

土壤分类专家系统设计

汪善勤^{1,2}, 周 勇¹, 张甘霖³, 舒 宁²

(1. 农业部亚热带土壤资源与环境重点实验室, 武汉 430070; 2 武汉大学遥感信息工程学院, 武汉 430070; 3 中国科学院南京土壤研究所, 南京 210008)

摘 要: 土壤系统分类以定量化指标划分土壤类型, 为土壤自动分类系统的建立提供了可能。为了模仿专家的思维过程实现土壤自动分类, 该文着重研究了如何采用面向对象的方法表达专家经验规则、诊断层和诊断特性知识, 在此基础上提出了土壤分类决策树的构造方式和推理方法; 并从土壤系统分类专家的知识表示和推理机的构建两个方面, 提出了土壤分类专家系统 (SCES) 的体系结构。GIS 作为管理空间信息和提供空间分析方法的平台, 也被有效地集成到系统中, 以实现土壤自动分类的目的。

关键词: 土壤系统分类; 专家系统; 地理信息系统

中图分类号: S15

文献标识码: A

文章编号: 1002-6819(2005)ZK-0039-05

0 引言

土壤分布具有较强的规律性, 而土壤分类系统则用定性过程或定量指标描述了这些规律, 从而为土壤自动分类提供了理论基础。到目前为止, 还没有比较完整的土壤自动分类系统出现。20 世纪 60 年代, 研究人员通过比较土壤分类与生物领域的数值系统分类法, 找出土壤分类的特点, 在计算机上实现土壤数据分析和分类^[1,2]。1975 年美国土壤系统分类建立以后^[3], 人们开始运用计算机技术、数据库管理技术、空间分析模型以及遥感等新技术, 与非层次型数值分类法一起用于分析和组织区域土壤数据, 并建立了检索信息系统^[4,5], 同时在比较土壤剖面的应用中取得了成效^[6]。20 世纪 90 年代初, 模糊分类方法被 Burrough 等人用于土壤分类, 提高了计算机分析的精度^[7]。Fisher 等用 Turbo Prolog 语言开发了文字界面的土壤系统分类的应用程序, 用户使用时需要预先判断诊断层和指出诊断指标数据之间的差异^[8,9]。由于这些系统没有利用专家经验和知识, 还不能称为真正的专家系统。1996 年, Galbraith 等使用 Net Weaver 工具, 用语义网络表达系统分类的知识体系, 开发了有机土纲、火山灰土纲、灰土纲和氧化土纲的检索专家系统^[10,11], 此系统加入了专家的经验, 能够按照检索系统完成推理。但是, 系统中专家经验所占比重小, 只在指标数据较少或缺乏的情况下, 才按照专家规则推理, 结果的可信度受到影响。近年来, 地理信息系统 (GIS)、全球定位系统 (GPS) 和遥感 (RS) (简称 3S) 技术, 成为获取土壤属性和空间信息的有效手段。尤其是利用 GIS 建立土壤类型分布与地理要素的空间分析模

型, 为实现土壤自动分类提供了新的技术^[12,13]。

我国 20 世纪 70 年代末建立了中国土壤分类系统, 90 年代又建立了中国土壤系统分类。两种分类在我国的土壤信息系统研制中都起着非常重要的作用^[14-16]。钟骏平等在 Galbraith 基础上开发了干旱土纲的专家系统^[17,18], 杨胜天等运用遥感技术, 获取植被覆盖度及类型, 结合地形地貌、地质等景观数据, 在已经确定土壤类型的情况下, 根据专家规则, 获取土壤分布状况, 实现了土壤制图计算机化^[19]。同时我国一些地区已经建立了土壤信息数据库^[20-22], 为建立中国土壤分类专家系统打下了基础。

以土壤系统分类的方法作为基础, 模拟专家判断的过程, 综合地形地貌、水文地质、土地利用方式和地带性植被分布等信息, 缩小检索范围, 再用定量指标分析确定土壤类型, 是设计土壤分类专家系统的指导思想。本文将从土壤分类专家的知识表示和推理机的构建两个方面, 探讨土壤自动分类专家系统的体系结构; 探索基于 GIS 和 ES 的土壤自动分类系统的体系结构。

