第 13 卷 第 2 期
1987 年 3 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 13, No. 2
Mar., 1987

树型分类器的研究

郑勤奋 吴健康 王文涛
（中国科学技术大学）

摘 要

本文从理论上给出了树分类器的数学定义，探讨了分类树的优化和性能评价函数这两个关键问题．从而提出了在卫星多光谱数据分类中切实可行的优化树分类器设计方法，包括合理的性能评价函数、新型的分类树数据结构和搜索的优化算法．最后给出了应用于森林分类的实验结果．

一、引 言

在遥感应用的各种分类问题中，目前广泛采用一次判决的(单层的)最大似然率分类器及集群分类器等分类方法．应用中，这些传统的分类方法暴露出它们在特征选择，运算效率，处理多种类、多来源数据等方面的缺陷．树分类器（Decision Tree Classifier）是一种具有自适应控制功能的分类系统，可以克服上述缺陷．图 1 是一个树分类器的简单例子．其中每一个非终止节点（用圆圈标记并标注以大写字母)都是一个单层分类器，终止节点(以三角形标记并标注以数字）表示分类结果．输入样本首先被送到根节点，经由某条路径所对应的一系列判决，到达终止节点，从而得出分类结果．由于树分类器是一个用树表示的最优分类系统，其中最优分类树的选择、各节点所对应单层分类器的优化设计等问题，在理论和实现上都有相当的难度，一直是不少学者注重研究并试图解决的问题[1-5]．美国普渡大学的 C. L. Wu 和 P. H. Swain 较早地将树型分类器应用于遥感数据处理[1]，提出了一个比较完整的树分类器的设计方法．本文则是在上述工作的基础上，总结出树分类器的数学描述，改进性能评价函数和优化树的搜索方法，从而形成一个比较切实可行的树分类器设计方法．

