均正确识别率 RT(%)
模糊分类程序	5×(1) ($A_l = 2$)	88.70
SF1	3×(1) ($A_l = 2$) 2×(6)	89.15
SF2	2×(1) ($A_l = 2$) 3×(6)	91.20
SF3	1×(1) ($A_l = 2$) 4×(6)	93.40
SF4	5×(6)	95.90
Bayes 统计分类程序		96.84

的基本要求.

2. 统计模糊分类方法及其性能

对有条件的分量采用(6)式所示的隶属函数,对不具备条件的分量采用(1)式所示的隶属函数,按(3)式构成各个类别的隶属函数(暂时只考虑等权情况).这就是所谓统计模糊分类方法.

在总共五个模式分量中,让其中的 n 个采用(6)式分量隶属函数,其余 $(5-n)$ 个采用(1)式分量隶属函数,随 n 的不同构成 SF1 到 SF4 四个分类程序.表 2 中列出了模糊分类程序、四个 SF 程序以及 Bayes 统计分类程序(每类用 500 个,总共用 2000 个检验样本,算出混淆矩阵及平均正确识别率)对数据集 I 的主要分类性能.不难看出,随着使用(6)式所示隶属函数的分量数目的增加,分类性能逐步提高.

为验证这种方法的可靠性,用上述各种程序对数据集 II 和 III 进行了分类试验,所得结果与表 2 结果有相同的规律.

三、模糊统计分类方法

1. 模糊统计分类方法

最小均方误差意义上的最小距离分类器使用下列鉴别函数:

$$g_i(X) = w_{i1}x_1 + w_{i2}x_2 + \cdots + w_{id}x_d - w_{id+1}. \quad i = 1, \cdots, K. \tag{7}$$

其中权重系数 w_{il} 按总的均方映射误差最小的准则由训练样本学习而得[4].不难发现,(7)式的统计分类器鉴别函数与(3)式的模糊分类器隶属函数有类似的线性叠加形式,主要差别在于前者是模式分量本身的加权和,而后者是模式分量隶属函数的加权和.这使笔者想起,可以利用模糊隶属函数对模式性质的猜测与估计,同时采用统计分类器在一定最佳准则下对权系数矩阵的学习结果,达到减少训练样本数目和提高分类性能的目的.

这相当于用诸模式分量的隶属函数代替模式分量本身作为分类器的输入,也就是让每个模式分量 x_l 在输入(直接的)最小均方误差意义上的最小距离分类器之前,都先经过一相应于模糊隶属函数的非线性变换 f_l,如图 2 所示.称这种分类方法为模糊统计方法.

