

文章编号: 1007 4619(2006) 03 0332-07

基于分类规则挖掘的遥感影像分类研究

库向阳¹, 薛惠锋¹, 雷学武², 汤国安³

(1. 西北工业大学 自动化学院, 陕西 西安 710072; 2. 西安地质矿产研究所, 陕西 西安 710054;

3. 南京师范大学 地理信息科学江苏省重点实验室, 江苏 南京 210097)

摘要: 分析了目前遥感影像的统计分类、神经网络分类及基于符号知识的逻辑推理分类方法的优缺点。以 GIS 为平台, 构建了多源空间数据库, 将数据挖掘的思想和方法引入遥感影像分类中, 提出了面向分类规则挖掘的遥感影像分类框架。针对遥感光谱数据及其他空间数据的特点, 定义了连续属性样本分类概念和分割点评价指标, 提出了一种新的连续属性样本分类规则挖掘算法。选择一个试验区, 采用该算法分别对遥感光谱数据、遥感光谱和 DEM 数据相结合的数据进行分类规则挖掘、遥感影像分类和分类精度比较。结果表明: (1) 该算法具有较高的分类精度; (2) 加入 DEM 等与分类相关的其他空间数据可以提高遥感影像的分类精度。通过挖掘分类规则进行遥感影像分类, 扩展了基于知识的逻辑推理分类方法中知识获取渠道, 提高了分类规则获取的智能化程度。新的连续属性样本分类规则挖掘算法, 扩展了归纳学习算法对连续属性样本分类的适应性。

关键词: 遥感影像分类; 地理信息系统; 多源空间数据库; 数据挖掘; 分类规则

中图分类号: TP393.3 文献标识码: A

The Remote Sensing Image Classification Research Based on Mining Classification Rules on the Spatial Database

SHE Xiangyang¹, XUE Hui feng¹, LEI Xuewu², TANG Guoan³

(1 College of automation, northwest polytechnic university, Shaanxi Xi'an 710072 China,

2 Xi'an Institute of Geology and Mineral Resources Shaanxi Xi'an 710054 China,

3 Jiangsu province key Laboratory for Geographical Information science Nanjing Normal University Jiangsu Nanjing 210097 China)

Abstract The advantages and disadvantages in remote sensing image classification methods at present are analyzed including maximum likelihood classification (MLC); Artificial neural networks (ANN) classification; Logic reasoning classification based on symbol and rules. In order to advance remote sensing image classification, multi resource spatial database is set up by means of GIS, the thought and concept of data mining are imported, and the remote sensing image classification frame is brought forward. The concept of classification and the index of judging splitting point on continuous valued attributes samples are defined, a new algorithm is proposed which mines classification rules on the continuous valued attributes spatial database. A trail area is selected, the two projects are put in practice which are based on Spectra Data, combining of Spectra data and DEM data respectively mining classification rules, classifying remote sensing image by the way of the algorithm, and comparing their classification accuracy. Trail outcome shows that: (1) the classification accuracy of the algorithm is good; (2) integrating DEM and other correlative data into the spatial database can advance classification accuracy. The way of mining classification rules being applied to remote sensing image classification broaden the channel of knowledge discovery in logic reasoning classification based on symbol and

收稿日期: 2004-10-08; 修订日期: 2005-02-23

基金项目: 国家高新技术研究发展计划资助项目(2001AA130023); 国家自然科学基金资助项目(40271089)。

作者简介: 库向阳(1968—), 男, 西北工业大学博士生, 从事空间数据挖掘和智能决策、地理信息系统及遥感技术应用、区域发展决策等方面研究。发表论文 10 余篇, 参编专著或教材 2 部。E-mail: xiangyangshd@sohu.com

rules advance the capability of discovering classification rules automatically. The new algorithm mining classification rules on the continuous valued attributes enlarge the adaptability of inducing learning algorithm to the classification of the continuous valued attributes samples.

