

基于波段选择的 MODIS 全国土地覆盖分类

赵德刚^{1,2}, 占玉林³, 刘翔⁴, 刘成林², 庄大方⁵

(1. 连云港市规划市政设计研究院, 连云港 222001; 2. 南昌大学建筑工程学院, 南昌 330031;

3. 中国科学院遥感应用研究所遥感科学国家重点实验室, 北京 100101;

4. 北京东方泰坦科技股份有限公司, 北京 100083; 5. 中国科学院资源环境科学数据中心, 北京 100101)

摘要: 以 MODIS 多光谱和多时相数据为输入参数进行了全国土地覆盖分类研究。从试验区 2007 年 MODIS 8 d 数据的合成影像(MOD 09)中提取 EVI、NDWI 和 NDSI 3 个指数,并将其作为特征波段与原有的 7 波段(B1~B7)形成 10 波段影像。以统计分类 J-M 距离平均值和 SVM 分类总精度为标准评价不同波段对土地覆盖分类的贡献。在全国范围内,选择贡献最大的 EVI、B7 和 B4 这 3 个波段的月合成值,并分别对其作 PCA 变换,选取各 PCA 变换后的前 3 个波段进行分类运算。研究结果表明,在没有其他辅助信息的境况下,基于 MODIS 贡献最大的前 3 个波段结合多时相信息能够在中分辨率区域土地覆盖分类中取得较好的分类结果,其精度为 78.04%。

关键词: MODIS; J-M 距离; 土地覆盖分类; SVM

中图分类号: TP 79 **文献标识码:** A **文章编号:** 1001-070X(2010)03-0108-06

0 引言

土地覆盖及其变化是全球环境变化过程中的重要因子,也是全球变化和碳循环模拟、气候模拟等研究的重要内容^[1]。中国的土地覆盖变化是全球变化研究的一个重要组成部分。

当前已有的全国大尺度土地覆盖产品的分辨率仅局限于 1 000 m,研究较多的是小区域的高分辨率土地覆盖分类,而中等分辨率的过渡产品还相当缺乏。中等分辨率成像光谱仪(MODIS)为大区域中等分辨率土地覆盖自动分类提供了机遇^[2],在很多区域尺度的土地覆盖研究中得到了应用。如恒河和印度河盆地的灌溉区覆盖图的绘制^[3],北非的地貌图^[4],大黄石生态系统的区域土地覆盖制图^[5]和我 国黄河盆地的土地覆盖分类^[6]等。

MODIS 的多光谱和多时相构成了多维光谱数据。然而,光谱和时相的增加并不能使土地覆盖分类的有效信息量成比例地增加^[7],这还取决于数据之间的相互独立性,往往最佳组合的低维特征空间就能区分所有土地覆盖类型^[8]。而且在相同训练样本前提下,过量维数的输入数据会出现休斯现象并降低分类运算速度。对于中国的大区域土地覆盖分类,提取最佳光谱信息组合,减低输入特征维数具有很大意义。

本文先以植被覆盖度较高的山东省为研究区,基于 J-M 距离平均值和土地覆盖分类总精度,对单一时相的 10 波段影像对土地覆盖分类的贡献进行研究,然后在全国范围内,对光谱贡献最佳的月合成波段进行 PCA 变换,进行全国土地覆盖分类研究。

1 研究方法

1.1 地物光谱可分性准则

地物光谱可分性的理论约束是贝叶斯误差,但这在实践中很难实现^[7]。实际应用常采用如基于概率统计的距离(如贝叶斯距离)、基于熵值的距离(如 K-L 距离)、基于相关性或几何性的距离(如 Fisher 比值)等。很多统计分离距离(如巴氏距离、J-M 距离、变形散度或马氏距离)都有在两类别分类问题中作为选择适当特征子空间标准的报道^[9]。怎样结合这些两类别判别距离获得多类别分类的整体分离指标还不是很明了,最直接的方法就是计算所有类对间距离的平均值。

本文选用 J-M(Jeffries-Matusita)距离作为准则来度量各类别之间的可分性,它不需要假定地物的正态分布,具有较好的通用性。J-M 距离的计算公式为:

