

# 三峡库区岩性空间分布规律挖掘

王贤敏 牛瑞卿

(中国地质大学地球物理与空间信息学院 武汉 430074)

**摘要** 我国广大南方地区,岩石上部覆盖着较厚的土壤和茂密的植被,因此在南方地区开展的遥感岩性分析研究较少,尚无成熟的方法可循。三峡库区属于南方高植被覆盖区域,岩性分析比较困难。针对三峡库区地形复杂、地质灾害频繁、土壤植被发育的情况分析和挖掘出岩性空间分布的规律;通过将遥感影像与地质图叠加,选择各地层边缘的像素点,分析其中心、正上方、正下方、正左方、正右方、左上方、右上方、左下方、右下方9个方向上的岩性,基于概念格算法和规则提取,挖掘出三峡库区嘉陵江组二段  $T_{1j}^2$ 、嘉陵江组三段  $T_{1j}^3$ 、巴东组一段  $T_2b^1$ 、巴东组二段  $T_2b^2$ 、大冶组  $T_1d$  等地层的岩性邻近规则和空间分布规律。

**关键词** 遥感影像,岩性分析,知识发现,空间分布

**中图分类号** TP391.4, TP309.2 **文献标识码** A

## Spatial Distribution Rule Mining of Lithology in Three Gorges

WANG Xian-min NIU Rui-qing

(Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China)

**Abstract** In the south area of china, there covers thick soil and flourish vegetation on the top of the rocks, so there is little research on lithology analysis by remote sensing in the south area, and there is also no mature methods on this aspect. Three Gorges is the south area which is thickly covered by vegetation and the lithology analysis is very difficult. Three Gorges possesses the characters of complicated terrain, frequent geologic disasters and full-grown vegetation and soil. We mined the spatial distribution rules of lithology aiming at the area of Three Gorges. We piled the remote sensing image with the geological graph, chose the pixel points on the edges of various stratum, analyzed the lithology in the adjacent 9 directions and based on the algorithm of concept grid mined the adjacent rules and spatial distribution rules of the lithology in the stratum of Jia Second Section Jia Ling River Group  $T_{1j}^2$ , Jia Third Section Jia Ling River Group  $T_{1j}^3$ , Ba First Section Ba Dong Group  $T_2b^1$ , Ba Second Section Ba Dong Group  $T_2b^2$  and Da Ye Group  $T_1d$ .

**Keywords** Remote sensing image, Lithology analysis, Knowledge discovery, Spatial distribution

遥感影像真实地记录了岩石光谱辐射特征及其形态特征。岩性分析是遥感地质填图中一个重要而且难度较大的领域。三峡库区地形复杂,地质灾害频繁,因此对三峡库区进行遥感岩性分析的研究显得尤其重要。

目前国内外关于遥感岩性分析的研究集中在植被稀少、人为影响较弱、岩石裸露率高的地区<sup>[1-5]</sup>,而很少在土壤植被较发育、岩石露头少的地区开展岩性分析研究<sup>[6]</sup>。我国广大的南方地区气候温暖湿润,岩石上部覆盖着较厚的土壤和茂密的植被,在影像上表现的通常是土壤和植被的信息,岩性信息很弱,给遥感岩性分析带来了很大的困难。目前,在南方地区开展的遥感岩性分析研究较少,尚无成熟的方法可循<sup>[4,5]</sup>。三峡库区属于南方高植被覆盖区域,岩性分析比较困难,笔者基本没有查到关于三峡库区岩性分析的报道。

近年来发展了一系列遥感岩性信息提取方法<sup>[4]</sup>,有混合光谱分解、NAPC(Noise-adjust Principal Components Trans-

form)、基于主成分的对对应分析、光谱角度填图、匹配滤波、相关波段吸收深度分析等,并且引入了神经网络方法、小波变换方法、分形方法等。但目前采用的遥感岩性分析方法大多数智能性较差,主要依靠目视解译进行岩性判读和分析,较少结合专家领域知识和数据挖掘领域的方法<sup>[3-5,7,8]</sup>。

