

# 分层分类器设计及实验结果

熊范纶

(中国科学院合肥智能机械研究所)

## 摘 要

本文介绍了笔者在分层分类器设计中决策规则选择、决策树的骨架结构设计及有效特征选择采用的一些方法,并将几种分类方法进行了比较,所提出的大值类法具有实用价值。以白细胞样本进行的试验结果表明,所采用的方法是有效的。

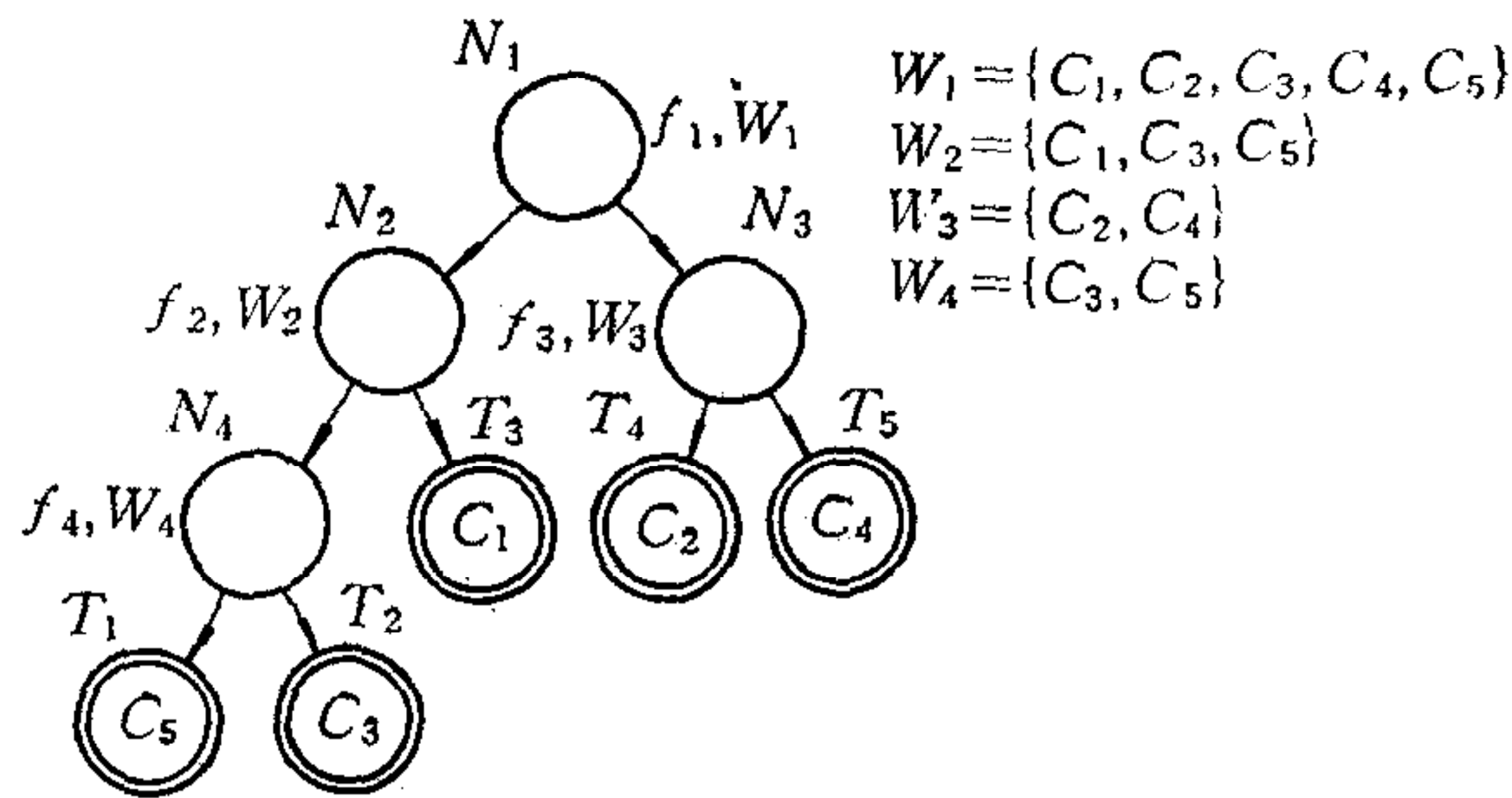
在统计模式识别方法中,对于象生物学细胞、遥感图象等属于多类多特征的模式,在训练样本不甚充分的情况下,采用经典的分类方法(如最大似然法、Fisher法等)常遇到较大的困难。近些年来,一种称为分层分类的方法被发展起来,它们对解决这种多类多特征的模式分类比经典分类(可称为一步分类)方法更为有效,更有吸引力<sup>[1-6]</sup>。

这种方法的基本思想是将样本分层次进行分类。首先选择某些特征将最容易分开的类集进行分类,然后在每个类集中又各自选择合适的特征,将容易分开的类集再进行分类。这样一层一层地分下去,直至最后将样本分到某个类别为止。

