

# 基于空间数据发掘的遥感图像分类方法研究\*

邸凯昌<sup>1</sup> 李德仁<sup>2</sup> 李德毅<sup>3</sup>

(1 国土资源部航空物探遥感中心遥感部,北京市学院路29号,100083)

(2 武汉测绘科技大学校长办公室,武汉市珞喻路129号,430079)

(3 中国电子系统工程研究所,北京市万寿路6号,100036)

**摘要** 采用数据发掘技术从GIS数据库和遥感图像中发现知识,用于改善遥感图像分类。提出了两种实施空间数据归纳学习的途径:在空间对象粒度上学习和直接在像元粒度上学习。分析了两种粒度学习的特点和适用范围,同时提出了一种归纳学习与传统图像分类法的结合方式。用北京地区SPOT多光谱图像和GIS数据库进行土地利用分类的试验证明,归纳学习能较好地解决同谱异物、同物异谱等问题,显著提高分类精度,并且能够根据发现的知识进一步细分类,扩展了遥感图像分类的能力。

**关键词** 数据发掘;知识发现;土地利用分类;归纳学习;学习粒度

**分类号** P208;TP751;TP18

遥感与GIS的集成是近年来遥感和GIS领域研究的热点,两者的集成主要用于以下两个方面:①遥感数据作为信息源更新GIS数据库;②GIS数据支持遥感图像处理和分析。两个方面相辅相成,使得GIS数据不断得到更新。

用GIS数据作为辅助数据用以改善遥感图像的分类,是遥感图像分类实践中经常采用的策略<sup>[8]</sup>。以往的研究主要是把GIS作为辅助数据用于训练区的选择、分类数据预分层、分类结果后处理,或作为附加“波段”用于分类过程,一般是采用统计或人机交互的方式实现,辅助数据利用的自动化和智能化程度不高。当分类器不严格要求数据具备某种统计特性时,将辅助数据作为附加“波段”是一种简单可行的方式。但当分类器要求数据必须具备一定的统计特性时,多数辅助数据往往又不满足统计分类所要求的统计特性,附加“波段”的方法就不合适了。

为了使遥感分类能够利用专家知识并具有逻辑推理的能力,使之更接近人类专家的思维,有的学者尝试用专家系统方法进行遥感图像分类,但经常遇到知识获取的“瓶颈”问题<sup>[1]</sup>。传统的知识获取方法通常是一个漫长而反复的过程,并且难免有遗漏,给实用化带来困难。

实际上,已建立的GIS数据库中就隐藏着大量的可供遥感图像分类用的知识,这些知识中有些属于“浅层知识”,如某一地区有无河流、道路的

最大和最小宽度等,这一般通过GIS的查询功能就能提取出来;还有一些知识属于“深层知识”,如空间位置分布规律、空间关联规则、形态特征区分规则等,它们并没有直接存储于空间数据库中,必须通过运算和学习才可能挖掘出来。

空间数据发掘和知识发现(spatial data mining and knowledge discovery, SDMKD)是从空间数据库中提取隐含的、用户感兴趣的空间的和非空间的模式和普遍特征的过程。空间数据发掘技术在遥感和GIS领域的两个主要应用是<sup>[5,9]</sup>GIS智能化分析和知识驱动的遥感图像解译。

SDMKD为遥感图像分类的知识获取开辟了一条新途径。Eklund等用C4.5归纳学习算法从TM图像和地理数据中提取知识,用于土壤盐度分析<sup>[2]</sup>;Huang等用C4.5算法从GIS数据和SPOT多光谱图像数据中提取知识进行湿地分类<sup>[1]</sup>。他们均把地理数据按图像像元大小转换成栅格形式进行学习。数据发掘特别是其中的归纳学习技术在空间数据库的实施途径,以及归纳学习与传统图像分类法的结合方式,是值得进一步深入研究的问题。

## 1 归纳学习及其在空间数据库的实施途径

可用于空间数据发掘的方法有很多种<sup>[4,5,9]</sup>,

收稿日期:1999-06-03.

