

地理信息系统支持下的知识获取及其在遥感影像植被分类中的应用研究*

刘卫国 龚建华 方红亮

(中国科学院地理研究所资源与环境信息系统国家重点实验室 北京 100101)

摘 要 本文根据植被类型分布与地理环境因子的关系,在地理信息系统和遥感技术支持下,通过 GIS 叠加、统计分析操作,建立植被分布与年积温、降水量、海拔高度、土壤类型等环境因子的定量化知识向量表。综合应用所得到的地学知识向量表和植被光谱特征值进行分类试验,得到研究区的植被分布图。文章以贺兰山地区为例,详细介绍该方法的应用。

关键词 遥感影像, 植被分类, 定量化知识, 地理信息系统

1 引 言

运用遥感技术进行植被定量化监测已经有了长足进展。但是遥感数据处理中也存在几个问题:(1)某些地区遥感数据的缺失;(2)影像局部有云、积雪或阴影。这些问题成为遥感数据在植被分类中有效应用的瓶颈。解决这几个问题,一方面可以应用多时相数据或多源遥感数据,另一方面可以运用地学知识来辅助分类^[1]。

运用地学知识的方法进行辅助分类,以往已有大量的实践。但大都采用询问地学专家对知识形式化处理,建立知识库与推理机的方法。即采用规则来表达专家在目视解译过程中所运用知识和经验^[1]。该方法的不足之处是随地学专家的“经验”差别造成分类的精度不同。

遥感数据经常与数字高程数据、降水量、温度数据、土壤类型等复合,对植被进行分类研究。大量实验证明这些辅助数据能够提高植被分类精度,并在一定程度上消除云、积雪等造成的影响^[2-4]。

Skidmore 基于数字高程数据(坡度、坡向、方位等数据)建立专家系统并对桉树类型进行分类,获得了比传统分类方法(最大似然等)更高的精度^[2]。但该文作者仅仅考虑到数字高程数据与植被分布的相关关系,对其它地学辅助因子如土壤类型、温度、降水量等未加考虑。Cibulaane Nyquist, Ernst and

Hoffer 等人的工作虽然分别考虑到土壤类型、温度、降水量等对植被种类分布的影响,但他们仅仅对单因子影响分别考虑,尚未对多因子作综合分析^[3,4]。

本文以贺兰山实验区为例,在数字高程模型、降水量、年积温和土壤类型等空间背景信息的支持下,综合研究了植被分布与地学环境因子之间的内在联系,建立了植被空间分布模型,成功地对植被进行了分类提取。

2 植被分布知识定量化方法

任何植物群落的存在,都是与它所存在的环境条件有着密切的关系,地球表面各地环境条件的差异是导致植物群落具有各式各样的类型及其分布特点的重要原因^[5,6]。遥感影像植被分类可以利用植被的电磁波谱特征、遥感影像的空间结构信息、多时相信息以及关于植被生长分布的地学知识和规律。运用地学知识来辅助分类,一般的研究者通过询问地学专家,建立规则库,运用推理机对研究区域的植被进行分类^[7-9]。本文用来辅助遥感影像分类的地学知识是从植被分布与地学因子的时空相互关系中获取的。研究表明^[5],植被分布与区域的土壤类型、降雨量、海拔高度、年积温等地学环境因子之间存在着密切关系。因此,通过分析区域的土壤类型、降雨量、海拔高度、年积温得到关于植被分布的定量知识。