Key words remote sensing image classification; geographic information system; multi source spatial data; data mining; classification rules

1 引言

遥感影像分类可分为基于密度分布函数的统计分类、神经网络分类及基于符号知识的逻辑推理分类等^[1-2]。基于密度分布函数的统计分类采用 Bayes 分类器来实现, 假定样本数据服从 Gauss 正态分布, 按待分像元与已知模式的相近程度进行分类。它具有容易对分类结果进行解释、可通过 Bayes 理论与其他先验知识进行融合、计算相对简单、实施方便等优点。但是训练样本的选择和参数估计直接影响分类结果, 由于对样本数据的正态分布要求, 难以结合其他空间数据进行遥感影像分类^[3]。神经网络具有良好的容错能力和自适应性, 对模式先验概率分布没有要求, 可融合多种类型数据进行遥感影像分类, 特别是当数据资料明显偏离假设的高斯分布时, 其优势更为突出。学者们将多层感知机 (MLP)、反向传播 (BP) 算法、基于径向基函数 (RBF) 等神经网络应用于遥感分类并取得了良好的分类结果, 但难以克服神经网络本身的缺陷: (1) 难以对结果作解释; (2) 不能完全解决局部最小问题; (3) 中间隐含层数和隐含节点数无法科学地确定。另外, 当数据维数增大时, 神经网络在判别相似类别的差异时, 容易造成误分。基于符号知识的逻辑推理分类方法是在传统地学分异规律的基础上, 通过对地学知识进行符号化表达和形式化逻辑推理的过程, 来实现信息的判别, 一定程度上能真实地反映地学分布规律。它基于遥感影像数据及其他空间数据, 通过专家经验总结、简单的数学统计和归纳方法等, 获得分类规则并进行遥感图像分类^[4-5]。它对数据没有正态分布的要求, 分类规则易于理解, 分类过程符合人的认识过程, 但是目前落后的知识获取手段限制了该方法的应用。针对多源空间数据 (遥感数据和其他空间数据), 一些学者将地学知识 (规则) 与神经网络等分类方法相结合, 采用 D-S 证据理论或模糊推理进行遥感图像分类, 或者将归纳学习与统计方法相结合确定模式的最后类别, 提高了遥感影像分类精度, 但是人为的割裂了光谱数据

与其他空间数据的内在关系, 增加了推理的复杂程度^[6-9]。随着 GIS 技术和数据挖掘方法的完善, 为直接针对多源空间数据挖掘分类规则、进行遥感影像分类提供了新的契机。但是目前数据挖掘和知识发现算法大都适用于离散化数据, 而多源空间数据大多为数量化数据, 特性更接近连续属性。为了扩展基于符号知识的逻辑推理分类方法中知识获取的渠道, 本文将数据挖掘方法引入遥感影像分类中, 以 GIS 为平台, 构建多源空间数据库, 提出面向分类规则挖掘的遥感影像分类框架。针对遥感光谱数据及其他空间数据的特点, 定义了连续属性样本分类概念和样本分割点评价指标, 提出了连续样本分类规则挖掘算法。选择一个试验区, 采用该算法分别对遥感光谱数据、遥感光谱和 DEM 数据相结合的数据进行分类规则挖掘、遥感影像分类和分类精度比较。

2 基于 GIS 面向分类规则挖掘的遥感影像分类框架

遥感影像中每一个像元的亮度代表该像元中地物平均辐射值, 它是地物的类型、纹理、状态、表面特征及特定电磁波综合作用的结果。为了克服遥感影像中的“异物同谱”和“同物异谱”现象, 提高影像分类的精度, 一个有效的途径是集成多源、多维空间数据构建空间数据库, 选取相关图层作为分类特征, 组成空间信息决策表, 引入数据挖掘方法获得分类知识, 进行遥感影像分类。GIS、RS 和数据挖掘为这一想法提供了强大的技术支持。

GIS 是空间数据采集、管理、分析、建模和可视化的工具, 空间数据管理、空间分析是 GIS 区别于其他信息系统的特有功能。因此, 面向遥感影像分类, GIS 作为一个空间数据管理和分析的平台担当如下角色: (1) 将与分类相关的空间数据作为一个“波段”与遥感光谱数据进行配准、叠加、集成, 构建一个多源空间数据库; (2) 为分类规则挖掘提供数据预处理和数据准备; (3) 显示分类结果和进行分类精度评价。

