

基于遥感信息的哈夫曼优化树 在山地土壤资源调查中的应用 ——以浙江省龙游县为例

沙晋明¹, 李小梅²

(1 福建师范大学地理所, 福建 福州 350007; 2 福建师范大学环境科学研究所, 福建 福州 350007)

摘要: 应用遥感手段进行土壤资源调查时, 可以得到土壤的母质类型、土地覆盖现状、地形地貌等特征信息。在二叉树模型(哈夫曼优化树)的支持下, 根据土壤特征信息可以对区域土壤的类、属、种进行逐级识别, 识别精度相对较高。

关键词: 二叉树; 哈夫曼优化树; 土壤资源; 遥感

中图分类号: F301.2 文献标识码: A

利用遥感图像可以获得区域内土壤成土因素的各种动态信息, 但如何充分利用这些数据信息, 使它与地面上的土壤对应起来。这需要将未知土壤单元的成土因素特征与区域内的目标土壤类型特征, 按一定的判别规则逐一分析比较, 来得到未知单元的土壤类型。对土壤类型的这种判别亦可理解为用一定的求解方法对答案进行搜索, 求解方法就是抽象出来的一系列具有层次性的土壤成土因素组合。

Philip H. Swain 等人在 1987 年以地球资源类别为例设计提出了信息树的思想^[1], 他们认为在每个交叉点(即节点)上采用不同的类型提取算法^[2], 但在技术上比较复杂, 所以并没有具体实现。本研究针对土壤类型划分的特点, 对经过修正后的土壤叠加综合图的属性库, 进行土属筛选分析, 构造土壤类型识别二叉树, 即哈夫曼优化树, 采用自适应地选择特征, 来解决土壤多特征多模式类型的识别问题。

1 研究区自然概况

研究区为我国东南地区的浙江省龙游县, 该县位于 28°44'10"N ~ 29°17'15"N, 119°14'1"E, 属于中亚热带气候。由于新构造运动的影响, 龙游县的地形特点为: 南北高, 中间低。南部属于仙霞岭余脉, 山体连绵, 谷峡坡陡。北部属于千里岗山脉, 衢江由西向东

穿过, 形成了中央的河谷盆地。因此该县南部主要为山区, 北部分布大量的丘岗和少量山地, 这种地貌格局决定着土壤成土因素的分异和物质能量的转换, 使本县的土壤类型具有明显的层带状分布规律。该县属于土壤分布带中的红壤带, 在山区的垂直带谱上广泛分布有黄壤、水稻土、潮土、紫色土、石灰岩土、山地草甸土和粗骨土等。

2 二叉树设计

2.1 二叉树

树, 就是不含有回路的连通图。在计算机科学和信息科学中, 树可以作为数据和信息的结构^[2], 也可以表示为一种搜索过程。二叉树作为树的一种, 在信息检索和搜索方面, 占有独特的优势。二叉树是一个有限的节点集合, 它或者为一空集合, 或者由一个根结点及两棵不相交的二叉树组成。此两棵二叉树分别称为根结点的左子树和右子树(图 1)。

每个二叉树由 n 个节点组成 ($n=0$), 节点分内节点(非终止节点)和外节点(终止节点或叶节点)。二叉树的左右子树又分别都是一棵二叉树。在二叉树的定义中, 带有递归的含义, 因而整个树的构造的子树的构造存在一致性。

收稿日期: 2001-12-18。

基金项目: 浙江省教委科研项目“县级红壤资源遥感调查计算机辅助解译技术研究”(19990516)、浙江省自然科学基金“基于三就观模型的土壤遥感计算机辅助解译技术研究”(499043)。

作者简介: 沙晋明(1964-), 男, 博士, 副教授, 现从事土壤遥感到 GIS 研究。

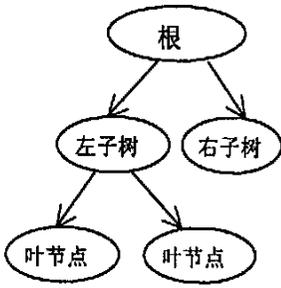


图 1 树结构示意图

Fig. 1 Signal graph of tree structure

2.2 二叉树设计框架

土壤类型识别二叉树设计也是从上三个原则出发,结合二叉树自身特性以及土壤成土因素特点,主要考虑以下方面:二叉树的树特征参数选取,包括内节点、外节点(叶节点)个数和叶通路长度等;内节点上数据子集的特征意义;二叉树分叉原则和判别条件。