收稿日期: 2009-11-19; 修订日期: 2009-12-28

基金项目: 国家高技术研究发展计划(编号: 2006AA12Z142、2006AA12010106)、对地观测技术国家测绘局重点实验室经费资助项目(编号: 200806)、中国科学院知识创新工程青年人才领域前沿项目专项项目、中国科学院遥感应用研究所遥感科学国家重点实验室和投资项目遥感动态监测与管理信息系统(国家 863 计划,编号: 2006AA120107)资助项目共同资助。

$$J_{ij} = \sqrt{2[1 - \exp(-B_{ij})]} \quad (1)$$

式中, B_{ij} 为巴氏 (Bhattacharyya) 距离,即

$$B_{ij} = \frac{1}{8} (M_i - M_j)^T \left[\frac{V_i + V_j}{2} \right]^{-1} (M_i - M_j) + \frac{1}{2} \ln \frac{|(V_i + V_j)/2|}{\sqrt{|V_i||V_j|}} \quad (2)$$

式中, M_i 和 M_j 分别是类别 i 和类别 j 的样本平均向量; V_i 和 V_j 是相应的的矩阵样本协方差。

J-M 距离具有收敛性,其判别标准如下:当 $0.0 < J_{ij} \leq 1.0$ 时,两类别之间不具备光谱可分性;当 $1.0 < J_{ij} \leq 1.8$ 时,两类别之间具有一定的光谱可分性,但存在较大程度的重叠;当 $1.8 < J_{ij} < 2.0$ 时,两类别之间具有很好的光谱可分性。

1.2 土地覆盖分类方法

为了检验光谱特征数据对土地覆盖类型分离的能力,使用整体分类精度作为指标进行评价。为了避免分类输入特征数据空间的改变影响分类器的表现,必须选择合适的分类器。支持向量机 (Support Vector Machines, SVM) 从结构上是基于核的监督分类器,它的表现不受特征维数影响,因此选择 SVM 作为本文的分类方法。

SVM^[10,11]是由 AT&T Bell 实验室的 Vapnik 等人提出的一种基于统计学习理论的机器学习技术。它是在线性分类器的基础上,通过引入结构风险最小化原理、最优化理论和核方法演化而成的。它只由少数的支持向量所确定,计算的复杂性取决于支

持向量的数目,而不是样本空间的维数,这在某种意义上避免了“维数灾难”,避免了“过学习”。

2 数据处理

2.1 MODIS 数据

选用 2007 年全国的 MODIS 8 d 数据合成的 MOD 09A1 和 MOD 09Q1 两种产品作为基础数据。首先对原始图像进行重采样,分辨率统一到 250 m。另外我们从 B1 ~ B7 中计算得到了地表植被生物特性的增强型植被指数 (EVI),即

$$EVI = \frac{2.5(B2 - B1)}{B2 + 6.0B1 + 7.5B3 + 1.0} \quad (3)$$

水体指数 (NDWI),即

$$NDWI = (B4 - B2)/(B4 + B2) \quad (4)$$

土壤亮度指数 (NDSI),即

$$NDSI = (B1 - B4)/(B1 + B4) \quad (5)$$

式中, $B1$ 、 $B2$ 、 $B3$ 、 $B4$ 分别为 MODIS 的 1、2、3、4 波段反射率。

把这些提取的特征指数作为附加特征波段加入到原始波段中。

2.2 土地覆盖类别

本文采用的土地覆盖分类体系包括 6 大陆地生态系统 22 个子类别。该分类系统是针对于我国土地覆盖实际情况,基于国际 IGBP 土地覆盖分类体系和我国现有分类体系建立的新的土地覆盖遥感分类系统^[12]。各类别定义如表 1 所示。

表 1 各类别定义及样点数

Tab. 1 Description of the land cover classes and respective number of collected samples

