

# 基于知识发现和决策规则的盐碱地 遥感分类方法研究

布和敖斯尔

(中国科学院遥感应用研究所, 北京 100101)

E-mail: buheaosier@263.net

**摘要** 提出了基于知识发现和决策规则基础的盐碱地 GIS 和遥感分类的方法。这种方法的特点是把盐碱地分类的地质专家思想和区域专家的思想应用到 GIS 数据挖掘中, 并把从 GIS 数据库中发现的知识, 按一定的规则应用到盐碱地分类的决策中。经过对华北平原黑龙港地区盐碱地分类实例的研究证明, 该研究方法不仅简化了数据运算过程, 节省数据运算时间, 而且会减少或避免分类过程中人为误差的产生。

**关键词** GIS 数据库知识发现方法 规则基础 数据挖掘 盐碱地分类

## 0 引言

光学遥感数据, 比如 Landsat TM 数据主要利用了地物物理反射、辐射能而感知地物。所有物体都具有发射、反射、吸收和散射电磁波辐射的特征, 并且可以把所有物体发出的电磁波的传输简单理解为直线形式,

$$\lambda = c/f$$

(其中,  $\lambda$  为波长;  $c$  为电磁波传输速度;  $f$  为频率) 地物发出的电磁波的不同性质和特性都由其波长 ( $\lambda$ ) 所决定。根据地物发出的电磁波辐射波长的不同, 能够区分出不同的地物类型, 即根据地物实测分光反射率 (spectral reflectance) 特性曲线的特征, 区分出不同的地物类型。比如水体在近红外波段有较强的反射, 土壤在可见光和短红外波段有较强的反射, 而水在红外波段几乎没有反射, 是全部吸收。植物叶面因叶绿素含量和叶面含水量的不同其反射率也不同。岩石也因其类型的不同其反射率特性也不同, 区分岩石种类时常用到的波长为  $1.3\mu\text{m} \sim 3.0\mu\text{m}$  间的短红外区。盐碱土壤与众不同, 由于盐碱类土壤的地表积盐程度随着土壤含水量 (soil moisture) 变化而发生季节性变化, 其反射率也有明显的季节性变化。盐碱土壤地区的“水-盐”动态控制着盐碱类土壤的反射率。

我们在华北平原沧州地区进行了观测, 发现从每年的冬季到 3 月份为止, 该地区的盐碱土壤地表积盐, 看上去呈现出大面积灰白色, 随着地下水的大量蒸发, 土壤盐分聚集到了土壤表层。反映到遥感图象上, 由于其反射率高而显示出浅颜色, 此时盐碱土壤图象上的灰度值最大; 但从每年的 4 月份开始, 随着农田灌溉和地表“洗盐”过程, 又把表土中的盐分压到地下去了。此时, 盐碱土壤的颜色较暗, 大面积呈现出暗黑的颜色, 反映到遥感图象上, 由于其反射率低而显示出暗色, 此时盐碱土壤图象上的灰度值最小。由此看来, 利用光学遥感数据, 比如 Landsat TM 数据对盐碱地进行分类必须有专家参与, 建立一套基于知识发现和决策规则的盐碱地遥感分类方法, 更有助于提高分类的精度。

## 1 盐碱地分类中的知识发现和决策树

一般认为盐碱地都分布于沿海、滨海低平原地区, 河流两边的低洼地, 排水不畅、常年积水或季节性积水的地区。但这只是一般性知识或笼统的概念, 根据这种命题无法准确地地区分出盐碱地来。为了准确划分盐碱土壤分布区, 必须首先要建立严格的知识发现和逻辑规则及决策树。这种用于划分盐碱土壤的知识发现、数据挖掘和决策规则, 是基于专家知识和比较完备的 GIS 数据库<sup>[1]</sup>。因为是基于遥感数

据对盐碱地进行分类,因此还要依赖于遥感图象的分类结果,把遥感图象初步分类结果输入到 GIS 数据库中,记录不同类型土壤和其它地物的灰度值(或反射率值),用于推理和决策。