1 土壤分类专家系统知识表示

土壤分类专家系统中使用的知识可分为两类: 一类是专家知识和经验, 另一类是分类体系中的诊断层和诊断特性。前者有一定的随意性; 后者体系结构严谨, 因此采取不同的方法表示两类知识。

1.1 分类检索体系的设计

土壤系统分类的检索系统具有明显的层次性体系结构。土纲作为最高土壤分类级别, 根据主要成土过程或影响主要成土过程的性质, 以主要诊断层和诊断特性的出现与否作为划分的依据, 因此土纲可作为独立实体对象看待; 从亚纲到土系, 在拥有上一层土纲所有的诊断层和诊断特征之外, 划分依据则以不同级别中影响差异最大的诊断特性为主。每一个诊断层和诊断特性也可作为独立对象处理, 在分类体系中作为主要的判断依据, 可以构成子体系。由此, 采用面向对象形式描述的分类体系, 由土类的层次结构与诊断层和诊断特性子体系构成 (图 1)。

收稿日期: 2003-03-28 修订日期: 2003-08-28

基金项目: 国家自然科学基金资助项目 (40271056); 国家高新技术研究发展计划 (863) 子课题 (2002AA 2Z4071-12)

作者简介: 汪善勤 (1974-), 男, 湖北汉川市人, 讲师, 主要从事地理信息系统与遥感应用、土地资源调查与评价、土壤地理等领域的研究。武汉市 华中农业大学资源与环境学院, 农业部亚热带土壤资源与环境重点实验室, 430070

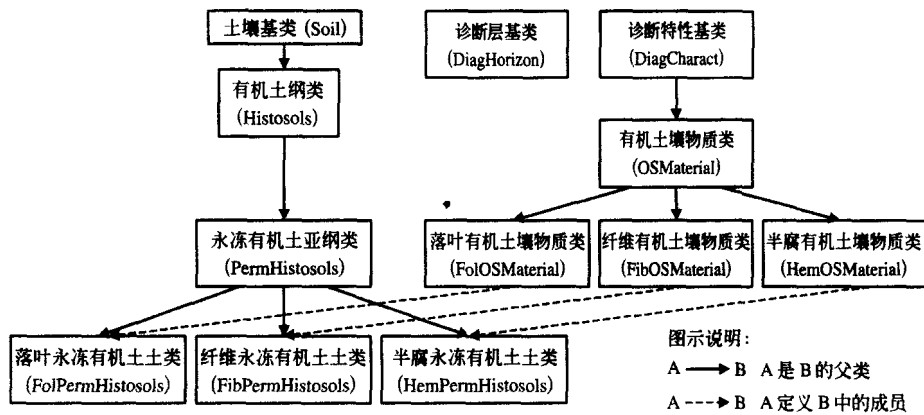


图1 土壤系统分类中类的定义结构框图

Fig 1 Classification definition framework of types in soil taxonomy

1.2 诊断层和诊断特性的表示方法

诊断层类由DiagHorizon 基类派生, 诊断特性由DiagCharact 基类派生, 这两个基类定义所有诊断层和诊断特性的公共属性和函数(图2)。诊断层和诊断特性知识可划分为二值判断(是或否)、数值范围、逻辑组合(和与或)、推断、曲线描述等5种形态, 均定义为类的属性和方法, 实现知识的表达和应用。土壤类中定义诊断层或诊断特性的对象成员, 建立土壤类与诊断层或诊断特性之间的联系(图1)。

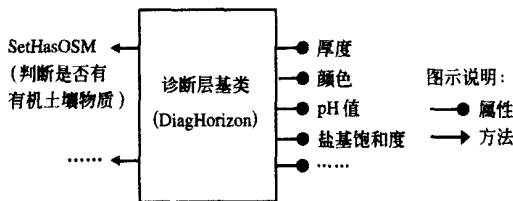


图2 诊断层基类定义

Fig 2 Classification definition of diagnosed horizon

1.3 专家知识与经验的表示方法

母质、地形与地貌、气候、生物和时间五大成土因素, 是专家鉴别土壤时需要获取的重要信息。在本系统中, 模拟专家的判断过程, 以信息流为中心, 组织专家规则。实际操作前, 系统需要尽可能多的获取五大成土因素的信息, 判断样本的土纲类型, 并依此指导下一级的分类过程。另外, 在诊断层和诊断特性中也反映了这五大成土因素, 如果缺乏土壤理化数据或短期内无法得到时, 专家规则可作为替代规则完成推断过程。

