

基于 GIS 的中国土壤分类专家系统设计*

汪善勤^{1,2} 周 勇^{2†} 张甘霖³

(1 武汉大学遥感信息工程学院, 武汉 430072)

(2 华中农业大学资源与环境学院, 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依据; 地理信息系统(GIS)的应用扩展了土壤空间信息表达的能力, 为实现土壤分类自动制图提供了手段。本文着重模仿专家的思维过程, 探讨如何实现土壤自动分类。(1)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, 并在此基础上提出了决策树的构造方式和推理方法;(2)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、土壤自动分类与地理信息融合的方法和基于GIS和专家系统(ES)的土壤自动分类检索系统的框架结构。

关键词 土壤系统分类; 面向对象; 决策树; 专家系统; 地理信息系统

中图分类号 X144 **文献标识码** A

我国已经建立了较为完整的土壤系统分类体系, 并在量化和标准化方面达到了较高的程度^[1]。但是, 由于土壤分类专业性强, 理论方法大多只被高等院校和科研院所从事土壤分类教学和研究的专家所掌握, 使用范围窄, 难于推广和应用到其他领域。而且使用检索系统时, 需要充足的指标数据, 才能确定类型归属。在实际的操作中, 获取地形地貌、水文地质、土地利用方式和地带性植被分布等信息后, 结合专家的判断来缩小检索范围^[2]。建立土壤自动分类专家系统一方面可以汇集土壤分类专家的宝贵经验, 解决土壤类别鉴定的复杂问题; 另一方面能够进一步完善土壤系统分类的知识体系。

土壤类型的自动识别在土壤科学研究、土地利用规划、农业生产实践、环境工程设计及教育等行业都有着重要的研究和应用价值, 成为近年来土壤学及相关领域的研究热点之一^[3]。1990年代中后期, Galbraith等以美国土壤系统分类检索系统为基础, 建立了土壤分类专家系统^[4,5]。钟俊平、蒋平安等按照 Galbraith的方法, 建立了中国干旱土的专家分类系统^[2,6]。这类系统应用知识规则和决策树, 组织和表达检索系统的知识体系, 能够自动完成检索过程。但由于检索系统要求判定数据全面而且准

确, 在实际操作时会因为数据不全而无法实施。

随着地理信息系统、地统计学、模糊数学、遥感等技术与方法在土壤学中应用的不断深化, 土壤类型自动识别有了更先进的分析技术和数据获取手段。Bogaert分别从土壤图和试验测定中获取的精度不同的数据, 用贝叶斯最大熵值法估计空间范围内任意地点的土壤属性, 与地统计学相比, 该方法具有很强的土壤学理论基础, 减少了随机性推断的误差^[7]。各种数学模型与GIS结合分析土壤类型的空间分布也有报道, 如基于Markov链的二维模型、Logistic模型、神经网络模型等^[8-10]。有报道根据土壤类型分布与环境要素之间的紧密关系, 建立了基于知识的模糊土壤制图系统(SoLIM), 将专家系统、模糊数学和GIS融为一体, 为土壤类型识别提供了新方法^[11-14]。在SoLIM中, 分别实现了基于规则和案例的推理模型, 前者假设影响土壤分布的环境要素具有独立性; 后者以案例为基础, 消除了获取详细土壤-环境要素关系的困难和独立性假设的局限^[15]。遥感数据一直是研究土壤属性空间分布的重要数据源。研究发现, 土壤表面光谱特征能够反映出有机质、质地、CEC等的空间分布情况。一般从遥感数据中很难直接获取单纯的土壤表面辐射能量, 通过应

* 国家自然科学基金项目(批准号:40271056)资助

† 通讯作者, E-mail: yzhou@mail.hzau.edu.cn

作者简介: 汪善勤(1974~), 男, 博士研究生, 讲师。主要研究方向为地理信息系统与遥感应用、土地资源调查与评价

收稿日期: 2004-09-13; 收到修改稿日期: 2005-03-14

用植被指数、土壤线等方法从图像中提取裸土像元,或去掉部分植被辐射能间接得到裸土信息^[16]。Chang 等多时相遥感数据的亮温值和土壤湿度数据推断土壤质地^[17]。Hengl 将地下水埋深、坡度、湿度等 9 种要素用于地形监督分类^[18]。这些技术与方法为获取调查区内的土壤信息提供了途径。这些信息和母质、气候、地形、土地利用等空间信息,共同作为土壤自动分类的判断数据,为实现土壤类型的自动识别打下了基础。

