

引用格式:Zhao hui,Wang Yunjia. Research on the Factors Affecting the Classification Accuracy of ETM Remote Sensing Image Land Cover/Use[J]. Remote Sensing Technology and Application,2012,27(4):600-608. [赵慧,汪云甲. 影响 ETM 影像土地利用/覆盖分类精度因素的研究[J]. 遥感技术与应用,2012,27(4):600-608.]

影响 ETM 影像土地利用/覆盖分类精度因素的研究

赵 慧^{1,2},汪云甲^{1,2}

(1. 中国矿业大学环境与测绘学院,江苏 徐州 221116;

2. 中国矿业大学国土环境与灾害监测国家测绘局重点实验室,江苏 徐州 221116)

摘要:训练样本量、辅助数据和分类法是影响土地利用/覆盖分类精度的 3 个主要因素,通过找到这 3 个因素的最佳组合方式以提高分类精度,分别在 25%、50%、75%、100% 样本量下,加入 NDVI、DEM 和纹理均值特征作为辅助数据,比较了分类回归树、支持向量机、最大似然法 3 种分类法的效果,探讨了训练样本、辅助数据以及分类技术对土地利用/覆盖分类精度的影响。结果表明:支持向量机总体分类精度较高,在相同样本量和没有有效辅助数据的情况下,SVM 可以获得最佳的分类结果,总体分类精度在 85% 以上;在进行分类时,加入 NDVI 和纹理均值特征使分类回归树分类精度提高了 2.82%,说明该方法对有效辅助数据的加入较为敏感;在获取的训练样本集有限而可获取有效的辅助数据时,应优先考虑利用分类回归树进行土地利用/覆盖分类。

关 键 词:决策树;支持向量机;训练样本量;土地利用/覆盖分类

中图分类号:TP 79 **文献标志码:**A **文章编号:**1004-0323(2012)04-0600-09

1 引 言

近 20 a 来,遥感技术在实用方面得到迅速发展,广泛应用于各种国土资源调查与环境评价及灾害监测中。土地利用/覆盖分类是国土资源调查与环境评价及灾害监测的前提。如何从遥感影像中获取高精度的土地利用/覆盖分类,是目前研究的热点,也是迫切需要解决的难题之一。

影响土地利用/覆盖分类精度的因素除了数据源本身的质量和分辨率外,还和分类法、辅助数据以及训练样本有关。国内外学者针对单个因素对土地利用/覆盖分类精度的影响做了大量研究。针对不同分类法对土地利用/覆盖分类精度的影响,Muchoney 等^[1]利用 MODIS 数据对美国中部进行土地利用/覆盖分类,比较了决策树分类法、神经网络法、最大似然法 3 种分类方法的效果,结果显示决策树分类法精度最高;惠文华^[2]对 TM 影像进行了 SVM 分类实验,并与 MLC 结果进行了比较,结果显示 SVM 分类精度

明显优于 MLC。另外,一些学者通过加入辅助数据来改善分类效果,邵晓敏等^[3]通过决策树分类法结合 NDVI 和纹理均值特征提取了荒漠植被信息,总体分类精度由最大似然分类法基于光谱信息的 79.17% 提高到 86.71%,说明纹理分析的方法能够弥补基于光谱分类方法的不足;Peddle 等^[4]研究表明多光谱图像、纹理信息以及 DEM 综合应用可以显著提高分类精度和地物识别能力;张锦水等^[5]对 IKONOS 影像利用 SVM 的分类方法,复合光谱、纹理和结构信息等多源数据信息进行分类,并与最大似然法和单源数据(光谱)SVM 分类结果进行比较,多源 SVM 分类精度最高。也有学者研究了训练样本量对分类精度的影响,阎静等^[6]利用神经网络法提取了湖北省早稻种植面积,并对不同样本点训练出的结果进行了比较分析;朱秀芳等^[7]使用不同量的训练样本,分别对光谱角制图、平行六面体、马氏距离、最小距离、最大似然和支持向量机 6 种方法进行了实验,探讨了训练样本对农作物种植面积测量精度的影响。但是目前尚无

收稿日期:2011-12-18;修订日期:2012-06-08

基金项目:资源三号卫星立体测图技术与应用示范项目(2011BAB01B06-06);江苏高校优势学科建设工程资助项目(PAPD,SA1102)。

作者简介:赵 慧(1987—),女,河北泊头人,硕士研究生,主要从事遥感土地利用/覆盖分类研究。Email:charis_zh@yahoo.cn。

研究与分析讨论以上 3 种因素对土地利用/覆盖分类精度的协同影响。

本文旨在探讨分类法、辅助数据及训练样本量与土地利用/覆盖分类精度之间的关系问题,从而找出训练样本量、辅助数据以及分类技术与土地利用/覆盖分类精度之间的对应关系,得出 3 种因素在土地利用/覆盖分类应用中的最佳组合方式,为提高土地利用/覆盖分类的精度提供可借鉴的理论和实验依据。