*中国科学院院长特别支持项目资助:基于地面特征的遥感专题信息提取研究。

收稿日期:1997-12-22;收到修改稿日期:1998-02-17

2.1 植被分布知识获取及分类原理

利用地理信息系统与遥感影像处理系统,获取植

被类型的定量知识表达及其分类的技术流程见图 1。

首先对研究区域的 TM 遥感影像进行预处理,然后选取一定数量的样本进行最大似然法监督分类,获

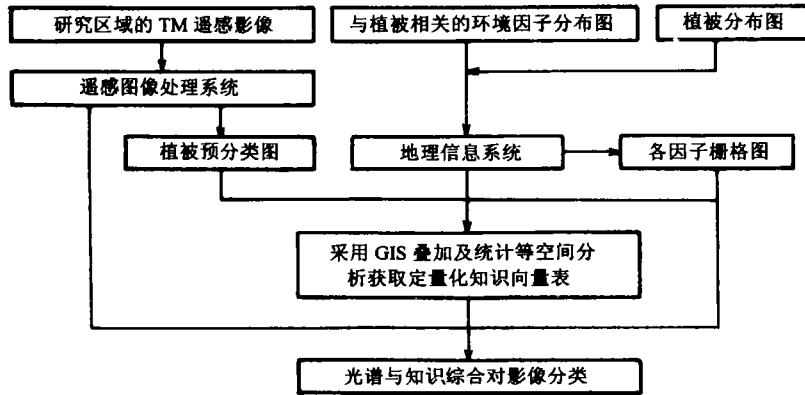


图1 植被分布知识获取及分类流程

Fig.1 Flow-line of extraction of knowledge and classification

取遥感影像植被预分类图。另一方面应用地理信息系统把同区域的植被分布图及与植被生长与分布相关地理环境因子如土壤类型、年积温、降水量、海拔高度图输入到 GIS 数据库,并与遥感影像数据进行匹配。将植被分布图及地理环境因子图用 GIS 的叠加、统计分析等功能获取植被与土壤类型、温度、降水、海拔等因子的定量化知识向量表。最后,运用遥感影像光谱数据与定量化知识向量表对影像进行综合分类,得到新的植被分类图。

2.2 植被分布与海拔、温度、降水量关系分析

与植被相关联的 4 种环境因子中,除土壤因子外,其他 3 种因子均可定量化描述。故在设计方法求解各环境因子与植被因子关系时,土壤是一种方法,而对其他 3 种因子则采取另外不同的算法。

我们应用地理信息系统的空间分析工具来定量

化分析植被分布变化与海拔等因子的关系,即将经栅格化的各因子分布图与栅格化后的植被分布图叠加分析,并统计记录每一种植被落入该因子的每一分布值内的数量 S_{ijk} (图 2)。其中 S_{ijk} 为统计数量; i 为植被类型; j 为环境因子类别; k 为环境因子值, $k \in$ 该环境因子值域。

将统计结果以图的形式描述(图 3):横轴代表定量环境因子值(海拔、温度和降水量),其值由低到高均匀分布,如海拔因子的值为 1990—3550m,相应横轴的值域为 1990—3550;纵轴为植被类型的分布数量 S_{ijk} 。根据每一植被类型在每一值上的分布数量进行连续曲线制图,得到描述此种植被在不同因子值域上的分布状况^[5]。