生态系统类型	土地覆盖类型	编码	含义
森林生态系统	常绿针叶林	1	指郁闭度 > 30%, 高度 > 2 m 的常绿针叶天然林和人工林
	常绿阔叶林	2	指郁闭度 > 30%, 高度 > 2 m 的常绿阔叶天然林和人工林
	落叶针叶林	3	指郁闭度 > 30%, 高度 > 2 m 的落叶针叶天然林和人工林
	落叶阔叶林	4	指郁闭度 > 30%, 高度 > 2 m 的落叶阔叶天然林和人工林
	混交林	5	指郁闭度 > 30%, 高度 > 2 m 的针阔混交天然林和人工林
	密闭灌丛、矮林	6	指郁闭度 > 60%, 高度 ≤ 2 m 的灌丛和矮林
	稀疏灌丛、矮林	7	指郁闭度 ≤ 60%, 高度 ≤ 2 m 的灌丛和矮林
草地生态系统	草甸草原	8	指覆盖度在 40% ~ 75%, 以草本植物为主的各类草地
	典型草原	9	指覆盖度在 20% ~ 40%, 以旱生草本为主的草地
	荒漠草原	10	指覆盖度在 5% ~ 20%, 以强旱生植物为主的草地
	高寒草原	11	指覆盖度在 20% ~ 40%, 以高海拔寒旱生植物为主的草地
农田生态系统	一年一熟	12	一年一熟粮作和耐寒经济作物
	一年两熟	13	一年两熟或两年三熟旱作
	一年水旱	14	一年水旱、单季稻连作或两熟粮作
	双季稻	15	双季稻连作及一年三熟作物
城市生态系统	建筑用地	16	包括城镇居民点、工矿、建设用地
湿地、水体生态系统	沼泽湿地	17	植被覆盖度高的湿生草地以及地势平坦低洼、排水不畅、长期潮湿积水且表层生长湿生草本植被的土地
	近海湿地	18	各种近海及海岸的浅海水域、海滩、珊瑚礁、红树林沼泽、河口水域、三角洲湿地等
	水体	19	陆地上各种淡水湖、咸水湖、水库以及河流
	冰雪	20	冰川、雪被
荒漠生态系统	荒漠、裸地	21	植被覆盖度在 5% 以下的荒漠及戈壁、裸露石山等无植被地段
	沙漠	22	沙地、流动沙丘

2.3 采样方法

本研究使用的采样数据除了江苏境内的部分实地采样点,在全国其他区域内缺乏足够的地面实际样本。为了保证样本的代表性,以中国科学院地理科学与资源研究所制作的由 2000 年 TM 影像数据解译生成的全国 30 m 土地利用分类图,并结合遥感影像目视解译选出各类有代表性的样点,按 7:3 的比例把总样本分成训练样本和验证样本。

3 实验结果与分析

3.1 MODIS 光谱对土地覆盖的贡献

选择山东省作为试验区主要是考虑该地区的植被覆盖度较高,具有很好的代表性。选择较易区分土地覆盖类别的 4 月份影像,把 EVI、NDWI 和 NDSI

这 3 个特征指数作为附加光谱波段加入原有的 7 波段中,构成 10 个光谱波段。EVI 综合了近红外波段 B1、红波段 B2 和蓝波段 B3 共 3 个波段信息,克服了土壤背景的影响和 NDVI 的植被高覆盖区易饱和、植被低覆盖区受土壤背景影响较大、对大气衰减的去除不彻底等缺陷^[13]。因此选择 EVI 作为区分土地覆盖类型的最佳波段,然后逐一加入其他波段。每次加入的波段以 J-M 距离的平均值为标准,在相同维数的不同组合中,某一加入波段的 J-M 距离平均值最大的波段被认为是土地覆盖分类贡献较大的波段,由此排列出其他 10 个波段对土地覆盖分类的贡献大小。使用 SVM 对不同维数的最佳组合特征空间进行分类,得出不同维特征空间的分类精度。波段贡献大小排列顺序和分类总精度如表 2 所示。

表 2 不同特征空间的 J-M 平均值和分类总精度

Tab.2 J-M mean distances and overall classification accuracy of different feature spaces

维数	1	2	3	4	5	6	7	8	9	10
	EVI	B7	B4	B6	B3	NDWI	B2	NDSI	B1	B5
J-M 平均值	—	1.627	1.778	1.872	1.921	1.941	1.966	1.974	1.986	1.988
分类总精度/%	42.12	73.94	82.27	87.39	88.94	89.92	91.21	91.82	92.27	92.54

实验中没有计算 EVI 一维空间的 J-M 距离,仅对其作了分类精度的评估。从表 2 中可以看出,只有 EVI 的一维特征空间的分类总精度与其他相比很低,只有 42.12%。一方面,这可能是由于分类器的问题,因为 SVM 在把原始一维输入空间的复杂决策边界向更高维特征空间的线性边界的转换能力上还有所欠缺^[14];另一方面,只有 EVI 一个特征指数的情况下,试验所选择的时相其 EVI 对分类贡献