本研究针对中国土壤分类专家系统(SCES)的设计和实现方法,以中国土壤系统分类(CST)的检索系统为骨架,以土壤发生学为基础,将专家知识和检索规则进行有效的组织和融合,达到自动识别土壤类型的目的;在 GIS 和专家系统平台上,构建土壤自动分类专家系统的体系结构;同时研究土壤类型判定的知识表达、推理机和模型库的建立与融合方法。

1 知识表示

逻辑表示法、规则表示法、框架表示法、语义网络表示法、面向对象表示法等是表达专家系统的常用方法。本文建立的知识库中包含两类知识,一类是 CST 的检索系统,另一类是土壤发生学理论和专家经验。前者具有层次性体系结构,包括对各级类型划分的定义、鉴定指标描述、特征分类等,因此采

取规则表示法、框架表示法和对象表示法的混合模式来描述;后者多以逻辑判断为主,可采用产生式规则的形式来表达。

1.1 知识库初始化

知识库的建立过程包括初始化、补充、修改和求精等。CST 检索系统中的各级土壤类型、诊断层和诊断特性是确定土壤类型的基础,可用来初始化 SCES。土纲、诊断层和诊断特性有独立的定义,按照对象表示法,建立类层次结构。土纲作为最高土壤分类级别,根据主要成土过程产生的或影响主要成土过程的性质划分,有不同的诊断层和诊断特性。土壤分类层次中,具有明显的继承关系,定义基类 Soil 派生出土纲,土纲派生出亚纲直到土系;子类开放式继承父类所有特征。诊断层和诊断特性分别建立类层次结构,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。诊断层、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、曲线描述等五种形态,均定义为类的成员变量和成员函数,实现知识的表达。土壤类中定义了划分时的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。因此,SCES 的初始知识库由土类的层次结构与诊断层和诊断特性子体系构成,采用面向对象形式构造基本框架。部分示例见图 1。

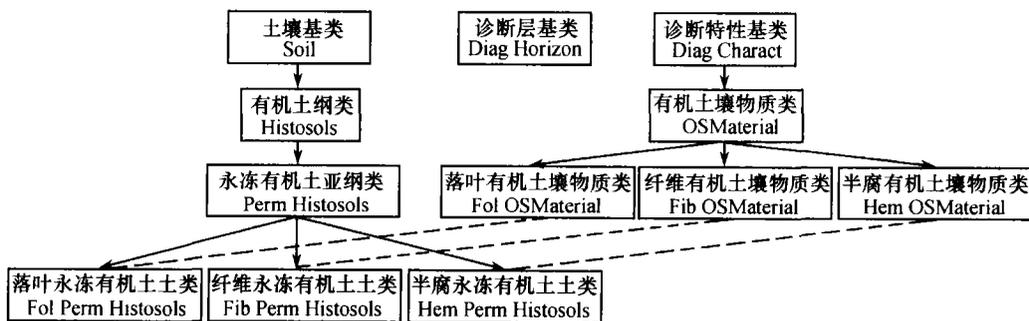


图 1 中国土壤系统分类中类的定义结构框图
Fig. 1 Definition of types in the Chinese Soil Taxonomy

1.2 专家知识与经验的表示方法

专家的知识与经验对土壤分类有着不可替代的作用。土壤发生分类中的母质、地形地貌、气候、生物和时间五大成土因素,在实际工作中,尤其是在野外调查中被当作鉴别土壤的主要依据。在本系统中,将发生学中的有关知识划分到专家知识和经验