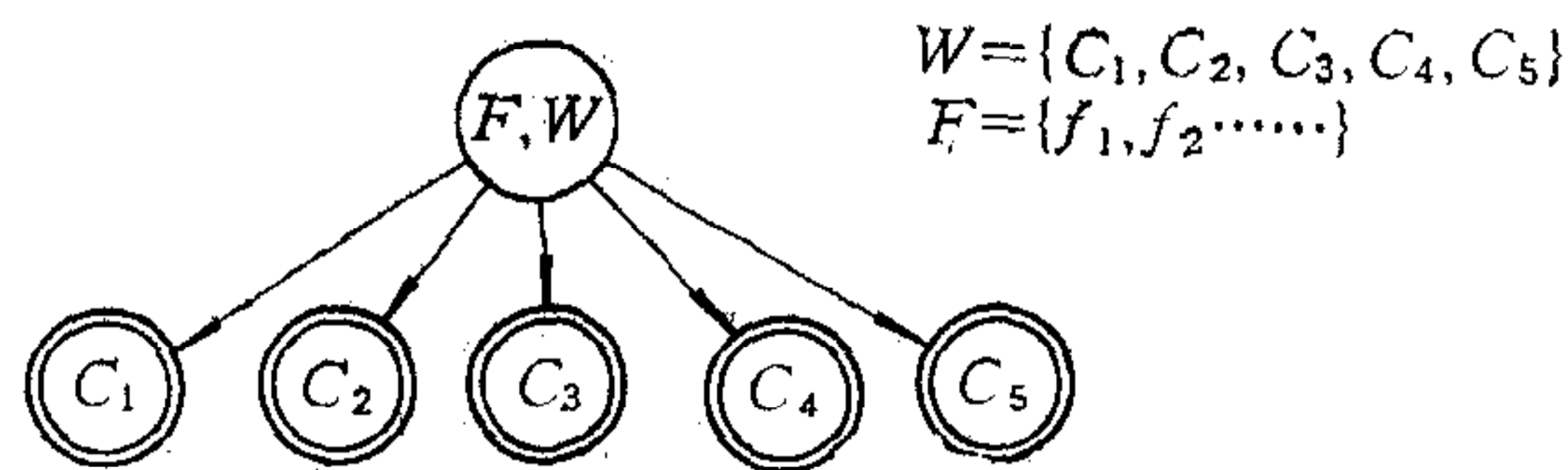
## 一、分层分类与一步分类

分层分类器常采用一种称为决策树的结构,所以又常称它为树分类器,如图1(a)所示。树由非终节点 $N_1-N_n$ 与终节点 $T_1-T_m$ 组成。每个非终节点含有两个以上的类别,在决策规则作用下,它可扩展到下层两个以上的子节点。终节点是样本分类的结束,每个终节点只含有一个类别。为简化起见,这里采用二叉树型式,即一个非终节点只扩展为两个子节点。图中 $N_1$ 称为根节点,样本由此开始分类。选择特征 $f_1$ 进行测量,根据某决策规则样本被分至 $W_2$ 类集或 $W_3$ 类集。若被分至 $W_3$ 类集,即去 $N_3$ 节点。再测量特征 $f_3$ ,按决策规则去下层两子节点中一个,直至样本被分到某个终节点 $T_i$ ,判别该样本属于某类 $C_i$ 为止。

可以看出,分层分类方法与一步分类方法(如图1(b)所示)不同,用一步分类法分类时,同时对所有被选择的特征进行测量,不管其中有些特征对区分某些类别的样本是否有用。分层分类法则不必为区分所有类别而去选择一个统一的最佳特征集,而只须在每层测量很少的(甚至一个)特征,分开几个类集,且每层的分类判据可以比较简单。一步分类法对所有类均采用一个统一的分类判据,当统计分布特性复杂时,所选分类判据常常不很



(a) 分层分类



(b) 一步分类

图 1 分层分类与一步分类结构

理想。由于分层分类方法将比较容易分开的类集先分类，因此在高层的分类风险可以较小。在以下各层也是尽量挑选那些较易分开某些类别的特征，因此分类精度可以得到改善，且分类器的设计比较直观，便于人机交互。当然，在最高层如果出现错误分类，则在下层是不可纠正的，这在一步分类中同样也是难以避免的。

## 二、分层分类器的设计

由于分类是逐层进行，呈树状结构，因此与一步分类器相比，分层分类器的设计比较复杂。它的设计包括决策规则选择、树骨架结构设计和特征选择等方面。由于它们之间存在着一定的约束关系，因此，整个分类器的最佳化不能简单取决于它们的各自最佳化，而必须在总体上进行考虑。Kulkarni 曾在理论上证明了能得到一个最佳的解<sup>[3]</sup>，但实际上难以实现。笔者采用了一些比较实际的设计方法，获得了满意结果。下面阐述本分类器设计中采用的实际方法。

### 1. 决策规则的选择

本分类器的决策规则采用具有最小误分类风险的贝叶斯决策方法。假定样本中有  $C_1, C_2, \dots, C_m$  共  $m$  个类， $\mathbf{x}$  为样本特征矢量，在获得样本特征  $\mathbf{x}$  的条件下属于类  $C_k$  的后验概率为  $P(C_k/\mathbf{x})$ ，则该样本分往  $C_j$  类的误分类风险为

$$r(C_j/\mathbf{x}) = \sum_{k=1}^m c_{jk} \times P(C_k/\mathbf{x}). \tag{1}$$

$c_{jk}$  为误分类代价。通常，当  $j = k$  (正确决策) 时，取  $c_{jk} = 0$ ； $j \neq k$  (错误决策) 时，取