基于 GIS 面向分类规则挖掘的遥感影像分类框

架如图 1 所示,可分为:(1)多源空间数据库的建立;(2)数据采样,构建信息决策表;(3)分类规则的挖掘与测试;(4)遥感影像分类与分类精度评定。

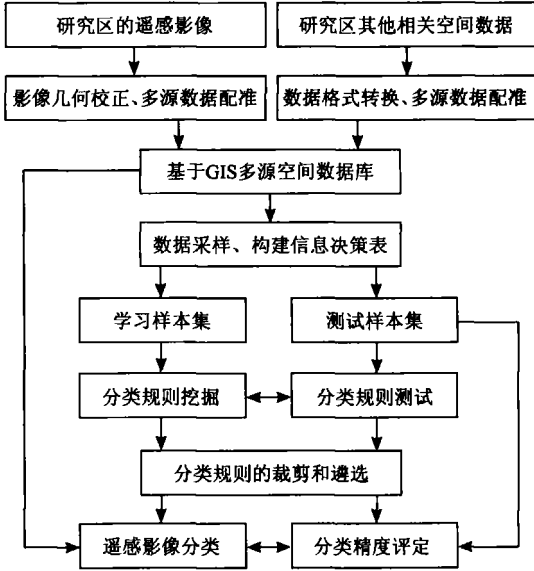


图 1 基于 GIS 面向分类规则挖掘的遥感影像分类框架
Fig 1 Remote image classification frame based on GIS in order to mine classification rules

- (1)多源空间数据库的建立: 遥感影像经几何校正、特征提取、多源数据配准等处理后,输入空间数据库。其他空间数据经数字化输入、矢量数据预处理(编辑、校正等)、数据格式转换(矢量数据转换成栅格数据)输入空间数据库。
- (2)数据采样、构建信息决策表: 使用遥感影像处理软件或 GIS 软件在遥感影像图层,采用计算机随机选点、人工判读影像类别或人工选点并判读影像类别、采集分类样本。由分类样本的属性特征构建空间信息决策表,并将样本分为学习样本和测试样本两类。
- (3)分类规则挖掘、测试和选择: 基于学习样本集使用分类规则挖掘算法挖掘分类规则,然后基于测试样本集对分类规则进行评价、选择。
- (4)遥感影像分类与分类精度评定: 使用有效分类规则进行遥感影像分类,对测试样本集进行分类,计算分类精度。

3 连续空间数据分类的基本问题

3.1 连续属性样本分类定义

设空间信息决策表 $S = \langle U, R, V, f \rangle$, 其中,

$R = C \cup \{d\}$ 是属性集合,子集 $C = \{a_1, a_2, \dots, a_k\}$ 和 $\{d\}$ 分别称为条件属性集和分类属性集。条件属性集对应多源空间数据库中的图层,分类属性集为空间数据中栅格或多边形对应的地物类型。 $U = \{x_1, x_2, \dots, x_n\}$ 是有限的对象集合,即论域,对应空间数据中的栅格或多边形单元。属性 a_i 的值域 V_{a_i} (栅格或多边形单元在图层 a_i 上的取值范围)上的一个分割点记为 (a, c^i) ,则在值域 $V_{a_i} = [l_{a_i}, r_{a_i}]$ 上的任意一个分割点的集合 $\{(a, c_1^i), (a, c_2^i), \dots, (a, c_{a_i}^i)\}$ 定义了 V_{a_i} 上的一种分类 p_{a_i} 。

$$\begin{cases} p_{a_i} = \{ [c_1^i, c_2^i), [c_2^i, c_3^i), \dots, [c_{a_i-1}^i, c_{a_i}^i] \} \\ l_{a_i} = c_1^i < c_2^i < c_3^i < \dots < c_{a_i-1}^i < c_{a_i}^i = r_{a_i} \\ V_{a_i} = [c_1^i, c_2^i) \cup [c_2^i, c_3^i) \cup \dots \cup [c_{a_i-1}^i, c_{a_i}^i] \end{cases} \quad (1)$$

3.2 候选分割点集定义

在空间信息决策表 $S = \langle U, R, V, f \rangle$, $R = C \cup \{d\}$ 中,候选分割点集生成过程如下:

- (1)由小到大排列属性值 $a_i(x)$,得到集合 $V = \{v_1^i, v_2^i, \dots, v_{n_{a_i}+1}^i\}$,其中, $x \in U$, $n_{a_i} + 1$ 为属性 a_i 不同取值个数。式中, $v_j^i < v_{j+1}^i, j = 1, 2, \dots, n_{a_i}$ 。
- (2)设 v_j^i, v_{j+1}^i 为属性值 $a_i(x)$ 序列中的两相邻值,则分割点 $c_j^i = (v_j^i + v_{j+1}^i) / 2$ 属性 a_i 相应的分割点子集 $p_{a_i} = \{c_1^i, c_2^i, \dots, c_{a_i}^i\}$ 。式中, $c_j^i < c_{j+1}^i, j = 1, 2, \dots, n_{a_i} - 1$ 。
- (3)信息表中所有属性形成的分割点集为:

$$p = \{p_{a_1}, p_{a_2}, \dots, p_{a_k}\}, a_i \in C (i = 1, 2, \dots, k), k \text{ 为条件属性个数。}$$