的范畴,建立与系统分类的对应关系,形成专家规则。系统根据五大成土因素的信息,对待定土壤的土纲级别做出大致判断。

专家知识与经验的表达实际上是将专家的思维过程规则化。专家经验与知识通过问卷调查的形式获取,然后改写成产生式规则。这些规则大部分在

基类 Soil 中描述,作为确定土纲类别范围的依据。例如在我国,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 notlessthan(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 = isPossibly(土壤样本, 有机土)

```

2 推理过程

在上述知识结构体系中,类层次结构中的派生

或继承关系是知识描述的主体,而专家经验和知识采用产生式规则表达。CST 检索系统需要完备的指标数据才能够完成检索过程,而且专家知识规则中也需要较全的五大成土因素信息,因此有必要在推理开始前,对数据的完整性进行检验,保证推理的顺利进行,并能够在数据确实或错误的情况下给出提示。

2.1 专家经验知识的推理过程

专家经验规则大部分在土壤基类中定义。专家推理过程通常有明显的倾向性,表现为专家经验规则可以排序,即可确定出规则的重要性。因此基于专家规则的推理过程,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,以有机土的推断为例,将专家经验的形式定义细化为最小决策树(图 2)。

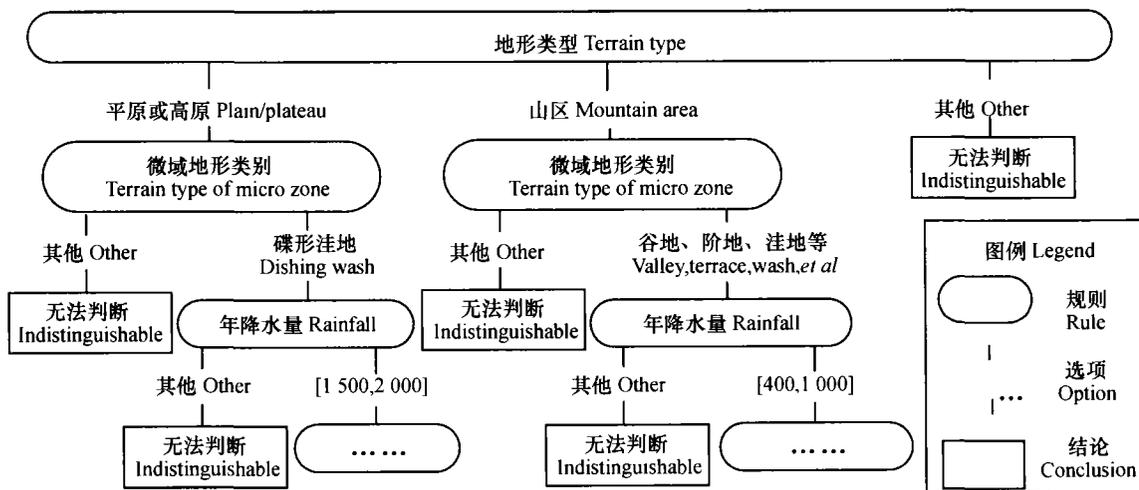


图 2 判断有机土的最小决策树(部分)
Fig.2 The least decision tree for identifying Histosols(part)

2.2 诊断层和诊断特性的推理过程

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 3 中所示的决策树。其中,非终端节点表示其孩子节点的逻辑关系,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和深度优先搜索以或节点为根节点的子树。

2.3 数据的完整性检验

检验数据的完整性分为两步进行。用户提供的数据可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。首先需要为每一个决策树建立必要条件集,用户界面根据此集合,提示用户必须输入各项数据或信息。而且此集合也为判断推理是否能够继续进行提供了依据。第二步在搜索过程中检验充分数据的完整性。决策树中的节点需要增加一个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,则其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,表示该节点的孩子节点的数据是完整的。

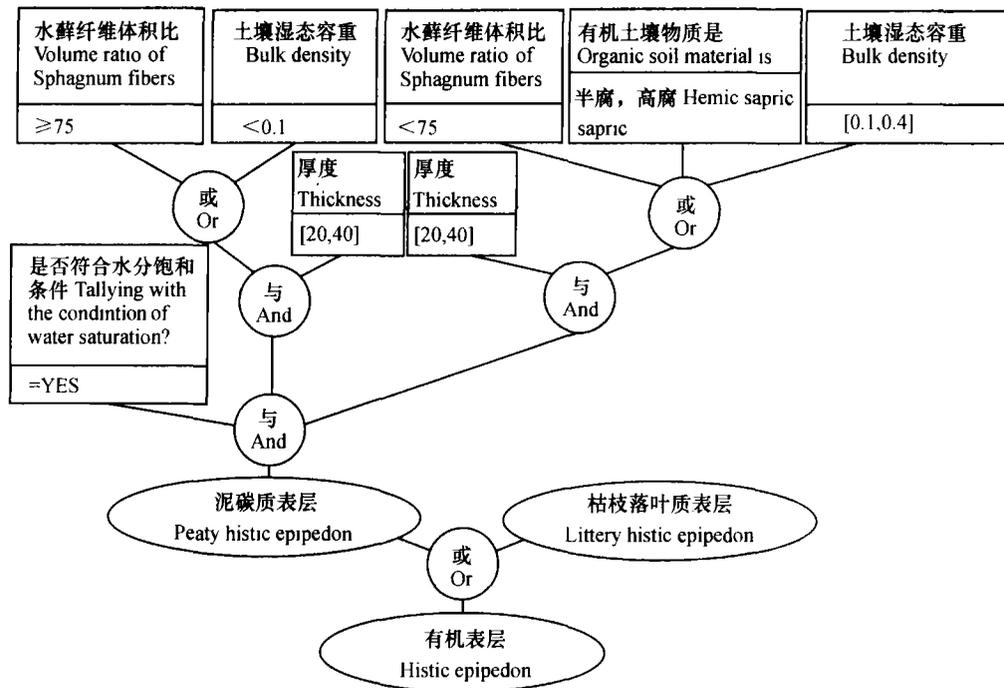


图3 有机表层判定树(部分)

Fig.3 Decision tree of Histic epipedons(part)

2.4 系统的推理流程

综合以上三个过程,获取用户信息后,整个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。执行步骤(5);

(5) 子类判断步骤(3)的结果,如果:

a. 数据不完整或不一致,将信息传递给父类,并结束推理;

b. 数据完整且一致,记录步骤(4)的推理结果,执行步骤(6)。

(6) 调用数据分析函数,如果数据充足,执行步骤(1);否则,结束推理。

3 GIS 与土壤分类专家系统结合

土壤剖面点的母质、气候、地形地貌的信息,可看

作空间点位信息,从纸质地图也可得到。为了获得区域信息,就需要使用 GIS。GIS 作为建立空间数据库、分析空间信息和图像显示的工具,已经被广泛的使用^[8]。如果与土壤专家系统结合,就可以模拟土壤调查与制图的全过程。理想的过程是:获取调查区域的母质、地形等宏观信息,获取采样点的理化分析数据;所有信息输入 GIS,以栅格格式存储;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,采用理想的扩散方程计算出各栅格的数据;然后,运用专家系统判断出土壤类型。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,可以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。

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。因此,在野外调查时,需要借助全球定位系统(GPS)获取准确的定位信息,并以点覆盖层的形式输入 GIS 保存。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。这样,根据已经记录的剖面定位坐标,就可提取出综合覆盖层中对应栅格中的所有数据。重新组织这些数据,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,因此可为 ES 使用。以上过程可以通过松散的集成形式实现,但为了提高整个系统的可维护性和完整性,在开发时,将

