

文章编号: 1000 0585(2001)05-0637-07

# 遥感影像分类与地学知识发现的集成研究

王 雷, 冯学智, 都金康

(南京大学城市与资源学系, 南京 210093)

**摘要:** 遥感与地学之间存在着数据与知识上巨大的互补性。本文通过地面类型数据将遥感影像分类与地学知识发现结合起来: 用遥感数据驱动发现地学知识, 用地学知识解释、确认、检验遥感分类结果, 并使用统计值和分布谱来定量化表达地学知识, 形成一体化的遥感地学分类系统。

**关 键 词:** 影像分类; 地学知识发现; 分类精度评价

**中图分类号:** TP79 **文献标识码:** A

## 1 问题的提出

遥感和地学是两个关系非常密切的学科, 在这个日益数字化的时代里, 他们在各自的领域里都面临着一些问题和挑战。

对于遥感而言, 地物光谱特征的研究, 已经为这门学科的研究打下了理论基础, 但是由于光谱信号的传播过程受到种种因素的干扰, 图像又来源于不同的传感器、不同的大气条件、不同的时相, 使得关于地物光谱特征的知识无法有效地运用于图像解释中。在实际的图像处理过程中, 无论是监督分类还是非监督分类的方法, 都需要大量的地面知识和图像知识, 将光谱知识用于图像解释是一项复杂的工作, 这种知识的不可重用性严重地制约了遥感的发展和实用化。“我们已被信息所淹没, 但正在忍受缺乏知识的煎熬。”<sup>[1]</sup>对于地学而言, 随着计量革命的发展, 地学研究已逐步向定量化迈进。在长期受到数据匮乏制约之后, GIS 和遥感的发展给地学研究带来了一线曙光。遥感在地面属性特征的获取上有着无可比拟的优势, 但在遥感光谱数据和地学数据之间仍存在着一个瓶颈, 这使得遥感数据的地学反演成为一个研究热点。由于受到数据的制约, 地学规律的发现和定量化描述需要很大的代价和周期, 许多知识处于定性描述的阶段。

综上所述, 在遥感和地学之间, 存在着数据和知识上巨大的互补性, 有没有一种方法, 可以用地学的知识来解释遥感的数据、用遥感的数据来发现地学的知识? 本文在这种思路的基础上, 尝试将遥感影像分类与地学知识发现结合起来, 成为一体的遥感地学分类系统 (图 1)。

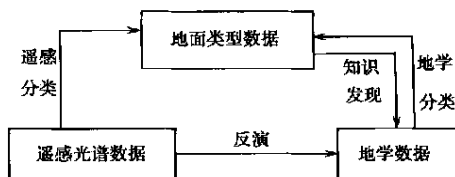


图 1 遥感地学的数据流

Fig. 1 The dataflow of RS & Geography

收稿日期: 2001-04-24; 修订日期: 2001-08-30

基金项目: 中德合作“江宁土地利用与可持续发展”(SILUP)项目资助

作者简介: 王雷 (1977-), 男, 山东鱼台县人, 南京大学 2000 级博士研究生。

## 2 方法的思考

### 2.1 现有方法的概述

在遥感分类的发展中,将光谱数据与地面数据结合起来以提高分类精度已成为共识。只有全面掌握系统的物质与能量信息才能做到对遥感信息的正确解释<sup>[2]</sup>。在地面数据与遥感数据集成分类的方法中,主要有三类:一是信息复合的方法<sup>[2~5]</sup>,将高程、气温、降水等地学数据及次生数据作为一个波段与遥感光谱数据复合在一起参与分类,通过增加初始信息量的办法提高分类精度;这种方法会改变数据在特征空间中的分布,有时会造成数据的压缩和拉伸,增加对结果解释的困难。二是基于规则判断的决策树分类方法<sup>[6~8]</sup>,这种方法适于遥感单要素提取及地学数据分类,光谱数据往往以单变量的形式出现;这种方法的关键在于地物类型阈值的确定,确定阈值的方法有样本观测、经验知识及基于信息熵的方法<sup>[9]</sup>,其中地学数据的阈值具有地学知识的含义。三是影像分类与规则判断相结合的办法,较简单的方法使用地学数据及相关规则对分类结果进行后处理以提高分类精度,较复杂的则将地学规则判断集成在遥感分类过程中<sup>[10]</sup>。