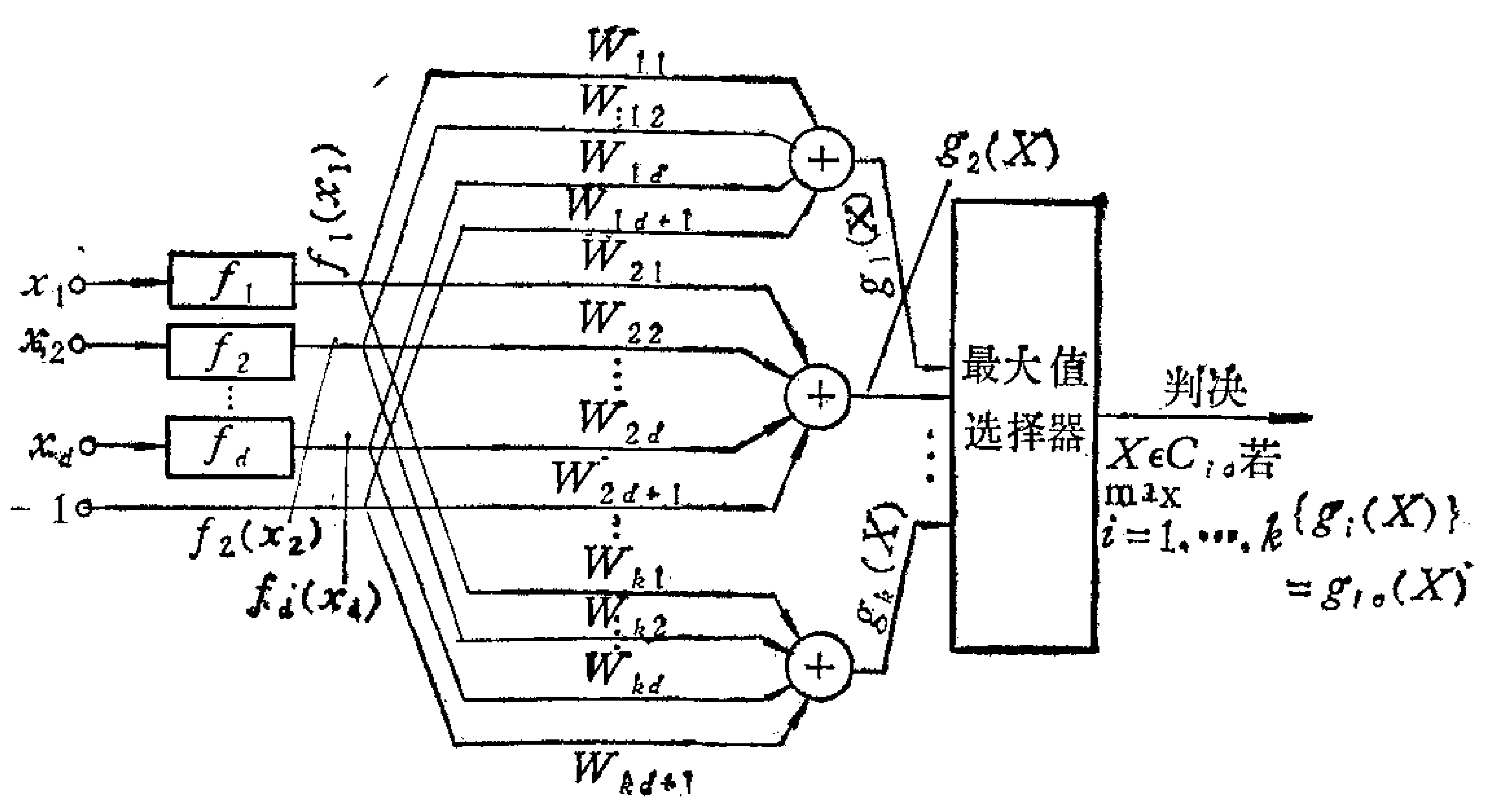

图 2 经过非线性变换的 K 类最小均方误差意义上的最小距离分类器

选用如下具有分量隶属函数意义的变换函数：

$$f_l(x_l)=\frac{1}{1+A/(x_l/x_{0l})^{E_l}},\quad l=1,2,\cdots,d. \tag{8}$$

其中自变量 x_l 是输入模式矢量的第 l 个分量； x_{0l} 是第 l 个模式分量在正常状态下的平均值；A 和 E_l 都是参数．要求

$$f_l(x_{1l})=\frac{1}{1+A/(x_{1l}/x_{0l})^{E_l}}=\alpha\quad(0.5<\alpha<1), \tag{9}$$

$$f_l(x_{0l})=\frac{1}{1+A}=1-\alpha. \tag{10}$$

(9)式中 x_{1l} 是第 l 个模式分量在异常状态下的平均值．由(9)和(10)式可以确定(8)式中的两个参数：

$$A=\frac{\alpha}{1-\alpha}, \tag{11}$$

$$E_l=\frac{2\log\dfrac{\alpha}{1-\alpha}}{\log(x_{1l}/x_{0l})}. \tag{12}$$

合适的 α 值需经学习后确定． 通过训练样本估计该分量在正常和异常状态下的均值 x_{0l} 和 x_{1l} 后，就能确定变换式中的参数．

2. 学习过程及分类性能

图 2 分类器的学习过程由以下三步组成：(1)将训练集既作为训练集又作为检验集，用试探法寻找最佳的 α 值，最佳准则是平均正确识别率 RT 最大．(2)用训练集估计 x_{0l}、x_{1l}，并求出 A 和 E_l．(3)估计权重系数矩阵．