3.3 候选分割点评价

不同条件属性和同一条件属性的不同分割点的分类能力不同,而且它们的分类能力具有互补性和相关性。在条件属性空间上对样本分割的目的就是把类别一致的样本划分成一子类,每一子类正好对应一个分类规则,而分割点数的多少和优劣决定了算法的效率、规则的简单性和个数。故把分割点对样本的分辨能力作为分割点选择的依据。本文选取文献 [10] 中的指标来评价候选分割点的分类能力。

将能够被分割点 c_m^i 区分开的实例对的个数定义为 $W^X(c_m^i)$, c_m^i 为属性 a_i 上第 m 个分割点。 $1 \leq m \leq n_{a_i}$, n_{a_i} 为属性 a_i 的分割点总数, $X \subseteq U$ 是由分割点 c_m^i 可以分开的实例集合, U 为实例全集。

分类属性值为 $j(j = 1, \dots, r, r \text{ 为分类数})$ 的实例

中,属于集合 X 且属性 a_i 的取值小于分割点 $c_m^{a_i}$ 的实例个数为:

$$f_j^X(c_m^{a_i}) = |\{x | x \in X \wedge [a_i(x) < c_m^{a_i}] \wedge [d(x) = j]\}| \quad (2)$$

分类属性值为 $j(j=1, \dots, r, r$ 为分类数) 的实例中,属于集合 X 且属性 a_i 的取值大于分割点 $c_m^{a_i}$ 的实例个数为:

$$f_j^X(c_m^{a_i}) = |\{x | x \in X \wedge [a_i(x) > c_m^{a_i}] \wedge [d(x) = j]\}| \quad (3)$$

所以有:

$$f^X(c_m^{a_i}) = \sum_{j=1}^r f_j^X = |\{x | x \in X \wedge [a_i(x) < c_m^{a_i}]\}| \quad (4)$$

$$r^X(c_m^{a_i}) = \sum_{j=1}^r f_j^X = |\{x | x \in X \wedge [a_i(x) > c_m^{a_i}]\}| \quad (5)$$

从而可以得到:

$$W^X(c_m^{a_i}) = f^X(c_m^{a_i}) \times r^X(c_m^{a_i}) - \sum_{j=1}^r f_j^X(c_m^{a_i}) \times f_j^X(c_m^{a_i}) \quad (6)$$

$W^X(c_m^{a_i})$ 值越大, 则说明分割点 $c_m^{a_i}$ 对样本的分类能力越强。

4 连续空间数据分类规则挖掘

4.1 基本思想

设连续属性样本为 X^0 , 候选分割点集为 H^0 , 最优分割点集为 P^0 (此时为空集)。在样本 X^0 的属性空间对候选分割点集 H^0 中分割点的分类能力进行评价, 选取 $W^X(c^0)$ 值最大分割点 c^0 将样本分为两个等价类 X_1^1 和 X_2^1 , X_1^1 子集为 $a_i(x)$ 大于 c^0 的样本集, X_2^1 子集为 $a_i(x)$ 小于 c^0 的样本集。向 P^0 集合分别增添 $+c^0$ 和 $-c^0$ 两个分割点, 形成与 X_1^1 和 X_2^1 相对应的最优分割点集 P_1^1 和 P_2^1 (“ $+c^0$ ”表示相应的样本集属性取值 $a_i(x)$ 大于所选分割点值 c^0 , “ $-c^0$ ”表示相应的样本集属性取值 $a_i(x)$ 小于所选分割点值 c^0)。在候选分割点集 H^0 中删除分割点 c^0 , 形成与 X_1^1 和 X_2^1 相对应的候选分割点集 H_1^1 和 H_2^1 。 P_1^1 和 P_2^1 记录划分该等价类此时已选的所有最优分割点集合, H_1^1 和 H_2^1 记录下一步分类时可供选择的分割点集合。第二次划分在第一次划分的基础上, 重新对每一等价类相应的候选分割点的分类能力进行评价, 生成相应的最优分割点集 $P_{11}^2, P_{12}^2, P_{21}^2, P_{22}^2$, 候选分割点集 $H_{11}^2, H_{12}^2, H_{21}^2, H_{22}^2$ 和等价类子集 $X_{11}^2, X_{12}^2, X_{21}^2, X_{22}^2$ 。依此类推, 当某一层的某一个等价类子集的属性一致时, 相应的最优分割点集经过整理, 便形成相应的分类规则, 并将该

