

文章编号:1671-8860(2008)11-1166-04

文献标志码:A

基于多项 Logit 模型的土地覆被分层分类方法研究

饶 雄¹ 张景雄^{1,2} 丁翼星¹

(1 武汉大学遥感信息工程学院,武汉市珞喻路129号,430079)

(2 武汉大学测绘遥感信息工程国家重点实验室,武汉市珞喻路129号,430079)

摘要:探讨了一种利用多项 Logit 模型分层提取土地覆盖专题信息的方法。考虑客观存在的异物同谱现象,构建分层分类体系,针对不同层的地物类别选取不同的预测变量构建多项 Logit 模型,分步骤地提取各地物类专题信息。将此方法应用于美国蒙大拿州中部地区的土地覆盖专题信息提取,结果表明,该方法较常规的使用同一组特征变量构建单一模型一次性地划分所有地物类的方法在总体分类精度上有了明显改善。

关键词:多项 Logit 模型;逐步回归;分层分类

中图法分类号:P237.4; TP753

常规的遥感分类一般采用一次性分类的策略,在同一分类体系方案下,利用单一的分类标准仅通过一次分类过程提取所有地物类的专题信息。但是,由于客观世界的复杂性和遥感影像本身的空间分辨率以及“同物异谱”和“异物同谱”现象的客观存在,在遥感影像分类过程中,某一特征组合可能只能较好地识别部分目标地物类别,而很难甚至不可能适合识别所有地物,按照同一分类标准对影像进行分类往往很难达到令人满意的分类效果^[1]。如果仍然采用这种分类策略,会因某些地物类别的分类精度不高而导致整体分类精度降低。

随着地学空间分析理论的不断丰富与完善,空间回归技术作为空间分析的重要工具被广泛地应用于地学领域的各个分支。广义线性模型作为一种有效的统计回归工具已经在资源环境、地理建模和生态学等方面得到了应用^[2-5]。本文探讨了利用广义线性模型——多项 Logit 模型分层提取土地覆盖专题信息的方法,以期获得比使用同一模型一次性区分所有地物类的方法更优的分类精度。

1 Logit 模型

二项 Logit 模型是一种常用的广义线性模型,主要用于二元响应变量的概率建模。响应变量 Y

是一个服从二项分布的随机变量,其公式为:

$$\ln[E(Y)] = \ln\left(\frac{p}{1-p}\right) = \alpha + \mathbf{x}'\boldsymbol{\beta} \quad (1)$$

式中, α 是回归方程的截距; \mathbf{x} 是由 q 个预测变量(自变量)的观测值 x_1, x_2, \dots, x_q 组成的向量; $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_q]^T$ 是回归模型系数向量; $E(Y)$ 是响应变量 Y 的期望值; $g()$ 是一个单调且可二次微分的联系函数; p 表示概率。

多项 Logit 模型是二项 Logit 模型的更一般的形式,在多元响应变量的类别之间不存在序次关系的情况下,可以采用多项 Logit 模型来预测类别,其公式为:

$$\ln\left[\frac{P(Y=i | X=\mathbf{x})}{P(Y=r | X=\mathbf{x})}\right] = \alpha_i + \mathbf{x}'\boldsymbol{\beta}_i, \\ i = 1, 2, \dots, k, i \neq r \quad (2)$$

式中, i 表示不同的类别; r 表示参照类,可以事先任意设定一个类别,不影响模型的预测结果。为了便于模型求解,将参照类对应的回归系数向量 $\boldsymbol{\beta}_r$ 设置为零向量。 $P(Y=i | X=\mathbf{x})$ 是度量样本 \mathbf{x} 属于已知类别 i 的后验概率。对于有 k 个类别的非序次响应变量,多项 Logit 模型中便有 $k-1$ 个 Logit 方程,拟合出 k 个概率:

$$P(Y=r | X=\mathbf{x}) = \frac{1}{1 + \sum_{j=1, j \neq r}^k e^{\alpha_j + \mathbf{x}'\boldsymbol{\beta}_j}}$$