2 研究区及其数据

2.1 研究区概况

本文以大同矿区为研究区。大同矿区位于山西

省大同市西南,地理坐标 $112^{\circ}49'32''\sim 113^{\circ}09'30''\text{E}$, $39^{\circ}52'\sim 40^{\circ}10'\text{N}$,地跨大同、朔州两市。它是全国八大煤炭生产基地之一,以“煤质好、灰粉低”著称,适用于机车和重工业。

研究区地势西高东低,大部分是高山和起伏较大的丘陵地区。属温带大陆性季风气候,年平均气温 6.5°C 。自然景观属灌丛山地,植被属于草原区,自然植被以针茅属为主的草本植物且自然植被覆盖度低于 30%。

2.2 数据源及预处理

数据源为研究区的 ETM 影像(1 200 行 \times 1 400 列)。影像成像时间为 2000 年 7 月 1 日,正值

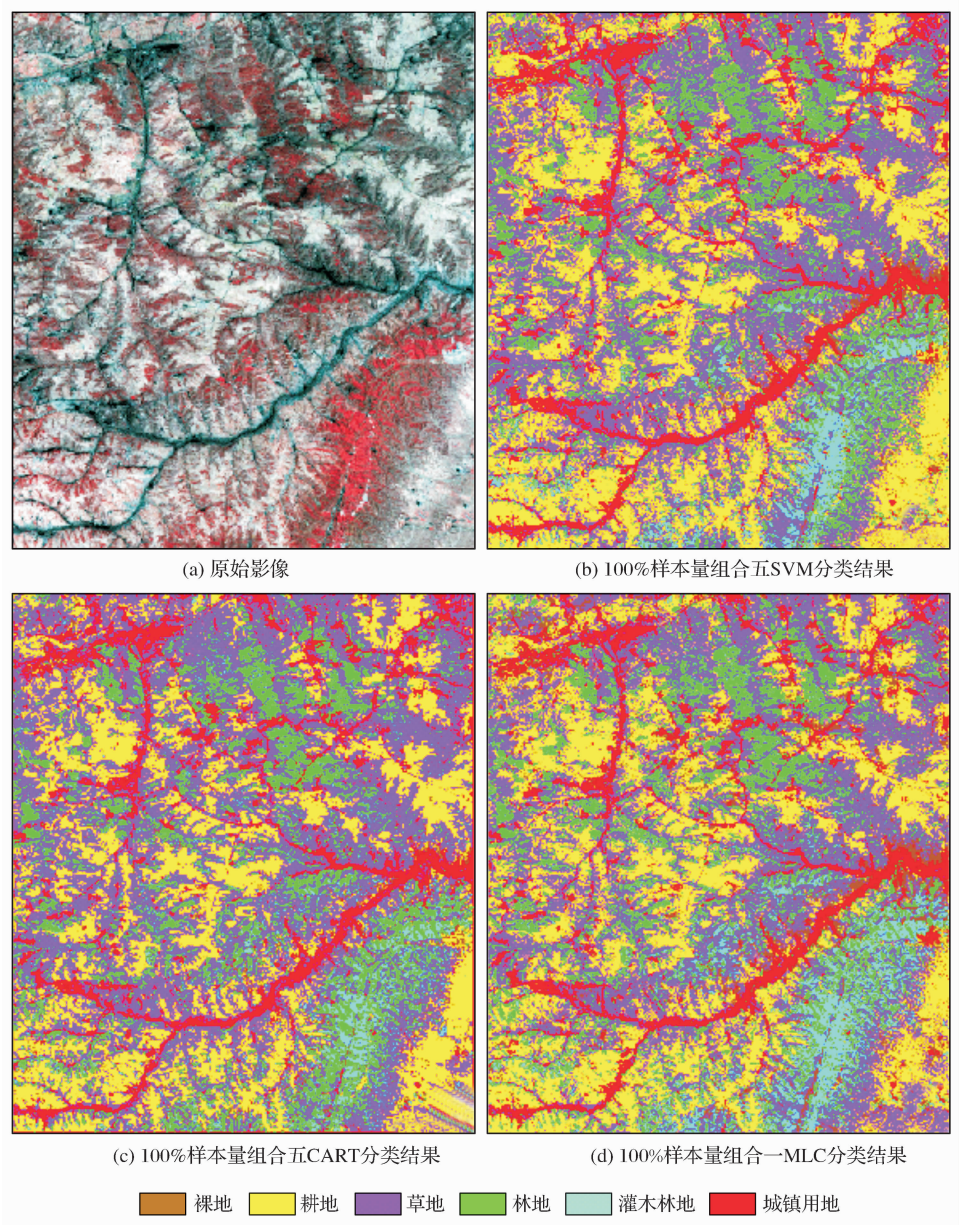


图 1 研究区 ETM 假彩色影像及部分实验结果

Fig. 1 ETM false color image of study area and parts of classification

研究区植被生长季节,影像较清晰,成像效果较好,植被的特征差异可以辨别。首先将影像裁剪得到研究区的多光谱数据,然后进行辐射校正、影像融合(利用 Pan sharpening 方法将多光谱与全色波段进行融合)以及几何精校正等数据预处理。影像融合将多光谱影像由 30 m 分辨率提高到 15 m。