子集删除。当某一层的等价类子集的集合为空或子集中的样本为不相容样本, 则停止。这一过程如图 2 所示。

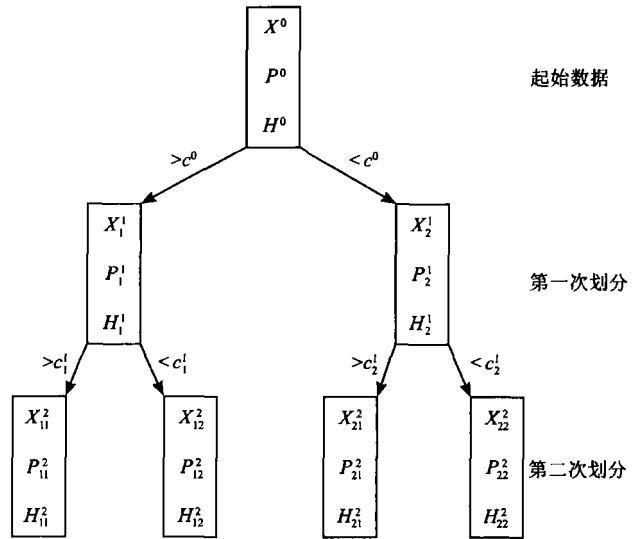


图 2 样本划分思路

Fig 2 Process classifying samples

4.2 分类规则挖掘算法

算法: 由多源空间数据直接挖掘分类规则。

输入: 样本属性值。

输出: 分类规则。

方法:

步骤 1: 随机将样本分为训练样本 X^0 和测试样本 Y^0 ;

步骤 2: 由训练样本 X^0 计算候选分割点集 H^0 , 并对其分类能力进行评价;

步骤 3: 选择最优分割点 c^0 , 将训练样本 X^0 划分为两个等价类 X_1^1 和 X_2^1 , 生成相应的最优分割点集 P_1^1 和 P_2^1 和候选分割点集 H_1^1 和 H_2^1 ;

步骤 4: 判断该层每一个等价类子集的分类属性是否一致。若一致, 由相应的最优分割点集产生分类规则, 并将该等价类子集删除。否则, 重新评价该等价类子集相应候选分割点集;

步骤 5: 为每一等价类子集选择相应的最优分割点, 进行等价类划分, 并生成相应的候选分割点集和最优分割点集;

步骤 6: 判断该层等价类子集集合是否为空或不相容样本。若是, 转向下一步, 否则转向步骤 4;

步骤 7: 使用测试样本 Y^0 对所有挖掘的分类规则进行测试。当某一测试样本 $x(x \in Y^0)$ 的分类预测错误时, 将 x 增加到训练样本集 X^0 中, 并从 Y^0 中

删除, 转向步骤 2。当测试样本集中不存在分类预测错误时, 输出分类规则;

步骤 8 终止。

5 遥感影像分类试验

5.1 试验数据处理流程

数据处理流程如下: (1)用地理信息系统 ArcGIS 8.3 或 ERDAS MAGNE 软件将遥感光谱数据和 DEM 数据分别作为一个图层, 进行数据配准, 构建多源空间数据库; (2)在多源空间数据库中, 对每一地物类型采用人工选点(按照均匀且同类样本差异较大的原则选点, 以确保样本有足够代表性)并判读影像类别进行采样, 将样本在各图层的取值输出为 *.text 文件; (3)用 Matlab 读取 *.text 文件, 并将样本分为学习样本和测试样本; (4)用 Matlab 语言编制本文算法的相应程序, 挖掘分类规则和进行分类规则测试; (5)依据有效分类规则, 编制相应应用 Matlab 程序进行遥感影像分类和分类精度评定。