土壤类型识别二叉树的根节点为区域内土壤类型的全体,如龙游县所有土类集成: {红壤、黄壤、山地草甸土、紫色土...},二叉树的左子树(内节点)为土壤某种成土因素或几种成土因素组合取值为“否”的土壤类型子集(取值为0),右子树或者为外节点,或者为土壤成土因素取值为“是”的土壤类型子集(取值为1)。内节点上的特征数据是土壤成土因素集合中的元素值,它的判别规则全部为“是否”类型,不存在模棱两可的结论。左右子树的并集一定是其父节点(或根节点)集合,左右子树的交集为空。

2.3 哈夫曼树

在许多应用中,常常将树上的节点赋上一个有意义的实数,把这个实数称为该节点的权,节点的带权路径长度规定为从树根节点到该节点之间的路径长度与该节点上权的乘积。那么,树的带权路径长度定义为树中所有叶节点(外节点)的带权路径长度之和,记为

$$WPL = \sum_{i=1}^n w_i l_i$$

式中 n 为叶节点的数目, w_i 为叶节点的权值, l_i 为根到叶节点的路径长度。

哈夫曼(Huffman)树又称最优二叉树。它是 n 个带权叶节点构成的所有二叉树中,带权路径长度 WPL 最小的二叉树。哈夫曼(Huffman)树的构造算法为:

1. 将与 n 个权值 $\{w_1, w_2, \dots, w_n\}$ 个节点构成 n

棵二叉树,它们组成森林 $F = \{I_1, I_2, \dots, I_n\}$,其中每棵二叉树 T_i 都只有一个权值为 w_i 的根节点,其左右子树均为空。

2. 在森林 F 中选出两棵根节点的权值最小的树作为一棵新树的左、右子树,且置新树的附加根节点的权值为其左右子树上根节点的权值为其左右子树上根节点的权值之和。

3. 从 F 中删除这两棵树,同时把新树进入 F 中。

4. 重复上步骤,直到 F 中只有一棵树为止,此树就是哈夫曼树。

哈夫曼树的应用很广,哈夫曼编码就是其中的一种。本研究根据哈夫曼树的构造思想研究土壤类型识别,并在此基础上提出土壤类型值化代码。

3 二叉树应用

3.1 土壤类型识别

在土壤类型识别体系中,如土类的划分是依据这种土类相对于其他土类的明显特征来确定它的类型。土类以下的土壤亚类的划分,也是在该土类集合内,将成土因素有明显差异的划分为几个亚类。因此在土壤类型划分过程中,不同成土因素或者成土因素中不同的特征属性值所起的作用不同。一种土壤区别于其他土壤的“明显特征”,很显然在土壤类型划分过程中起着重要的作用。将标志土壤类型的“明显特征”提取出来,赋予权重(权重值的大小决定于它在集合中的“特征明显程度”),就可以构造哈夫曼树。反过来,可以利用已构建的哈夫曼树来实现未知土壤类型的识别。

对未知土壤单元而言,要确定它的归属,完全可以从它的成土因素信息出发,沿着土壤类型识别二叉树,从根节点到叶节点逐步搜索,就可以快速准确地识别出它的归属。二叉树上每个内节点的判别规则,可以依据本集合内元素的“明显特征”确定。在此原则下,本文借助于遥感监测信息对龙游全县的所有土壤单元,进行类型识别,最后可以做到对土属的划分。以水稻为例来说明它的亚类划分的二叉树实现(如图2)。

以此类推,利用二叉树提取模型,采用简单易行的“明显特征”提取策略,可以依次构建土类(图3)、土壤亚类(图2,4)和土属层次(图5)上的二叉树。最后将不同土类、土壤亚类、土属二叉树,组合成一棵土壤类型识别巨型树(图6)。对龙游的所有地块