专家规则以产生式规则形式为主, 大部分定义在土壤基类Soil中, 作为确定土纲类别范围的依据。考虑到仅用一种成土因素无法缩小土纲的搜索范围, 需要清楚地确定某类或某几类土纲的决策因素集。例如, 中国的有机土的分布与地形地貌和气候有很大关系, 主要出现在地形部位低洼、气候冷湿的地区, 改写成产生式规则的形式, 则描述如下:

IF is(地形, in(平原, 高原, 山区))

AND is(微域地形, in(碟形洼地, 谷地, 阶地, 洼地,))
 AND 年平均气温 < 10
 AND is(年降水量, between(400, 2000))
 AND // 剖面构型, 有机表层颜色等其他判断依据
 THEN
 value = is Possibly (土壤样本, 有机土)

2 推理过程

2.1 专家经验知识的推理过程

专家规则集中定义在土壤基类中, 另外在各级子类中也有部分定义。通常, 专家推理过程有明显的倾向性, 表现为专家规则可以排序, 即可确定出规则的重要性。因此, 基于专家规则的推理过程, 以构造最小决策树, 采用深度优先搜索方法实现。按照最小决策树的构造思想, 可将前面的专家经验的形式定义细化为推断有机土的最小决策树(图3)。

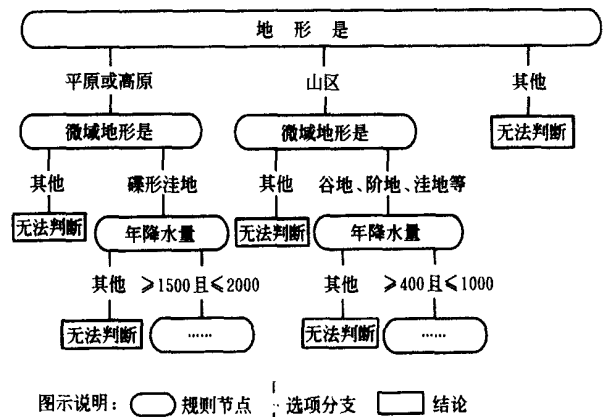


图3 判断有机土的最小决策树(部分)

Fig 3 Least decision tree for deducing organic soil (part)

2.2 诊断层和诊断特性的推理过程

对于诊断层或诊断特性规则中的条件, 相互之间存

在与和或两种逻辑关系。因此建立的决策树与专家经验决策树有所不同。以有机表层为例, 则建立如图4中所示的决策树。

其中, 非终端节点表示其与子节点的逻辑关系, 叶子节点表示各项指标的条件值。针对不同的逻辑节点采取不同的搜索策略: 宽度优先搜索以“与”节点为根节点的子树; 深度优先搜索以“或”节点为根节点的子树。

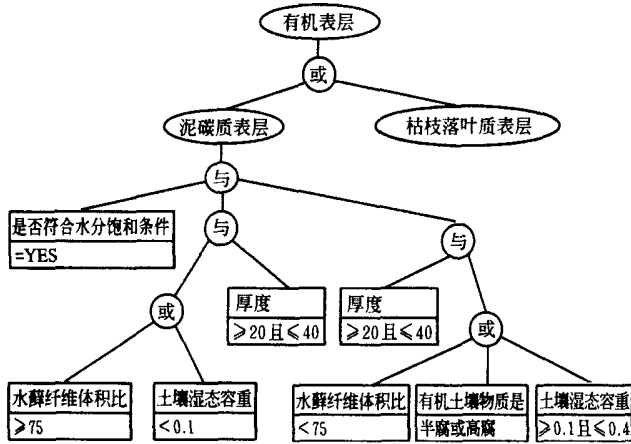


图4 有机表层决策树(部分)

Fig 4 Decision making tree of organic soil horizon (part)

2.3 数据的完整性检验

检验数据的完整性分为两步进行。样本数据划分为必要数据和充分数据。第一步检验必要数据的完整性。首先为每一个决策树建立必要条件集, 此集合也可作为判断推理是否能够继续进行的依据。第二步在搜索过程中检验充分数据的完整性。决策树的节点增加信息域记录该节点的访问情况。当推理过程终止时, 判断终止节点, 如果是“与”节点, 则其子节点中所表示的条件部分成立, 信息域中置1, 表示数据不完整; 如果是“或”节点, 则所有子节点表示的条件都不成立, 信息域同样置1。对已经搜索的逻辑节点, 信息域均置0, 表示该节点的子节点数据是完整的。