试验选取宁夏青铜峡地区 500 像元 × 500 像元、6 波段 TM 遥感图像和对应 1:25 万比例尺的 DEM 数据。按照数据处理流程得到 6 类样本: C₁—河流、C₂—居民地、C₃—农田、C₄—沙滩、C₅—山地、C₆—水田, 共 2815 像元。选取样本的 1/2(1408 像元)作为训练样本, 其余 1/2(1407 像元)作为测试样本。

5.2 遥感影像分类试验

方案一: 基于遥感光谱数据, 程序经过 9 个循环, 挖掘出 45 个分类规则, 使用有效分类规则对测试样本集进行分类, 得到分类误差矩阵, 如表 1 所示。

表 1 基于遥感光谱数据的分类误差矩阵表

Table 1 Error matrix of classifier based on spectra data

实际 \ 分类	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	Σ
C ₁	169	0	0	1	0	0	170
C ₂	1	317	1	1	17	1	338
C ₃	1	2	218	0	0	1	222
C ₄	2	0	1	237	0	0	240
C ₅	0	16	0	0	208	0	224
C ₆	0	0	0	0	0	210	210
Σ	173	335	220	239	225	212	1404

注: 分对率 = 85.9%; Kappa = 0.84

方案二: 基于遥感光谱和 DEM 数据, 程序经过 6 个循环, 挖掘出 18 个分类规则。选择灵敏性、特效性、精度作为分类规则的评价指标^[11], 灵敏性 (sensitivity > 2%) 和精度 (precision > 75%) 的有效分类规则如下:

规则 11: $[14 \leq a_6(x) < 73.5] \wedge [36 \leq a_4(x) < 107] \wedge [94.5 < a_2(x) \leq 151] \Rightarrow$ 河流 (94.3%, 100%, 100%)

规则 12: $[14 \leq a_6(x) < 73.5] \wedge [107 < a_4(x) \leq 181] \wedge [85.5 < a_3(x) \leq 191] \wedge [18 \leq a_5(x) < 92.5] \Rightarrow$ 河流 (2.3%, 99.9%, 80.0%)

规则 21: $[91 < a_6(x) \leq 178] \wedge [81 \leq a_7(x) < 84.5] \Rightarrow$ 居民地 (93.7%, 99.9%, 99.7%)

规则 22: $[73.5 < a_6(x) \leq 178] \wedge [84.5 < a_7(x) < 86.5] \wedge [60 \leq a_2(x) < 117.5] \wedge [105.5 < a_1(x) \leq 143] \Rightarrow$ 居民地 (4.5%, 100%, 100%)

规则 31: $[14 \leq a_6(x) < 73.5] \wedge [107 < a_4(x) \leq 181] \wedge [49 \leq a_3(x) < 85.5] \wedge [60 \leq a_2(x) < 76.5] \Rightarrow$ 农田 (95.0%, 100%, 100%)

规则 32: $[14 \leq a_6(x) < 73.5] \wedge [107 < a_4(x) \leq 181] \wedge [49 \leq a_3(x) < 85.5] \wedge [76.5 < a_2(x) \leq 151] \wedge [86.5 < a_1(x) \leq 143] \Rightarrow$ 农田 (3.6%, 100%, 100%)

规则 41: $[14 \leq a_6(x) < 73.5] \wedge [107 < a_4(x) \leq 181] \wedge [85.5 < a_3(x) \leq 181] \wedge [92.5 < a_5(x) \leq 203] \Rightarrow$ 沙滩 (15.0%, 99.8%, 94.7%)

规则 42: $[73.5 < a_6(x) < 91.0] \wedge [81 \leq a_7(x) < 84.5] \wedge [98.5 < a_2(x) \leq 151] \Rightarrow$ 沙滩 (83.8%, 99.9%, 99.5%)

规则 51: $[73.5 < a_6(x) \leq 178] \wedge [84.5 < a_7(x) < 86.5] \wedge [117.5 < a_2(x) \leq 151] \Rightarrow$ 山区 (72.0%, 100%, 100%)

规则 52: $[73.5 < a_6(x) \leq 178] \wedge [84.5 < a_7(x) < 86.5] \wedge [60 \leq a_2(x) < 117.5] \Rightarrow$ 山区 (6.2%, 100%, 100%)

规则 53: $[73.5 < a_6(x) < 152.5] \wedge [84.5 < a_7(x) < 86.5] \wedge [117.5 < a_2(x) \leq 151] \Rightarrow$ 山区 (20.9%, 100%, 100%)