2.3 提取标准样本集

通过人工目视解译结合 Google Earth 高分辨率遥感影像及实地踏勘,将研究区土地利用/覆盖类型分为耕地、林地、灌木林地、裸地、草地和建设用地(包括城镇居民点、农村居民地、道路、工矿建设用地)6 种地物类型。分析各类地物的光谱特征,并进行标准样本集的提取,为保证标准样本集的质量,在提取过程中遵循以下 5 个原则^[1]:

(1) 每个样本区在局部范围内具有代表性,且

- 其面积应大于 8×8 个像元。
- (2) 样本的面应包含在类别较大的同类地物图斑范围内,样本的边界距离图斑的边界至少为 4 个像元。最好不要将整个图斑选为样本,以避免误配准和几何校正时产生的偏差导致提取的样本不纯。
- (3) 当同类地物背景差异太大时,可设为子类,分类后根据实际情况再将其合并为一类。
- (4) 在地物类型复杂多样的区域,样本数据需多选。
- (5) 在样本提取过程中,保证只有漏分,没有错分。
- 按照以上标准共提取 5 095 个样本点,将其以比例 8 : 2 随机分成两部分,为保证分类技术是在充足样本量的情况下进行,将 80% 部分作为训练样本,20% 部分作为测试样本。标准样本集、测试样本集以及不同百分比的训练样本数量见表 1。

表 1 各样本集的样本数量(AOI/Pixels)
Table 1 Sample size (AOI/Pixels)

地物类型	标准样本集	不同百分比的训练样本量				测试样本集
		100%	75%	50%	25%	
裸地	18/137	14/93	10/65	7/50	3/39	4/44
耕地	123/1 550	98/1 236	73/918	49/553	24/287	25/314
草地	105/805	84/619	63/454	42/310	21/150	21/186
灌木林地	75/862	60/609	45/471	30/298	15/188	15/253
建设用地	79/773	63/616	47/492	31/287	15/179	16/157
林地	90/968	72/796	54/612	36/382	18/162	18/172
总计	490/5 095	391/3 969	292/3 012	195/1 880	96/1 005	99/1 126

3 实验与分析

3.1 实验过程与结果

本文以大同矿区为研究区,数据源使用目前应用广泛的 ETM 遥感影像,辅助数据选取常用的 NDVI 和 DEM 数据,由标准样本集分层随机提取以 25% 为增量的不同百分比(25%~100%)的 4 组样本数据,配合不同特征组合的辅助数据(表 2),选

用目前较为常用的遥感影像分类算法:分类与回归树 CART(Classification and Regression Tree)、支持向量机 SVM(Support Vector Machine)和最大似然法 MLC(Maximum Likelihood Classification),分别进行多次试验,具体流程见图 2。在对比分析每次试验结果的精度的基础上,综合探讨了在土地利用/覆盖分类应用中,3 种影响因素的最佳组合方式。

表 2 辅助数据的不同特征组合
Table 2 Different feature combination of auxiliary data

组合	特征组合
组合一	band1, band2, band3, band4, band5
组合二	band1, band2, band3, band4, band5, NDVI
组合三	band1, band2, band3, band4, band5, t_1, t_2, t_3, t_4, t_5
组合四	band1, band2, band3, band4, band5, DEM
组合五	band1, band2, band3, band4, band5, NDVI, t_1, t_2, t_3, t_4, t_5

注:band1、band2、band3、band4、band5 分别表示各个波段信息;DEM、NDVI 分别表示数字高程模型和归一化植被指数; t_1 、 t_2 、 t_3 、 t_4 、 t_5 分别表示相应波段的纹理均值特征

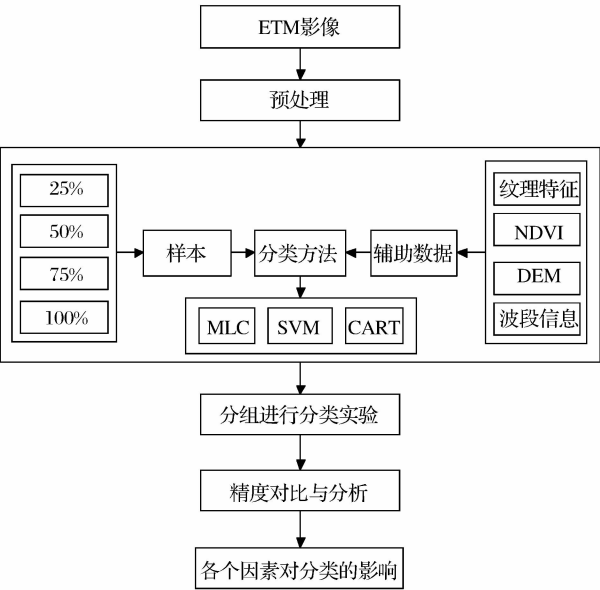


图 2 实验流程

Fig. 2 Flow chart of the experiment

本文旨在探讨影响土地利用/覆盖分类精度的 3 个因素,即训练样本量、辅助数据以及分类法对分类精度的协同影响,进行了多次实验,根据实验目的的

不同将这些实验分为两组:第一组实验旨在探讨训练样本量及分类技术对分类精度的影响;第二组实验旨在探讨辅助数据能否在单纯利用光谱信息的基础上有效改善混分现象从而提高分类精度。最后将两组实验结果进行综合讨论,得到以上 3 个因素对土地利用/覆盖分类精度的协同影响及其最佳的组合方式。