以上三类方法都是着眼于分类方法及提高分类精度,此外,利用遥感分类的结果对地学数据复合统计,可以发现与表述地学知识和规律<sup>[2,8,11]</sup>,这种知识又可以作为信息复合分类中的地学数据的类中心<sup>[2]</sup>。本文就是在这种思路的基础上,将遥感非监督分类与地学知识发现结合起来。

在现有的各种分类方法中,一个普遍性的问题在于:分类过程周期长效率低,对遥感分类的结果精度的评价方法复杂。本文通过将遥感数据与地学知识结合起来,以改进遥感数据分类过程及地学知识的发现与表达。

### 2.2 本文的方法思路

本文提出了一种快速有效地进行遥感影像分类及相关地学知识发现的思路,具体方法是:首先对影像进行较大类别数的非监督分类,利用未知类别与地形数据复合,通过对其相关性的分析,判断类别所代表的地物类型及其分类特性,同时确定了地物类别与地学因子之间的相关性。

在光谱数据空间中,我们用地物光谱知识来解释数据;在地学因子的数据空间中,我们用地学知识来解释数据。要将遥感光谱数据与地学知识结合起来,地面类型数据是一座桥梁。首先通过遥感数据分类将非语义信息转化为地物类型语义信息,再使用这种语义信息在地学数据的非语义信息中进行知识发现,利用统计值和图谱的形式来发现和表达地学规律,最后利用所发现的知识对遥感分类的结果进行判定,并根据数据与知识的一致性对分类结果进行评价。这里所说的知识发现,不是分类过程中从数据到类型的类型发现,而是从类型到数据的相关关系发现。

地物光谱特征 景观地形分布特征

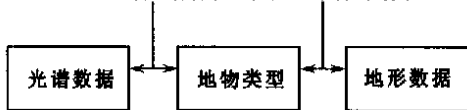


图 2 数据间的知识逻辑关系

Fig. 2 The knowledge relation between data

在我们的实例中,利用遥感景观分布与地形因子之间的相关关系作为知识发现的目标。地物光谱特性是遥感光谱数据与地物类型之间相关的理论基础,遥感景观的地形分布特征则表达了地

物类型与地形之间的相关关系。通过三类数据与两种相关的连锁互动关系，最终完成遥感数据到地学知识的知识发现过程和从地学知识到遥感数据的数据解释过程（图 2）。

3 实践范例

实验采用南京近郊江宁县 SPOT XS 卫星影像数据，数据获取时间为 1998-8-10；经校正后地面分辨率为 22m。地形数据来自 1:5 万地形图，其等高线经 ARC/INFO TOPOGRID 过程生成 DEM，并产生相应地面分辨率的坡度、坡向数据。分类过程如下。

3.1 SPOT 图像的分类

为了提高对植被的分类效果，首先生成 NDVI 比值图像，将其同 SPOT XS 三个波段复合在一起，参与分类。用 ISODATA 非监督分类方法将图像分为 20 类。

然后通过观察类中心的光谱响应值，将 20 类分为两组：植被类型组和非植被类型组。植被组包括各种类型的植被，非植被组包括水、居民地、岩石、荒地；植被组主要受植被（作物）地带性的控制，非植被组则受地形的影响。

事实证明，植被指数对区分类型组十分有效，地物类型在 ndvi 上的排列顺序是：水、岩石、居民地、荒草地、植被，用 ndvi 顺序可以很容易将植被与前四组分开。在本例中，1-9 属于非植被类型组，10-20 属于植被类型组。

表 1 光谱数据非监督分类类型的地形相关统计值

Tab 1 The statistical value of the unsupervised classification result overlayed with topographical data

地形特征	类别码	高程 (H)				坡度 (S)			
		最大值	均值 m	方差 s	众数	最大值	均值 m	方差 s	众数
A	1	144	12. 2	13. 2	7	34	1. 9	1. 9	1
A	2	174	17. 9	16. 5	9	69	2. 2	2. 6	1
B	3	262	28. 6	36. 4	9	71	5. 7	7. 3	1
C	4	242	22. 0	23. 7	9	75	3. 7	5. 3	1
C	5	245	22. 2	21. 1	9	72	3. 3	4. 6	1
C	6	246	25. 6	22. 3	9	73	3. 6	4. 6	1
A	7	234	19. 0	19. 1	9	67	2. 7	3. 8	1
B	8	263	30. 5	33. 3	9	78	5. 9	7. 1	1
C	9	249	27. 1	23. 0	9	72	3. 7	4. 5	1
D	10	326	69. 1	59. 9	9	80	11. 9	8. 3	1
E	11	266	31. 3	29. 7	9	72	4. 9	5. 8	1
D	12	327	63. 5	58. 6	8	81	10. 7	8. 5	1
F	13	233	24. 8	16. 9	20	71	2. 7	3. 1	1
E	14	298	33. 3	37. 4	9	80	5. 5	6. 7	1
E	15	227	23. 6	17. 4	9	76	2. 6	3. 2	1
E	16	293	24. 7	33. 7	8	81	4. 8	6. 7	1
E	17	279	22. 2	27. 4	9	80	3. 5	5. 2	1
F	18	217	24. 1	15. 6	19	78	2. 5	2. 7	1
F	19	221	23. 8	13. 7	20	65	2. 3	2. 3	1
F	20	240	26. 4	21. 7	20	78	3. 1	3. 7	1