\*国家自然科学基金、测绘遥感信息工程国家重点实验室开放研究基金及教育部博士点专项基金资助项目,编号49631050、WKL(97)0302及0901/992501。

归纳学习是最重要的一种方法,它旨在从大量的经验数据中归纳抽取一般的判定规则和模式。归纳学习的算法有很多种,最著名的是 ID3 系列,包括 ID3、C4.5、C5.0。

ID3 系列算法是一种决策树学习算法,采用分治(divide and conquer)的策略,用属性值对例子集逐步划分,直到一个结点仅含有同一类的例子为止。核心算法是 ID3。决策树算法的关键是分类属性的选择, ID3 系列算法采用熵来选择属性,效果非常理想。

设训练例子集  $PN$  含有  $p$  个正例和  $n$  个反例,则一个例子属于正例集  $PE$  的概率为  $p/(p+n)$ ,属于反例集  $NE$  的概率为  $n/(p+n)$ 。一个决策树可以看作正、反例集的消息源,产生这些消息所需要的期望信息为:

$$I(p, n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

设属性  $A$  取值为  $\{A_1, A_2, \dots, A_r\}$ , 它们将  $PN$  分成  $r$  个子集  $\{PN_1, PN_2, \dots, PN_r\}$ 。设  $PN_i$  含有  $p_i$  个正例,  $n_i$  个反例,则  $PN_i$  子树所需要的期望信息为  $I(p_i, n_i)$ ,而以  $A$  为根的树所需要的期望信息为其各子树所需要的期望信息的加权平均值,即  $E(A) = \sum_{i=1}^r \frac{p_i + n_i}{p+n} I(p_i, n_i)$ ,因此,按  $A$  分支的信息增益为:  $gain(A) = I(p, n) - E(A)$ 。ID3 检查所有的候选属性,选择增益最大的属性  $A$  作为根结点,形成树。然后对子树  $PN_1, \dots, PN_r$  以同样的方法处理,递归形成决策树。

ID3 算法的优点是分类和测试速度快,适合于大数据库的学习。其缺点是决策树的知识表示没有规则好,特别对一棵大的分类树,其意义很费解,且该算法只能处理离散属性。C4.5 在 ID3 的基础上增加了将决策树转换为等价的产生式规则的功能,解决了连续取值的数据的学习问题,并可以分类多个类别。C4.5 算法是归纳学习中最流行的算法。C5.0 是 C4.5 的进一步改进版本,它对大数据库的处理更快,并增加了 Boost 等技术。因此,本文以 C5.0 算法为基础研究归纳学习在空间数据库的实现。

C5.0 算法要求输入数据的每一个元组由若干个条件属性和一个类别值属性组成,条件属性值可以是离散值或连续值,类别必须是离散值。对于一般的数据库,只要把数据库的记录作为元组,字段作为属性,便可以应用 C5.0 算法发现知识。

在空间数据库中实施归纳学习,首先要确定

把什么作为学习的元组,也就是空间数据库学习的粒度问题。学习的粒度分为两种,即在空间对象粒度上学习和直接在像元粒度上学习。空间对象可以是图形数据库中的面、线和点对象,也可以是遥感图像中经过处理和分析得到的面特征(如均质区多边形)和线特征(如边缘线)。像元主要指遥感图像的像元,也指栅格图形的单元。以空间对象为学习粒度,可以充分利用空间对象的位置、形态、空间关联等特征,得到对象的分类规则,可用于 GIS 智能化分析和决策支持,也可用于遥感图像分类。这样的规则用于遥感图像分类时,必须先用其他分类方法形成线特征和面特征,才可以进一步应用规则分类。以像元为学习粒度,可以充分利用像元的位置、多光谱值、高程、坡度等具体而详细的信息,得到的分类规则精确,适用于图像分类,不便直接用于 GIS 智能化分析和决策支持,但可以作为它们的中间过程。两种粒度的学习也各有缺点,像元粒度的学习无法利用形态特征,很难利用空间关联等信息;空间对象粒度的学习难以利用对象内部更详细的信息,如以多边形为学习粒度就难以利用其内部精确的高程值、坡度值等,而只能用平均值或其他典型值。两种粒度的学习要根据应用选用或结合实用。

确定了学习的粒度即学习的元组后,需要确定学习的属性。在一般的关系数据库中,学习的属性直接取自字段或经过简单的数学或逻辑运算派生出学习用的属性。空间数据库中图形的几何特征和空间关系等一般并不直接存储于数据库中,而是隐含在多个图层的图形数据中,需要经过 GIS 专有的空间运算和空间分析才能得到归纳学习用的属性。比如,要确定某空间对象所处的高程带,需要应用叠置分析。