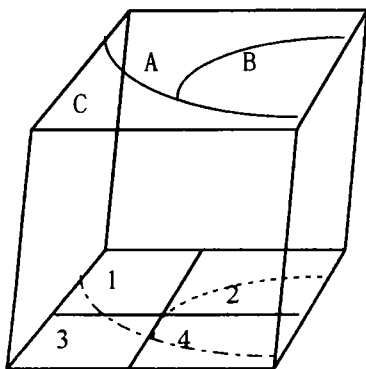


图2 定量因子与植被分布关系分析方法图

Fig.2 Analysis method of the relations between vegetation and factors

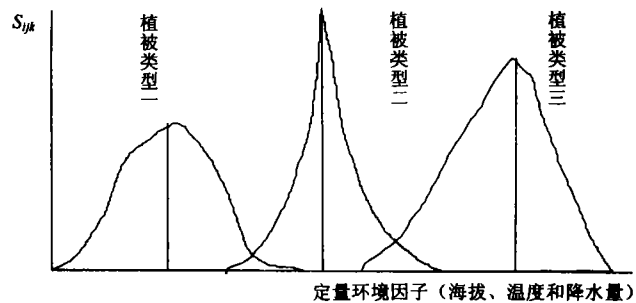


图3 植被分布与定量环境因子关系

Fig.3 Relations between vegetation and quantitative factors

2.3 植被类型分布与土壤类型关系的分析

植被的分布与土壤有着密切的关系。但要定量地描述某种土壤类型对某种植被类型的适宜程度很困难。本文引入 Bayes 决策^[10]和模糊数学中的隶属

度的概念^[11] 试图解决定量化描述问题。

将栅格化植被类型分布图与栅格化土壤分布图进行叠加分析,计算每种植被类型在每种土壤类型上的分布数量。以每种植被在不同土壤类型的分布数量表示该植被在该土壤类型上的生长适宜程度(值越大,适宜性越强)。数据经标准化后用于构造植被类型对土壤类型的隶属函数 $\mu(x)$ 。

$$\mu_{ij}(x) = \frac{S_{ij}}{\sum_{j=1}^m S_{ij}} \quad (1)$$

$$\sum_{j=1}^m \mu_{ij}(x) = 1 \quad (2)$$

公式(1),(2)中, S_{ij} 表示在第 j 种土壤上第 i 类植被的分布数量, $i \in (1, 2, \dots, n)$, $j \in (1, 2, \dots, m)$; n 表示植被类型总数, m 表示土壤类型总数。

运用 Bayes 决策判别规划^[10]来定量化描述土壤类型对植被的适宜性^[2]。

$$P(V_a | T_j) = \frac{P(T_j | V_a) P(V_a)}{\sum_{i=1}^n P(T_j | V_i) P(V_i)} \quad (3)$$

公式(3)中, $P(T_j | V_a) = \mu_{ij}(x)$, ($i = V_a$)代表植被类型 V_a 生长于土壤类型 j 条件概率密度。 $P(V_a)$ 代表第 V_a 种植被类型的先验概率,用植被预分类图中 V_a 类型植被所占面积的百分比表示。 $P(V_a | T_j)$ 代表后验分布概率,即已知某点土壤类型 j ,则该点的植被类型是 V_a 的可能性。该值越大,则该点的植被类型越趋于 V_a 。

2.4 综合知识定量化表达

植物生长分布是多种因素共同促进的结果,单因子分析不能全面反映植被分布变化状态和趋势,只有将各种因子综合考虑,才能更有效地反应植被分布变化的内在规律,找到各种因子共同作用的环境效应^[5]。

研究表明,各种植被类型分布沿环境变化具有明显的规律性并呈近似正态分布^[5]。正态分布的峰值是该类的聚集重心,应用与峰值相对应的环境因子值对植被类型进行定量描述最具有代表性。取每种植被类型的分布峰值所对应的海拔高度、降水量、年积温值作为该类的初始划分代表点,便构成了包括海拔高度、降水量、年积温等3个因子的集成环境因子向量空间, $S_{V_i}(C_i, R_i, E_i)$ 。考虑到土壤与植被关系的表达与其他因子与植被关系的表达不同,在空间向量表达式 $S_{V_i}(C_i, R_i, E_i)$ 中为土壤留出一维空间,形成

$$S_{V_i}(C_i, R_i, E_i, S_i) \quad (4)$$

考虑到遥感光谱与地学知识的综合运用进行植被分类,将原有光谱信息按波段在空间向量表达式中增加维数,形成由综合信息构成的空间向量

$$S_{V_i}(C_i, R_i, E_i, S_i, G_{i1}, \dots, G_{im})$$

式中: C_i 为第 i 类植被最适宜生长温度值, R_i 为第 i 类植被最适宜生长降水量值, E_i 为第 i 类植被最适宜生长海拔值, S_i 为第 i 类植被适宜生长土壤的后验概率, G_{im} 为第 i 类植被在 M 波段的特征值。