### 3.2 分类结果的地形特征判定

(1) 以每一类型数据为模板, 提取高程及坡度数据, 并计算取值范围、均值、方差、众数等统计值得表 1; 并对每一类生成其高度和坡度的相关直方图。

(2) 利用这些地形相关统计值和相关直方图, 对两组进行地形特征的识别确认。在第一组和第二组中分别可以识别出三种地形特征类型, 分别用 A、B、C、D、E、F 表示。

A 类: 包括 1、2、7 三个类别, 其特点是高程和坡度的均值和方差都很低,  $H_m < 20m$ ,  $S_m < 3^\circ$ 。应为低地与水相关的地形类型。

B 类: 包括 3、8 两类, 特点是  $H_m$ 、 $S_m$ 、 $H_s$ 、 $S_s$  均很大,  $H_m > 28m$ ,  $S_m > 5.6^\circ$ ,  $H_s > 30$ ,  $S_s > 7$ ;  $H_s$ 、 $S_s$  的值为第一组中最大, 说明 H、S 的分布离散, 应为平原和山地的混合类型。

C 类: 包括 4、5、6、9,  $22m < H_m < 27m$ ,  $3^\circ < S_m < 4^\circ$ , 其值居中;  $H_s$ 、 $S_s$  亦介于 A、B 两类之间; 应为平原、岗地类型

D 类: 包括 10、12,  $H_m$  达到 63–70m,  $S_m$  达到  $10-12^\circ$ , 为各类中均值和方差最大者, 应为山地类型。

E 类: 包括 11、14、15、16、17, 这几类是以平原为主的类型, 个别类包括一些山地类型。

F 类: 包括 13、18、19、20, 这几类是与岗地相关的类型。与 E 类的主要差别在于: E 类的 H 峰值 (众数) 出现于 10m 左右, 而 F 类的峰值出现于 20m 左右。

D、E、F 三种高植被类型在光谱上是较难确认分辨的, E、F 两类的地形统计特征又有很多混淆之处, 但从地形相关直方图上可以明显的区分其特征。

图 3 中的连续小峰值是由于等高线生成 DEM 过程中, 等高线的标记高程值所造成的, 小峰间的距离即代表 10m 等高距; 从直方图的总体趋势仍可以反映出相关地形特征。

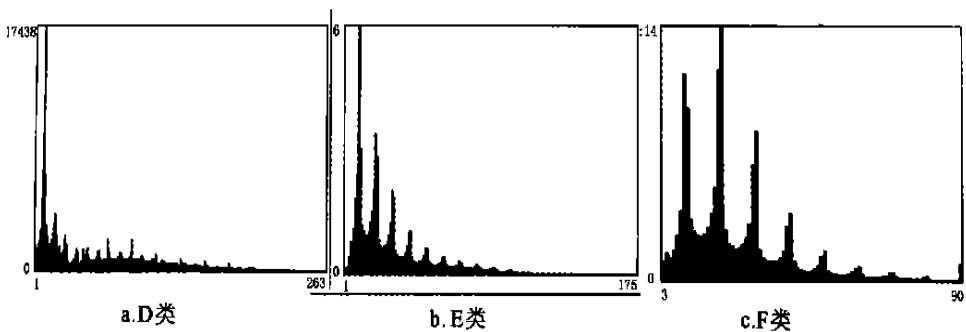


图 3 植被类型的典型高度直方图

Fig. 3 The typical histogram of vegetation distribution on height

D 类的高程直方图存在 50~100m 的波峰, 其坡度在  $10^\circ$  左右出现峰值; 这在整个区域背景和其他各类型中都是很特殊的。E 类和 F 类的高程直方图在峰值和形态上都有明显的差别。

