

# 遥感数据的粗糙集表示及分类

刘 斌, 陆 华, 刘国涛

(成都理工大学信息管理学院, 成都 610059)

**摘 要:** 由于粗糙集理论对不确定性的描述是相对客观的, 且在无需先验信息的情况下提供了严格地处理数据分类问题的数学方法, 结合遥感数据的不确定性, 就能对遥感数据进行分类处理。文章通过对遥感数据的处理, 经过数据离散化等一系列数据处理后得到约简属性, 再根据 Fisher 线性判别对遥感数据进行监督分类, 得到分类结果。可以得出粗糙集理论应用于遥感数据分类具有良好的应用前景。

**关键词:** 粗糙集算法; 遥感数据; 图像分类

**中图分类号:** TP391.4

**文献标识码:** A

遥感图像分类是遥感图像计算机解译中的基础工作, 遥感图像分类的质量对于遥感图像的计算机解译有着非常重要的意义, 实际应用的许多方面会涉及到这一问题, 此外, 在一些新的领域和前沿, 如基于图像内容的数据检索、基于图像内容和对象的压缩等, 遥感图像分类是其必不可少的环节。并且, 图像分类在土地利用、农作物估产、环境检测、工程勘测、交通规划管理、基础设施布局、自然灾害监测以及各种资源的调查普查等方面发挥了十分重要的作用<sup>[1-2]</sup>。所以研究遥感分类过程中的不确定性也就具有十分重要的意义<sup>[3]</sup>。本文以实际遥感数据做基础, 通过粗糙集理论应用在遥感数据的分类上。

## 1 研究区概况及数据源

本例的研究区选择岷江上游的毛儿盖地区, 其范围是: 东经  $103^{\circ}00' - 103^{\circ}30'$  北纬  $30^{\circ}20' - 32^{\circ}40'$ , 研究区面积约为  $1\ 760\ \text{km}^2$ , 跨松潘县和黑水县的部分区域。区域内水系发育较好, 较大的水系有贯穿整个研究区的毛儿盖河、热务沟等, 该区域植被与土壤的形成与发育复杂多样, 沿高程有明显的垂直地带性。其植被分区上属于泛北极植物区中国喜马拉雅植物亚区横断山脉地区的一部分, 植被垂直带分布明显, 表现出明显的温性森林、干旱河谷灌丛、亚高山森林、亚高山草甸和灌丛等生态类型。本研究区包含了岷江上游区域典型的植被、土

壤和地貌等特征。本例的原始图像是岷江上游毛儿盖地区的遥感 ETM 图像, 它是  $1\ 654 \times 1\ 324$  像素的图像, 为了更清晰地显示和分析分类结果, 本文选取原图中  $100 \times 100$  像素的图像进行分类, 同时截取了本地区植被分布与 ETM 数据反演的植被含水率的叠图中包含图 1 (a) 相应地区的图像作为分类结果分析的参考, 如图 1 (b) 所示, 本地区的植被类型主要有天然草地、旱地、灌木林过渡带、灌木林和常绿林水浇地等种类。2000 年研究区遥感本底调查植被分类为水浇地、灌木林、灌草过渡带、天然草地、河流和裸岩。

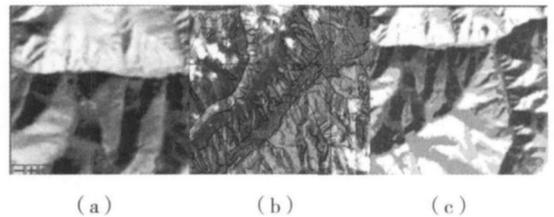


图 1 研究区原始图像

## 2 样本的选取与离散化

### 2.1 样本的选取

通过遥感软件 env i 对该地区遥感图像进行聚类, 类别选定为 6 个类别, 分类结果如图 1 (c)。对遥感影响各个不同类别地物进行采样, 可以得到一个  $400 \times 400 \times 7$  的像素值表, 为了方便运算, 取  $100 \times 100 \times 7$  范围内的

像素值, 并对采样数据建立一个决策表, 选取 60 个样本, 每一类选取 10 个, 决策表形式表达的遥感影像知识表达系统示例见表 1。表 1 中, 遥感图像有 7 个波段, 分成 6 类, 条件属性有 7 个 (7 个波段), 决策属性有 1 个。lno 和 cno 分别代表提取的像素值的坐标, c1, ..., c7 是条件属性, c 是决策属性, 代表分类类别: 1 代表水浇地, 2 代表灌木林, 3 代表灌草过渡带, 4 代表天然草地, 5 代表河流, 6 代表裸岩。