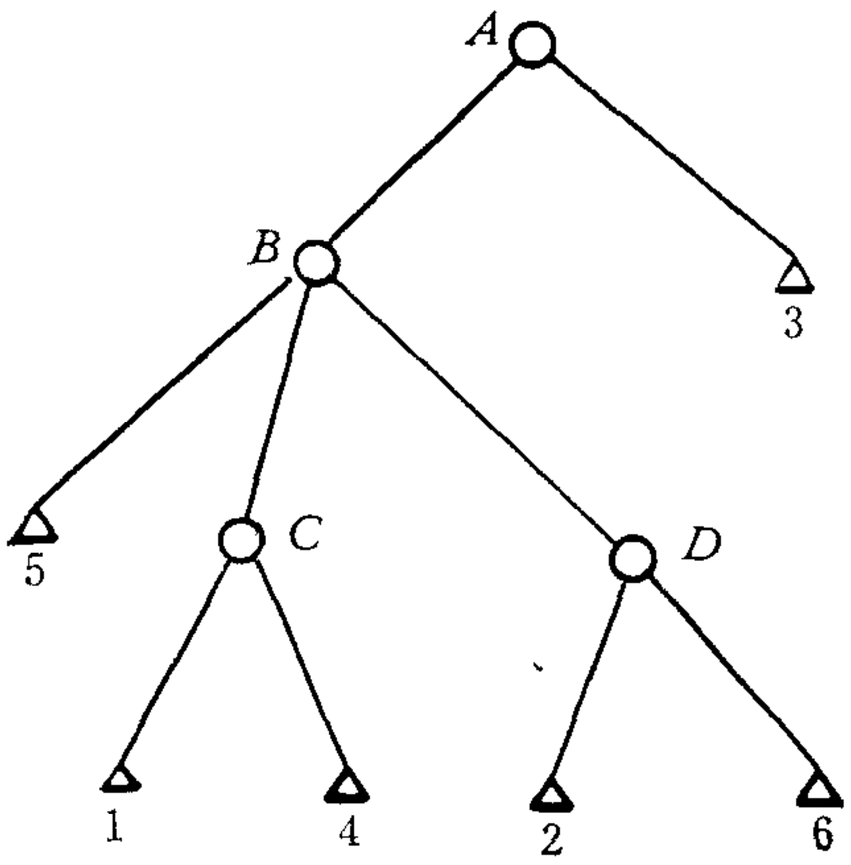

图 1 树分类器示意图

本文于 1985 年 1 月 11 日收到．

二、树分类器的数学定义

设M为样本全体的集合，$\varphi=\{1,2,\cdots,K\}$为要分的K类结果，$\pi=\{\varphi_1,\varphi_2,\cdots,\varphi_N\}$为$\varphi$的一个分割，满足：1) $\varphi_i\neq\phi$, $i=1,2,\cdots,N$; 2) $\bigcup\limits_{1\leqslant i\leqslant N}\varphi_i=\varphi$; 3) $\varphi_i\neq\varphi_j$, $i\neq j$; 4) $N>1$。则分类过程就是$M\to\pi$的一个映射，$f: M\to\pi$。显然f是由定义映射的有关参数和结果集合唯一确定的，记为$f=\eta(D,\pi)$。其中D是表示$M\to\pi$映射的参数；π是φ的一个分割，表示分类结果集合，称为分类的索引集。例如对单层ML分类器$\eta_{ML}(D,\pi)$，由于是单层分类，要求每个φ_i都是终止结果，故φ_i只能包括单个元素。由π所满足的条件可推出$\pi=\varphi$，这时的D就包括各类的均值、方差和先验概率等信息。

对于树分类器，由于采用的是多层分类，每一次分类就对应于一个节点，所以树分类器可以抽象地看作为满足一定条件的节点集合。定义如下：

定义一。 $M\to\pi$的分类器$\eta(D,\pi)$称为节点。

定义二。 设$T(M,\varphi)$是由M和φ生成的所有节点集合$\eta_i(D_i,\pi_i),\pi_i=\{\varphi_{i,1},\varphi_{i,2},\cdots,\varphi_{i,N_i}\}$；$\eta_j(D_j,\pi_j),\pi_j=\{\varphi_{j,1},\varphi_{j,2},\cdots,\varphi_{j,N_j}\}$为$T(M,\varphi)$的两个元素。若存在$1\leqslant K\leqslant N_i$，使得$\bigcup\limits_{1\leqslant l\leqslant N_j}\varphi_{j,l}=\varphi_{i,K}$，则称$\eta_i$为$\eta_j$的父节点；$\eta_j$为$\eta_i$的子节点。

定义三。 仅包含单一元素的集合称为单点集。

定义四。 对应的索引集完全由单点集构成的节点称为简单节点。

定义五。 设$\beta\subset T(M,\varphi)$为一节点集合，$\eta\in\beta$，当在$\beta$中不存在$\eta$的父节点时，称$\eta$为$\beta$的根节点。

定义六。 设$\beta\subset T(M,\varphi)$为一节点集合，当满足：1) 在$\beta$中有且仅有一个根节点；2) β中任一节点至多只能有一个父节点，则称β为一个树分类器。

定义七。 设$\beta\subset T(M,\varphi)$为一树分类器，若$\beta$中任一节点的任一非单索引元素都对应有$\beta$中的一个子节点，则称$\beta$为完全树分类器。

由于实际应用中要求的都是完全树分类器，以下提到的树分类器均指完全树分类器。下面给出分类过程的数学描述。

定义八。 设$\beta\subset T(M,\varphi)$为一树分类器，对任一$x\in M$，按下述迭代规则产生的$\beta$中节点的一个有序子集$S_\beta(x)$，称为$\beta$对$x$的唯一决策序列。

1) $S_\beta(x)$的第一个元素为β的根节点。

2) 设$\eta_j(D_j,\pi_j),\pi_j=\{\varphi_{j,1},\varphi_{j,2},\cdots,\varphi_{j,N_j}\}$为$S_\beta(x)$的第$j$个元素。$\nu(D,\pi)\in\beta,\pi=\{\varphi_1,\cdots,\varphi_N\}$。若$\eta_j$判决$x$到$\varphi_{j,p}$，且$\bigcup\limits_{1\leqslant q\leqslant N}\varphi_q=\varphi_{j,p}$，取$\nu$为$S_\beta(x)$的第$j+1$个元素，若不存在$\beta$中的节点满足这一条件，则决策序列终止。

这里，$S_\beta(x)$就是树分类器β对样本x进行分类所经过的节点序列，对应着对x的一系列分类操作。$S_\beta(x)$的生成过程就是树分类器β对x的分类过程，$S_\beta(x)$序列中最末一个节点的判决所对应的索引元素就是x的最后分类结果。

三、树分类器的设计

原则上讲,最优分类树应是在广泛意义下的,其约束条件应尽量少,只要满足: 1) 每一节点的分枝数不超过该节点所包含的类数; 2)同一节点不应包含两个完全相同的子节点. 除此之外,不应对特征子集的选择,节点的分裂形式和判决准则进行限制. 然而, 上述原则导致了问题的复杂性. 在具体实现时, 人们往往在不影响性能或很少影响性能的前提下, 加进某些限制, 使问题得以简化. 在这里, 考虑到卫星多光谱数据可以近似地认为满足高斯分布, 故在所有非终止节点上,采用最大似然分类判决准则.