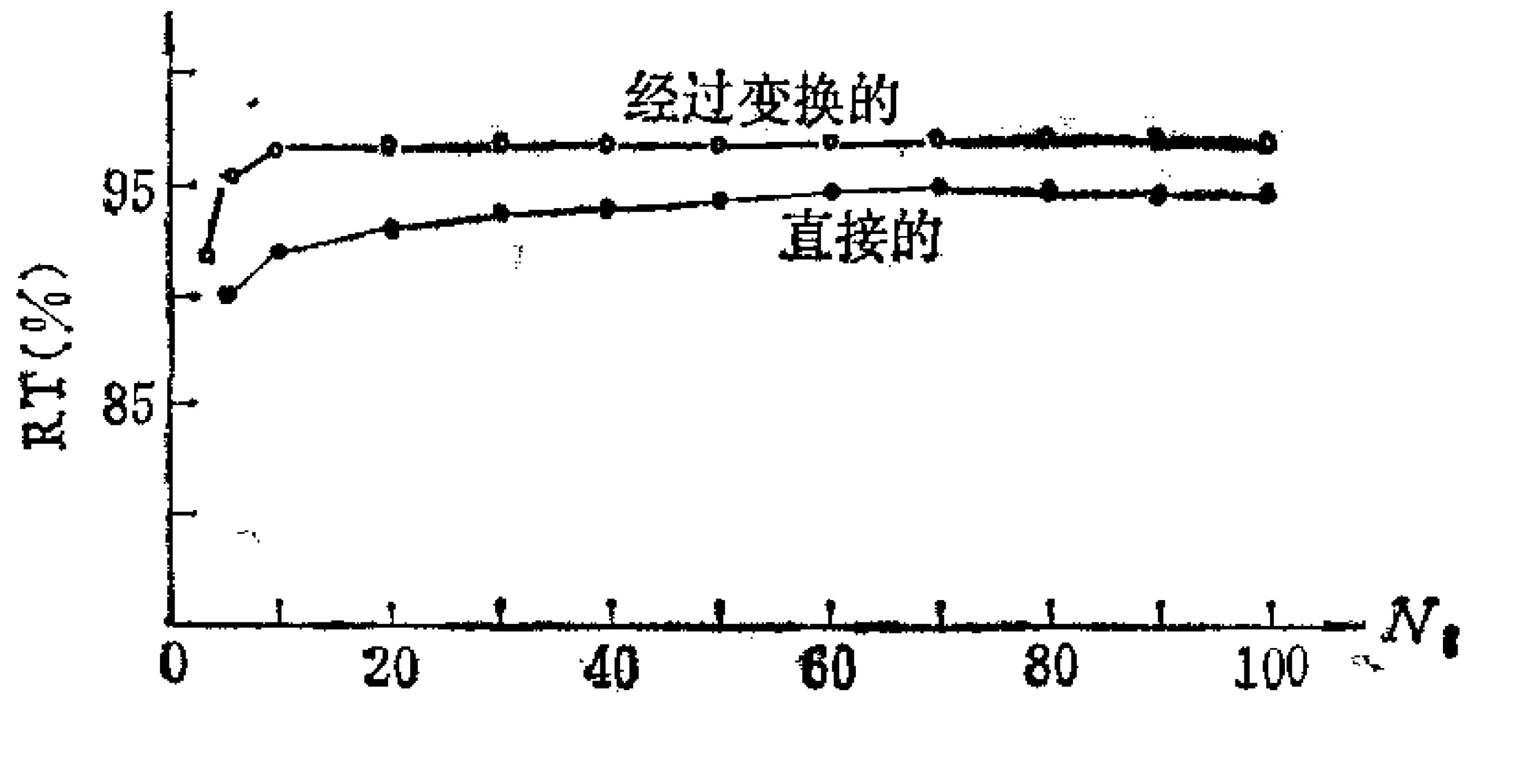

图 3　对数据集 I 两种分类器的性能与训练样本数的关系

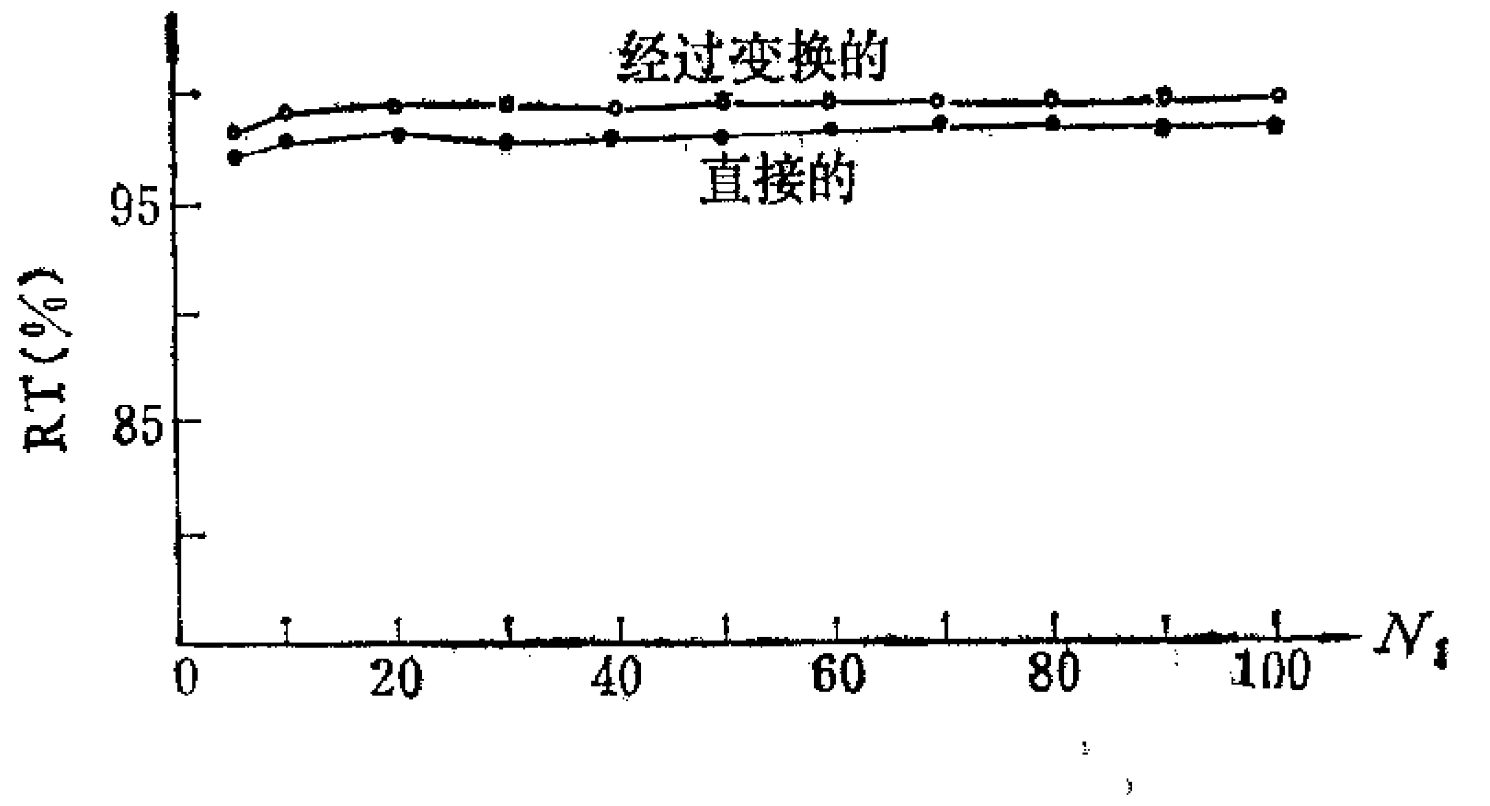

图 4　对数据集 II 两种分类器的性能与训练样本数的关系

对通过学习确定了各个参数值的分类器,用数据集 I 中每类 500 个样本检验其性能.所得平均正确识别率 RT 与训练集中每类样本数 N_1 的关系如图 3 所示. 作为比较,将前述直接的最小均方误差意义上的最小距离分类器的性能曲线也画在同一图上. 由图 3 可见,对于经过变换的分类器,(1)当训练样本数达到每类 $N_1=10$ 时,RT 就已达到稳定状态;(2) RT 的稳定值为 96.66%,非常接近于最佳的 Bayes 分类器所达到的性能极限(96.84%).

用数据集 II 和 III 进行检验的结果见图 4 和图 5. 两者变换后的性能都优于直接法的结果,且接近各自 Bayes 分类器的水平.