GIS 提供的外部函数和 ActiveX 控件封装到用 Visual C++ 建立的 GIS 和 ES 应用程序接口(GEApi)类中,系统运行时,由它的实例负责从 GIS 中接收数据,并驱动数据库引擎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 4)。临时文件使用完毕,仍然由 GEApi 实例负责提示用户保存或驱动数据库引擎删除该文件。

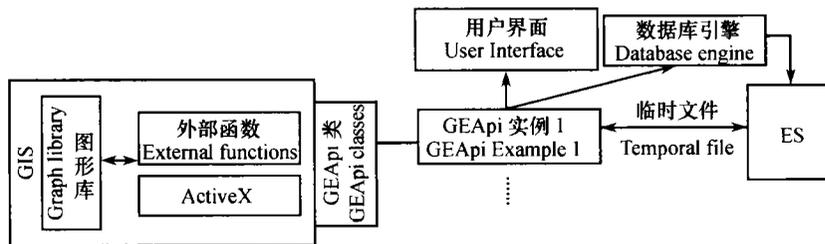


图 4 GIS 与 ES 间的数据传递

Fig.4 Data transformation between GIS and ES

专家系统的推断结果也以临时文件的形式提供给 GEApi 实例,此文件可以是普通的文本文件,也可以数据库文件的形式存储。不管采用哪种格式,都需要通过 GEApi 实例分析和重构数据,生成新的可被 GIS 使用的文件,然后调用 GIS 的外部函数或驱动 ActiveX 控件,将土壤类型信息写到综合覆盖层文件中。如果 ES 无法完成推理时,则向 GEApi 实例发出错误信息,并由它提示用户。

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四部分组成(图 5)。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,相应用户的请求。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,交给检索系统相应;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。

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。这个结果可能不是惟一的,系统可以在推理过程中,计算出结果成立的概率,提供给用户。

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) 对用户输入的空间数据进行组织与管理。
- (2) 根据土壤调查区域的地形图、母质分布图、土地利用图、降水分布图等,提取剖面点的相关信息。
- (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。解释器驱动需要提供剖面的空间点位信息。

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。这可以由管理员完成,或由系统自动更新。

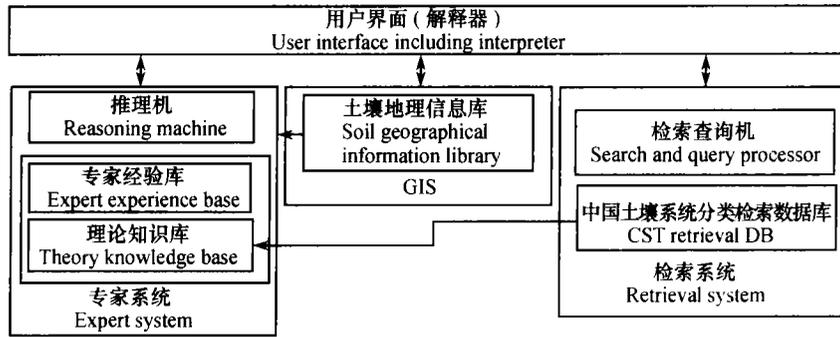


图5 系统数据流图

Fig.5 Diagram of system data flow

5 小 结

1) 以 ES 为核心并结合 GIS 的方法建立土壤自动分类专家系统(SCES),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,并融合了土壤发生学的知识,作为系统分类的基础,提高了专家系统的效率和应用范围。此系统不仅可应用于土壤研究,还可为土地持续利用规划、资源环境管理、农业生产过程提供重要的辅助决策信息,如土壤质地、土层厚度、排水和持水能力等。

2) 土壤自动分类系统的理论以中国土壤系统分类为核心,符合土壤分类国际化的要求,便于国际交流。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) 土壤的类型众多,到目前为止,大多数类似的系统都只针对几种土壤建立了专家知识库。为了提高系统的信息容量,一方面需要建立所有土壤类型的专家知识库,另一方面需要从积累的大量剖面数据中提取知识。这些工作如果完全用手工进行,会大大降低系统的更新能力。因此,有必要研究土壤知识的自动生成系统,尤其需要应用数据库的规则发掘技术。这也是完善系统功能时需要进一步研究的内容。

参 考 文 献

[1] 龚子同等.中国土壤系统分类-理论方法实践.北京:科学出版

社 1999 1~904. Gong Z T, *et al.* Theory, Methodology and Application of Chinese Soil Taxonomy (In Chinese) Beijing: Science Press, 1999. 1~904

- [2] 钟骏平, 闵勇, 蒋平安. 中国土壤系统分类检索及数据库系统. 新疆大学学报, 1999, 22(3): 215~218. Zhong J P, Ming Y, Jiang P A. Chinese Soil Taxonomy retrieval and database system (In Chinese). Journal of Xinjiang University, 1999, 22(3): 215~218