$c_{jk} = 1$ . 考虑到概率密度函数  $f(\mathbf{x}/C_k)$  的计算较困难, 因此, 假定所用特征是类条件统计独立的, 即可以

$$f(\mathbf{x}/C_k) = \prod_{i=1}^n f(x_i/C_k). \quad (2)$$

其中  $x_i$  是特征矢量  $\mathbf{x}$  的第  $i$  个特征值. 假定样本共有  $n$  个特征, 经推导得到

$$r(C_j/\mathbf{x}) = \frac{\sum_{\substack{k=1 \\ k \neq j}}^m P(C_k) \times \prod_{i=1}^n f(x_i/C_k)}{\sum_{k=1}^m P(C_k) \times \prod_{i=1}^n f(x_i/C_k)}. \quad (3)$$

式中  $P(C_k)$  为先验概率, 可由训练样本集估算出

$$P(C_k) = \frac{C_k \text{ 类样本数}}{\text{总样本数}}. \quad (4)$$

$f(x_i/C_k)$  的估算方法是将  $C_k$  类样本的特征  $i$  的测量值按其变化范围分成  $N$  个状态, 则频率  $F(C_k, i, 1)$  到  $F(C_k, i, N)$  即是  $f(x_i/C_k)$  的离散值.

## 2. 树的骨架结构设计

本分类器树的骨架结构采用“聚结”方法进行设计. 用相似测量方法将相近的类逐层归并上去, 最后得到整个分类树的骨架. 为使树的根节点附近的错误分类降低到最小, 采用了一个反映误分类风险的类间关系矩阵. 聚类之间的相似性用一个与其决策规则(贝叶斯决策)相对应的关系系数来度量<sup>[5]</sup>. 首先求出  $C_j, C_i$  两类之间关于某特征  $i$  的关系系数  $R_i(C_j, C_i)$ ,

$$R_i(C_j, C_i) = \frac{\sum_{l=1}^N \min\{P(C_j) \times F(C_j, i, l), P(C_i) \times F(C_i, i, l)\}}{\sum_{l=1}^N \max\{P(C_j) \times F(C_j, i, l), P(C_i) \times F(C_i, i, l)\}}. \quad (5)$$

$R_i(C_j, C_i)$  的值在 0 到 1 之间, 且  $R_i(C_j, C_i) = R_i(C_i, C_j)$ . 当  $C_j = C_i$  时,  $R_i(C_j, C_i) = 1$ . 该特征  $i$  的所有类间关系系数组成一个关系矩阵. 在该矩阵中找出除 1 外最大的关系系数, 则得到该层能相互合并的两个类, 由这两个类归并成为一个新类. 新类的先验概率及类条件概率(以频率表示)分别取为

$$P(C_{new}) = P(C_j) + P(C_i), \quad (6)$$

$$F(C_{new}, i, l) = \frac{P(C_j) \times F(C_j, i, l) + P(C_i) \times F(C_i, i, l)}{P(C_j) + P(C_i)}. \quad (7)$$

再按 (5) 式算出最大关系系数, 又得到一个新类. 重复进行上述步骤, 最后得到该特征的聚结树状图.

特征不同, 得到的树状图也不相同. 由于此时完整的树骨架还未建立, 还不知道哪些特征在分类中起作用. 为此在设计中采用折衷处理方法, 对每个特征选用一个权系数  $W_i$ , 对那些估计将来在分类中有较好判别能力的特征选用较低的权, 从而得到所有特征的总类间关系系数

$$R(C_j, C_i) = \sum_{i=1}^n W_i \times R_i(C_j, C_i). \quad (8)$$

同单个特征方法一样,每次有两个类或类集被归并,重复进行以上步骤,直至得到总的聚结树状图。

### 3. 特征选择

同一步分类法一样,在决策树各非终节点上选择合适而有效的特征来进行分类是一个重要问题。如前所述,本分类器在每个非终节点上只选择一个特征参加分类。但在样本特征数较多的情况下,选择最佳特征组的计算量依然是巨大的。为此,可采用“分枝估界”方法。

在分类路上特征统计独立的假设下,树结构特征分配的性能指标  $M$  可以用正确识别的概率来表达<sup>[3]</sup>,即

$$M = \sum_{i=1}^m P(C_i) \times \prod_{N_d} P_i(C_i/N_d), \quad N_d \in S(C_i). \quad (9)$$

式中  $P(C_i)$  为  $C_i$  类的先验概率,共  $m$  个类。 $N_d$  是从  $C_i$  类终节点到根节点所经路上的节点, $S(C_i)$  是这些节点的集合,例如图 1 中, $S(C_3) = \{N_1, N_2, N_4\}$ 。 $P_i(C_i/N_d)$  是在节点  $N_d$  测量特征为  $i$  的样本属于  $C_i$  类的正确决策概率。在特征选择过程中,当找到一个具有高于  $M$  值的特征组时,可将小于这  $M$  值的所有其它特征组否决掉,从而迅速找到一组有效特征。

## 三、分类方法

决策树设计好后,需考虑怎样使得分类速度最快,效率最高。本分类器分别采用了以下几种分类方法:

### 1. 状态空间搜索方法

图 1(a) 决策树结构可以被看作是一个状态空间图,决策分类可以用状态空间搜索方法来实现。本实验采用  $S$  可采纳及  $B$  可采纳两种策略进行搜索。它们类似于  $A^*$  控制策略,是有序搜索<sup>[3,7]</sup>,只是在估价函数中增加了一项到达终节点的决策风险估计。这两种策略按照模式分类的特点采取两种不同的决策风险估计。 $S$  可采纳策略的风险估计是取到达终节点的那条路上的特征测量的函数,且利用一个决策风险的低界函数。 $B$  可采纳策略的风险估计不仅取到达终节点那条路上的特征测量的函数,而且取其它路上所有特征测量的函数。它除了采用与  $S$  可采纳相同的风险低界函数外,还采用了一个风险上界函数。

### 2. 决策矢量方法

该方法如图 2 所示。 $d_i$  是非终节点  $N_i$  上进行决策分类的决策矢量,它含有与特征测量相同的状态数。在二叉树中, $d_{ij}$  取 0, 1 两个值。 $d_{ij} = 0$  代表样本在节点  $N_i$  的被选特征  $f_i$  的测量值处于  $j$  状态,则分往左子节点; $d_{ij} = 1$  时,则分往右子节点。

与上面特征选择相似,它也要选择有效特征进行分类,需从数目巨大的决策矢量组中

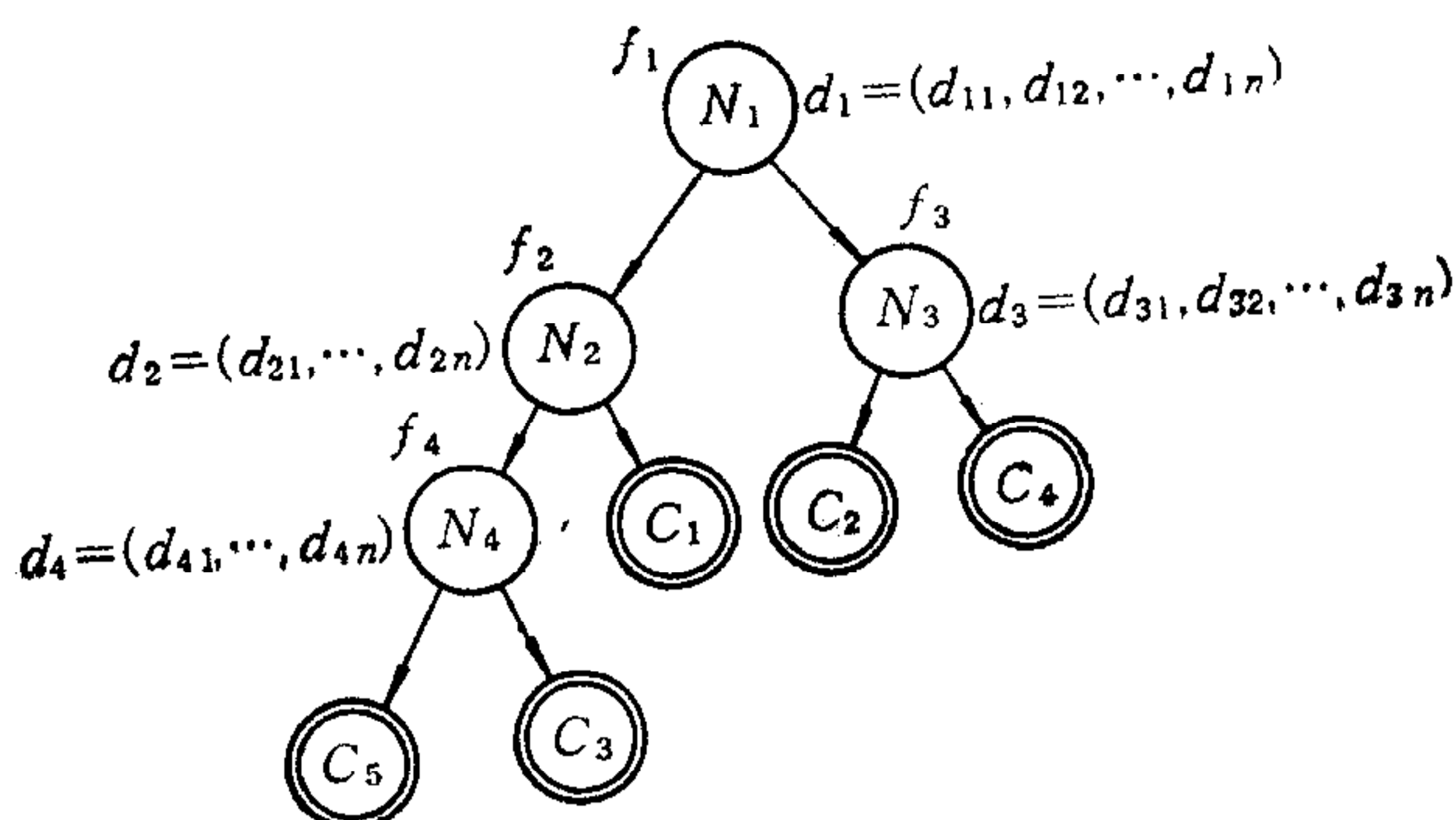


图2 决策矢量分类方法

选取一组最佳的决策矢量组。为减小计算量,用以下两种方法进行计算:

### (1) 动态规划方法

采用动态规划、聚类及分枝估界方法来寻求这组最佳决策矢量<sup>[3]</sup>。实际上,这时的计算量仍然很大。

### (2) 大值类方法

这是笔者提出的一种最简单的选择决策矢量的方法。在每一非终节点  $N_i$  上,计算该节点所含各类  $C_k \in W(N_i)$  在所选特征  $f_i$  的状态  $j$  时的  $P(C_k) \times f(x_i = j/C_k)$  值,选取其中具有最大值的那个类  $C_m$ 。

若  $C_m \in W_0(N_i)$ , 则  $d_{ij} = 0$ ; 若  $C_m \in W_1(N_i)$ , 则  $d_{ij} = 1$ 。  $W(N_i)$  代表节点  $N_i$  的类集。  $W_0(N_i)$ ,  $W_1(N_i)$  分别是节点  $N_i$  的左、右子节点中的类子集。  $P(C_k)$  是  $C_k$  类的先验概率。  $f(x_i = j/C_k)$  表示在节点  $N_i$  上测量特征  $f_i$  的值落在  $j$  状态的概率密度函数。

## 四、实验结果与分析

为检验上述分层分类器设计方法,运用一个白细胞样本集进行了实验。该样本集共有 1379 个包含六类 32 个特征的白细胞。每类样本数为分叶核细胞 464 个,杆状核细胞 197 个,淋巴细胞 368 个,单核细胞 184 个,嗜酸细胞 104 个,嗜碱细胞 62 个。32 个特征反映白细胞的几何、颜色、光密度、纹理、形状等参数。

分类实验是用旋转法与重代法进行的。旋转法是将 1379 个白细胞样本均匀分成四组,然后依次用四个组中的三个组作为训练集设计分类器,另一组作为测试集。重代法则将全部样本既作为训练集又作为测试集。

图 3 和图 4 是分别用旋转法和重代法得到的分类器树结构。树骨架用聚结法得到。节点内的数字为节点号,非终节点外的数字是用分枝估界技术选择的有效特征号(32 个特征用号码 1—32 代表)。

表 1 列出整个白细胞样本在图 3 图 4 的决策树上采用  $S$  可采纳和  $B$  可采纳状态空间搜索方法进行分类的结果,可以看到运用  $B$  可采纳策略的分类精度比用  $S$  可采纳策略要