单元按此方法进行土壤类型识别, 类型提取精度达到 90%。具体搜索时, 从根节点出发, 自上而下逐步进行, 先分出土类, 再根据条件找到亚类, 最后找到土属。

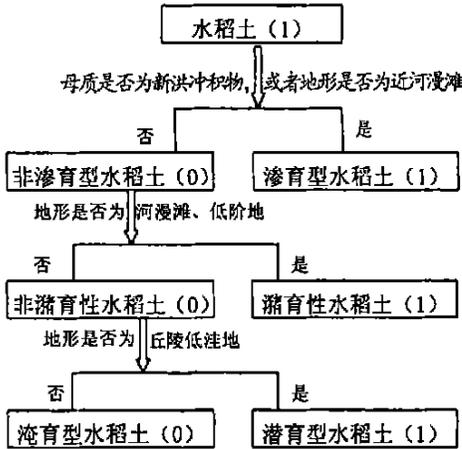


图 2 水稻土的亚类划分二叉树

Fig. 2 Two branches tree to classify paddy soil

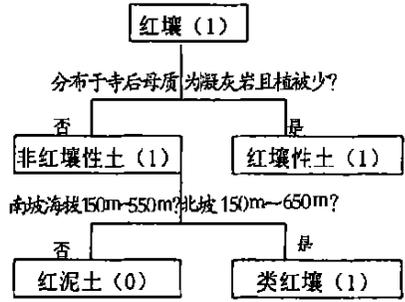


图 4 红壤亚类二叉树构型

Fig. 4 Two branches tree

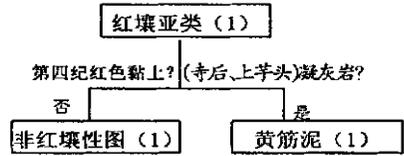


图 5 土属二叉树构型

Fig. 5 Two branches tree model of soil

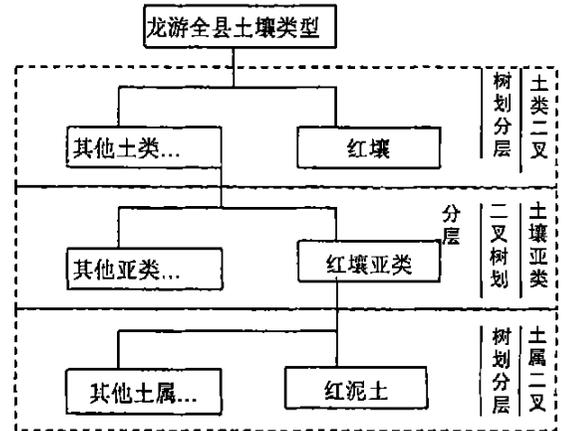


图 6 土壤类型提取巨型树

Fig. 6 Large tree to pick out soil types

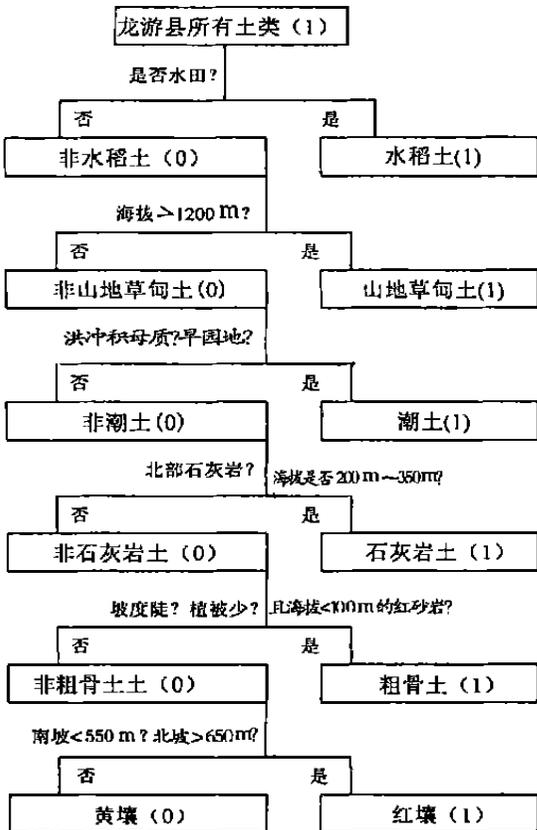


图 3 土类二叉树构造

Fig. 3 Two branches tree of soil groups

3.2 土壤类型的数值化编码

根据以哈夫曼算法构建的土壤类型识别二叉树, 可以对龙游所有土壤类型进行编码^[5]。编码规则为: 根节点的值假定为 1, 根据满足判别条件的一枝取值为 1, 不满足条件的取值为 0 的假设, 由上向下计算每一个外节点的数值代码(二进制), 可以称之为此外节点相对于本树根节点的值。

以水稻土亚类划分二叉树为例, 相对于水稻土而言, 渗育型水稻土的编码为: 11(十进制为 3), 潜育性水稻土的编码为: 101(十进制为 5), 潜育型

水稻土的编码为: 1001(十进制为 9), 而淹育型水稻土的编码为: 1000(十进制为 8)。

将水稻土亚类二叉树移植到上层的土类二叉树上, 按上述方法又可以计算各个水稻土亚类在整棵树上的代码或数值。相对于龙游土壤这个根节点而言, 渗育型水稻土的编码为: 111(十进制为 7), 渗育性型水稻土的编码: 1101(十进制为 13), 潜育型水稻土的编码: 11001(十进制为 25), 而淹育型水稻土的编码: 11000(十进制为 24)。同理, 将红壤亚类土属二叉树移植到红壤亚类二叉树上, 再将它们移植到土类二叉树上, 这样就可以得到土属相对于根节点的数值代码。红泥土的编码为: 100000010010, 黄筋泥的编码为: 100000010011。