$$P(Y = i | X = \mathbf{x}) = \frac{e^{a_i + \mathbf{x}'\beta_i}}{1 + \sum_{j=1, j \neq i}^k e^{a_j + \mathbf{x}'\beta_j}},$$

$$i = 1, 2, \dots, k, i \neq r \quad (3)$$

分别表示待分类样本属于各已知类别的后验概率。

利用多项 Logit 模型进行分类的实质是用高阶的多项式模拟地物类别属性与地理空间数据之间的关系。本文采用多项 Logit 模型模拟土地覆盖类型与其在遥感影像上表现的波谱特征以及地形因子之间的非线性相关关系,它可以明确地表达出影像上某一像元属于各已知类别的概率。假定土地覆盖类型服从多项分布,将其设定为模型的响应变量,把遥感影像、地形因子及其变换形式作为预测变量来构建模型,逐像元预测其所属的地物类别。应用多项 Logit 模型进行遥感影像分类时,首先利用训练样本拟合出 $k-1$ 组模型回归系数,然后将待分类样本的特征向量代入拟合后的模型,得到的 k 个概率值分别代表该样本属于已知各类的概率,将其类别标记为所属概率最大的那一类。

2 实验分析

2.1 数据源及预处理

研究区域位于美国蒙大拿州中部,北纬 $46^{\circ} 29' \sim 48^{\circ} 23'$,西经 $107^{\circ} 52' \sim 111^{\circ} 08'$ 。3 座山脉分别坐落于研究区的东南、南及北部,地形错综复杂,土地覆盖类型丰富。实验选择的遥感数据是 Landsat TM 多波段数据,轨道号为 P38/R27,过境时间是 1996-07-12,空间分辨率为 30 m。同时,实验还采用了地形数据,包括高程、坡度、坡向,空间分辨率为 30 m。数据均已投影到同一坐标系统中并配准。

已知类别样本的采集使用的是分层随机抽样法。实验过程中共选用了 2 434 个已知类别样本点,其中 1 215 个作为训练样本,用于拟合多项 Logit 模型的回归系数,其余的 1 219 个作为检测样本,用于评定模型的预测精度。选择 TM1 ~ TM5, TM7, MNDVI, Elevation, Slope, Slope/Aspect 作为多项 Logit 模型建模中可供选择的预测变量。

2.2 实验结果及分析

构建多项 Logit 模型时,首先将所有可供选择的预测变量引入模型,然后采用逐步回归技术(向前-向后)自动筛选变量,在每一步经 F 检验选择进

入模型的变量,而后再进行 F 检验,看是否需要剔除某些预测变量。这一过程一直进行到既没有变量需要引入,也没有变量需要剔除为止。调整模型以获得最佳分类精度,将分类精度最高的一类或几类从待分类样本中抽取出来,对余下的数据重复上述步骤,分步骤地提取各地物类专题信息。

由于预测变量的选择及其组合方式是影响多项 Logit 模型预测精度的主要因素,本文从研究目标地物类的光谱特征差异入手,首先绘制各地物类的灰度均值曲线(见图 1),计算均值曲线的相关系数,将相关系数大于 0.95 的曲线对应的地物类合并,最终将 17 类合并成 7 个大类,构成分类体系中的第一层地物类,把 17 个地物类作为第二层(见表 1)。针对不同层分别建模预测地物类别,其中第一层的总体分类精度达到 95.89%,Kappa 系数为 0.949,在第一层分类结束后分别对其中的 3 个大类 B_1 、 B_3 和 B_5 进行细分,最终提取出所有目标地物类。