从上面的地形相关分析还可以看出, A、B、C 三类的差别主要体现在统计值上, 反映了地形导致的分布差异。D、E、F 三类的差别可以从直方图形态上容易的区分开, 反

映了农业土地利用的地带性规律。

3.3 地物类型的判定与分类精度评价

(1) 地物类型的判断：综合每类的地形判断，光谱类中心特性，以及对图像结构的观察，对分类结果进行地物类型的确认。

在地形特征 A 中，类别 1 是典型的水类型，类别 2、7 在地形上属水类型，光谱上则介于水与其它类型之间，应为傍水的地物类，以及与水相关的混合象元类。初步判定 2 为水边荒地，7 为傍水植被。

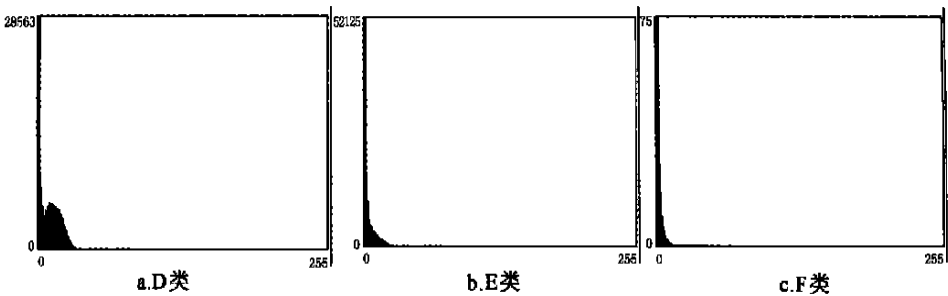


图 4 植被类型的典型坡度直方图

Fig. 4 The typical histogram of vegetation distribution on slope

在地形特征 B 中，类别 3 是典型的岩石类，在平原区为大型的工矿企业，山区则为矿场裸岩；8 是草地类型，初步判定为平原荒草与山地荒草的混分类。

在地形特征 C 中，4、5 为居民地类型，6、9 为荒草地类型。从地形上来看，4、5 的高程均值明显小于 6、9 的高程均值。

根据宁镇丘陵地带的植被与农业作物地带性特征：山区为林地类型，岗地为旱地类型，平原冲谷为水田类型，可以将 D 类判断为林地，E 类判断为水田类型，F 类判断为旱地类型。

(2) 混分的确认：在遥感中，同谱异物是影响分类精度的一个重要因素，判断和评价混分类型也是一项困难的工作。在这里，利用类型的相关地形统计方差，可以发现混分类别并对混分的程度做出判断。

在相关地形因子的统计值中，方差较大说明数据离散程度较高，这种现象有两方面的解释。一是该类型在地形因子空间中的分布比较离散，在本例中，B、F 两类即属此例。B 类中各类的  $H_s > 30$ ， $S_s > 7$ ，在第一组各类中是最大的，因为岩类和草类在平地、山地均有广泛分布。F 类中各类的  $H_s > 58$ ， $S_s > 8$ ，在第二组各类中是最大的，这是因为林地分布于山区，其高程和坡度的变化范围远比平原类型更大。

二是该类型中出现同谱异物的混分现象，而这两类的地形分布特征不同。E 类中的类型即属此种情况。E 类是平原水稻为主的类型，其高程和坡度的均值和方差应该较小，但 E

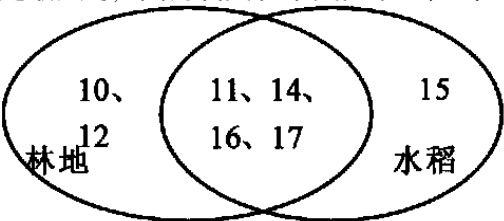


图 5 林地与水稻的光谱混分示意

Fig. 5 The mixed decision of forest and rice

类中 11、14、16、17 的均值和方差都较大 ( $H_s > 27$ ,  $S_s > 5$ ); 通过空间分布及光谱上的进一步研究, 可以发现这些类别处于水稻和林地的光谱混合区。而类型 15 的方差较小, 说明其类型的内聚性好, 处于水稻的非光谱混合区 (图 5)。