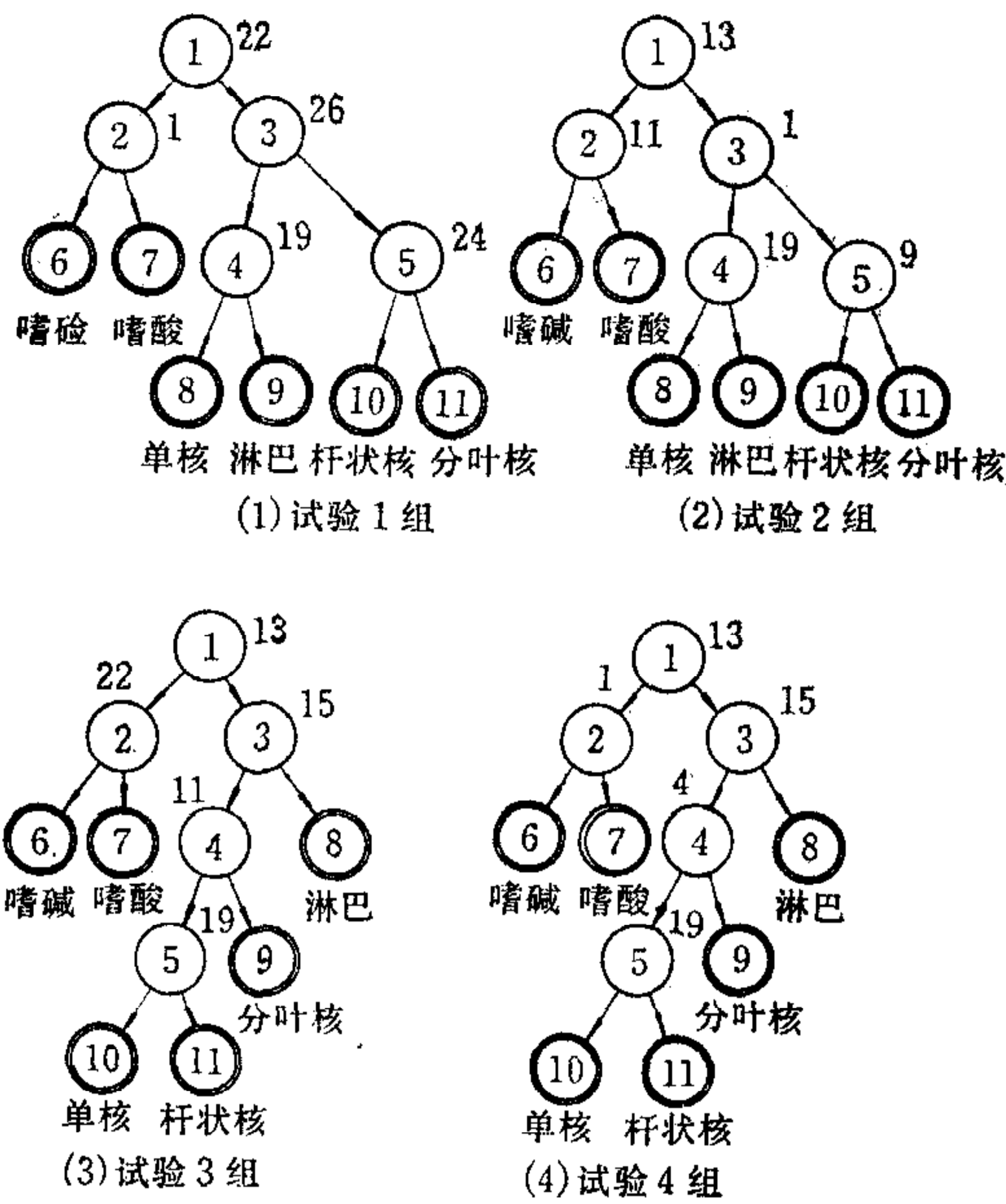


图 3 用旋转法得到的分类器树结构

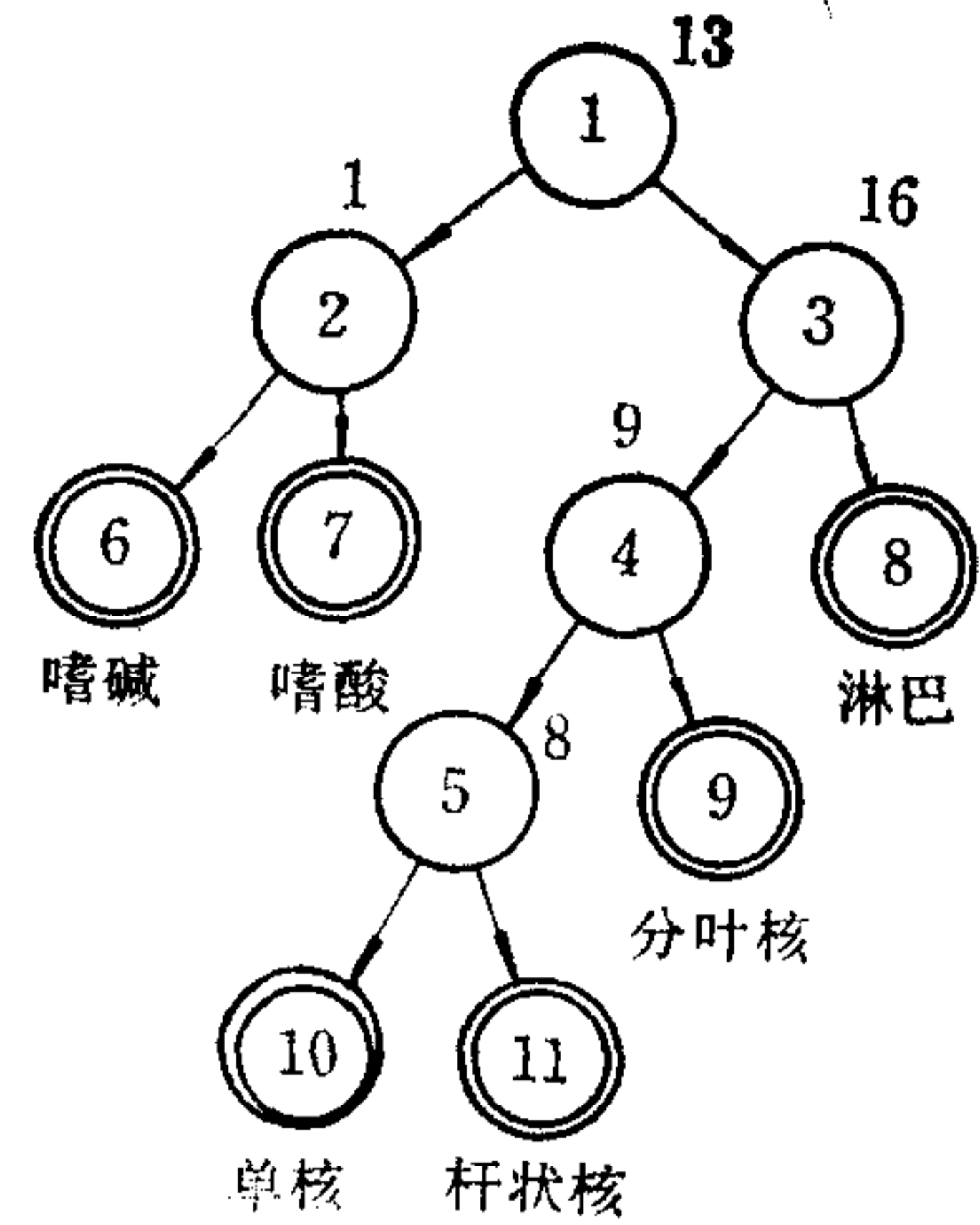


图 4 用重代法得到的分类器树结构

高些，原因是  $B$  可采纳策略是对所有搜索路径上的全部被选特征进行测量，表 2 为  $B$  可采纳策略运用重代法得到的分类结果。用旋转法所得到的四个组分类结果与它类似。由表 2 可以看到，淋巴、单核、嗜酸、嗜碱以及分叶核五类细胞的分类结果比较理想，杆状核

表 1 状态空间搜索方法的分类精度(百分比)