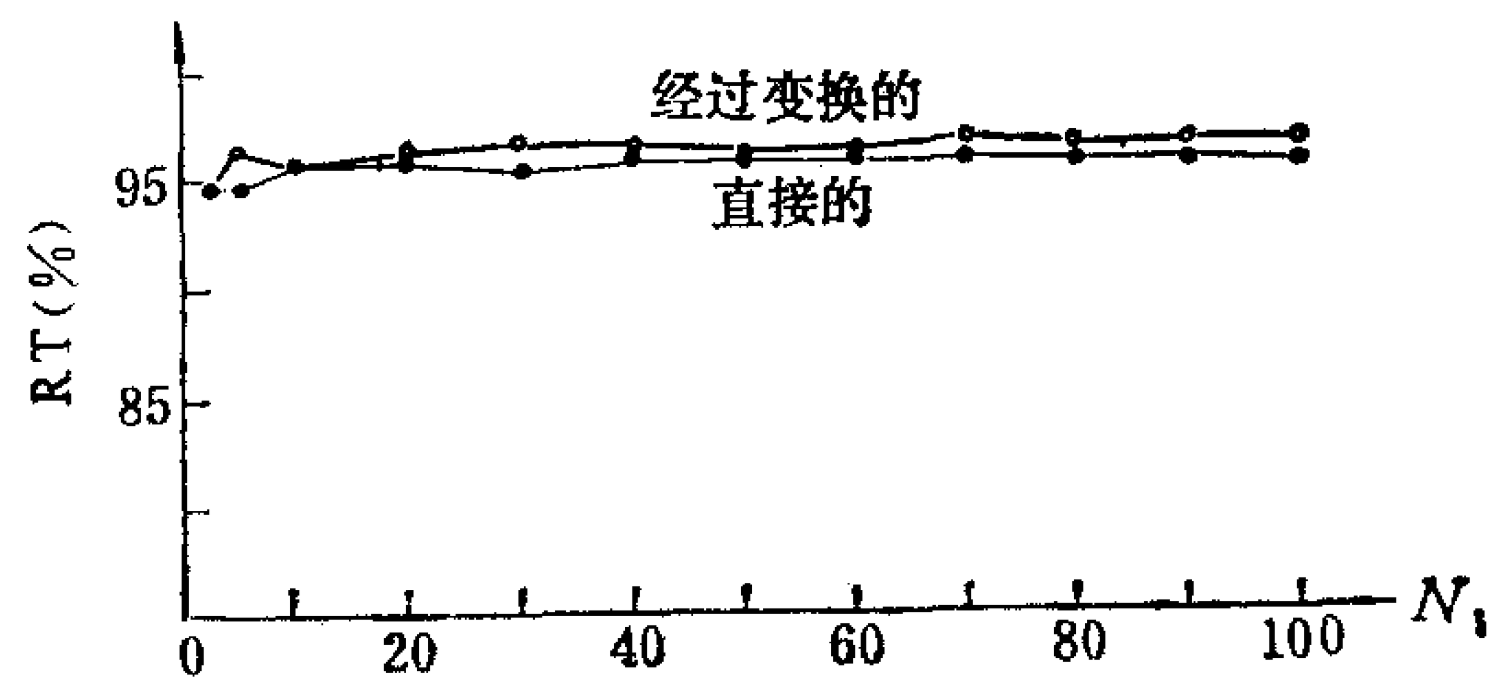

图 5 对数据集 III 两种分类器的性能与训练样本数的关系

参 考 文 献

[1] Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press, 1972.
[2] Chang, S. S. L., On Fuzzy Algorithm and Mapping, in «Fuzzy Automata and Decision Process», M. M. Gupta Ed. Amsterdam, North-Holland, 1977.
[3] Chang, S. S. L., On a Fuzzy Algorithm and Its Implementation, *Trans. IEEE*, SMC-8 (1978), No. 1, 31—32.
[4] Ahmed, N., Rao, K. R., Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, 1975.

STATISTICAL-FUZZY METHOD AND FUZZY-STATISTICAL METHOD FOR PATTERN RECOGNITION

TAO DUCHUN
(Institute of Acoustics, Academia Sinica)

ABSTRACT

The statistical-fuzzy method and the fuzzy-statistical method for pattern recognition are developed in this paper on the basis of the discussion on the generalities, differences and the respective suitable scopes of statistical approaches and fuzzy approaches to pattern recognition. The statistical-fuzzy method is to adopt in a fuzzy classifier the membership functions which make full use of the statistical information of the pattern components, so that the performance of the classifier is better than that of common fuzzy classifiers. The fuzzy-statistical method is to replace the pattern components by their fuzzy membership functions as inputs in a classifier which is based on the statistical method. From the results of the classification experiments made with the data sets given in this paper, it can be seen that the classification performance of this method can approach the optimal level of the Bayesian classifier with quite a small training set.