用于盐碱土壤分类的 GIS 数据库内容应包括:(1)土地利用与土地覆盖分类数据;(2)高程数据;(3)地下水水位(等值线)数据;(4)土壤类型数据;(5)土壤质地数据;(6)土壤水分状态数据;(7)植被指数;(8)河流水系分布数据;(9)地物反射率(灰度值)数据;(10)居民点分布与人口密度数据等。用于盐碱土壤分类的知识发现中[肯定]和[否定]权重相等,在[肯定]和[否定]判断处决策树要分叉,但是肯定判断或否定判断树又最终结点处的结论(类型)可能是同一个类型,也可能是不同的类型,这取决于储存的知识库,或在人机对话系统中取决于专家知识。常用知识发现程序设计中,或者通过多次否定来达到[肯定],发现事物;或者多次肯定来达到最终的[肯定],发现事物;或者是基于二叉分支[是]和[否]的多叉树模式方法。

从图 1 可以得出这样的结论,树-1 的最终结论为类型 11(该类型被定为非盐碱土壤区域),那么它是从否定(排除)决策达到肯定的例子。树-1 是按照地形高程来划分盐碱类土壤的例子。这里用高程是否小于 15 米来绝对地判断是否盐碱土壤分布区,高程值大于 15 米,被绝对地直接划到类型 11(非盐碱土壤区域),说明在知识库中发现盐碱土壤都分布于海拔 15 米以下的低平原地区。那么,是否在海拔 15 米以下的低平原地区都是盐碱土壤?肯定不是的。在树-1 中应用地下水水位来进一步区分是否盐碱土壤。从知识库中发现地下水水位大于 5 米的地方并没有盐碱土壤分布,从而判定在海拔 15 米以下的低平原地区,如果地下水水位大于 5 米,那么就没有盐碱土壤,这一区域被划到类型 11(非盐碱土壤区域)。

从图 1 树-2 的知识发现决策中我们可以得出这样的结论,树-2 的最终结论为类型 22(该类型被定为盐碱类土壤区域),那么它是从肯定(排除)决策达到肯定的例子。树-2 是通过注重土壤类型来区分盐碱类土壤的,在 GIS 数据库中,土壤类型-135001 为潮土类土壤的分布区域,但在仔细分析该决策树后发现,盐碱土壤还分布在非潮土类土壤类型区域。在非潮土类土壤类型区域,如果满足植被指数(NDVI)值小于 0.111 且同时要满足大于 -0.333 的话可以判定为盐碱类土壤区域。同样在潮土类土壤的分布区域也不一定都是盐碱土壤,即潮土不等于盐

碱土壤。在潮土类土壤中 Landsat TM 数据的第五波段上的灰度值大于 136,便可以判断为盐碱类土壤。说明在潮土类土壤中,分光反射率大的土壤可以判定为盐碱类土壤,因为潮土普遍呈现暗黑颜色,该土壤反射率都比较低。