并不是最好的。

从表中的光谱排列顺序可以看出:进入第 2 和第 3 位的是短波红外波段 B7 和绿波段 B4,这两个波段处于光谱的两端,相关性很弱,而且也都不包含在 EVI 中。随着光谱波段的加入,各类别之间的可分性不断增加,最后趋于平稳,分类总精度也具有相同的变化趋势。对比 J-M 平均值和分类总精度两个指标随维数的变化图(图 1)更直观。

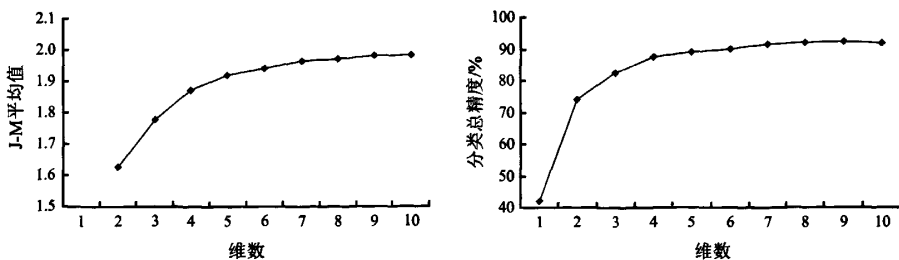


图 1 J-M 平均距离(左)和分类总精度(右)随特征维数变化曲线

Fig.1 J-M mean distances and overall classification accuracy progression when increasing the number of input feature spaces

从图 1(左)中可以得出,特征空间增加到 3 维后,J-M 距离的平均值达到 1.8 以上,各类别间达到了较好的可分性。以后数据维数的加入对 J-M 距离平均值的贡献越来越小,这是由于后面的光谱

对类别分离的贡献不断减小和 J-M 距离的收敛性所导致的。图 1(右)中的分类精度在前 3 个波段已经达到了 80% 以上,前 4 波段达到 87.39%,比全部光谱波段参加分类的结果(92.54%)相差了

5.15%。因此可以选择 EVI、B7 和 B4 这 3 个波段为全国土地覆盖的输入波段。

3.2 基于波段选择的全国土地覆盖分类

从 MODIS 8d 数据合成的 MOD 09A1 和 MOD 09Q1 两种产品中提取 EVI、B7 和 B4 这 3 个波段的月合成数据。EVI 采用最大值合成法,B7 和 B4 选择每个月中云最少的一期。分别对 EVI、B7 和 B4 的月合成数据进行主成分变换,把有用的信息压缩到几个主成分中,这样很少的波段就可以替代变换之前的多波段数据而不损失过多的信息,大大减少了数据处理的时间。另外,PCA 变换还可以去除部分低质量数据的噪声^[15]。经过 PCA 变换的前 3 个波段占各波段信息量如表 3 所示。

选择 3 个波段的前 3 个 PCA 主成分共 9 个波段进行全国的土地覆盖分类。

表 3 不同输入参数的前 3 个主成分信息比
Tab.3 The first three principal components of different parameters