构造一个多项 Logit 模型和一个单隐含层结构的多层感知器网络(MLP)分别对数据进行一

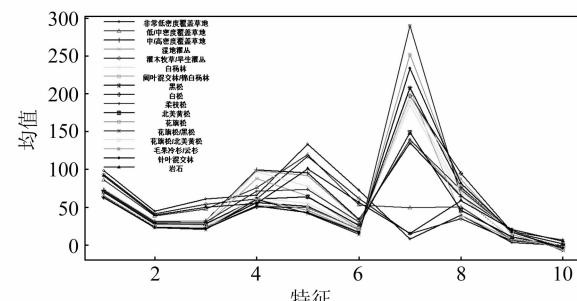


图 1 各地物类的灰度均值曲线

Fig. 1 Mean Gray Value Curves of Land Cover Types

表 1 分层分类体系

Tab. 1 Hierarchical Classification Scheme

大类(第一层)	小类(第二层)
B_1	S_1 非常低密度覆盖草地 S_5 灌木牧草/旱生灌丛
B_2	S_2 低/中密度覆盖草地
B_3	S_3 中/高密度覆盖草地 S_4 湿地灌丛 S_6 白杨林
B_4	S_7 阔叶混交林/棉白杨林
B_5	S_8 黑松 S_9 白松 S_{10} 柔枝松 S_{12} 花旗松 S_{13} 花旗松/黑松 S_{14} 花旗松/北美黄松 S_{15} 毛果冷杉/云杉 S_{16} 针叶混交林
B_6	S_{11} 北美黄松
B_7	S_{17} 岩石

次性分类。经过反复实验,选定一个包含10个结点的输入层、24个结点的隐含层、17个结点的输出层的MLP网络结构。表2列出了基于多项Logit模型的分层分类与一次性分类精度对照表。图2是17种地物类专题图叠合在一起的土地覆盖分类专题图。

本文方法考虑到多项Logit模型的结构对其预测能力的显著影响,针对不同层的地物类分别构建模型,降低了判别空间的维数,减少了模型拟合复杂度。实验结果表明,基于多项Logit模型的分层分类算法是有效的,比常规的使用同一组变量构建单一模型一次性地划分所有地物类的方法在总体精度上有了明显改善。

致谢:感谢美国蒙大拿大学数学学院的Brian M. Steele教授为本文提供宝贵的实验数据。

表2 分类精度对照表
Tab. 2 Comparison of Classification Accuracy of Two Schemes

地物类别	多项 Logit 模型一次性分类		多项 Logit 模型分层分类		MLP 一次性分类	
	用户精度/%	生产者精度/%	用户精度/%	生产者精度/%	用户精度/%	生产者精度/%
非常低密度覆盖草地	70.00	65.63	81.82	84.38	83.33	62.50
低/中密度覆盖草地	93.68	96.22	95.16	96.67	94.70	96.22
中/高密度覆盖草地	50.00	33.33	67.57	52.08	64.10	52.08
湿地灌丛	69.42	77.06	80.00	84.40	73.55	81.65
灌木牧草/旱生灌丛	77.59	77.59	91.23	89.66	79.37	86.21
白杨林	57.14	40.00	60.00	90.00	83.33	50.00
阔叶混交林/棉白杨林	93.38	95.92	94.08	97.28	93.96	95.24
黑松	56.85	63.85	58.78	66.92	65.52	73.08
白松	60.00	75.00	58.82	62.50	60.00	75.00
柔枝松	77.14	71.05	77.27	89.47	78.38	76.31
北美黄松	92.56	94.12	97.44	95.80	93.10	90.76
花旗松	63.58	70.07	66.67	70.75	73.20	76.19
花旗松/黑松	45.45	35.71	50.00	28.57	50.00	42.86
花旗松/北美黄松	66.67	16.67	48.57	16.67	57.14	33.33
毛果冷杉/云杉	36.84	28.00	38.46	20.00	43.48	40.00
针叶混交林	52.17	36.36	39.29	33.33	65.22	45.45
岩石	97.47	93.90	97.53	96.34	97.59	98.78
总体精度/%	76.62		81.10		80.02	
Kappa	0.741		—		0.779	

参 考 文 献

- [1] 赵英时. 遥感应用分析原理与方法 [M]. 北京: 科学出版社, 2002.
- [2] 朱源, 康慕谊. 排序和广义线性模型与广义可加模型在土地覆盖类型与环境关系研究中的应用 [J]. 生态学杂志, 2005, 24(7): 807-811.
- [3] Moisen G G, Jr Edwards T C. Use of Generalized Linear Models and Digital Data in a Forest Inventory of Northern Utah [J]. Journal of Agricultural, Biological, and Environmental Statistics, 1999, 4: 372-390.