2.4 系统的推理流程

综合以上三个过程, 系统的推理流程如下:

- 1) 根据可供判断的初步信息和专家经验规则确定下一步进入的子类, 并将数据传递给子类;
- 2) 构造子类的诊断层或诊断特性类成员, 赋予相关数据;
- 3) 调用诊断层或诊断特性类的检验函数, 对数据的完整性和一致性进行检验。数据完整且一致, 则执行步骤4); 否则, 执行步骤5);
- 4) 调用诊断层或诊断特性的规则推理函数, 对实验数据进行判断。记录推理结果, 执行步骤6);
- 5) 调用数据分析函数, 如果数据充足, 执行步骤1); 否则, 执行步骤6);
- 6) 结束推理。

3 GIS 与土壤分类专家系统结合

在土壤分类专家系统中, 使用GIS建立空间数据

库、分析空间信息和图像显示。空间数据库中包含样点区的母质、地形等信息; 采样点的空间定位信息和理化分析数据的点位图。在此基础上, 以母质、地形地貌、水文等专题图为底图, 以采样数据为输入参数, 以克里金插值为基础, 建立各种土壤理化数据的空间分布模型。专家系统利用这些数据和模型就可以推断出非采样点图斑的土壤类型。就目前的研究来看, 很难找到求解精度高、适用性广的扩散方程。在实际过程中, 先通过专家系统判断得到采样点的土壤类型; 然后以母质、地形地貌等分布图为底图, 采用模糊分类的方法得到相同土壤类型的区域分界。为了实现此过程, 系统设计时重点解决以下问题。

3.1 GIS 与ES 之间的数据传递

GIS 与ES 之间传递的数据可分为两类, 一类是GIS 中存储的土壤母质、地形地貌、剖面描述以及剖面点的空间定位信息。另一类是ES 的推断结果, 其中包括土壤类型或者无法完成推理的错误信息。

首先, 剖面的综合信息由GIS 从图形库中提取。各种图件均以栅格文件的形式存储, 然后对所有图件进行叠加, 生成包含所有剖面信息栅格化的综合覆盖层文件。根据已经记录的剖面定位坐标, 提取出综合覆盖层中对应栅格中的所有数据, 生成一个临时数据库文件, 该文件的格式、结构、字段名和属性与ES 中的数据库文件保持一致或兼容。以上过程以紧密集成的形式实现, 将GIS 提供的外部函数和ActiveX 控件封装到用GIS 和ES 应用程序接口(GEApi)类中, 系统运行时, GEApi 负责从GIS 中接收数据和生成临时的数据库文件, 然后向ES 发送数据分析请求消息(图5)。

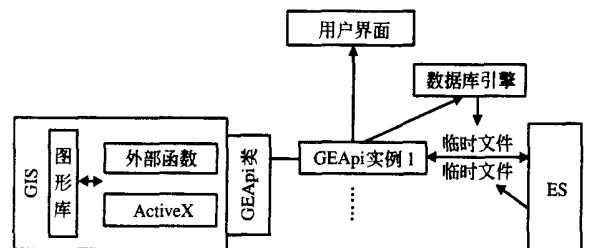


图5 GIS 与ES 间的数据传递

Fig 5 Data transmission between GIS and ES

专家系统的推断结果及时反馈给GEApi, 结果经过分析和重构后, 可被GIS 使用, 土壤类型信息将被写到综合覆盖层文件中。如果ES 无法完成推理时, 则向GEApi 发出错误信息。

3.2 土壤类型空间分布模型

按照发生分类学的观点, 土壤分布由母质、气候等五大成土因素决定。而系统分类理论认为土纲的划分主要由土壤的理化指标(诊断层和诊断特性)决定, 从亚纲到土系则主要以气候、土壤颜色、成土过程、新生体等加以区别^[7, 8]。考虑到土壤制图的范围, 不同的制图比例尺和精度, 这些因素的选择也将有所不同。如大比例尺的制图, 系统分类中的因素为主, 兼顾发生分类的母质