- [3] 赵其国, 叶方. 信息化与农业现代化. 土壤学报, 2004, 41(3): 449~455. Zhao Q G, Ye F. Information technology and agro-modernization (In Chinese). Acta Pedologica Sinica, 2004, 41(3): 449~455
- [4] Galbraith J M, Bryant R. A functional analysis of soil taxonomy in relation to expert system techniques. Soil Science, 1998, 163(9): 739~747
- [5] Galbraith J M, Bryant R B, Ahrens R J. An expert system for soil taxonomy. Soil Science, 1998, 163(9): 748~758
- [6] 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. 新疆大学学报, 1999, 22(2): 100~105. Jiang P A, Zhong J P, Li B G. The compute aided method for identifying diagnostic horizon and characteristics in Soil Taxonomy (In Chinese). Journal of Xinjiang University, 1999, 22(3): 215~218
- [7] Bogaert P, Dimnitri D'Or. Estimating soil properties from thematic soil map: the Bayesian maximum entropy approach. Soil Science Society of America Journal, 2002, 66(5): 1492~1500
- [8] Li W D, Zhang C R, Burt J E, *et al.* Two-dimension Markov Chain simulation of soil type spatial distribution. Soil Science Society of America Journal, 2004, 68(5): 1479~1490
- [9] Campling P, Gobin A, Feyen J. Logistic modeling to spatially predict the probability of soil drainage classes. Soil Science Society of America Journal, 2002, 66(4): 1390~1401
- [10] 沈掌泉, 周斌, 孔繁胜, 等. 应用广义回归神经网络进行土壤空间变异研究. 土壤学报, 2004, 41(3): 471~475. Shen Z Q, Zhou B, Kong F S, *et al.* Study on spatial variety of soil properties by generalized regression neural network (In Chinese). Acta Pedologica Sinica, 2004, 41(3): 471~475
- [11] Zhu A X, Band L E. A knowledge-based approach to data integration for soil mapping. Canadian Journal of Remote Sensing, 1994, 20(4): 108~118
- [12] Zhu A X, Band L E, Dutton B. Automated soil inference under fuzzy

- logic. *Ecological Modeling*, 1996, 90:123 ~ 145
- [13] Zhu A X. Mapping soil landscape as spatial continua: the neural network approach. *Water Resources Research*, 2000, 36(3):663 ~ 677
- [14] Zhu A X, Hudson B, Burt J, *et al.* Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal*, 2001, 65(5):1 463 ~ 1 472