3 基于知识的遥感影像植被分类

海拔高度、降水量、年积温和土壤图均以栅格形式存储,并且4张图所描述区域完全相同。4张环境因子分布图与遥感影像叠加进行分类(图4)。对应图中每一点,从不同的图中均可以得到不同的因子值和不同波段的光谱值构成综合信息空间向量, $S_{V_p}(C_p, R_p, E_p, S_p, G_{p1}, \dots, G_{pm})$, p 表示空间中任意点。在分类过程中,各因子采用相同的权重计算每一点 $S_{V_p}(C_p, R_p, E_p, S_p, G_{p1}, \dots, G_{pm})$ 到各种植被类型的重心点 $S_{V_i}(C_i, R_i, E_i, S_i, G_{i1}, \dots, G_{im})$ 的欧氏空间距离 S_{D_i} 。

$$S_{D_i} = \|C_j - C_i\| + \|R_j - R_i\| + \|E_j - E_i\| + \|S_j - S_i\| + \dots + \|G_{jm} - G_{im}\| \quad (5)$$

公式(5)中, $\|G_{pm} - G_{im}\|$ 等均为点到点的欧氏空间距离。而 $\|S_p - S_i\|$ 的计算则略有不同。例如,从土壤图中读出 p 点的值为土壤类型 k ,从植被类型对土壤类型的后验概率函数得到第 k 类型土壤对 i 类植被后验概率值 $P(V_i | T_k)$ 。 $P(V_i | T_k)$ 越大,说明 k 型土壤越适应第 i 种植被的生长。欧氏空间表现为距离很近。因为 $P(V_i | T_k) \leq 1$,所以, $\|S_p - S_i\|$ 的计算转化

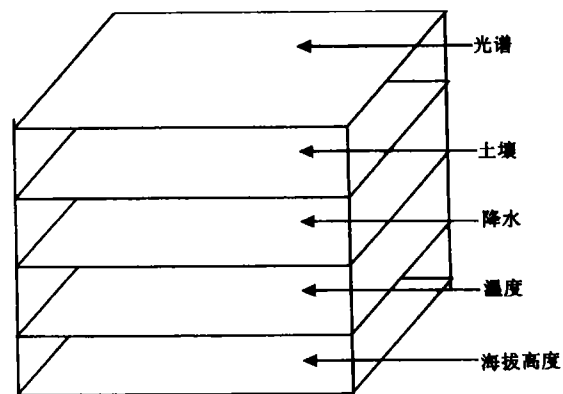


图4 基于知识的分类

Fig.4 The classification based on knowledge

为 $|1 - p(V_i | T_k)|$ 。最终该点到各种类型植被的距离表现为:

$$S_{D_i} = [C_p - C_i]^2 + (R_p - R_i)^2 + (E_p - E_i)^2 + (1 - P(V_i | T_k))^2 + \dots + (G_{pm} - G_{im})^2 \quad (6)$$

采用最小距离法来确定植被类型归属(即计算此点的综合环境指标更适合哪种植被类型)。

4 实例研究

以贺兰山为例,应用地理信息系统进行基于知识的遥感影像植被分类。根据贺兰山现有的资料,我们只采用海拔高度、降水量、年积温、土壤4种因子进行分析。

4.1 研究区域状况

贺兰山位于银川平原和阿拉善高原之间,北纬 $38^{\circ}21' - 39^{\circ}22'$, 东经 $105^{\circ}49' - 106^{\circ}41'$, 呈北北东—南南西走向,南北长约250km,东西宽20—40km,海拔一般为2000—3000m,主峰位于分水岭西侧的内蒙古境内。贺兰山作为国家自然保护区,近些年来人为破坏较少,自然环境也未有明显变化^[12]。我们以1988年贺兰山森林资源分布图、土壤图、年积温图、降水分布图和数字高程模型做为知识获取源。由于自然条件变化较小并且植被生长分布的确遵从一定自然规律,这种规律一般在较短的时期内不会发生变化^[13,14],这样便确保了依靠1988年的各因子分布图所获取的知识依旧适用于现在的分类系统。试验所用数据为1994年9月23日TM影像。