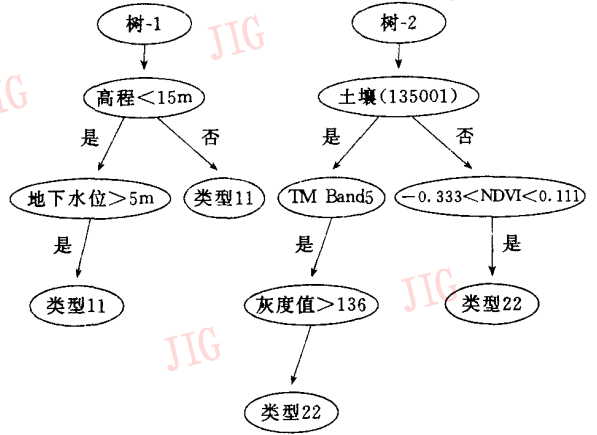


图 1 基于遥感数据的盐碱土壤分类知识发现与决策树结构例子

决策树判定取决于知识发现,即依赖知识库和专家思想。知识发现更多地依赖于一个完备的 GIS 空间的和属性的数据库。表 1 是用于盐碱土壤分类的 GIS 数据库数据类型。点-线-面数据中,点和面数据多以栅格形式储存,而线(道路、排水渠、引水渠等)数据多以矢量形式储存到数据库中。坡度数据来源于 DEM 坡度、坡向计算和 50 年代黑白航空相片套合计算;地质数据只包括土壤质地和地层断面;土地利用/土地覆盖类型数据直接从遥感图象的判读中得到;植被指数(NDVI)从 NOAA AVHRR 每周接收数据和 Landsat TM 波段 3 和波段 4 数据的运算中得到;其它数据主要来自数字化产品。数字化产品来自底图比例尺为 1:100000 的测绘成果。其中,土壤类型数据是具有多层面、多级 GIS 数据,用户编码为 6 位数,包含的信息有:土类、亚类、质地、水热、地形等。DEM 的等高间距为 1 米,坡度分为 7 级。地下水水位把点数据插值后绘制成地下水水位等值线图,并以测点为中心,换算成 1KM×1KM 栅格。彩色图版 I 中的图 1 为知识发现 GIS 数据库中的矢量数据应用的一个例子,彩色图版 I 中的图 1 所示的是在土壤类型数据上叠合了河流水系矢量数据,从中可以分析出不同土壤类型分布在河流两侧哪个部位,与河流的距离如何等,这些数据是经过数字化获得的。

表 1 知识发现 GIS 数据类型

数据类型	数据获取	数据格式
土壤类型	数字 GIS coverage	栅格/矢量
高程	数字 GIS coverage	DEM/栅格
地下水位	数字 GIS coverage(钻探/钻空数据)	栅格/网格点
地质	数字 GIS coverage(钻探/钻空数据)	栅格/矢量
坡度	数字 GIS DEM coverage	DEM/栅格
土壤水分(TDR 测量)	数字 GIS coverage 测量数据(点数据)	栅格/网格点
土地利用	遥感调查	栅格/矢量
Landsat TM 数据(7 个波段)	1987 年 3 月/1996 年 3 月	栅格
地表引水/排水渠	数字 GIS coverage	矢量
植被指数(NDVI)	遥感调查	栅格

## 2 知识发现规则基础(rule base)的建立

从多种约束条件判定一种盐碱土壤类型,这是知识发现系统(knowledge base system)的基本单元。但是一个正常运行的比较完整的知识发现系统,需要构筑多种规则基础(rule base),以神经网络判断事物的原理,把专家的知识应用到逻辑推理和准确决策上<sup>[1,2]</sup>。

Rule A:soil type in {135001,135100,127005,127001}  
 Slope $\leq$ 2  
 Elevation $\leq$ 15  
 Groundwater $\leq$ 5  
 TM Band-5(D. C.) $>$ 190  
 Land use types {11,12,13,15,16,17}  
 → Class 22(High Salinity)

从规则 A(Rule A)的构筑中可以看出,规则构筑依赖 GIS 数据库和专家思想,判断速度的快慢取决于规则结构的合理构筑,这会直接影响系统运行效率。规则 A 因为是从土壤类型入手构筑规则来判定是否重盐碱土壤,因此判定区域大,需要多种约束条件的辅助判定,系统运算需要一定的时间。盐碱土壤的形成与土壤类型并没有直接的关系,而与其它条件却有直接的关系,因此如果把土壤类型作为起始点,则增加了系统循环次数。在盐碱土壤分类中把土壤类型只作为辅助判定条件即可,它不能成为判断是否盐碱土壤的限制性因子。在华北平原盐碱土壤的生成发育过程中,微地形(高程、坡度等)、地下水位、土地利用、河流水系分布等因子起着重要作用,甚至成为盐碱土壤的生成发育的限制因子。规则

B 直接从限制因子入手构筑了判定是否重盐碱化土壤的知识规则。

Rule B:elevation $\leq$ 15  
 Groundwater $\leq$ 5  
 Slope $\leq$ 2  
 TM Band-5(D. C.) $>$ 190  
 → Class 22(High Salinity)

从规则 B 的构筑中可以看到,直接从盐碱土壤的生成发育的限制因子之一的高程(地形)开始,把盐碱土壤搜索的范围缩小为高程小于 15 米且地下水位小于 5 米的坡度小于 2 米的范围,在这个范围内,如果满足了在 Landsat TM5 波段灰度值大于 150,则可认为是重盐碱化土壤分布区域。这一规则的构筑,缩小了判断范围,提高了判断的准确性概率。决策树与规则基础有着紧密的关系,合理的规则基础会减少决策树的分叉,从整体来说会减少总决策树的数量<sup>[1,3]</sup>。