表 1 决策表

lno	cno	c1	c2	c3	c4	c5	c6	c7	c
11	8	78	81	119	76	73	110	63	1
11	9	75	80	110	47	73	106	63	1
11	10	72	76	98	8	70	102	61	1
11	11	65	62	86	31	70	99	61	1
11	12	59	49	81	60	75	95	65	1
11	13	59	53	97	66	80	91	66	1
11	14	62	73	103	47	79	91	65	1
11	15	68	75	102	47	75	91	64	1
11	16	72	71	101	19	73	92	62	1
11	17	68	70	98	11	71	92	60	1
40	29	62	66	86	145	74	126	64	2
40	30	72	83	118	164	75	124	66	2
40	31	72	78	87	147	74	120	64	2
40	32	62	60	55	149	74	120	63	2

### 2.2 遥感数据的离散化

根据等频率划分, 给定的参数 6 把 60 个对象分成段, 每段中有 10 个对象, 将这个属性在所有实例上的取值从小往大进行排列, 然后平均划分为 6 段, 即得到断点集, 每相邻两个断点之间所包含的属性值的个数是相等的。经过等频率划分处理后得到等频率断点如图 2 所示。得到的离散数据是一个不可分辨矩阵, 由此可以得到分辨矩阵见表 2。

0	0	0	0	0	1	1	1	1
66.5	73.5	93.5	127.5	170.5	74	100.5	134	152.5
1	2	2	2	2	2	3	3	3
178.5	99.5	171.5	211.5	231	249.5	105	148	189.5
3	3	4	4	4	4	4	5	5
206.5	248	74.5	93.5	121.5	149	186.5	111	118.5
5	5	5	6	6	6	6	6	
123	129.5	167	63.5	76	101.5	120.5	160.5	

图 2 等频率断点图

表 2 分辨矩阵

c1	c2	c3	c4	c5	c6	c7	c
2	1	1	0	0	0	0	1
0	0	0	0	1	0	1	1
1	0	1	0	0	0	0	1
0	0	0	1	0	3	1	2
1	1	1	2	1	3	1	2
0	0	0	2	0	2	0	2
2	2	2	1	3	3	3	3
2	2	2	1	4	3	3	3

### 3 属性约简

基于分辨矩阵的启发式算法步骤如下:

- (1) 令约简后得到的属性集合等于条件属性集合, 即  $reduct = R$ 。
- (2) 计算分辨矩阵 M, 并找到所有不包含核属性的属性组合  $S$ 。
- (3) 将所有不包含核属性的组合表示为析取范式的形式, 即  $P = \bigwedge \{ \bigvee a_{i_j} \mid i = 1, 2, \dots, s, k = 1, 2, \dots, m \}$ 。
- (4) 将转化的析取范式的形式按属性重要度的计算公式计算属性的重要度。
- (5) 选择其中属性最小的属性  $a$ , 使得  $reduct = reduct - \{a\}$ 。
- (6) 根据  $p_1/p_0 < \alpha$  判断约简是否成立。取  $\alpha = 0.05$  若成立, 删除因条件属性约简而引入的冗余样本和不一致样本, 若  $i = i + 1$  再删除一个剩下的属性重要性最小的, 再来判断, 直到判断条件不成立, 则恢复约简该属性前的样本数据, 结束约简<sup>[4]</sup>。

最终得到的约简结果为 {c4 c5}, 正确度可以达到 100%, 长度为 2。可以从结果中看出约简为条件属性 {c4 c5} 时其拟合精度达到 100%, 满足评审的要求。

### 4 分类结果

#### 4.1 Fisher线性判别

假设有一集合 X 包含 N 个 d 维样本  $x_1, x_2, \dots, x_n$  其中  $N_1$  个属于  $w_1$  类的样本记为子集  $X_1$ ,  $N_2$  个属于  $w_2$  类的样本  $X_2$ 。若对  $x_n$  的分量作线性组合可得标量:

$$y_n = w^T x_n, n = 1, 2, \dots, N_i$$

这样便得到 N 个一维样本  $y_n$  组成的集合, 并分为两个子集  $y_1$  和  $y_2$ 。

#### 4.2 7 个波段的分类结果

根据 Fisher 线性判别<sup>[5]</sup>, 对遥感数据监督分类的分类系数见表 3。根据 7 个条件属性可以看出, 选取 60 个分类样本数据, 剩下 9 940 个待分类数据, 对应 6 个类别分别有 (2 755 820 1 691 868 1 959 1 847), 将遥感数据分成 6 类, 可以得到一个分类统计, 第一类占 0.277, 第二类占 0.082, 第三类占 0.170, 第四类占 0.087, 第五类占 0.197, 第六类占 0.186, 分类精度达到了 98.3%。将分类结果通过图像转化得到的图像如图 3 所示。

#### 4.3 第四和第五属性分类结果

根据 Fisher 线性判别, 对遥感数据进行监督分类的分类系数见表 4。以第四和第五波段为条件属性可以看出, 选取 60 个分类样本数据, 剩下 9 940 个待分类数据, 对应 6 个类别分别有 (2 755 820 1 691 868 1 959 1 847), 将遥感