形式概念分析理论,也称作概念格理论,是一种用数学的形式化语言来反映人形成概念的过程的集合理论模型,非常适合于发现数据中潜在的概念,能够对遥感影像数据挖掘的知识规则在理论上进行很好的解释。但目前国际上,将形式概念分析应用于空间数据挖掘的相关研究还很少,且没有结合空间数据的特点<sup>[9]</sup>。

本文针对三峡库区地形复杂、地质灾害频繁、土壤植被发育的情况分析和挖掘出岩性空间分布的规律;通过将遥感影像与地质图叠加,选择各地层边缘的像素点,分析其中心、正上方、正下方、正左方、正右方、左上方、右上方、左下方、右

到稿日期:2008-07-02 返修日期:2009-03-20 本文受国家自然科学基金项目(40672205),国家863高新技术研究发展计划(2007AA12Z160),中国地质大学优秀青年教师科学基金(CUGQNL0813)资助。

王贤敏(1978-),女,博士后,主要研究方向为空间信息隐藏技术、空间数据挖掘,E-mail:wangmin10291@sina.com;牛瑞卿(1969-),男,副教授,主要研究方向为天空地一体化地球观测信息融合与可视化、人类工程活动与岩土体变化遥感检测技术等。

方 9 个方向上的岩性;基于概念格算法和规则提取,挖掘出三峡库区嘉陵江组二段  $T_{1j}^2$ 、嘉陵江组三段  $T_{1j}^3$ 、巴东组一段  $T_2b^1$ 、巴东组二段  $T_2b^2$ 、大冶组  $T_{1d}$  等地层的岩性邻近规则和空间分布规律。通过对岩性空间分布规律的研究,能够为三峡库区岩性的智能分类和解译提供重要的信息。

### 1 形式背景构造

本文选择三峡库区湖北恩施巴东县作为研究区域。该区域的主要地层有:嘉陵江组二段  $T_{1j}^2$ ,呈现蜂窝纹理状,主要由灰岩、白云岩、溶崩角砾岩构成;嘉陵江组三段  $T_{1j}^3$ ,主要由含泥质灰岩夹白云岩灰岩及灰岩,中部夹燧石团块构成,局部见蠕虫状、条带状构造;巴东组二段  $T_2b^2$ ,该地层上部是紫红色页岩、砂质页岩夹薄一中厚层石英粉砂岩,具交错层理,下部是砂质粘土岩夹中厚层钙、泥质胶结的石英粉砂岩、泥岩灰岩及含铜砂岩;巴东组一段  $T_1b^1$ ,泥页岩以氧化色为主,岩石破碎状、块状层发育,底部发育白云岩,页岩页理发育,偶夹泥质粉砂岩;大冶组  $T_{1d}$ ,主要由灰、黄灰色钙质页岩夹薄一中厚层泥质灰岩构成。将三峡库区湖北恩施巴东县地区的遥感影像与地质图叠加,选择各地层边界处的点,对其进行 8 领域空间邻近关系的空间分布规律分析,分别以 8 领域方向为属性项,以具体的地层为属性值,得到如表 1 所列的多值属性背景表。

表 1 多值属性背景表

中心位置	属性							
	正上方	正下方	正左方	正右方	左上方	右上方	左下方	右下方
$T_{1j}^2$	$T_{1j}^2$	$T_{1j}^2$	$T_{1j}^2$	$T_{1j}^2$	$T_{1j}^2$	$T_{1j}^2$	$T_{1j}^2$	$T_{1j}^2$
$T_{1j}^3$	$T_{1j}^3$	$T_{1j}^3$	$T_{1j}^3$	$T_{1j}^3$	$T_{1j}^3$	$T_{1j}^3$	$T_{1j}^3$	$T_{1j}^3$
$T_2b^2$	$T_2b^2$	$T_2b^2$	$T_2b^2$	$T_2b^2$	$T_2b^2$	$T_2b^2$	$T_2b^2$	$T_2b^2$
$T_1b^1$	$T_1b^1$	$T_1b^1$	$T_1b^1$	$T_1b^1$	$T_1b^1$	$T_1b^1$	$T_1b^1$	$T_1b^1$
$T_{1d}$	$T_{1d}$	$T_{1d}$	$T_{1d}$	$T_{1d}$	$T_{1d}$	$T_{1d}$	$T_{1d}$	$T_{1d}$