由于树的形式各种各样,很难用变量空间描述. 因此,不能用数学的手段优化. 目前一般都采用启发式算法逐层建立合适的树结构. 这里采用前向剪枝搜索法 (Guided Search with Forward Pruning). 该算法是通过在每一层上最优来寻求全局的最优(严格讲是准最优). 下一层的优化在上一层次的基础上进行. 其实现方法概括如下: 设各类均为高斯分布,对给定的训练样本集合求出各类的分布参数, 即均值矢量和协方差矩阵,然后将每一类都看成特征空间中由其中心点所代表的、由其分布参数所确定的点集,进行生成树的操作:

1) 初始化,置当前节点为根节点,设初值;

2) 对于当前节点,取出其内容,作如下操作:

a) 对所有候选特征子集,做:

计算各类对可分离度,

聚类得到候选分类器结构,

计算该候选分类器性能评价函数;

b) 选择性能评价函数最好的候选分类器作为本节点的分类器.

3) 取尚未运算过的非终止节点为当前节点,重复 2) 的运算. 若无未运算过的非终止节点,一个分类树已产生完毕,程序结束.

下面就设计过程中的一些关键问题,作比较详细的说明.

1. 分类树的数据结构

分类树是从上到下逐步生成的,树的分枝数和深度都可以任意变化.这使得数据结构中常用的树表示法因占用太多内存而不适用. 这里我们设计了一个先宽后深的一维数组编码表示法数据结构. 它从树根开始逐层从左到右搜索并存入一维数组. 对非终止节点,仅记下其直接子节点的个数,对终止节点, 记下其对应类号的负数以区别于非终止节点. 这种数据结构生成简单,存贮紧凑,树的编码顺序同分类树的生成过程一致. 它不易改动,这对于表示分类树不成任何问题. 因为在分类树设计过程中,仅最低一层节点需要变动,一旦分类树设计完毕用于分类时, 它并不需改动. 这种数据结构的另一缺点是,要查找某一节点必须从根节点开始,逐层推下来. 这恰好与树分类器的分类过程一致,自然不成问题. 图 2 是图 1 所示分类树的编码表示. 这里示出的仅是供检索用的部分, 实际的数据结构还应存贮各节点分类所采用的特征、各子节点的先验概率、均值矢量、协方差矩阵的逆阵等数据. 在分类树设计过程中产生树的编码表示, 以及在分类过程中对它进行检索以指导分类,都只要借助于几个辅助指针便可以方便地实现[7].

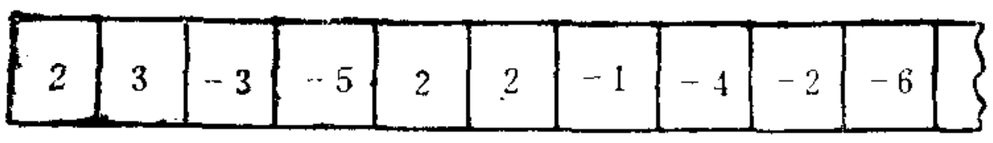

图 2　图 1 中分类树的编码表示

2. 类间可分性度量和误差概率的估计

在分类树设计过程中，作聚类分析和计算性能评价函数，都要用到类间可分离性测度。这里采用发散性度量。即：对两个正态分布 $N(M_1, \Sigma_1)$ 和 $N(M_2, \Sigma_2)$，其可分性测度为

$$D_T = 2000 \cdot [1 - \exp(-D/8)].$$

其中　$D = \frac{1}{2}\mathrm{tr}[\Sigma_1 - \Sigma_2][\Sigma_1^{-1} - \Sigma_2^{-1}] + \frac{1}{2}[M_1 - M_2]'[\Sigma_1^{-1} + \Sigma_2^{-1}][M_1 - M_2]$.