图 1 为在空间数据库实施归纳学习的流程图。空间数据库学习样本一般用随机抽样的方法选取,当数据量不是很大时也可以取全部数据。经过确定学习粒度和属性,将学习数据组织成类似于关系数据表的形式,输入到 C5.0 算法中进行归纳学习。C5.0 的输出为决策树和等价的产生式规则,由于产生式规则更便于阅读和应用,宜选用产生式规则输出。

## 2 归纳学习用于遥感图像分类

目前最常用的遥感图像分类方法是以贝叶斯分类方法为代表的统计方法。贝叶斯分类法又称最大似然法,具有坚实的理论基础,在各类光谱数

据满足统计正态性假设的条件下,理论上能获得最小的分类误差。但贝叶斯分类法无法解决同谱异物、同物异谱的问题。

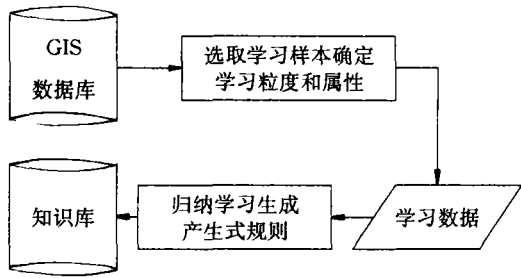


图1 空间数据库归纳学习流程图

Fig. 1 Flow Diagram of Inductive Learning in Spatial Database

对于一般的多光谱遥感图像数据,贝叶斯分类法通常能较好地地区分大的地类,如水体、居民地、绿地等。土地利用调查则需要更细的分类,如需要把水域进一步细分为河流、湖泊、水库和坑塘等,这可以认为是同谱异物问题;需要把绿地分为菜地、果园、林地、草地等,这既有同谱异物问题,也有同物异谱问题。要细分水域,必须利用形态特征和分布规律,而要细分绿地必须利用分布规律,也要利用不同绿地间微弱的光谱差异。因此,本文从土地利用 GIS 数据和高程数据获取两类知识,细分水域的知识学习粒度确定为多边形;细分绿地的知识学习粒度确定为像元。

在像元粒度的学习中,笔者提出一种归纳学习与贝叶斯分类法相结合的方法。先用贝叶斯分类法进行初始分类,保留分类得到的像元属于各个类别的概率值,用概率值、位置和高程等构成学习的属性,进行归纳学习。由于类别概率值由像元的光谱值和类别的光谱统计参数计算而来,这样就既利用了像元的光谱信息,又利用了类别的光谱统计信息。用同样的样本,分别用原始的像元光谱值和像元类别概率值与位置、高程属性一起进行归纳学习,用像元类别概率值的学习精度高于用原始的像元光谱值的精度,说明这种归纳学习与贝叶斯分类相结合的方法是有效的。

在多边形粒度的学习中,以 GIS 数据库中多个类别水域的多边形数据为学习对象,经过学习得到普遍的形态特征和分布规则。贝叶斯分类与之平行地进行,取概率值最大的类别为结果类别,对分类结果图像上的水域经过栅格矢量转换成均质区多边形,并提取位置和形态特征属性,运用学习得到的规则进行推理,将这些水域多边形细分为河流、湖泊、水库和坑塘等。这种归纳学习

与贝叶斯分类的结合方式是一种松散的结合。

图 2 为归纳学习用于遥感图像分类的工作流程图。从图中可以看出,在整个过程中,GIS 数据库扮演了重要的角色。

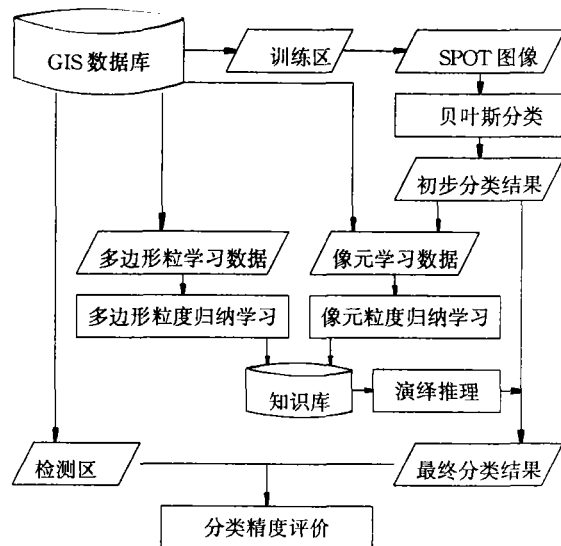


图2 应用归纳学习的遥感图像分类流程图

Fig. 2 Flow Diagram of Remote Sensing Image Classification with Inductive Learning