### 2 增量式概念格算法

本文采用增量式的概念格构造算法构造概念格。增量式概念格算法只扫描数据库一次,在给定的概念格  $L$  中增量式地插入新事务  $T$ ,从而产生新格  $L'$ 。增量式构造概念格是在给定原始背景  $K=(O,A,R)$  所对应的初始概念格  $L=(CS(K),\leq)$  以及新增对象  $o^*$  的情况下,求解形式背景  $K^*=(O \cup \{o^*\},A,R^*)$  所对应的新的概念格  $L^*=(CS(K^*),\leq)$ 。增量式概念格算法根据初始概念格中的每个节点和新增对象之间的关系,形式化地定义概念格节点的类型。

#### 1) 不变节点

如果格节点  $C$  满足  $C$  的内涵  $Intent(C) \cap f(o^*) = \emptyset$ ,  $f(o^*)$  为新对象  $o^*$  的属性集,则  $C$  为不变节点。不变节点是新格  $L^*$  中保留的原格  $L$  中的节点,节点的内涵与外延均不发生变化;

#### 2) 更新格节点

如果一个格节点  $C$  满足  $Intent(C) \subseteq f(o^*)$ ,则  $C$  称为一个更新节点。更新节点的内涵不变,外延增加:  $Extent(C) = Extent(C) + 1$ ;

#### 3) 产生子格节点

如果某个格节点  $C=(O,A)$  满足如下两个条件:

(1) 令  $Intersection = Intent(C) \cap f(o^*)$ ,在  $L$  中不存在某个节点  $C_1$ ,满足  $Intent(C_1) = Intersection$ ;

(2) 对于  $L$  中与  $C$  具有偏序关系的节点  $C_2 > C$ ,均有  $Intent(C_2) \cap f(o^*) \neq Intersection$  成立。

则  $C$  为产生子格节点。

#### 4) 新生格节点

对于新格  $L^*$  中的任何一个节点  $C$ ,如果  $L$  中不存在另外一个节点  $C_1$  满足  $Intent(C_1) = Intent(C)$ ,则节点  $C$  称为新生格节点。

增量式概念格构造算法在数据库中每读取一条记录,就生成一个事务集  $T$ ;从概念格  $L$  的顶层节点开始搜索,根据层数的降序,将事务集  $T$  与格  $L$  中的节点依次比较和做集合运算,从而在概念格中动态生成和插入新节点,同时更新节点的父子关系。

### 3 三峡库区岩性空间分布规律挖掘

岩性规则采用形如  $A \Rightarrow B$  产生式规则的表达式,其中  $A$ ,  $B$  为属性集。它的直观含义是,在数据库中具有属性集  $A$  的对象可能也具有属性集  $B$ 。每条规则带有两个参数:支持度和置信度。

#### 3.1 寻找频繁封闭项集

对于封闭项集,如果同时满足最小支持度阈值,则为频繁封闭项集<sup>[12]</sup>。对于构造的概念格,如果概念格节点满足最小支持度阈值,则为频繁概念格节点。每个频繁概念格节点的内涵即为一个频繁封闭项集。

#### 3.2 生成频繁封闭项集的产生子

根据频繁项集的产生子来产生规则。产生子可以认为决定该概念节点所对应的概念是该概念的本质属性,或者说是产生该频繁封闭项集的子项集,即内涵缩减集。

产生子的计算方法如下:

1) 计算每个频繁封闭格节点的内涵集的所有非空子集。如果某个子集是其父节点内涵的子集,则删除。

2) 对于每一个子集,如果存在另一个子集是该子集的真子集,则删除该子集。

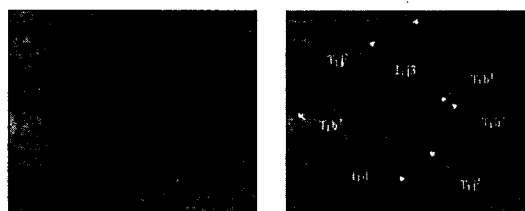
经过以上处理后剩下的就是该频繁封闭节点的产生子,根据这些产生子计算关联规则。

#### 3.3 生成岩性规则

分析概念格节点之间的包含与被包含关系,生成岩性规则。在概念格中,从上到下进行计算,对于两个相邻的具有父子关系的频繁封闭节点,根据其对应的频繁封闭项集及其产生子生成岩性规则。对于两个相邻的具有父子关系的频繁封闭节点,其对应的频繁封闭项集及其产生子为父节点  $f_1, G_{f_1}$ ,子节点  $f_2, G_{f_2}$ ,则产生的规则为  $r: g \Rightarrow (f_2 - g) | g \in G_{f_1}$ 。

### 4 三峡库区岩性空间分布规律挖掘实验

本文采用三峡库区湖北恩施巴东县的一块 TM 子影像 (411×340) 作为实验遥感影像,如图 1(a) 所示。其与地质图叠加以后的图像如图 1(b) 所示,其中各地层线用箭头标识。



(a) 三峡库区 TM743 波段影像 (b) 叠加地质图的图像

图 1 三峡库区遥感影像和指标图像

本文基于增量式概念格算法,选取 70 个对象,按照表 1 给出的多值属性背景表,生成的三维概念格如图 2 所示。取最小支持度阈值  $\text{minsupport}=0.1$ ,最小置信度阈值  $\text{minconfidence}=0.8$ ,挖掘出岩性空间分布规律空间邻近规则 250 条。由于篇幅所限,仅列出其中 10 条规则,如表 2 所列。

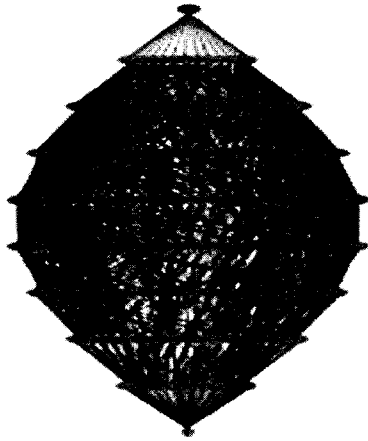


图 2 生成的概念格

表 2 岩性空间分布规律

规则编号	规则	支持度	置信度
1	右上方位置 $T_1b^2 \rightarrow$ 左下方和右下方位置均为 $T_1b^1$	0.17	0.92
2	右下方位置 $T_1j^2 \rightarrow$ 右上方位置 $T_1j^3$	0.14	0.83
3	正上方位置 $T_1b^2 \rightarrow$ 右下方位置 $T_1b^1$	0.18	0.92
4	正上方位置 $T_1b^2 \rightarrow$ 正下方位置 $T_1b^1$	0.18	0.92
5	中心位置 $T_1j^2$ ,右下方 $T_1j^2 \rightarrow$ 右上方 $T_1j^3$	0.12	0.81
6	中心位置 $T_1j^2$ ,左下方位置 $T_1j^3 \rightarrow$ 左上方位置 $T_1j^2$ ,正下方 $T_1j^3$	0.14	0.9
7	左下方和正下方位置均为 $T_1j^2 \rightarrow$ 右上方和正上方位置均为 $T_1j^3$	0.14	0.9
8	中心和左下方位置均为 $T_1j^3 \rightarrow$ 右上方和正上方位置均为 $T_1b^1$	0.12	0.9
9	中心位置 $T_1j^3$ ,右上方位置 $T_1b^1 \rightarrow$ 左下方位置 $T_1j^3$ ,正上方位置 $T_1b^1$	0.12	0.9
10	左下方位置 $T_1j^3$ ,正上方位置 $T_1b^1 \rightarrow$ 右上方位置 $T_1b^1$ ,正下方位置 $T_1j^3$	0.12	0.81