Rule C:TM Band-5(D. C.) $\geq$ 136  
 Spectral reflectance $>$ 48%  
 $-0.333 < NDVI < 0.111$   
 Land use types {11,12,13,15,16,17}  
 → Class 21(Salinity soil)

规则 C 构筑的是从 Landsat TM 图象反射率(灰度值)(第 5 波段)、地物分光反射率实测值和植被覆盖(第 3、4 波段)情况去区分盐碱土壤的例子。从规则 C 的构筑中看到,如果满足 Landsat TM 第 5 波段灰度值大于等于 136,地物分光反射率实测值大于 48%且植被指数大于 $-0.333$ 和小于 $0.111$ ,那么只要不是人工地面(这里人工地面土地利用用户编码为 14),便可判断为轻盐碱化土壤(类型 21)。

从以上例子可以看出,可以用多种方法或途径发现盐碱土壤,但是一个完整的盐碱土壤分类规则系统必须要考虑所有可能的途径,即充分挖掘数据库和知识库中的所有潜在资源<sup>[4]</sup>。

### 3 华北平原黑龙港地区的盐碱地分类

我们用两种不同的方法对华北平原黑龙港地区的盐碱地进行了分类,其一为直接应用遥感手段,从光学遥感数据去分类盐碱土壤的传统方法;其二为应用知识发现和构筑规则基础的 GIS 综合方法分类盐碱土壤。第一种方法,选用了研究区 1996 年 3 月的 Landsat TM 数据,经过初步的大气和几何纠正后,首先对 Landsat TM 7 个波段数据进行主成分分析,根据分析结果选择第 3、4、5 波段数据进行了彩色合成,应用监督分类方法,对合成后的图象进行了分类,分类结果如彩色图版 I 中的图 2 所示。从分类结果看,这种分类方法比较清楚地反映了不同程度的盐碱化土壤在华北平原黑龙港地区的分布情况。但这种分类过程需要很长时间的人机交互过程才能完成,分类结果中参杂着人为的主观判断因素,时常有遗漏或对细小部分的忽略<sup>[5]</sup>。

彩色图版 I 中的图 3 是根据规则 A 和规则 C 的原理分类的结果。彩色图版 I 中的图 3 根据 Landsat TM 图象第 5 波段的反射率值,应用 Landsat TM 图象第 3、4 波段数据计算的 NDVI 值,在土壤类型、土地利用、坡度、地下水位等约束条件下,区分出轻、中盐碱化土壤分布区域。从知识发现 GIS 数据库中得出,华北平原黑龙港地区盐碱土壤分布在浅色草甸土、沼泽化浅色草甸土、滨海草甸性土壤上面,坡度小于 2,高程小于 12 米,地下水位在 3 米至 5 米之间,土地利用上属于旱地、草地区,但有些水稻田周围也有盐碱地的分布,在河流两侧的低洼地上分布着大面积的盐碱地。NDVI 值 0~0.111 之间,Landsat TM 图象第 5 波段灰度值大于 135,分布在 136~243 之间,地面分光反射率值 48%~59% 之间。

### 4 结论

决策树和规则基础的合理性、有效性、快捷性等不仅取决于知识发现 GIS 内容,而且与专家思想

(逻辑推理、逻辑规则的构筑)也有很大的关系。比如就拿盐碱土壤分布区域的划分为例,在知识发现 GIS 中的居民点的分布、人口密度等间接指标(辅助信息)对盐碱地的区分有很大的帮助,因为过去居民点都集中分布在土质比较好的、便于耕作的土地上;相比之下盐碱地收成不好,投入大产出少,不便于耕作,是人烟稀少的荒废地区。也就是说,居民点密度、人口密度与盐碱地有相逆分布。这些辅助信息对盐碱地分类很有帮助。

知识发现和决策规则方法把过去单一的判断、主观判断更加合理化、逻辑化和客观化了,充分吸收了区域研究专家思想和区域工作经验,这些符合地学研究规律。在 GIS 庞大的数据库中按一定的规则基础挖掘自己有用的信息,按一定的决策路线,进行判断决策,提高了数据库使用效率。本研究提出了用知识发现和决策规则对盐碱地进行分类的遥感方法。这种方法是基于遥感和 GIS 技术手段,并充分挖掘 GIS 数据库潜在资源,又吸收了区域研究专家的思想,能够在有效地对研究对象进行多角度立体分析的基础上,对目标进行综合决策。在盐碱地分类研究中的应用表明,这种方法准确识别目标的精度取决于知识发现 GIS 数据库内容、决策树结构及决策树起始点的选择,构筑合理的规则基础等因子。其中知识发现 GIS 数据库内容是基础,而决策树、规则基础取决于区域专家对目标的认知水平。