当 D_T 在 1000—2000 时，可以用下面的表达式来估计分类误差概率[6]：

$$\varepsilon_{12} = 0.32 \cdot (1 - D_T/2000).$$

对 N 类问题，总误差概率为

$$\varepsilon = (2/N)^{0.7} \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} P_i P_j \varepsilon_{ij}.$$

其中 P_i 为类 ω_i 的先验概率。

3. 聚类算法

树优化中的子节点生成是采用图论中无监集群聚类方法完成的[1]，其基本过程是：首先计算各类对的类间距离 D_{Tij} 形成类间距离矩阵，而后根据距离测度将类的顺序进行调整，以使得邻近的类排在一起，接着对调整过的距离阵作用上门限 T_h 得到二值距离矩阵，最后对二值距离矩阵进行搜索得出聚类方案。这里对不同的类组间允许有重复的类，以适用于分类树的需要。

4. 分类树的性能评价函数

在分类树的自动设计中，树结构的优化是通过比较各候选方案的评价函数实现的。评价函数应能反映分类的错误率和运行时间这两个因素。这里采用错误率与运行时间的加权和作为评价函数

$$E(d) = -T(d) - K\varepsilon(d).$$

其中 d 表示节点；T 为分类时间；ε 为错误率；K 为加权因子，代表了用户对运行时间和错误率的相对关心程度，每项前面的负号是为了符合习惯上认为评价函数值越高性能越好而设的。

考虑到多层分类器父节点的性能应同其子节点有关，将子节点取最不利的情况，并将时间项改用相对于单层分类器的百分数表示，代入误差概率和运行时间计算式得：

$$E(d_i) = \frac{1}{M(M+1)N_i}\left\{P_i[M(M+1)N_i - m_i(m_i+1)c_i] - \sum_{j=1}^{c_i}[P_j^i m_i(m_i+1)N_j^i]\right\} + K\left\{P_i\varepsilon_0(d_i) - P_i\varepsilon(d_i)\right.$$

$$- \sum_{j=1}^{c_i} [P_i^j \varepsilon_0(d_i^j)] \Big\}.$$

其中M为可利用的特征数；N_i为节点d_i包含的类数；P_i为节点d_i的先验概率；m_i为节点d_i分类所采用的特征数；c_i为节点d_i的子节点数；P_i^j为d_i的第j个子节点的先验概率；N_i^j为d_i的第j个子节点包含的类数；K为运算时间和误差率的相对权因子；$\varepsilon_0(d_i)$为d_i进行单层一次判决时的误差率；$\varepsilon(d_i)$为d_i进行树分类时的误差率；$\varepsilon_0(d_i^j)$为d_i的第j个子节点的误差率.

误差概率和先验概率由下述途径求得：

设某一节点有n类被分入m个子节点，用一$n \times m$的关系矩阵 $A = [a_{ri}]$ 表示

$$a_{rj} = \begin{cases} 1, & \text{当类 } \omega_r \text{ 属于第 } j \text{ 个子节点;} \\ 0, & \text{其它.} \end{cases}$$

从树分类过程知：一个属于类ω_r的样本被分入第j子集点的概率Q_{rj}为：1）当类ω_r不属于第j子节点时，应为类ω_r与第j子节点代表类的错误率之和；2）当类ω_r仅属于第j子节点时，为类ω_r在当前节点的先验概率；3）当类ω_r属于第j子节点同时还属于另一子节点k时，将类ω_r在当前节点的先验概率按照类ω_r对j，k两个子节点中心的距离分成两部分，属于子节点j的部分即为所求. 上述思想可用下式表示：

$$Q_{rj} = \begin{cases} e_{rj}, & \text{当 } a_{rj} = 0; \\ P_r, & \text{当仅有一个 } j \text{ 满足 } a_{rj} = 1,\ j = 1,2,\cdots,m; \\ \dfrac{D_{Trk}}{D_{Trj} + D_{Trk}} P_r, & \text{当 } a_{rj} = 1 \text{ 且 } a_{rk} = 1. \end{cases}$$

其中

$$e_{rj} = \begin{cases} 0, & \text{当子节点包含类 } \omega_r \text{ 时;} \\ \sum\limits_{l \in q} \left(\dfrac{2}{N}\right)^{0.7} P_r\, \varepsilon_{rl}\, P_l, & \text{其它.} \end{cases}$$

$$\varepsilon_{rl} = 0.32(1 - D_{Trl}/2000), \qquad l \in q.$$

q为第j子节点的代表类集.