由于篇幅所限,本文仅取挖掘出的规则中的几条进行解释。

规则:右下方位置  $T_1j^2 \Rightarrow$ 右上方位置  $T_1j^3$ 。该规则的含义是:若南东方向的地层是嘉陵江组二段,则可推断北东方向的地层必为嘉陵江组三段;

规则:正上方位置  $T_1b^2$ ,正左方位置  $T_1b^1 \Rightarrow$ 右上方位置  $T_1b^2$ ,右下方位置  $T_1b^1$ ,左下方位置  $T_1b^1$ ,正下方位置  $T_1b^1$ 。

(上接第 205 页)

[7] 张玲,刘勇,何伟. 自适应遗传算法在车牌定位中的应用[J]. 计算机应用,2008,28(1):185

[8] Gao Li,Dai Shangping, et al. Using Genetic Algorithm for Data Mining Optimization in an Image Database[C]//Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). 2007

[9] Koperski K, Han J. Discovery of spatial association rules in

该规则的含义是:若正北方向的地层是巴东组二段,正西方向的地层是巴东组一段,则可推断北东方向的地层为巴东组二段,南东方向、南西方向、正南方向的地层均为巴东组一段。

可见挖掘出的规则很好地说明了地层的空间临近关系和空间分布规律,并且与实际情况相符,从而为岩性的智能解译提供了重要的判据。

**结束语** 本文针对三峡库区地形复杂、地质灾害频繁、土壤植被发育的情况分析和挖掘出岩性空间分布的规律;通过将遥感影像与地质图叠加,选择各地层边缘的像素点,分析其中心、正上方、正下方、正左方、正右方、左上方、右上方、左下方、右下方 9 个方向上的岩性;基于概念格算法和规则提取,挖掘出三峡库区嘉陵江组二段  $T_1j^2$ 、嘉陵江组三段  $T_1j^3$ 、巴东组一段  $T_2b^1$ 、巴东组二段  $T_2b^2$ 、大冶组  $T_1d$  等地层的岩性邻近规则和空间分布规律。通过对岩性空间分布规律的研究,能够为三峡库区岩性的智能分类和解译提供重要的信息和先验知识。下一步工作将研究挖掘出的规则应用于岩性的智能分类和解译。

## 参考文献

[1] Hunt G R. Spectroscopic Properties of Rocks and Minerals in Handbook of Physical Properties of Rocks[M]. Volume I. Boca Raton: CRC Press, 1982

[2] Clark R N, Roush T L. Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications[J]. Journal of Geophysical Research, 1984, 89(B7): 6329-6340

[3] Rowan L C, Simpson C J, Mars J C. Hyperspectral Analysis of the Ultramafic Complex and Adjacent Lithologies at Mordant [J]. Australia. Remote Sensing of Environment, 2004, 91(3): 419-431

[4] 赵建华,杨树锋,陈汉林. 基于分形纹理的遥感图像岩性识别方法[J]. 遥感信息, 2004(2): 1-4

[5] 黄颖端,李培军. 基于地统计学的图像纹理在岩性分类中的应用[J]. 国土资源遥感, 2003, 3: 45-49

[6] 马超飞,马建文. 应用多源数据提取高植被覆盖地区岩性信息—以湖南祁阳地区为例[J]. 地质科学, 2002, 37(3): 365-371

[7] 库向阳,薛惠锋,雷学武,等. 基于分类规则挖掘的遥感影像分类研究[J]. 遥感学报, 2006, 10(3): 332-338

[8] 孙庆先,方涛,郭达志,等. 空间数据挖掘技术中的划区效应及在矿山中的应用[J]. 煤炭学报, 2007, 32(8): 804-807

[9] 秦昆. 基于形式概念分析的图像数据挖掘研究[D]. 武汉: 武汉大学, 2004

geographic information databases[C] // Proc. of International Symposium on Advance in Spatial Databases, SSD, LNCS, vol. 951, Springer Verlag, 1995: 47-66

[10] Chen G, Wei Q. Fuzzy association rules and the extended mining algorithms[J]. Information Sciences, 2002, 147: 201-228

[11] 马超飞,刘建强. 遥感图像多维量化关联规则挖掘[J]. 遥感技术与应用, 2003, 18(4): 244-246