用 C5.0 算法进行归纳学习获取的知识为一组分类规则和一个缺省类别,每一条规则附有一个 0~1 之间的置信度值。如图 2 所示,在贝叶斯方法初步分类的基础上,根据分类规则通过演绎推理得到最终的分类结果。运用规则进行推理时所需的属性和数据格式与学习时相同,只是没有类别属性。在演绎推理中采用了如下策略:①若仅激活一条规则(某像元或均质区多边形的属性满足一条规则的条件),取规则的输出类别为结果类别;②若同时激活多条规则,则取置信度(总和)高的类别为结果类别;③若同时激活多条规则,且类别的置信度(总和)相同,则取学习时覆盖样本多的规则的输出为结果类别;④没有规则被激活时,取缺省类别为结果类别。

在基于空间数据发掘的遥感图像分类中,GIS 信息的利用与作为辅助数据的传统方法显著不同。传统方法直接利用 GIS 数据本身,而基于空间数据发掘的遥感图像分类利用的是从 GIS 数据中通过归纳学习挖掘出的知识。知识比数据更概括、精练、可靠和容易理解,更重要的是,多条规则可以表达复杂的非线性知识。因此,利用 GIS 数据中的知识比直接利用 GIS 数据本身可能对提高遥感图像分类的精度和可靠性更有效。

### 3 试验及分析

试验选用 1996 年北京地区的 SPOT 多光谱图像,原始图像大小为 2 412 像元×2 399 像元、3 波段,经过拉伸和精纠正后用作图像分类的原始图像,见图 3。纠正后图像大小为 2 834 像元×2 824 像元。所用 GIS 数据库为 1996 年以前的北京地区 1:10 万土地利用数据库;所用 GIS 软件为 ArcView 3.0a;遥感图像处理软件为 ENVI 3.0;归纳学习软件为 Rulequest,公司根据 C5.0 算法开发的 See5 1.10。同时还用 C++ 语言自行开发了部分数据处理和接口程序。

为便于对比分析,首先单独用贝叶斯分类法进行分类,把精纠正的 SPOT 图像调入 GIS 系统中与 GIS 图层叠加显示,人机交互地选取训练区和检测区。为保证训练区和检测区具有代表性,它们之间有一些公用的区域。将图像分成水体、灌溉水田、水浇地、旱地、菜地、果园、林地、居民地等 8 类,分

类的混淆矩阵见表 1。可以看出,水域、灌溉水田、水浇地、居民地、菜地分类精度较高。菜地呈亮绿色,与其他绿地区分较好;而旱地、果园、林地均呈暗绿色,光谱差异小,相互错分严重,分类精度低。另外,林地中的部分阴影错分成水域。

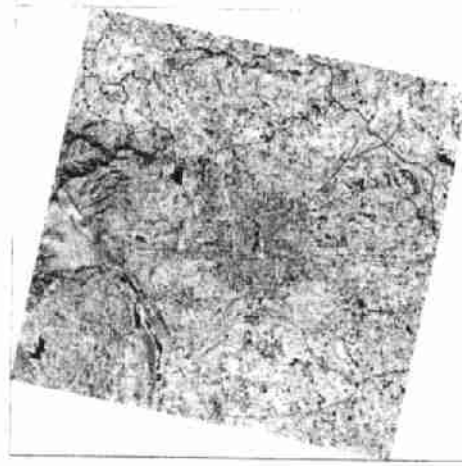


图 3 SPOT 多光谱图像(抽样显示)

Fig. 3 SPOT Multi-spectral Image (Resampled)