对于所采用的聚类算法，每一类至多被分入两个子节点，故上式给出的计算是完备的. 由它可进一步求出 P_j^i 和 $\varepsilon(d)$.

$$P_j^i = \sum_{r=1}^{n} a_{rj}\, Q_{rj}, \qquad j = 1,2,\cdots,m;$$

$$\varepsilon(d_i) = \sum_{r=1}^{n} \sum_{j=1}^{m} e_{rj} = \sum_{r=1}^{n} \sum_{j=1}^{m} (1 - a_{rj}) Q_{rj}.$$

计算 $\varepsilon_0(d)$ 时对应于 $A = I$，即 $m = n$，$a_{rj} = \delta_{rj}$，这时有

$$e_{rj} = \left(\frac{2}{N}\right)^{0.7} P_r \varepsilon_{rj} P_j,$$

$$\varepsilon_0(d) = \sum_{r=1}^{n} \sum_{j=1}^{m} e_{rj} = \left(\frac{2}{N}\right)^{0.7} \sum_{r=1}^{n} \left(P_r \sum_{j=1}^{m} \varepsilon_{rj} P_j \right).$$

对于 $\varepsilon_0(d_j^i)$，只要注意到用 Q_{rj} 替换 P_r 等，就可以由 $\varepsilon_0(d)$ 的计算式求出.

将上述性能评价函数与某些参考文献中所述的相比，我们去除了某些不合理的简化和假设，完全和合理化了运算时间的估算，校正了非终止节点先验概率的计算方法．反复实验表明，这样得出的性能评价函数每个细节的物理意义都很明确，并使得设计出的分类树的最优性大大提高．在性能评价函数的具体计算中，由于采用了查表方法，简化了运算，有效地缩短了运算时间．

四、分 类 实 验

为了验证前面所述树分类器的正确性和有效性，将其应用于南京茅山地区的森林分类．所使用的陆地卫星多光谱图像尺寸为 128×64，分 12 类．训练区数据是由林业调查部门提供的．表 1 列出了不同参数的树分类器的性能分析数据，其参考是一次判决的最大似然率分类器．因为它使用全部特征，应该计算量最大，精度最高．然而，第一组树分类器分类精度几乎与它相同，分类时间却只是它的四分之三．同时，还应注意到，在我们的实验中，由于数据条件的限制，特征维数只有4，还不足以显示树分类处理多维、多种类数据的优点．

表 1　树分类器的性能分析

组　号	T_n	K	相对节约时间	相对分类精度
1	1950	20	25.0%	99.6%
2	1500	20	53.0%	89.9%
3	1950	5	74.0%	84.9%
4	1500	5	78.0%	83.9%