第一组实验旨在探讨训练样本量及分类法对分类精度的影响。由于在相同样本量的情况下,相同分类技术也会因每次随机抽取实际得到的训练样本集和测试样本集不完全相同而产生不同的分类结果,为减少单次分类引起的随机误差,使在相同样本量下的分类结果趋于稳定,因此对标准样本集按比例 8 : 2 进行 5 次分层随机抽取。将得到的 5 个训练样本集作为 100% 训练样本集,75%、50%、25% 训练样本集是从 100% 训练样本集分别按相应百分比随机抽取而来,用于评定每次实验分类精度的测试样本集都选用首次抽取得到的 20% 测试样本集。然后利用 CART、MLC 和 SVM 对得到的 20 个训练样本(表 1)分别进行分类及精度评定,实验结果如表 3 所示。

表 3 组合一不同样本量下 CART、MLC、SVM 的总体分类精度

Table 3 The CART,MLC,SVM total classification accuracy of different sample quantity in group 1

样本量/%	分类技术	实验 1/%	实验 2/%	实验 3/%	实验 4/%	实验 5/%	均值/%
100	CART	85.73	83.30	85.55	81.19	84.87	84.13
	MLC	87.38	82.88	88.40	85.92	88.12	86.54
	SVM	87.69	86.47	88.71	85.73	89.88	87.70
75	CART	86.66	83.51	86.78	81.29	87.19	85.08
	MLC	87.49	85.41	88.71	84.69	88.49	86.96
	SVM	86.45	87.53	88.61	83.36	89.32	87.05
50	CART	84.18	84.78	86.57	80.06	86.72	84.46
	MLC	87.18	83.40	87.69	84.03	85.24	85.51
	SVM	85.01	87.53	86.88	82.51	89.04	86.19
25	CART	83.04	82.77	81.89	79.02	81.52	82.49
	MLC	86.35	78.96	85.35	84.78	86.26	84.64
	SVM	82.94	84.78	85.96	85.26	85.42	85.34

实验 5 中 SVM 在 100% 样本量的分类精度评价误差矩阵如表 4 所示。

表 3 中各实验的总体分类精度范围在 80%~90% 之间,究其原因,主要是由于影像成像效果好,实验区域地物区分明显,且样本是按照 2.3 中的样本提取标准进行选取,较为合理且充分,因此总体误分和错分现象比较少,各分类技术的精度普遍较高,这也说明样本选取标准具有合理性,一定程度上提高了分类精度。总体分类精度的均值以 100% 样本

量下 SVM 精度最高,为 87.70%。在本组 60 次实验中 100% 样本量下 SVM 实验 5 总体分类精度最高,表 4 列出了该实验结果的分类精度评价误差矩阵。分析表 4 发现,本组实验中存在的主要问题是林地与灌木林地、裸地与建设用地和耕地的混分现象比较严重,究其原因,主要是这些地物类型的光谱特征十分相似,导致裸地与建设用地和耕地的混淆;研究区林地与灌木林地交错分布,导致大量的林地被错分为灌木林地。

表 4 实验 5 利用 SVM 在 100%样本量下的分类精度评价误差矩阵

类型		测试样本						行和	生产者精度/%	用户精度/%
		裸地	耕地	草地	灌木林地	建设用地	林地			
训练样本	裸地	12	6	3		4		25	52.17	48.00
	耕地	1	287	12				300	93.18	95.67
	草地	6	13	147			8	174	86.98	84.48
	灌木林地		1		179		16	196	86.89	91.33
	建设用地	4	1	7		142		154	97.26	92.21
	林地				27		201	228	89.33	92.21
	列和	23	308	169	206	146	225		1 077	
总体精度:89.88% Kappa 系数:0.8729										

第二组实验旨在探讨辅助数据能否在单纯利用光谱信息的基础上有效改善混分现象从而提高分类精度。首先将 ETM 多光谱影像(band1~band5)与 DEM、NDVI 及纹理均值特征按照不同的特征进行

组合(表 2),利用 CART、MLC、SVM 分类方法分别输入上组实验得到的 5 组训练样本集进行实验并分别评定其精度,实验结果如表 5 所示。为避免数据冗余,表5仅列出每次实验在5个特征组合下总

表 5 不同特征组合在不同样本量下的总体分类精度

Table 5 The total classification accuracy of different feature combination and sample quantity

样本量/%	分类技术	组合一/%	组合二/%	组合三/%	组合四/%	组合五/%
100	CART	84.13	86.09	86.88	84.00	87.48
	MLC	86.54	86.43	87.34	86.71	87.03
	SVM	87.70	87.96	88.67	87.56	89.59
75	CART	85.08	86.70	87.68	83.85	87.90
	MLC	86.96	86.01	87.42	85.96	87.55
	SVM	87.06	86.94	88.49	86.49	88.33
50	CART	84.46	85.84	87.89	82.61	87.48
	MLC	85.51	85.71	87.24	85.12	87.27
	SVM	86.19	88.00	88.96	86.16	88.90
25	CART	82.44	84.02	82.60	81.80	83.90
	MLC	84.56	84.51	83.34	82.31	84.48
	SVM	85.26	84.83	86.14	83.29	86.14