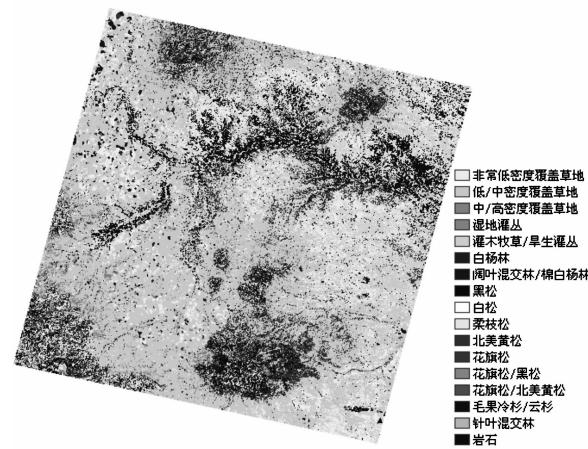


图2 多项 Logit 模型分层分类的土地覆盖专题图
Fig. 2 Land Cover Thematic Map Produced Via Hierarchical Multinomial Logit Modeling

- [4] Arellano-Neri O, Frohn R C. Image-Based Logistic Regression Parameters of Deforestation in Rondonia, Brazil [C]. IGARSS'01, Sydney, Australia, 2001.
- [5] Frohn R C. Comparison of a Landsat-Based Logistic Regression Model and Socioeconomic Simulation Model of Deforestation in the Amazon [C]. IGARSS'01, Sydney, Australia, 2001.

第一作者简介:饶雄,硕士生。主要从事遥感影像智能化处理算法的研究。

E-mail:raushon@126.com

Research and Application of Land Cover Hierarchical Classification Approach Using Multinomial Logit Models

RAO Xiong¹ ZHANG Jingxiong^{1, 2} DING Yixing¹

(1 School of Remote Sensing Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing,
129 Luoyu Road, Wuhan University, Wuhan 430079, China)

Abstract: Since there exists the objective phenomenon that different classes sharing the same spectral characteristics, a hierarchical categorical mapping approach is developed using generalized linear models. According to the similarities of spectral characteristics, different classes of high similar spectral characteristics are merged into the same one. Then a hierarchical modeling scheme is formed. Different predictor variables are chosen to build different multinomial Logit models to extract classes in different layers. Mask technique is employed to extract thematic maps step by step. Ultimately, all thematic maps are incorporated into a whole one. Experimental results show that the hierarchical modeling using generalized linear models is an effective approach to improve land cover mapping quality.

Key words: multinomial Logit models; stepwise regression; hierarchical classification

About the first author: RAO Xiong, postgraduate, majors in remote sensing data intelligent processing algorithm.

E-mail: raushon@126.com

(上接第 1145 页)

Precision and Error Analysis of LCD-Based Camera Calibration Method

ZHAN Zongqian¹ ZHANG Zuxun² ZHANG Jianqing²

(1 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: A LCD-based camera calibration approach using special image shooting method is presented. The results show that the approach is very effective and feasible. Because of the great difference in depth precision between the simulation and actual experimentation, several possible error resources are carefully analyzed. By the stepwise regression analysis using simulated and actual data, non-planarity of LCD is validated. And when the non-planarity of LCD is corrected in the course of LCD-based calibration, precision can be further improved.

Key words: camera calibration; liquid crystal display; stepwise regression analysis

About the first author: ZHAN Zongqian, Ph.D, lecturer, majors in photogrammetry and computer vision.

E-mail: zqzhan@sgg.whu.edu.cn