4.2 贺兰山地区植被分布与环境因子关系分析

4.2.1 植被分布与环境因子关系分析

我们运用2.2中介绍的方法,分析了贺兰山地区植被类型与海拔区间、降水量区间、年积温区间等的关系,从统计分析过程中,我们发现各种植被类型在环境因子上是近似正态分布。表1详细列出了10种植被类型在各环境因子区间的分布状况。

表1显示各种植被类型都分布在特定的因子区间内,并在某些点处呈现峰值。如亚高山草甸分布在海拔2789—3550m,并且在海拔3189m处分布最广(峰值处)。从表1中还可看出各种植被类型分布并非完全独立,而是交错分布。如年积温值在 $10^{\circ}\text{C} - 1618^{\circ}\text{C}$ 是亚高山草甸和云杉的共同分布区间。

4.2.2 植被分布与土壤类型关系分析

运用2.3中的方法,对研究区植被分布与土壤类型的关系进行分析,结果表明每种植被类型均有其最适宜生长土壤类型。表2表明每种植被类型在不同的土壤类型上分布后验概率(数据经标准化处理)。如表2中第1列表示已知某点的土壤类型为灰钙土,则该点植被类型是亚高山草甸的概率为0.11。第2列表示已知某点的土壤类型为钙质灰漠土,则该点的植被类型是山前苔地草原、荒漠化草原和荒漠的概率分别为0.114、0.843和0.091。在分类过程中,应用公式(7)描述的方法根据土壤类型来计算该点的植被类型时,显然该点到荒漠化草原的距离最小 $(1 - 0.843)$ 。如果只根据土壤来判断植被类型,则该点植被类型是荒漠化草原的可能性最大。

4.2.3 知识向量表

经过上述两步工作,我们分别得到了贺兰山地

表1 贺兰山植被与海拔、降水量和年积温的关系分析

Table 1 The relations between the vegetation and elevation, precipitation and annual accumulated temperature in Helan mountain

植被类型 \ 因子	海拔区间/m			降水量区间/mm			年积温区间/°C		
	最小	峰值处	最大	最小	峰值处	最大	最小	峰值处	最大
亚高山草甸	2789	3189	3550	362.7	410.1	445.0	10	161	1618
云杉	2240	2590	2989	293.1	337.3	385.9	10	1079	2095
油松	1940	2190	2490	256.5	289.5	325.6	760	1485	1641
针阔混交林	1891	2240	2640	253.5	293.1	341.0	420	1641	2753
灌丛	1341	1691	2090	185.1	228.4	277.3	1079	2454	3333
山地疏林草原	1541	1990	2490	209.3	265.3	325.6	760	1989	3151
山前台地草原	1092	1191	1891	159.0	168.1	253.5	1641	3216	3333
荒漠化草原	1491	1791	1940	204.3	228.4	256.5	1794	2454	3051
农田及其他	1092	1141	1241	159.0	161.3	173.4	2603	3303	3333
荒漠	1092	1141	1541	159.0	161.3	209.3	2454	3303	3333

表2 植被分布与土壤类型关系分析
Table 2 Analysis of the relations between vegetation distribution and soil type