体分类精度的平均值。

3.2 实验结果与分析

3.2.1 训练样本量及分类技术对分类精度的影响

图 3 是根据 3 种分类方法在各个样本量下 5 次实验的分类结果作出的总体分类精度随样本量的变化曲线,图 4 是 Kappa 系数随样本量的变化曲线。

分析总体分类精度和 Kappa 系数变化的均值曲线图(图 3(f)和图 4(f)),SVM 总体表现优于 MLC 和 CART。造成 CART 分类精度最低的原因是多方面的:其一,由于标准样本集中灌木林地与林地、裸地与建设用地和耕地的光谱特征十分相似,像元的分离度不高;其二,训练样本量充分的条件下,CART 生成的决策树结构庞大,出现了“过度拟合”

现象^[8],如在实验 5 中统计得到:75%样本量下得到的树结构共 106 个叶节点,100%样本量下共 166 个叶节点,导致样本量由 75%增加到 100%时总体分类精度呈下降趋势,分类效率降低。随着样本量的增加,SVM 分类精度和 Kappa 系数逐渐增长,最终趋于稳定,而 MLC 和 CART 呈先上升后下降的趋势,MLC 和 CART 分别在 75%和 50%样本量下达到峰值。样本量由 25%增加到 50%,CART 增长最快,SVM 增长缓慢。这说明 SVM 较 CART 和 MLC 对于样本量的变化具有良好的稳定性。SVM 的理论基础是基于小样本情况下机器学习规律的统计学习理论,以结构风险最小化为准则,在算法的收敛性、分类精度方面都具有较高的性能^[10],实验结