等因素。而进行小比例尺制图时,则以发生分类的因素为主。专家的经验可以有效地指导因素的选择。因此,需要建立与各种比例尺对应的选择因素的规则,建立因素集。在获取了相应的数据,并且通过专家系统得到剖面点的土壤类型,即可以这些剖面点作为种子点。采用克里金插值算法,求解出非剖面栅格点的土壤类型。

4 数据流程图结构

系统设计由用户界面(解释器)、专家系统、地理信息系统和检索系统4部分组成(图6)。用户界面除了提供输入数据和输出结果的表格和图形外,对用户输入的数据和信息由解释器来完成。对用户输入的数据进行分级,必要和非必要的条件,在非必要条件较少的情况下,系统将驱动专家系统和地理信息系统提供辅助信息,响应用户的请求。在非必要条件较多的情况下,可以将用户的请求转化为检索请求,交给检索系统;如果检索结果不确定,再交给专家系统完成,或提示用户补充相关信息。

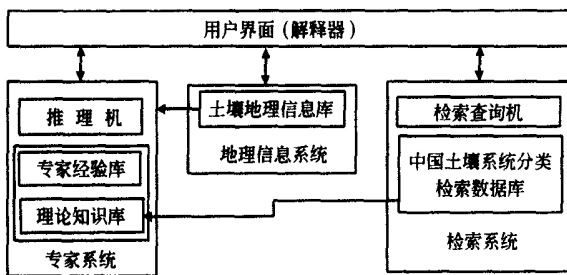


图6 系统数据流程图

Fig 6 Diagram of system data transmission

专家系统的推理机和知识库共同模拟土壤分类专家的行为。首先,推理机接受用户提出由解释器分类和验证的土壤剖面信息和试验数据,按要求从土壤地理信息库中提取相关的辅助信息;然后根据专家经验缩小目标范围;其次从理论知识库中搜索决策信息,构造最小决策树;最后得到分类的结果。这个结果可能不是唯一的,系统可以在推理过程中,计算出结果成立的概率。

地理信息系统中存储有关土壤的地理信息,包括土壤母质分布图、土地利用状况、区域降水分布图、其他地带性因素的分布图等,以空间数据的形式提供使用。此系统的功能包括:

- 1) 组织与管理空间数据;
- 2) 提取剖面点的地形、母质分布、土地利用状况、地下水位等信息;
- 3) 在解释器的驱动下,为专家系统提供辅助决策的空间信息;
- 4) 根据专家系统推理或检索的结果,以地形图、母质分布图、土地利用图为底图,模拟绘制土壤图。

检索查询机驱动检索数据库,搜索根据用户提供的数据和信息,采取一定的搜索策略,在检索数据库中查找相匹配的记录,并将结果返回用户。检索数据库是以中国土壤系统分类检索系统为基础建立的,采用关系数据库系统管理。同时,作为专家系统的知识库的一部分,

从中发掘规则,建立理论知识库。随着检索系统的完善和更新,对理论知识库进行同步修改和补充。

5 结语

1) 以专家系统为核心建立的土壤自动分类系统,能够充分利用多年积累的土壤调查资料,提高数据的利用率。系统采用面向对象的设计方法定义土壤类型,将专家经验和理论知识集中在对象中,并融合了发生分类学的知识,作为系统分类的补充,提高了专家系统的效率和应用范围。

2) 土壤自动分类系统的知识库能够随着系统分类的研究进展,独立的进行更新和完善。系统的各个部分可以独立设计和运行,也可以集成在一起,成为一个完整的应用软件,可扩充能力较强,能够以组件的形式提供数据服务。

3) 土壤自动分类系统引入了GIS,提出了土壤分类专家系统应用地理信息技术的方法,建立了一套较为合理的技术流程。土壤自动分类系统引入GIS的功能,使得土壤自动制图成为可能。

4) 由于土壤的类型众多,在应用数据库的规则发掘技术的基础上,建立土壤知识的自动生成系统,从积累的大量剖面数据中提取知识,有利于快速建立完整的专家知识库。

[参 考 文 献]