规则 61: $[14 \leq a_6(x) < 73.5] \wedge [36 \leq a_4(x) < 107.5] \wedge [60 \leq a_2(x) < 94.5] \Rightarrow$ 水田 (99.1%, 100%, 100%)

分类规则括号内依次为规则的灵敏性、特效性、精度值。a₁(x), a₂(x), a₃(x), a₄(x), a₅(x), a₆(x) 分别是 x 像元在 6 个波段的取值, a₇(x) 是 x 像元的

DEM 属性值。使用有效分类规则对测试样本集进行分类得到分类误差矩阵如表 2 所示。

表 2 基于遥感光谱数据和 DEM 数据的分类误差矩阵

Table 2 Error matrix of classifier based on spectra and DEM data

分类 \ 实际	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	∑
C ₁	169	0	0	1	0	0	170
C ₂	2	331	1	1	1	2	338
C ₃	0	1	218	0	0	0	219
C ₄	2	0	1	237	0	0	240
C ₅	0	3	0	0	224	0	227
C ₆	0	0	0	0	0	210	210
∑	174	335	220	239	225	212	1404

注: 分对率 = 93.6%; Kappa = 0.93

比较表 1 和表 2 可知: 方案二加入 DEM 数据后, 测试样本的分类正确率提高 7.7%。在表 1 中, 居民地和山地光谱属性数据存在交叉, 具有较大的错分率。但是居民地和山地的 DEM 数据有较大的差异, 因此加入 DEM 数据后, 表 2 中居民地和山地分类精度明显提高。

图 3 为试验区的假彩色合成影像。图 4(a) 为基于遥感光谱数据使用有效分类规则进行遥感影像分类的结果, 经统计漏分率为 11.0%。图 4(b) 为基于遥感光谱数据和 DEM 数据使用有效分类规则进行遥感影像分类的结果, 经统计漏分率为 5.2%, 漏分率降低 5.8%。



图 3 TM5 4 3 假彩色合成影像

Fig 3 TM5 4 3 pseudo color synthesize image

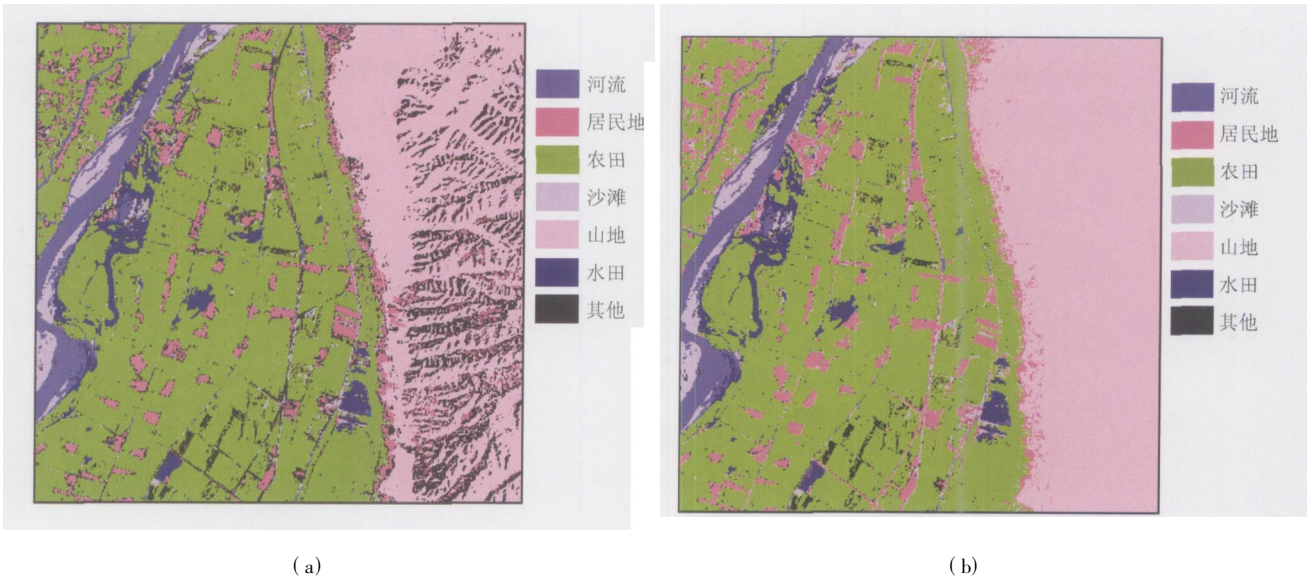


图 4 遥感影像分类结果

Fig 4 Remote image classifier outcome

6 结 论

本文以 GIS 为平台, 构建多源空间数据库, 将分类规则挖掘方法引入遥感影像分类中, 增强了基于

符号知识的逻辑推理分类方法的实用性。针对空间数据库的特性, 提出了一种新的连续样本分类规则挖掘算法, 扩展了归纳学习算法对连续型属性样本分类的适应性。实验表明: (1) 该算法是一种简单有效的有监督分类算法, 具有较高分类精度; (2) 该