表 1 贝叶斯分类混淆矩阵

Tab. 1 Confusion Matrix of Bayes Classification

分类类别	实 际 类 别								总和
	水域	灌溉水田	水浇地	旱地	菜地	果园	林地	居民地	
水域	3.900	0.003	0.020	0.013	0.002	0.021	2.303	0.535	6.797
灌溉水田	0.004	8.496	0.087	0.151	0.141	0.140	0.103	0.712	9.835
水浇地	0.003	0.016	10.423	0.026	0.012	0.076	0.013	0.623	11.192
旱地	0.063	0.480	0.172	1.709	0.361	2.226	2.292	1.080	8.384
菜地	0.001	0.087	0.002	0.114	3.974	0.634	0.435	0.219	5.465
果园	0.010	0.009	0.002	0.325	0.263	4.422	4.571	0.065	9.666
林地	0.214	0.006	0.000	0.271	0.045	1.354	15.671	0.642	18.202
居民地	0.132	0.039	0.127	0.080	0.049	0.168	0.839	29.024	30.459
总和	4.328	9.135	10.834	2.689	4.846	9.041	26.227	32.901	100.00
精度/%	90.113	93.010	96.204	63.580	81.994	48.913	59.754	88.217	

总精度 = 77.619 9%

Kappa 系数 = 0.747 4

通过对贝叶斯分类结果的分析可知,归纳学习主要用于两方面:细分水域并区分阴影和提高旱地、果园、林地的分类精度。前者采用多边形粒度的学习,后者采用像元粒度的学习。选用的 GIS 图层为土地利用层和等高线层,由于等高线和高程点很稀少,无法插值成 DEM,而把等高线层处理成高程带多边形层,分为小于 50m(lt50)、50m~100m、100m~200m、大于 200m(gt200)等高程带。

首先进行水域的学习,选取的属性有面积 area、地理位置 Xcoord 和 Ycoord(几何中心位置)、密集度(周长<sup>2</sup>/(4π 面积))compactness、所处高程带 height;类别属性值为河流 71、湖泊 72、水

库 73、坑塘 74 和阴影 99。用 604 个多边形进行学习,得到 10 条产生式规则,见图 4。学习精度为 98.8%,即 98.8% 的学习例子能被这 10 条规则正确分类。这些规则描述了空间分布规律、几何特征规律等,例如,第 1 条规则覆盖了 19 个例子,描述的是“若多边形密集度大于 7.190 882,并且位于小于 50m 这一高程带,则该多边形是河流”,可见密集度这一属性对识别河流起了关键作用。第 9、10 条规则为用高程识别阴影等等,应用这些规则识别阴影并对水域进行细分,仅识别阴影一项,就可以将林地的分类精度提高到 68.002%,同时减少了水域的包含误差。

对旱地、果园、林地的学习采用像元粒度,选

取的属性有图像坐标、高程带及分别属于旱地、果园、林地的概率等,输出为3个类别。从大量的样本中随机地选取1%(2 909个)学习,得到63条

规则,学习精度为97.9%,即97.9%的学习例子(2 848个)能被这63条规则正确分类。另外随机地选取1%样本进行检测,检测精度为94.4%。

Rule 1: (cover 19) compactness > 7.190 882 height = lt50 → class 71 [0.952]	Rule 6: (cover 213) Ycoord > 4 428 958 compactness ≤ 7.190 882 height = lt50 → class 74 [0.986]
Rule 2: (cover 5) Xcoord > 453 423.5 Xcoord ≤ 455 898.7 Ycoord > 4 414 676 Ycoord ≤ 4 428 958 compactness > 2.409 397 compactness ≤ 7.190 882 → class 72 [0.857]	Rule 7: (cover 281) Xcoord > 451 894.7 compactness ≤ 7.190 882 → class 74 [0.975]
Rule 3: (cover 33) Xcoord ≤ 455 898.7 Ycoord > 4 414 676 Ycoord ≤ 4 428 958 compactness ≤ 7.190 882 height = lt50 → class 72 [0.771]	Rule 8: (cover 38) area ≤ 500 000 height = 50~100 → class 74 [0.950]
Rule 4: (cover 4) area > 500 000 height = 50 100 → class 73 [0.667]	Rule 9: (cover 85) height = gt200 → class 99 [0.989]
Rule 5: (cover 144) Ycoord ≤ 4 414 676 compactness ≤ 7.190 882 height = lt50 → class 74 [0.993]	Rule 10: (cover 7) height = 100~200 → class 99 [0.778]
	Default class: 74
	-----
	Evaluation (604 cases):
	Errors 7(1.2%)