表 3 分类系数

类别 波段	1	2	3	4	5	6
c1	0.892	0.32	0.426	2.753	1.042	1.312
c2	-0.469	0.0166	-0.339	-1.678	-0.752	-0.952
c3	0.854	0.384	1.406	2.119	1.578	1.547
c4	-0.221	0.465	0.0539	-0.0528	0.192	0.17
c5	1.277	0.49	2.054	2.25	3.356	4.004
c6	1.105	1.548	1.123	1.504	1.554	1.364
c7	-1.537	-1.147	-1.938	-2.992	-3.503	-3.091
常数	-106.103	-140.367	-237.177	-471.634	-374.758	-525.695

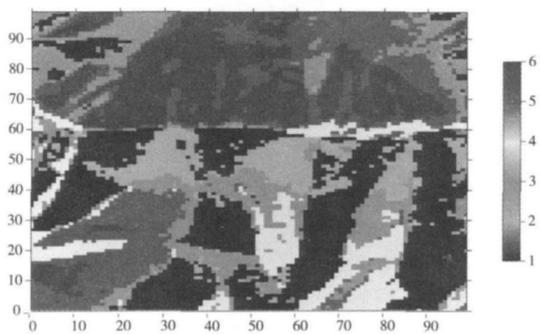


图 3 7个波段的分类图

数据分成 6类,可以得到一个分类统计,第一类占 0.24 第二类占 0.202 第三类占 0.108 第四类占 0.077,第五类占 0.177,第六类占 0.196 分类精度达到了 100%. 根据约简属性得到的约简结果如图 4所示。

表 4 分类系数

类别	1	2	3	4	5	6
c4	0.0166	0.525	0.351	0.625	0.582	0.573
c5	0.927	0.705	1.547	1.158	1.78	2.615
常数	-36.38	-71.024	-133.634	-132.957	-211.103	-373.095

### 5 结束语

本文着重介绍了基于粗糙集理论的遥感影像分类知识发现的方法流程:首先对遥感影像进行采样,对样本数据及遥感数据进行离散化;其次,对离散化后的采样

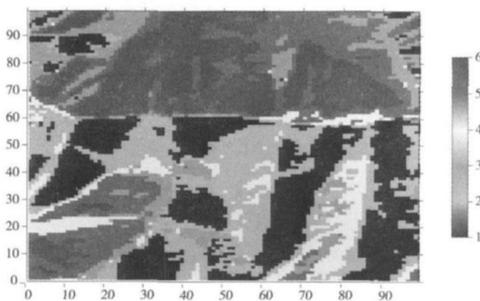


图 4 第四和第五波段分类图

数据进行属性约简。提出了一种基于分辨矩阵的启发式属性约简算法,大大提高了算法的效率,再次从约简后的知识中得出规则集,最后生成分类,对遥感数据进行分类。尽管粗糙集对知识的不完全处理有其优势,但由于粗糙集理论本身未包含不确定或者不精确原始数据的因素,单纯使用粗糙集理论,不一定能有效地描述不精确或不确定的实际问题,所以,在实际使用时,粗糙集方法常常需要与其他方法结合起来使用,互为补充。

### 参考文献:

- [1] 梅安新,彭望绿,秦其明,等. 遥感导论 [M]. 北京:高等教育出版社,2001
- [2] 倪金生,蒋一军,张富民. 遥感图像处理理论与实践 [M]. 北京:电子工业出版社,2008
- [3] 葛咏,王劲峰. 遥感信息的不确定性研究误差传递模型 [M]. 北京:地质出版社,2003
- [4] 简清明,曾黄麟,叶晓彤. 粗糙集特征选择和支持向量机在入侵检测系统中的应用 [J]. 四川理工学院学报:自然科学版,2009,22(5): 62-64.
- [5] Zhu William. Relationship among basic concepts in covering-based rough sets [J]. Information Sciences, 2009, 179: 2478-2486

## Classification and Presentation of Remote Sensing Data on Rough Set

LIU Bin, LU Hua, LIU Guo-tao

(School of Management of Information, Chengdu University of Technology, Chengdu 610059 China)

**Abstract** Because of rough set theory to the description of uncertainty is relatively objective, and provided a rigorous classification of data to deal with the issue of mathematical methods without a priori information. Combined with the uncertainty of remote sensing data, the classification of remote sensing data could be done. In the paper, through the processing of remote sensing data, for example, data discretization and so on, attribute reduction could be got. According to Fisher Linear Discriminant to supervised classification of remote sensing data, then the classification results could be gained. So the results show that rough set theory can be applied to the classification of remote sensing data, and has a good application prospects.

**Key words** rough set algorithm, remote sensing data, classification of image