算法可以判断属性重要性和进行条件属性约简,对于高光谱遥感影像分类的优势尤为明显;(3)简化了分类时的复杂推理;(4)将遥感光谱数据与其他相关空间数据相结合,可以提高遥感影像的分类精度。

参 考 文 献 (References)

- [1] Wilkinson G G. A Review of Current Issues in the Integration of GIS and Remote Sensing Data [J]. *Int J. Geographical Information Systems* 1996 10(1): 85—101
- [2] Luo J G, Zhou C H, Yang Y. Land Cover and Land Use Classification Based on Remote Sensing Intelligent Geo-Interpreting Model [J]. *Journal of Natural Resources* 2001, 16(2): 179—183 [骆剑承, 周成虎, 杨艳. 遥感地学智能图解模型支持下的土地覆盖 土地利用分类 [J]. 自然资源学报, 2001, 16(2): 179—183]
- [3] Jayantha E Hierarchical Maximum-Likelihood Classification for Improved Accuracies [J]. *IEEE Transactions on Geosciences and Remote Sensing* 1997, 36(4): 1122—1143
- [4] Shu H L, Mao Z Y. Knowledge Based Image Classification Approach Supported by a GIS [J]. *Acta Geodaetica et Cartographica Sinica* 1997, 26(4): 328—336. [术洪磊, 毛赞猷. GIS辅助下的基于知识的遥感影像分类方法研究——以土地覆盖 土地利用类型为例 [J]. 测绘学报, 1997, 26(4): 328—336.]
- [5] Zhou C H, Luo J G, Yang X M *et al*. The Geo-Interpreting and Analysis of Remote Sensing Image [M]. Beijing: Science Press, 1999 [周成虎, 骆剑承, 杨晓梅等. 遥感影像地学理解与分
- 析 [M]. 北京: 科学出版社, 1999]
- [6] Di K G, Li D R, Li D Y. Study of Remote Sensing Image Based on Spatial Data Mining Techniques [J]. *Journal of Wuhan Technial University of Surveying and Mapping* 2000 25(1): 42—48. [邱凯昌, 李德仁, 李德毅. 基于空间数据发掘的遥感图像分类方法研究 [J]. 武汉测绘科技大学学报, 2000 25(1): 42—48.]
- [7] Luo J G, Zhou C H, Yang Y. ANN Remote Sensing Classification Model and Its Integration Approach with Geoknowledge [J]. *Journal of Remote Sensing* 2001 5(2): 123—129 [骆剑承, 周成虎, 杨艳. 人工神经网络遥感影像分类模型及其与知识集成方法研究 [J]. 遥感学报, 2001 5(2): 123—129]
- [8] Derek R. Peddlea, David T. Ferguson. Optimization of multisource data analysis: an example using evidential reasoning for GIS data classification [J]. *Computers & Geosciences* 2002 28: 45—52
- [9] Ursula C. Benz, Peter Hofmann, Gregor Willhauck *et al*. Multi-resolution Object-oriented Fuzzy Analysis of Remote Sensing Data for GIS-ready Information [J]. *ISPRS Journal of Photogrammetry & Remote Sensing* 2004 58: 239—258
- [10] Wang G Y. Rough Set Theory and Knowledge Discovery [M]. Xi'an: Xi'an Jiao Tong University Press, 2001 [王国胤. Rough集理论与知识获取 [M]. 西安: 西安交通大学出版社, 2001.]
- [11] Han J W, Micheline Kamber, Fan M, Meng X F. Interpretation Data Mining Concepts and Techniques [M]. Beijing: China Machine Press, 2001 [Han Jawei, Micheline Kamber 著. 范明, 孟小峰译. 数据挖掘概念与技术 [M]. 北京: 机械工业出版社, 2001.]