- [15] Shi X, Zhu A X, Burt J, *et al.* A case-based reasoning approach to fuzzy soil mapping. *Soil Science Society of America Journal*, 2004, 68(3):885 ~ 894
- [16] Fox G A, Sabbagh G J, Searcy S W, *et al.* An automated soil line identification routine for remotely sensed images. *Soil Science Society of America Journal*, 2004, 68(4):1326 ~ 1331
- [17] Chang D H, Islam S. Estimation of soil physical properties using remote sensing and artificial neural network. *Remote Sensing and Environment*, 2000, 74:534 ~ 544
- [18] Hengl T, Rossiter D G. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. *Soil Science Society of America Journal*, 2003, 67(6):1 810 ~ 1 822

DESIGNING OF SOIL CLASSIFICATION EXPERT SYSTEM BASED ON GEOGRAPHICAL INFORMATION SYSTEM

Wang Shanqin^{1,2} Zhou Yong^{2†} Zhang Ganlin³

(1 College of Remote Sensing Information Engineering, Wuhan University, Wuhan 430072, China)

(2 College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China)

(3 Institute of Soil Science, Chinese Academy of Science, Nanjing 210008, China)

Abstract The Soil Taxonomy offers a theoretic basis for establishment of a soil classification expert system (SCES), which divides soil types according to quantitative soil indexes. The Geographical Information System (GIS) technology offers an effective tool for automatic mapping of soil classification. This paper concentrates on reasonable methods of developing SCES and automatic soil mapping by simulating thinking processes of soil classification experts and using the GIS technology. Firstly, object-oriented methods are used to express the experience and knowledge of soil diagnostic horizons and diagnostic characteristics of soil classification experts; and on such a basis, decision tree structures and consequence ways are discussed. Then a logic structure of the SCES, methods to combine geographic information with automatic soil classification, and the framework of an automatic soil classification searching system based on GIS and ES are brought forth from the angle of knowledge expression of soil classification experts and construction of an inference engine.

Key words Soil taxonomy; Object-oriented; Decision tree; Expert system; GIS