- [1] Rayner J H. Classification of soil by numerical methods [J]. Journal of Soil Science, 1966, 17: 79- 92
- [2] Russell J S, Moorr A W. Use of a numerical method in detemining affinities between some deep sandy soils[J]. Geodema, 1968, 1: 47- 68
- [3] Soil Survey Staff. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. U SDA - NRCS Agric Handbook 436[M]. U. S. Government Printing Office, Washington, D C. 1975
- [4] Gornden A D. Classification and assignment in soil science[J]. Soil Use and Management, 1987, 3: 3- 8
- [5] Dale M B, M cB ratney A B, Russell J S. On the role of expert system and numerical taxonomy in soil classification[J]. Journal of Soil Science, 1989, 40: 223- 234
- [6] Little L P, Ross D R. The Levenshtein metric, a new means for soil classification tested by data from a sand-pdzol chronoquence and evaluated by discriminant analysis[J]. Australian Journal of Soil Research, 1985, 23: 115 - 130
- [7] Burrough L, Friedmen J H, O lshen R A, et al. Classification and Regrssion Trees[M]. W adsworth, Belmont, C A. 1984
- [8] Fisher P F, Blachandr C S. STAX: A Turbo-Prolog rule-based system for Soil Taxonomy[J]. Computers and Geosciences, 1989, 15: 295- 324
- [9] Buol S W, Rebertus R A. Soil taxonomy: Keys to classification[C]. SM SS Computer Software Programs NO. 1. U SA D, U SDA -NRCS, Washington, D C. 1991
- [10] Galbraith J M, Bryant R B. A functional analysis of soil

- taxonomy in relation to expert system techniques [J]. Soil Science, 1998, 163(9): 739- 747.
- [11] Galbraith J M, Bryant R B, Ahrens R J. An expert system for soil taxonomy [J]. Soil Science, 1998, 163 (9): 748- 758
- [12] Zhu Z, Band L, Vertessy R, et al Derivation of Soil Properties Using a Soil Land Inference Model (SoL M) [J]. America Journal of Soil Science Society, 1997, 61: 523- 533
- [13] S De Bruin, W ielemaker W H, Molenaar M. Formalisation of Soil- landscape Knowledge through interactive hierarchical disaggregation [J]. Geodema, 1999, 91: 151- 172
- [14] 周 勇, 汪善勤, 王庆云, 等. 建立土地资源信息系统的若干问题与对策 [J]. 遥感学, 1999, 3(1): 71- 76
- [15] 全国土壤普查办公室. 中国土壤 [M]. 北京: 中国农业出版社, 1998
- [16] 龚子同, 等. 中国土壤系统分类理论、方法和实践 [M]. 北京: 科学出版社, 1999
- [17] 蒋平安, 钟骏平, 李保国. 土壤系统分类中诊断层和诊断特性的计算机辅助识别方法 [J]. 新疆大学学报, 1999, 22 (2): 100- 105
- [18] 钟骏平, 闵 勇, 蒋平安. 中国土壤系统分类检索及数据库系统 [J]. 新疆大学学报, 1999, 22(3): 215- 218
- [19] 杨胜天, 朱启疆, 李天杰. RS 和 GIS 支持下的土壤系统分类制图方法研究—以贵州省贵阳市为例 [J]. 土壤学报, 2001, 38(1): 41- 48
- [20] 周 勇, 王庆云. 湖北省土壤系统分类数据库系统的建立 [J]. 华中农业大学学报, 1996, 15(6): 540- 546
- [21] 王庆云, 徐能海, 周 勇, 等. 湖北省土系概要 [M]. 武汉: 湖北科学技术出版社, 1997.
- [22] 章明奎, 魏孝孚, 厉仁安. 浙江省土系概论 [M]. 北京: 中国农业科技出版社, 2000

Design of soil classification expert system

Wang Shanqin^{1,2}, Zhou Yong¹, Zhang Ganlin³, Shu Ning²

(1. Key Laboratory of Subtropical Soil Resources and Environment, Ministry of Agriculture of China, Wuhan 430070, China; 2. College of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430070, China; 3. Institute of Soil Science, Chinese Academy of Science, Nanjing 210008, China)

Abstract: The successful design of Soil Classification Expert System (SCES) depends on the soil taxonomy that classifies soil types using quantitative data. This paper presents an oriented-object based method to simulate the rational thinking of soil classification experts to make rational rules of expert knowledge, soil diagnostic horizon and diagnostic characteristic theory. The structures of soil classification decision trees and reasoning machine were also discussed to construct the framework of SCES. The GIS was also integrated into the system to provide geographic data and spatial analysis means to implement soil classification.

Key words: soil taxonomy; expert system; geographic information system (GIS)