土壤类型 植被类型	土壤类型									
	灰钙土	钙质灰 漠土	新积土	风沙土	粗骨土	潮土	泥炭土	灌溉土	亚高山 草甸土	灰褐土
亚高山草甸	0.11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.863	0.027
云杉	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.082	0.918
油松	0.0	0.0	0.0	0.0	0.0	0.0	0.332	0.0	0.0	0.668
针阔混交林	0.0	0.0	0.0	0.0	0.0	0.0	0.364	0.0	0.0	0.636
灌丛	0.0	0.0	0.0	0.0	0.638	0.0	0.301	0.0	0.0	0.061
山疏林草原	0.0	0.0	0.0	0.0	0.793	0.0	0.207	0.0	0.0	0.0
山前台地草原	0.0	0.114	0.776	0.0	0.1	0.0	0.0	0.0	0.0	0.0
荒漠化草原	0.0	0.843	0.076	0.0	0.091	0.0	0.0	0.0	0.0	0.0
农田及其它	0.0	0.0	0.165	0.0	0.0	0.154	0.0	0.691	0.0	0.0
荒漠	0.0	0.091	0.11	0.799	0.0	0.0	0.0	0.0	0.0	0.0

表3 知识向量表
Table 3 Knowledge vector expression of each kind of vegetation

植被类型 因子	植被类型									
	亚高山 草甸	云杉	油松	针阔混 交林	灌丛	山地疏林 草原	山前后 地草原	荒漠化 草原	农田及 其它	荒漠
海拔/m	3189	2590	2190	2240	1691	1990	1191	1791	1141	1141
降水/mm	410.1	337.3	289.5	293.1	228.4	265.3	168.1	228.4	161.3	161.3
温度/℃	161	1079	1685	1641	2454	1989	3216	2454	3303	3303
土壤类型	0.863	0.918	0.668	0.636	0.638	0.793	0.776	0.843	0.691	0.799

区植被分布与降水量、海拔、温度和土壤的关系。根据公式(4),可以构成每种植被类型的分布知识向量表(表3)。

4.3 基于知识的植被分类

经过以上工作,得到了研究区植被类型在海拔区间、降水区间、温度区间和土壤类型聚集重心。运用公式(6),综合考虑所得到的知识向量和植被光谱特征,用最小距离方法进行植被分类。所得到的植被分布图,总精度为81%。

传统的最大似然分类方法对研究区进行了植被分类,所得植被分布图精度为72%。这一结果表明综合运用地学知识与光谱信息进行分类的方法,比单纯运用植被光谱特征进行分类所得植被类别分布的精度有了较大的提高。

5 结 论

本试验运用地理信息系统的空间分析技术,综合分析了植被空间分布与高程、降水量、温度和土壤类型的地学关联性,证明植被分布与高程、降水量、温度和土壤类型之间是相互制约、相互影响的。

在建立植被分类定量化知识库方法上,作了较为成功的工作。该方法的特点是从GIS数据库中自动获取知识,并且采用知识向量表示方法,消除了基于规则知识获取时人为经验的影响。具有全自动、速度快、知识获取定量化等优点,可以弥补某些地区遥感数据损失或云、阴影等造成的影响。

本试验的研究区域——贺兰山地区是国家自然保护区,基本上为原始林和少量次生林,人为干扰破坏较少,植被生长自然规律近些年来基本上没有发生变化,这一事实保证了知识获取和表达的精确性。但对于内部自然环境发生较大变化的研究区域,如何确保知识表达的精确性还有待进一步的研究和探讨。

本试验是针对贺兰山地区森林资源调查而进行的方法探索。森林资源每5年或10年都要进行一次详细调查。依靠以往的调查数据和遥感影像,辅以少量的地面调查,基于知识进行分类即达到一般遥感影响分类达不到的精度,而且节省资源和时间。