这里的红壤亚类作为土属的根点时为 1, 而作为土类的叶节点为 0, 如何连接? 因为当红壤亚类作为土类树上取值为 0 的叶节点时, 可以将它延伸到下一级, 使

其在树上取值为 1, 这样就实现了对等连接。

由上可以得到这样的结论: 在二叉树上的每个土壤类型(外节点)都可以用数值或代码来表示, 而且不同的土壤类型在一定的类型提取规则下, 数值不同, 这样就可以实现土壤类型数值化。进一步可以在计算机上, 实现土壤类型的高效自动化判别。

3.3 土壤类型识别二叉树的计算机实现方法^[3]

基于以上对土壤类型识别二叉树的模型分析以及具体实现过程中特征选和判别条件的确定, 利用遥感影象获得的土壤单元的成土因素特征, 可以在数据库中直接实现土壤类型的二叉树判别。

土壤类型识别二叉树用动态链表作为它的存储结构如图 7 所示。图中 Data、Data1、Data2 为节点集合, 集合中每个元素用一个二维数组描述, 外节点(叶节点)的 Data 为具体的结果值。Link、Rlink 分别为左右指针, 指向左子树和右子树。

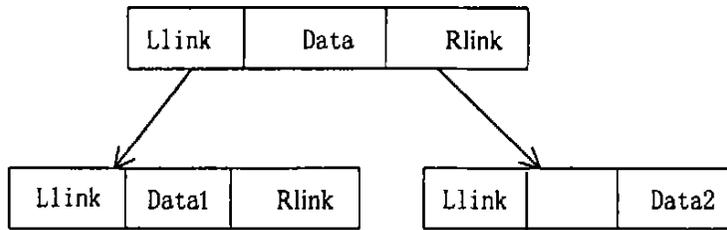


图 7 二叉树的链表存储结构

Fig. 7 Memory structure in linking the two-branches-tree

判别函数表(表 1)为每一步的判别特征, 即对三个成土因素中的特征选取。最后可以形成一个判别结果表(表 2)。根据这种思路, 可以用 Visual Foxpro 或其他高级编程语言实现。本研究在 Visual Foxpro 环境下, 精度可以达到 90%。

表 1 土壤二叉树的判别函数表

Table 1 Discriminant function of two-branches-tree for soil

判别序号	特征 1	特征 2
1	1	0	0 0...
2	0	1	0 0...
3	0	0	1, 0 0...

自动判别。

2. 土壤类型二叉树的设计同时使土壤类型数值化表达成为现实, 这种编码本身就可以反映出土壤类型的特征。对照土壤类型判别特征表, 从编码就可以恢复出某种土壤类型在二叉树上的位置, 并了解到这种土壤类型的其他特征。

3. 土壤类型二叉树是以土壤的“明显特征”为线索而构造的, “明显特征”的选择和微象是二叉树构造的关键所以需要丰富的专家知识和深厚的专业功底。

表 2 土壤二叉树的判别结果表

Table 2 Discriminant results of two-branches-tree for soil

未知单元	外节点
1	DataN
2	DataM
...	...

3 结论

1. 土壤类型的二叉树判别, 使多特征多模式的土壤类型划分清晰条理, 克服了以往土壤类型识别中指标多而且混乱不清的缺点, 易于在计算机实现

参考文献:

- [1] 杨世仁. 空间数据的三角形四叉树结构及其基本操作[J]. 环境遥感, 1995, 10(4): 269~273.
- [2] 王振宇. 树的枚举与算法复杂性分析[M]. 北京: 国防工业出版社, 1991, 1~13.
- [3] 舒贤林, 徐志才. 图论基础及其应用[M]. 北京: 北京邮电学院出版社, 1988, 86~94.
- [4] 徐孝凯. 数据结构简明教程[M]. 北京: 清华大学出版社, 1994, 109~112.
- [5] 曾小明. 逐步判别树型分类器的设计与应用[A]. 见方有清, 等. 遥感图像分析方法及其在林业中应用研究[C]. 南京: 南京林业大学, 1986 121~123.
- [6] Xiaping Jia and John A Richards. Progressive Two-Class Decision Classifier for Optimization of Class Discriminations. Remote sensing of Environment, 1998, 63(3): 289~293.

Application of Huffman Tree Based on the Information of Remote Sensing for Investigating Soil Resource

SHA Jin-ming¹ and LI Xiao-mei²

(1. Institute of Geography, Fujian Normal University, Fuzhou 350007 China;

2. Institute of Environmental Sciences, Fujian Normal University, Fuzhou 350007 China)

Abstract: When investigating soil resources, some information such as soil primitive rocks, land cover, land form and geomorphy can be drawn from images of remote sensing. According to the information, soil category, family and species may be gradually distinguished through the model of two branches tree.

Key words: two-branches-tree; huffman tree; soil resources; remote sensing