图4 归纳学习得到的水域细分规则

Fig. 4 Rules Discovered by Inductive Learning to Subdivide Waters

将两种粒度学习得到的规则对贝叶斯初步分类结果进行处理,得到最终的分类结果。由于贝叶斯分类无法对水域进行细分,为了便于与单独的贝叶斯分类结果对照比较,对水域细分规则仅用于区分阴影,得到的分类结果见图5。用与贝叶斯分类同样的检测区进行精度检测,混淆矩阵

和精度指标见表2。从表2中可知,旱地、果园、林地的分类精度,提高的幅度分别为6.231%、29.648%和32.056%,较好地解决了同谱异物和同物异谱的问题,精度提高效果显著。总的分类精度提高了11.2552%,Kappa系数提高了0.1245。

表2 归纳学习与贝叶斯分类相结合分类的混淆矩阵

Tab.2 Confusion Matrix of Bayes Classification Combined with Inductive Learning

类别	实际类别								总和
	水域	灌溉水田	水浇地	旱地	菜地	果园	林地	居民地	
水域	3.900	0.003	0.020	0.012	0.002	0.019	0.139	0.535	4.631
灌溉水田	0.004	8.496	0.087	0.151	0.141	0.140	0.103	0.712	9.835
水浇地	0.003	0.016	10.423	0.026	0.012	0.076	0.013	0.623	11.192
旱地	0.063	0.480	0.172	1.877	0.361	0.205	0.149	1.080	4.386
菜地	0.001	0.087	0.002	0.114	3.974	0.634	0.435	0.219	5.465
果园	0.009	0.009	0.002	0.210	0.263	7.102	0.470	0.065	8.131
林地	0.215	0.006	0.000	0.218	0.045	0.696	24.079	0.642	25.899
居民地	0.132	0.039	0.127	0.080	0.049	0.168	0.839	29.024	30.460
总和	4.328	9.135	10.834	2.689	4.846	9.041	26.227	32.901	100.00
精度/%	90.113	93.010	96.204	69.811	81.994	78.561	91.810	88.217	

总精度 = 88.875 1%

Kappa系数 = 0.871 9

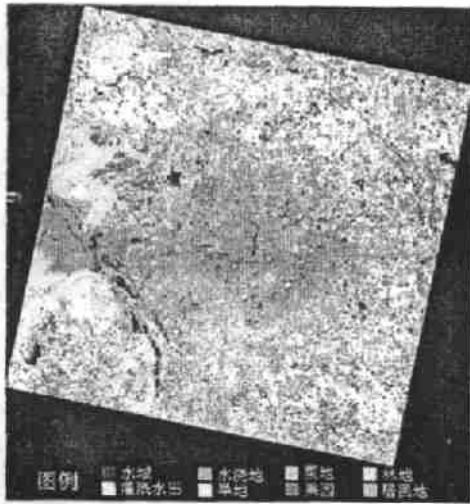


图5 归纳学习与贝叶斯分类相结合的图像分类结果(抽样显示)

Fig.5 Classification Result by Combining Bayes Method with Inductive Learning(Resampled)

## 4 结 语

GIS与遥感的智能化集成是一个困难的问题,其中的关键之一是知识的发掘和利用。归纳学习在空间数据库的实现及与传统图像分类法的结合在遥感与GIS的集成中具有理论和实用价值。下一步还需要进行多种数据发掘方法在遥感图像分类中的应用研究,特别是获取不确定性知识和不确定性推理的研究。另外,本文仅用SPOT图像进行了分类试验,归纳学习在其他多种遥感图像(如TM、SAR等)分类中的应用效果对比,也值得进一步研究。

## 参 考 文 献

1 Huang X Q, John R J. A Machine-Learning Approach to

Automated Knowledge-base Building for Remote Sensing Image Analysis with GIS Data. *Photogrammetric Engineering & Remote Sensing*, 1997, 63(10): 1 185 ~ 1 194

2 Eklund P W, Kirkby S D, Salim A. Data Mining and Soil Salinity Analysis. *Int. J. Geographical Information Sciences*, 1998, 12(3): 247 ~ 268

3 Quinlan J R. C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann, 1993

4 Li D R, Cheng T. KDG: Knowledge Discovery from GIS-Propositions on the Use of KDD in an Intelligent GIS. In: Proc. ACTES, The Canadian Conf. on GIS, 1994