### 参考文献

- 1 Eklund P W, Kirkby S D, Salim A. Data mining and soil salinity analysis. *INT J Geographical Information Science*, 1998,12(3): 247~268.
- 2 Van Beurden A U C J, Douven W J A M. Aggregation issues of spatial information in environmental research. *INT J Geographical Information Science*, 1999,13(5):513~527.
- 3 Taylor K, Walker G, Abel D. A framework for model integration in spatial decision support systems. *INT J Geographical Information Science*,1999,13(6):533~555.
- 4 Thomes I L, Allcock G M. Determining the confidence level for a classification. *Photogrammetric Engineering and Remote Sensing*, 1984,30:1491~1496.
- 5 Buheaosier C N Ng, Hidenori Takahashi, Masami Kaneko. The simulation of Yellow River channel long time change and classification of saline soil area in North China Plain using Landsat TM data. *ISPRS, Volume 32, Part 4W12, Dynamic and Multi-Dimension GIS*, 351~356.

## Classification of Saline Soil Based on Knowledge Discovery and Rule Base System Using Remote Sensing Data

Buheaosier

(Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101)

**Abstract** This paper presents the classification of saline soil method, based on knowledge discovery and decision supporting rule base systems using remotely sensed data and GIS. The geo-science expert thoughts and experiences to GIS data mining resulted from classification of saline soil and monitoring secondary soil salinization are introduced. In the study, we constructed perfection knowledge discovery rules which are applied to the North China Plain. The resulting rules are used to improve knowledge discovery system performance.

**Keywords** Knowledge discovery system, Decision support rule base, GIS & Remote Sensing, Saline soil

### 康柏推出全新大众化价格 PC

## ——Compaq Prosignia 321 “全功能”PC 上市

谁说国际名牌一定贵?日前,IT 巨擎康柏电脑公司率先挑战这一固有观点。从 1999 年 8 月 30 日起,售价仅为人民币 5499 元的全新 Compaq Prosignia 321“全功能”PC 攻占各大城市,全面加速国际级 PC 的大众化进程。

Compaq Prosignia 321“全功能”PC 不但实现价格大突破,还将与时代并进,率先试行英特网上大拍卖。已于 1999 年 9 月 1、2 日在著名的网易网站举办了“1999 世纪大拍卖”。

Compaq Prosignia 321“全功能”PC 拥有令人惊喜的超低价格,更具康柏卓越的超群品质。采用 Intel® Celeron™处理器 366MHz,配备 32MB SDRAM 内存、4.3GB 硬盘、14”彩显,预装 Microsoft® Windows® 98 中文操作系统,其低价位、高性能可使广大家庭用户和中小型企业用户轻松赶超信息化浪潮。

作为康柏家族新成员,Compaq Prosignia 321“全功能”PC 具有极高的性能价格比,还将通过“321”充分体现康柏独特的品牌理念,全面履行对用户的承诺。

“3”包括康柏的 3 个保证:保证为用户提供最优惠的价格、最出众的质量与最满意的服务。

“2”代表 2 项免费:售出一年内,康柏电脑公司对康柏 Prosignia 321“全功能”PC 实行免费零部件更换及免费送修服务,有别于一般低价 PC 不含服务维修的做法。

“1”则是 1 个大突破:突破“国际名牌电脑一定贵”的固有观念,以绝大多数用户都能接受的人民币 5499 元的价格面世,使更多的中国用户都能轻松拥有国际水准的康柏品牌电脑。

Compaq Prosignia 321“全功能”PC 另一大特色就是凭借康柏的雄厚实力推出前所未有的“5 天维修承诺”:一年免费维修期内,在北京、上海、沈阳、广州、成都、深圳 6 大主要城市,康柏电脑公司保证 5 天内解决电脑故障,如无法按时修复,康柏将另外提供一台台式机,使用户在维修期间,工作、生活如常。