分类策略	旋 转 法						重 代 法
	第一组	第二组	第三组	第四组	均 值	方 差	
$S$ 可采纳法	83.3	82.6	74.4	79.4	79.9	3.5	80.6
$B$ 可采纳法	83.6	83.7	77.3	82.3	81.7	2.6	85.1

表 2  $B$  可采纳搜索用重代法得到的分类结果

分类结果 原类名	分叶核	杆状核	淋 巴	单 核	嗜 酸	嗜 碱
分叶核	415	47	1	0	1	0
杆状核	110	77	6	1	3	0
淋 巴	0	6	357	4	0	1
单 核	0	2	4	175	1	2
嗜 酸	2	1	0	0	101	0
嗜 碱	0	1	6	5	1	49

细胞分类的错误率相当大，多数误分为分叶核。这是因为区分分叶核与杆状核细胞的主要根据是核形状的差异。实验中所选的特征 9、19、24 是反映核的特征的有关参数，如核面积、核的颈等，但用这类特征描述核的形状并不十分准确。

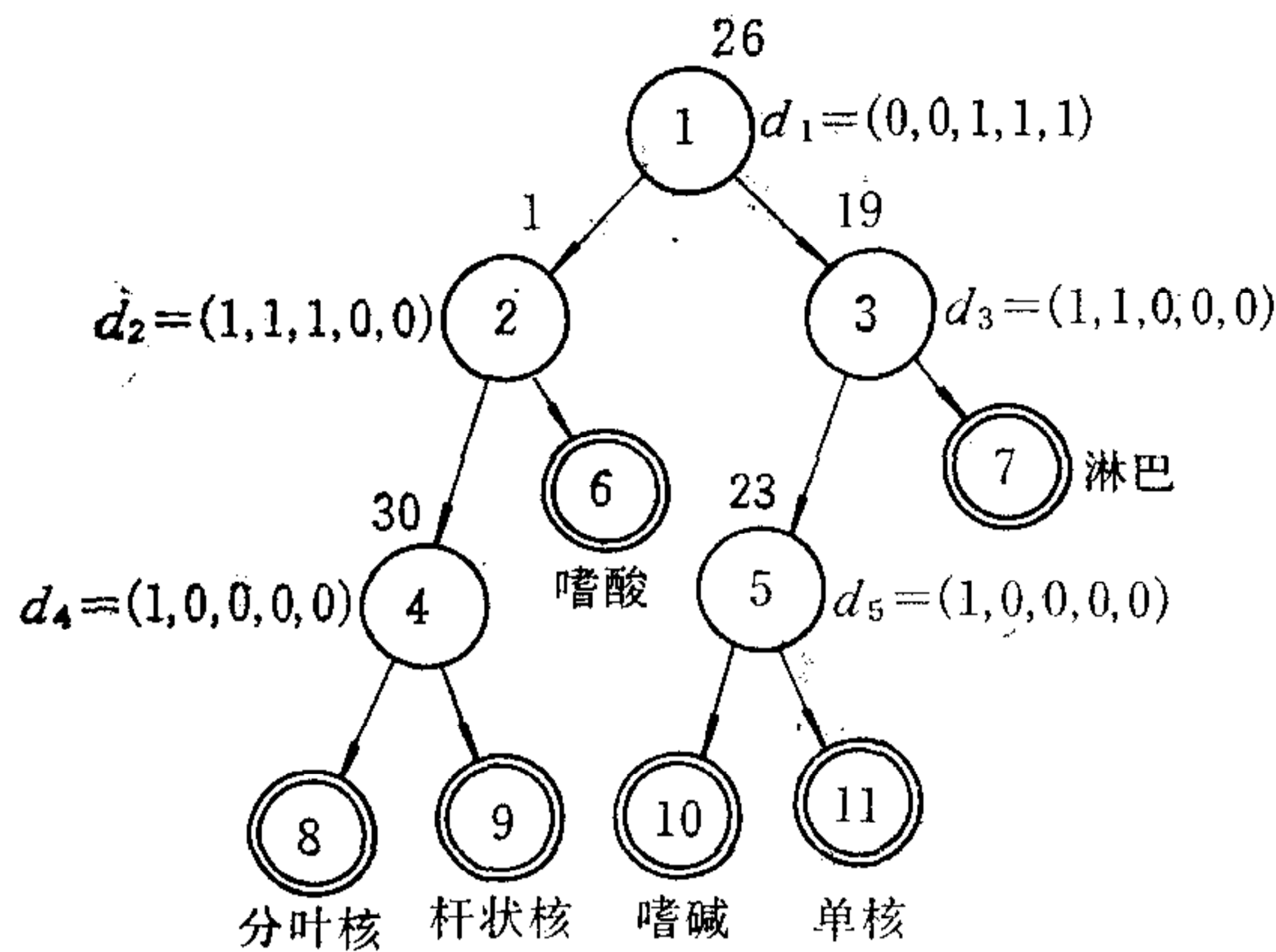


图 5 用决策矢量法分类的分类器树结构

图 5 是采用决策矢量方法进行分类的分类器树结构。树骨架是用美国马里兰大学一个人机交互模式分类系统 MIPACS 来建造的，特征选择采用分枝估界技术。每个特征选用五个测量状态，所以决策矢量也是五个状态。表 3 是这种方法用动态规划法及大值类法分类的结果。它们与搜索方法分类结果相近。但大值类方法决策矢量的确定是最简单的。由实验知道，大值类方法确定决策矢量所需时间只是动态规划方法的三分之一。由表 3 可以看出，当特征测量状态为十时，大值类法分类结果比动态规划法好。表 4 是图 5 中在各非终节点上用这两种方法得到的决策矢量。可以看到，除试验第 2 组节点 3 的决策矢量稍有不同外，其余皆相同。以上结果均反映了大值类法的有效性。