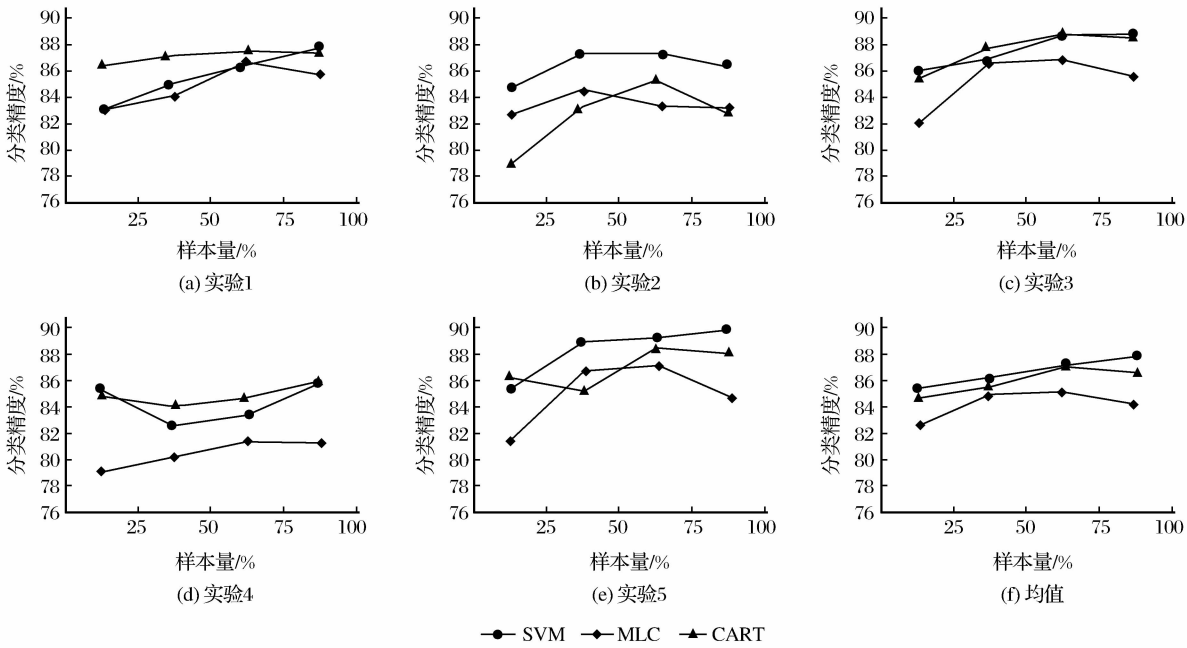


图 3 总体分类精度随样本量的变化曲线
Fig. 3 Curves of overall accuracy with sample sizes

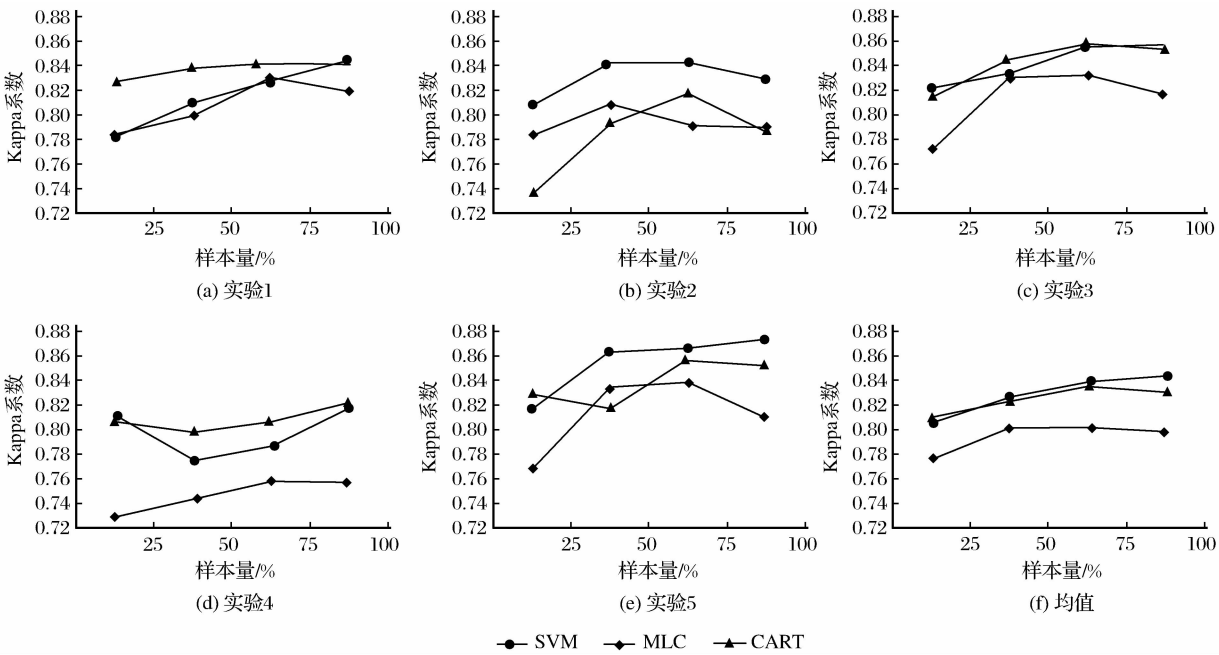


图 4 Kappa 系数随样本量的变化曲线
Fig. 4 Curves of Kappa coefficient with sample sizes

果与该理论一致。

在相同样本量下,5 次实验总体分类精度波动程度也是不一样的,在大样本量(100%)和小样本量(25%)下 SVM 的总体分类精度波动最小,在中等样本量(75%、50%)下 MLC 总体分类精度波动最小。这说明在样本量充足的情况下,利用 SVM 进行土地利用/覆盖分类可以得到较高的分类精度。

对于呈正态分布的训练样本,MLC 易于建立判别函数,有较好的统计特性^[9]。

3.2.2 相同训练样本量下不同辅助数据特征组合对分类结果的影响

根据表 5 及图 5,无论何种样本量下何种分类方法,加入 DEM 数据的分类效果最差,甚至对个别地物类型有反作用,本实验中 DEM 数据未得到有

效利用。其他辅助数据的加入对不同地物类型的提取效果都有不同程度的改善:NDVI 数据的加入,对于改善裸地与耕地、林地与灌木林地的混分现象有明显帮助,由于影像的成像时间正值研究区植被生长期,NDVI 是用来反映绿色植物生长状况和分布

应用最广的特征指数;纹理均值特征数据的加入对各样本量下的分类精度均有明显提高,改善了裸地与建设用地的混分现象;NDVI 和纹理均值特征的共同加入使 SVM 在 100%样本量下的分类精度达到最高。综上所述,在本实验中NDVI和纹理均值