致谢:感谢周成虎研究员、游先祥教授在本文完成过程中给予的宝贵意见和真挚的帮助。

参 考 文 献

- 1 徐冠华. 遥感信息科学的进展和展望(代序). 见徐冠华等主编. 遥感在中国. 北京:测绘出版社,1996,1—13.
- 2 Skidmore, A. K. An expert system classifies Eucalyptus forest types using Thematic Mapper data and a digital terrain model. *PE&RS*, 1989, 55(10):1449—1469.
- 3 Cibula W. G., Nyquist M. D. Use of topographic and climatological models in a geographical database to improve Landsat Mss classification for Olympic National park. 1987, *PE&RS*, 53(1):67—75.
- 4 Ernst C. L. R.M. Hoffer. Using Landsat Mss data with soil information to identify wetland habitats. *Satellite Hydrology, Proceeding of the 5th Annual William T. Pecora Symp.* 1979, 474—478.
- 5 刘卫国,吕鸣伦. 地理信息系统和遥感技术支持下的山地环境梯度分析方法研究. 1997, *地理研究*, 16(3).
- 6 Robert H. Whittaker 著. 王伯逊等译. 植物群落排序. 北京:科学出版社,1986.
- 7 T. A. Warner, D. W. Lecandowski, R. Bell. Rule-Based Geobotanical Classification of Topographic, Aeromagnetic, and Remotely Sensed Vegetation Community Data. *Remote Sensing Environment*. 1994, 50:41—45.
- 8 Haluk Certin, T. A. Warner, D. W. Lecandowski. Data Classification, Visualization, and Enhancement Using n-Dimensional Probability Density Functions (nPDF): AVIRIS, TIMS, TM, and Geophysical Applications. *ISPRS Journal of Photogrammetry & Remote Sensing*, 1993, 59:1755—1764.
- 9 周成虎,杜云艳,骆剑承. 基于知识的 AVHRR 影像的水体自动识别方法与模型研究. *自然灾害学报*,1996, 5(3).
- 10 边肇旗. 模式识别. 北京:清华大学出版社,1989.
- 11 赵振宇等. 模糊理论和神经网络的基础和应用. 北京:清华大学出版社,1996.
- 12 狄维忠. 贺兰山维管植物. 西北大学出版社,1987.
- 13 许朝斋,林之光. 贺兰山区气候若干问题. *地理学报*,1993, 48(2).
- 14 Liu Weiguo, Lu Minglun. Helan Mountain Environment Gradient Analysis and Plant Community Cartography. *Chinese Journal of Arid Land Research*, 10(2):115—122.

作 者 简 介

刘卫国,男,1971年生。1993年毕业于河北大学电子与信息工程系,获计算机工学学士学位;1996年毕业于北京林业大学获硕士学位。研究方向为GIS网络分析、交通地理信息系统、GIS与专业模型集成和遥感图像处理。已发表论文7篇。

Knowledge Extraction from GIS Database and its Application in Vegetation Classification

Liu Weiguo Gong Jianhua Fang Hongliang

(LREIS, Institute of Geography, CAS Beijing 100101)

Abstract There are many classification methods for thematic mapping using satellite remote sensing image. The accuracy of classification is limited and unsatisfied. Here, a knowledge-based vegetation classification is presented, which includes two main steps: extracting knowledge and classification based on knowledge.

The first step is that the knowledge is extracted by analyzing the relationship between the relative environment factors and each kind of vegetation supported by GIS. The distribution maps of several environment factors, including soil, elevation, temperature, precipitation, are selected and are overlapped respectively with vegetation distribution map of previous period. As far as the soil factor, which can not be described with numerical value, the Bayes theory was introduced to describe the probability of the classification. On the other hand, for the factors including elevation, temperature and precipitation described with numerical value, a new method was introduced to describe the classification probability. In the end, a vector expression is designed to describe the relationship between each kind of environment factors and each kind of vegetation through statistical method. Thus, the knowledge (the vector expression) is extracted.

The second step is that a method based on knowledge derived from above is designed to classify the vegetation for remote sensing image. The difference between this method and other methods is that the knowledge is brought into play as assumed several bands of spectrum. During the classification procedure, the contextual check was made to ascertain whether the pixels had been classified into the vegetation types or not that were ecologically valid for the grid cell being considered. After used the knowledge, the accuracy is increased obviously.

To summarize, the classification accuracy was increased about 9 percent using the knowledge-based method presented, compared to the supervised classification method that does not use knowledge.

Key words Remote sensing image, vegetation classification, Quantitative knowledge, GIS