混分类型中一般都有一个主要的地物类型, 可以体现其空间分布特征。在一种地物大类内部, 各个子类型之间比较其均值方差, 对混分类性和亚类型的判定比较可靠。

(3) 地物类型与坡向之间的关系: 利用已获取地物类型分别与坡向叠和, 生成坡向直方图, 可以观察地物类型与坡向之间的统计相关性。通过观察发现旱地、居民地与坡向之间的相关性比较明显; 旱地偏东向的频率较高, 居民地则是偏南向的频率较高。

## 4 结论

这种方法提高了遥感非监督分类的效率与可靠性, 有助于地学规律的发现与描述。

分类过程仅需要较少的光谱知识、定性的地面知识, 不需大量的研究图像即可完成。为类型的确认提供定量化的依据, 减少主观性认识的错误。对混合象元类及混分类型的确认有了定量化的方法。

相对于地学现象的不确定性和数据的不确定性而言, 用统计特征值和分布谱的方式描述地学规律是一种有效的方法。从类型到数据的相关关系发现, 可以把对规律的定性理解转向定量描述。用遥感分类中新的类型, 从新的地学数据资料中发现新的相关规律。有助于地学认知过程向全数字化更进一步。

总之, 遥感地学分类的方法是遥感数据和地学知识紧密的集合在一起; 把类型的分割和类型的确认、解释、评价结合在一起; 把类型的发现与知识的发现结合在一起; 把知识的图谱表达与类型的确认结合在一起。

## 参考文献:

- [1] Usama M. Fayyad, etc. From data mining to knowledge discover [M]. AAAI Press, 1997.
- [2] 刘卫国, 等. 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[J]. 遥感学报, 1998, 2(3): 234~ 240
- [3] 刘纪远, 等. 基于 GIS 的中国东北植被综合分类研究[J]. 遥感学报, 1998, 2(4).
- [4] Paul Treitz, Philip Howarth. Integrating Spectral, Spatial, and Terrain Variables for Forest Ecosystem Classification[J]. PE&RS, 2000, (3): 305~ 315.
- [5] Aplsit Elumnoh, Rajendra P Shrestha. Application of DEM data to Landsat Image Classification: Evaluation in a Tropical Wet Dry Landscape of Tailand[J]. PE&RS, 2000, (3): 297~ 303.
- [6] 张健挺. 决策树方法在地学空间数据挖掘中的应用[J]. 地理研究 1998, 增刊.
- [7] 布和敖斯尔. 基于知识发现和决策规则的盐碱地遥感分类方法研究[J]. 中国图形图像学报, 1999, 4(A)(11): 965~ 968.
- [8] 周成虎. 遥感影像地学理解与分析[M]. 北京科学出版社, 1999. 91~ 142.
- [9] 张健挺. 基于信息上的地学数据挖掘模型及其应用研究[D]. 中科院地理研究所博士论文, 1999.
- [10] 术洪磊. GIS 辅助下的基于知识的遥感影像分类方法研究[J]. 测绘学报 1997, 26(6).
- [11] 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[J]. 地理研究, 1997, 16(3): 63~ 69.
- [12] 陈述彭, 岳天祥, 励惠国. 地学信息图谱研究及其应用[J]. 地理研究, 2000, 19(4) 337~ 343.
- [13] 方洪亮, 黄绚. 地学应用中遥感图像处理若干问题分析[J]. 地理研究, 1997, 16(2): 96~ 103.

# On the integration between image classification and geographical knowledge discovery

WANG Lei, FENG Xue zhi, DU Jir kang

(Department of Urban and Resources Science, Nanjing University, Nanjing 210093, China)

**Abstract:** Great complementarity exists between remote sensing image data and geographical knowledge. This paper tries to unify the image classification and geographical knowledge discovery through ground classes data, i. e., to discover geographical knowledge with remote sensing data drive, to confirm, explain and evaluate image classification result with geographical knowledge, and to represent geographical knowledge with statistic value and distribution atlas. All these come to be an incorporated Remote Sensing and Geographic Classification System.

The steps of this method are as follows: firstly, to divide the image into relative big number(> 20) of classes using the unsupervised classification; then overlay these unknown classes with the DEM data and get some statistic values and distribution atlas for each class; finally use these values and atlases to name, explain and evaluate each class of the classification result. Meanwhile the correlation between the ground object type and the topographical data is acquired and expressed as well.

The example shows that this method makes the classification more efficient and reliable, and it is useful to express and discover the geographical knowledge. The conclusion is that, we can use other data to interpret the result of unsupervised classification, name and check each class, and at the same time, acquire the geographical knowledge from the pattern in the image data.

**Key words:** image classification; geographical knowledge discovery; precision evaluation