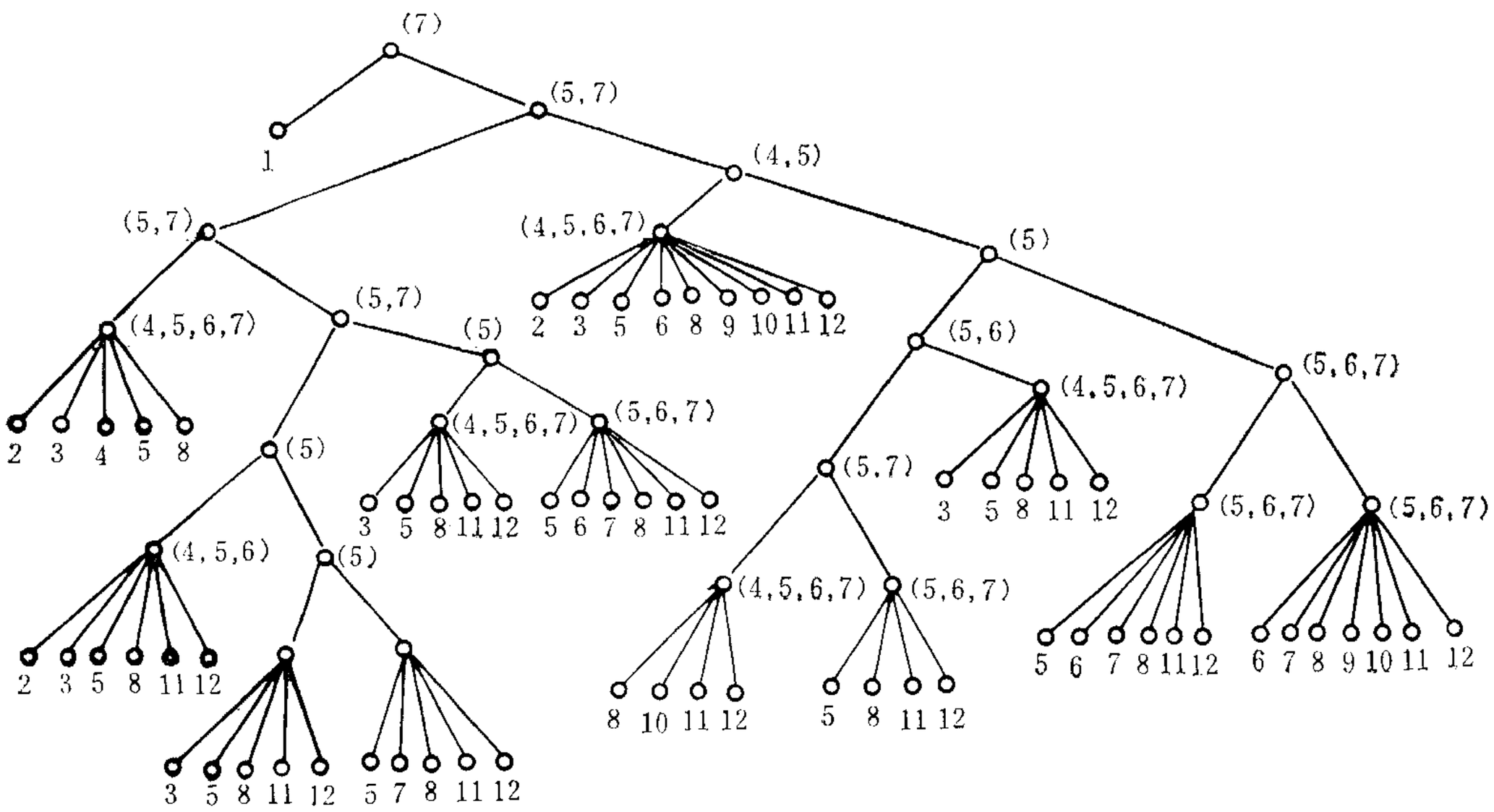

图 3　$T_n = 1500$，$K = 20$ 时生成的最优分类树结构

图 3 是 $T_n = 1500, K = 20$ 时的分类树结构．括号内的数字是特征波段号，即多光

谱中的 4,5,6,7 波段. 终止节点的标号即最终分类号. 在根节点, 只须用一个特征〈7 波段〉,就把第 1 类〈水〉与其余各类分开了. 由于其它各类均属植被,因此,含有植被信息最多的 5 波段特征出现在其它所有非终止节点上.

五、结 束 语

在本文所叙述的树分类器的研究中,我们将重点放在树分类器的性能评价函数上,由此产生了文中所述的性能评价函数的最后形式及各种参数的估算方法. 它和其它某些形式比起来,更完善而合理,保证了分类树的优化性能. 同时也应指出,由于条件限制,我们的实验是比较初步的. 我们将在卫星数据的多时相、多种类数据分类中做进一步试验,以不断改进和完善树分类器技术.

参 考 文 献

[1] Wu, C. L., Landgreba, D. A. and Swain, P. H., The Decision Tree Approach to Classification, Purdue University, West Lafayette, LARS inform. Note 09011740, 1974.

[2] Argentiero, P., Chin, R. and Beaudet, P., An Automated Approach to the Design of Decision Tree Classifier, *IEEE Trans. PAMI-4* (1983), 51—57.

[3] Hauska, H. and Swain, P. H., The Decision Tree Classifier: Design and Potential, Proc. 2nd Symp. on Machine Processing of Remotely Sensed Data, 1975, 142—147.

[4] Kulkarni, A. V., On the Mean Accuracy of Hierachical Classifiers, *IEEE Trans. Computer,* C-**27** (1978), 771—776.

[5] Moret, B. M. F., Decision Tree and Diagrams, *ACM Computing Surveys,* **14**(1982), 593—623.

[6] Swain, P. H., and King, R. C., Two Effective Feature Selection Criteria for multispectrual Remote Sensing, Proc. First Joint Conf. on Pattern Recognition, 536—540.

A STUDY OF DECISION TREE CLASSIFIER

ZHENG QINFEN WU JIANKANG WANG WENTAO

(*University of Science and Technology of China*)

ABSTRACT

In this paper, a mathematical definition of decision tree classifier is given. Two key problems——optimization of the decision tree classifier and performance measure are studied, so that an approach to the design of decision tree classifier with application to satellite multispectral data classification is proposed, including reasonable performance measure, a special data structure, and optimization algorithm for searching. Finally, experimental results of forest classification is also given.