5 Li D R, Di K C, Li D Y. A Framework of Spatial Data Mining and Knowledge Discovery. In: Proc. Int. Workshop on Image Analysis and Information Fusion (IAIF'97). Australia: Adelaide, 1997

6 李德毅. 归纳学习: 从数据库中发现知识. 第10届全国数据库学术会议, 沈阳, 1992

7 洪家荣. 归纳学习——算法·理论·应用. 北京: 科学出版社, 1997

8 张继贤, 赵俊江, 韩丽华, 等. 遥感信息源自动更新GIS数据库的方法与关键技术. 中国地理信息系统协会首届年会, 1995

9 邱凯昌, 李德仁, 李德毅. 空间数据发掘和知识发现的框架. 武汉测绘科技大学学报, 1997, 22(4): 328 ~ 332

邱凯昌, 男, 32岁, 博士生, 高级工程师, 现主要从事空间数据发掘和知识发现、遥感与GIS集成理论研究。代表成果: 空间数据发掘和知识发现的理论和方法; 从遥感图像中提取浅海水深信息等。

E-mail: kedi@public3.bta.net.cn

## Study of Remote Sensing Image Classification Based on Spatial Data Mining Techniques

DI Kaichang<sup>1</sup> LI Deren<sup>2</sup> LI Deyi<sup>3</sup>

(1 Center for Remote Sensing in Geology, 29 College Road, Beijing, China, 100083)

(2 Presidential Secretariat, WTUSM, 129 Luoyu Road, Wuhan, China, 430079)

(3 Institute of China Electronic System Engineering, 6 Wanshou Road, Beijing, China, 100036)

**Abstract** Data mining techniques are used to discover knowledge from GIS database and remote sensing image data in order to improve image classification. Two learning granularities are proposed for inductive learning from spatial data. One is spatial object granularity and the other is

pixel granularity. The characteristics and application scope of the two granularities are discussed. We also present an approach to combine inductive learning with conventional image classification methods, which selects class probability of Bayes classification as learning attributes. A land use classification experiment is performed in the Beijing area using SPOT multi-spectral image and GIS data. Rules about spatial distribution patterns and geometric features are discovered by C5.0 inductive learning algorithm and then the image is reclassified by deductive reasoning. Compared with the result produced only by Bayes classification, the overall accuracy increased 11 percent and the accuracy of some classes, such as garden and forest, increased about 30 percent. The results indicate that inductive learning can resolve spectral confusion to a great extent. Combining Bayes method with inductive learning not only improves classification accuracy greatly, but also extends the classification by subdividing some classes with the discovered knowledge.

**Key words** data mining; knowledge discovery; land use classification; inductive learning; learning granularity

---

DI Kaichang, male, 32, doctoral candidate, senior engineer. His research areas include spatial data mining and knowledge discovery, integration of remote sensing and GIS. His typical achievements involve theory and methods of spatial data mining and knowledge discovery; shallow water depth extraction from remote sensing images, etc.

E-mail: kodi@public3. bta. net. cn

---

(上接第 17 页) used for the approximation of the earth's gravity field, such as wavelet transform, multiple-input output method, etc. Overdetermined geodetic boundary value problem (OG-BVP), non-linear geodetic boundary value problem and stochastic boundary value problem should be widely investigated in future. ⑦ VLBI and SLR are also primary techniques for monitoring the earth rotating parameters, the movements of global/local or regional plates and the earth crust deformation. ⑧ It has been proved that most important meteorological results from GPS/MET using ground-based and spaceborne-based GPS techniques was that the agreement with numerical weather prediction (NWP) models is noticeably better in data dense than in data poor regions. This shows that GPS/MET results assimilated into NWP models can improve the weather forecast models. In addition, GPS/MET data will be helpful to study global/local climate changes. ⑨ GPS-altimetry can be used for sea surface or ice sheet topography coverage, but some corresponding theoretical basis and data processing methods should be widely investigated. ⑩ High precision static and real-time or near real-time kinematic GPS technology can be used for the health monitoring purpose of large structures such as large dam, tall buildings and long-span or cable-stayed bridges.

**Key words** modern geodesy; research achievements; progress

---

LIU Jingnan, male, 56, professor, Ph. D supervisor, member of the Chinese Academy of Engineering. His major research orientation include space geodesy and geodynamics. His typical achievements are the theory and scheme of high precision GPS data processing in China, etc.

E-mail: jnliu@wtusm. edu. cn