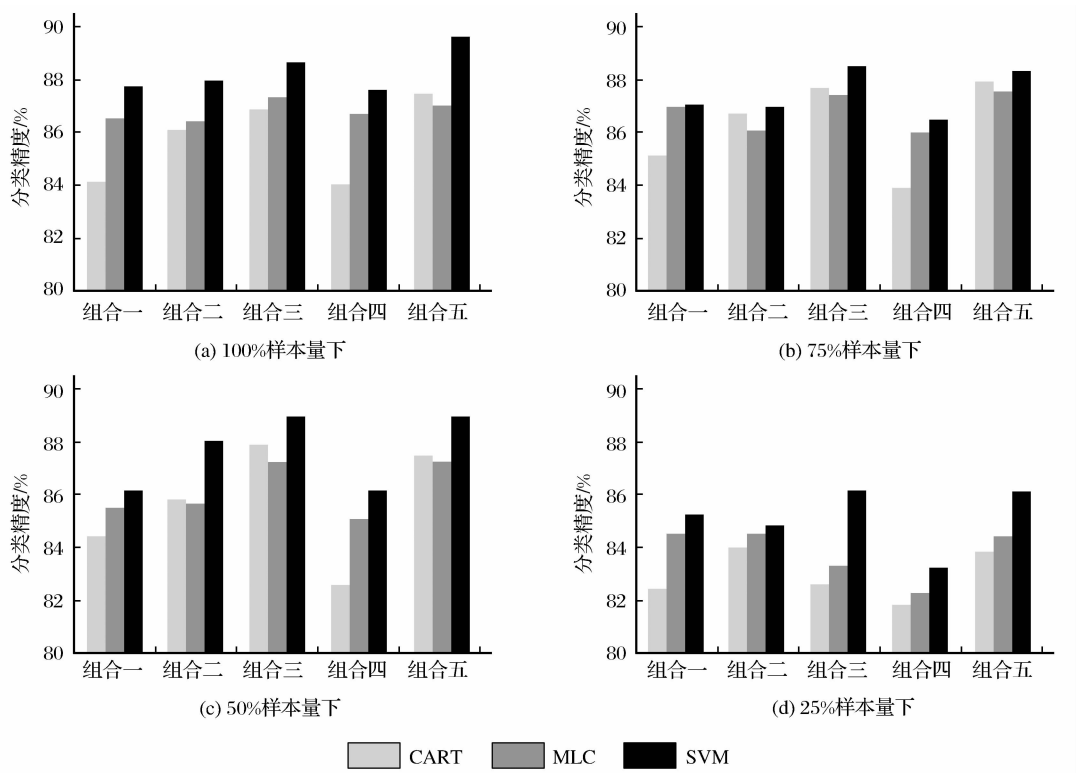


图 5 不同样本量下总体分类精度与特征组合关系柱状图

Fig. 5 The histograms of the total classification accuracy under different groups

特征数据是有效的辅助数据。

辅助数据对不同分类方法的影响:辅助数据的加入对 3 种分类方法 CART、MLC、SVM 都有不同程度的改善,其中以 CART 效果最为明显,SVM 次之。在 100% 样本量下:NDVI 数据的加入使 CART 的总体分类精度由组合一的 84.13% 提高到 86.09%,使 SVM 提高了 0.2%;加入纹理均值特征数据使 CART 总体分类精度提高了 2.7%,而 SVM 由 87.70% 提高到 88.67%,即提高了 0.97%。当样本量大于 25% 时,在特征组合二、三、五中 CART 的表现优于 MLC。这说明 MLC 处理多维复杂数据集时效果较差,而 CART 对多维复杂数据特征具有更好的弹性和鲁棒性,因此当源数据各维具有不同的统计分布和尺度,且用于分类的训练样本量适中,用决策树分类法能够获得理想的分类结果^[9]。

3.2.3 多因素共同作用对分类结果的影响

根据实验得到的大量数据,分析多个因素共同作用时,对分类结果的影响。当获取的样本量较大时,同时加入有效的辅助数据能够显著改善分类效果;观察图 5(a)与图 5(b),图 5(b)中组合一到组合五 CART 精度由 85.08% 增加到 87.90%,即加入 NDVI 和纹理均值特征使分类精度提高了 2.82%,而图 5(b)中组合一与图 5(a)中组合一的 CART 相比,分类精度仅提高了 0.85%,说明 CART 分类方法对有效的辅助数据最为敏感,且不适合于训练样本量较大的情况,为避免生成结构庞大的决策树而产生训练过度造成分类效率不高,用于 CART 分类的训练样本量能够反映真实地物的基本分布情况即可,加入合理的辅助数据的 CART 能够获得较高的分类精度;SVM 总体分类精度较高,在相同样本量和没有有效辅助数据的情况下,

SVM 可以获得最佳的分类结果。因此在进行分类时,样本量的增加和有效辅助数据的加入,能够明显提高各种分类方法的分类精度,同时选用 SVM 可以实现较理想的分类效果,为分类过程的进行提供了有效的指导。

4 结 语

本文结合 Google Earth 高分辨率遥感影像和实地踏勘,在提取标准样本集的基础上,构建不同样本量(100%、75%、50%、25%)的训练样本集,分别利用 CART、MLC 和 SVM 3 种分类方法,结合 DEM、NDVI 以及纹理均值特征辅助数据对研究区 ETM 影像进行了土地利用/覆盖分类,并对其精度进行了评定与分析,得出 3 种影响因素(训练样本量、分类技术、辅助数据)的最佳组合方式:

(1) 在不同样本量及不同特征组合下,SVM 具有最高的分类精度及 Kappa 系数。随着样本量的增加,SVM 分类精度及 Kappa 系数不断增长,并趋于平稳。在能够获取足够多的训练样本集时,应优先考虑利用 SVM 进行土地利用/覆盖分类。

(2) 加入合理的辅助数据是提高分类精度的有效手段,与单纯利用光谱特征数据进行分类相比,同时加入 NDVI 和纹理均值特征的 CART 分类精度提高幅度最大,对于 CART 辅助数据是比提高样本量更有效的手段。在获取的训练样本集有限而可获取有效的辅助数据时,应优先考虑利用 CART 进行土地利用/覆盖分类。

(3) 由于研究区 ETM 影像地形破碎,地物光谱特征混杂,仅仅加入 NDVI 和纹理均值特征并不能使地物类型界限分明,还会受到土壤背景、空间关联、形状等因素的影响,将更有效的辅助数据应用于土地利用/覆盖分类有待进一步研究。

参考文献(References):

[1] Muchoney D, Borak J, Borak H C, *et al.* Application of the MODIS Global Supervised Classification to Vegetation and Land Cover Mapping of Central America[J]. International Journal of Remote Sensing, 2000, (21): 1115-1138.