图 3 至图 5 中非终节点旁列出了分类时所选择的特征号。它们与 Kulkarni 所得结论以及化验师进行白细胞分类时选择特征的规律基本相似。

本试验是在 UNIVAC 1180 计算机上进行的。旋转法试验从树设计、特征选择到分类，计算机执行时间约一分钟。重代法因数据量大，需两分钟。

表3 决策矢量方法的分类精度(百分比)

测量状态数	决策矢量选择方法	旋 转 法						重代法
		第一组	第二组	第三组	第四组	均 值	方 差	
5	动态规划法	82.1	80.5	83.7	83.4	82.4	1.3	81.1
	大值类法	82.1	80.5	83.7	83.4	82.4	1.3	81.1
10	动态规划法	80.4	75.3	79.9	80.2	79.0	2.1	79.4
	大值类法	82.4	79.9	84.3	83.4	82.5	1.6	82.1

表 4 动态规划法与大值类法产生的决策矢量比较

决策矢量 选择方法	节点号	旋 转 法				重代法
		第一组	第二组	第三组	第四组	
动态规划法	1	0 0 1 1 1	与 第 一 组 结 果 相 同			同
	2	1 1 1 0 0				
	3	1 1 0 0 0				
	4	1 0 0 0 0				
	5	1 0 0 0 0				
大值类法	1	0 0 1 1 1	0 0 1 1 1	与 第 一 组 结 果 相 同		同
	2	1 1 1 0 0	1 1 1 0 0			
	3	1 1 0 0 0	0 1 1 0 0			
	4	1 0 0 0 0	1 0 0 0 0			
	5	1 0 0 0 0	1 0 0 0 0			

## 四、结 论

上述实验结果表明, 本文阐述的设计分层分类器的各种方法是可取的。在二叉树上运用简单的贝叶斯决策规则, 避免建立复杂的判别函数, 显示了分层方法的优点。用由类间关系矩阵产生树状图的聚结技术所建立的二叉树骨架结构, 虽然不是最佳的, 但它反映了样本数据的自然结构, 是可取的。

具有六类 32 个特征的白细胞, 最多只选用五个特征(不同非终节点可以选择相同特征)参加分类, 这可以大大节约特征抽取与测量上的花费。这在某些模式分类中, 是很有实际价值的。运用分枝估界方法选择有效特征, 符合化验师的选择规律, 是一个颇有效的特征选择方法。

本实验样本是一个较为典型的样本集, 几种分类方法的分类精度均达 80% 左右。若能抽取较好的区分分叶核与杆状核的特征, 分类效果定可进一步改善。动态规划方法选择最佳决策矢量组速度慢, 而大值类方法既简单, 速度又快, 且分类效果较好, 因此具有实用意义。

嗜酸、嗜碱等细胞以较少的样本作为训练集便能取得较好的分类效果, 无需用大量的训练样本, 这是分层分类又一个优于一步分类的地方。

本研究与实验工作得到美国马里兰大学模式分析实验室 L. N. Kanal 教授的指导, 在此表示感谢。

## 参 考 文 献

- [1] You K. C. and Fu K. S., An Approach to the Design of Linear Binary Tree Classifier, Proc. Symposium of Machine Processing of Remotely Sensed Data, Purdue University, 3A-1 to 3A-10, 1976.
- [2] Kulkarni, A. V. and Kanal, L. N., An Optimization Approach to Hierarchical Classifier Design,



- Proc. 3rd Intern. Joint Conf. on Pattern Recognition, San Diego, 1976.
- [ 3 ] Kulkarni, A. V., Optimal and Heuristic Synthesis of Hierarchical Classifiers, Ph. D. dissertation, University of Maryland, 1976.
- [ 4 ] Shi Qingyun, A Method for the Design of Binary Tree Classifiers, Proc. PRIP Conf., 1981.
- [ 5 ] Bernard, E. D. A. S., On Tree Structure Design and Hierarchical Classifiers, Vrije University, Amsterdam, 1980.
- [ 6 ] Xiong Fanlun, Dattatreya, G. R., and Kanal, L. N., Hierarchical Methods for Leukocyte Classification, Proc. IEEE International Conference on System, Man and Cybernetics, 1984, 403—407.
- [ 7 ] Nilsson, N. J., Principle of Artificial Intelligence, Tioga Publishing Co., 1980.

## HIERARCHICAL CLASSIFIER DESIGN AND ITS EXPERIMENTAL RESULTS

XIONG FANLUN

*(Hefei Institute of Intelligent Machines, Academia Sinica, China)*

### ABSTRACT

In this paper, a number of approaches for the design of hierarchical classifiers are described, which includes the selection of decision rules, the design of decision tree skeleton, the selection of efficient features. And a comparison of some classification methods is made. Experimental results for the classification of white blood cell samples are reported, which shows that these approaches are very attractive.