[2] Hui Wenhua. TM Image Classification based on Support Vec-

tor Machine[J]. Journal of Earth Sciences and Environment, 2006, 28(2): 93-95. [惠文华. 基于支持向量机的遥感图像分类方法[J]. 地球科学与环境学报, 2006, 28(2): 93-95.]

[3] Shao Xiaomin, Liu Yong. Deriving Vegetation Information in Ulan Buh Desert based on Texture[J]. Remote Sensing of Technology and Application, 2010, 25(5): 687-694. [邵晓敏, 刘勇. 基于纹理的乌兰布和沙漠地区植被信息提取[J]. 遥感技术与应用, 2010, 25(5): 687-694.]

[4] Peddle D R, Franklin S E. Image Texture Processing and Data Intergration for Surface Pattern Discrimination[J]. Photogrammetric Engineering and Remote Sensing, 1991, 57(4): 413-420.

[5] Zhang Jinshui, He Chunyang, Pan Yaozhong, *et al.* The High Spatial Resolution RS Image Classification based on SVM Method with the Muti-Source Data[J]. Journal of Remote Sensing, 2006, 10(1): 49-57. [张锦水, 何春阳, 潘耀忠, 等. 基于 SVM 的多源信息复合的高空间分辨率遥感数据分类研究[J]. 遥感学报, 2006, 10(1): 49-57.]

[6] Yan Jing, Wang Wen, Li Xiangge. Extracting the Rice Planting Area Using an Artificial Neural Network[J]. Journal of Remote Sensing, 2001, 5(3): 227-231. [阎静, 王汶, 李湘阁. 利用神经网络方法提取水稻种植面积—以湖北省双季早稻为例[J]. 遥感学报, 2001, 5(3): 227-231.]

[7] Zhu Xiufang, Pan Yaozhong, Zhang Jinshui, *et al.* The Effects of Training Sample on the Wheat Planting Area Measure Accuracy in TM Scale(iv): The Accuracy Response of Different Classifiers to Training Samples[J]. Journal of Remote Sensing, 2007, 11(6): 826-837. [朱秀芳, 潘耀忠, 张锦水, 等. 训练样本对 TM 尺度小麦种植面积测量精度影响研究(iv)-训练样本与分类方法间分类精度响应关系研究[J]. 遥感学报, 2007, 11(6): 826-837.]

[8] Zhao Ping, Fu Yunfei, Zheng Liugen, *et al.* CART-based Land Use/cover Classification of Remote Sensing Images[J]. Journal of Remote Sensing, 2005, 9(6): 708-716. [赵萍, 傅云飞, 郑刘根, 等. 基于分类回归树分析的遥感影像土地利用/覆盖分类研究[J]. 遥感学报, 2005, 9(6): 708-716.]

[9] Li Shuang, Zhang Erxun. The Decision Tree Classification and Its Application in Land Cover[J]. Area Research and Development, 2003, 22(1): 17-21. [李爽, 张二勋. 基于决策树的遥感影像分类方法研究[J]. 地域研究与开发, 2003, 22(1): 17-21.]

[10] Liu Jianghua, Cheng Junshi, Chen Jiapin. Support Vector Machine Training Algorithm: A Review[J]. Information and Control, 2002, 31(1): 45-50. [刘江华, 程君实, 陈佳品. 支持向量机训练算法综述[J]. 信息与控制, 2002, 31(1): 45-50.]

Research on the Factors Affecting the Classification Accuracy
of ETM Remote Sensing Image Land Cover/Use

Zhao hui^{1,2}, Wang Yunjia^{1,2}

(1. *School of Environment Science and Spatial Information, China University of Mining and Technology, Xuzhou 221116, China;*

2. *Key Laboratory for Land Environment and Disaster Monitoring of SBSM, China University of Mining and Technology, Xuzhou 221116, China*)

Abstract: Training sample size, auxiliary data and classification algorithm are the three main factors influencing on the land use and cover classification accuracy, through finding out the best combination of the three factors to improve the classification accuracy. Multispectral imagery and DEM from ETM satellite are applied to perform experiments under four sample sizes (100%, 75%, 50%, 25%), along with calculated NDVI layer and mean texture measure. By contrasting the classification accuracy produced using three different classification algorithms: Classification and Regression Tree(CART), Support Vector Machine(SVM) and Maximum Likelihood Classification(MLC), to discuss the influence of training sample, auxiliary data and different classification algorithms to the classification accuracy. The results indicate that the SVM can obtain the best classification result when there is no effective auxiliary data, and the total accuracy can reach to 85%, when CART is used to classification, as the additional NDVI and texture feature, the accuracy increased 2.82%, it is sensitive to auxiliary data, which can obviously improve the classification accuracy. While in the condition of limited training sample obtain the effective auxiliary data, the CART can achieve the highest classification accuracy.

Key words: CART; SVM; Training sample; Land use/cover classification