主成分	信息比/%
EVI	82.90
	9.80
	3.55
B7	81.50
	9.59
	4.22
B4	86.30
	7.53
	2.36

3.2.1 土地覆盖分类结果

基于 MODIS 对土地覆盖贡献最佳的前 3 波段结合多时相信息作为分类输入的全国土地覆盖分类图如图 2 所示。

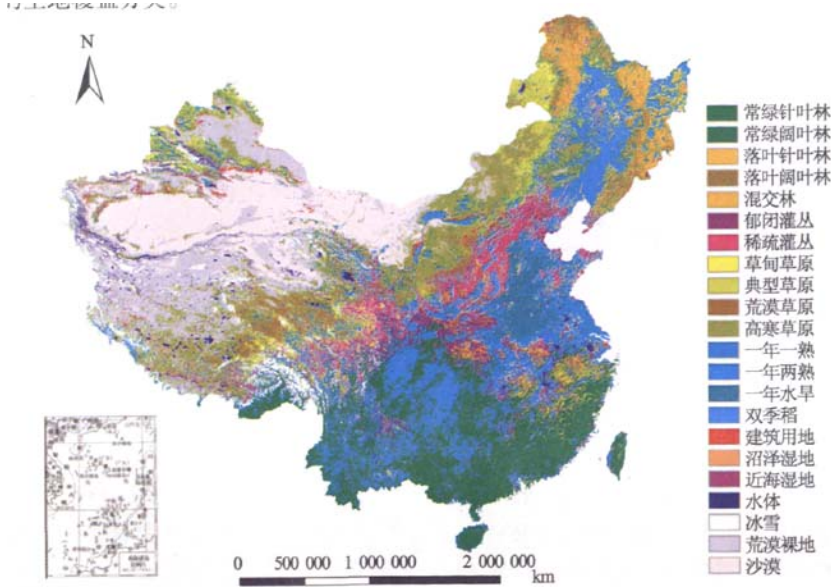


图 2 全国土地覆盖分类示意图
Fig.2 Land cover classification schematic map of China

图 2 描述了我国土地覆盖与植被覆盖类型的基本状况。全国大致以大兴安岭—吕梁山—青藏高原东南缘一线为分界线,以东的土地覆盖类型的植被覆盖度较高,以西的覆盖类型的植被覆盖度较低。

从分类结果图上统计得到我国森林覆盖面积为 $2.526 \times 10^8 \text{ hm}^2$, 占全国土地面积的 26.6%, 与我国林地面积^[16] 相差 7.1% 左右。草地分布于温带东部半湿润区向西部半干旱区和干旱区的大面积过渡区域,连续呈带状分布。我国的耕地分布很不均衡,地区间差异较大,其中东部集中了大部分的耕地。荒漠占据了中国的大半个西北,从分类结果图中统计出全国沙漠面积为 $2.22 \times 10^8 \text{ hm}^2$, 占全国土地面

积的 23.38%, 显示了中国大面积荒漠化态势。

3.2.2 土地覆盖分类精度评价

在分类图上随机选取 947 个样点进行精度评价。精度评价总精度为 78.04%, kappa 系数为 0.765。沼泽湿地和浅海湿地等面积相对较少的地物类别的生产精度相对都比较低,其中沼泽湿地的生产精度只有 66.67%。这可能有两个原因,第一是训练样本的选择,第二是由于某一光谱波段对特定类别(特别是所占比例较小的类别)的区分贡献较大,但是对所有类别的整体贡献不大而被舍弃,从而引起对应类别的精度下降。具体的分类结果生产精度和用户精度如表 4 所示。

表 4 分类精度

Tab.4 Accuracy assessment of different
land - cover types

土地覆盖类型	编码	生产精度/%	用户精度/%
常绿针叶林	1	72.41	72.41
常绿阔叶林	2	70.15	65.28
落叶针叶林	3	69.57	66.67
落叶阔叶林	4	74.36	63.04
混交林	5	72.41	67.74
郁闭灌丛	6	76.19	61.54
稀疏灌丛	7	80.95	62.96
草甸草原	8	67.31	87.50
典型草原	9	75.00	100.00
荒漠草原	10	77.05	88.68
高寒草原	11	88.37	79.16
一年一熟	12	73.85	80.00
一年两熟	13	83.33	62.50
一年水旱	14	70.45	63.27
双季稻	15	72.58	68.18
建筑用地	16	75.44	84.31
沼泽湿地	17	66.67	90.91
浅海湿地	18	68.00	85.00
内陆水体	19	84.72	84.72
冰雪	20	100.00	100.00
荒漠裸地	21	100.00	74.07
沙漠	22	100.00	100.00

4 结论

本文首先以山东省为研究区,以 J - M 距离和 SVM 分类总精度为标准,对 MODIS 不同光谱波段对土地覆盖分类的贡献进行研究;然后,在全国范围内,选择贡献最大的前 3 个光谱波段的 2007 年 12 个月的时间序列合成影像,分别作 PCA 变换进行信息增强与压缩处理,对提取的主要信息用 SVM 进行土地覆盖分类,得到了 2007 年全国土地覆盖分类图,其验证精度为 78.04%。

MODIS 的多光谱波段对土地覆盖分类的贡献不同。贡献最大的前 3 个波段 EVI、B7 和 B4 结合多时相信息的土地覆盖分类结果总精度几乎达到了使用所有光谱和时相的分类结果精度。在不使用任何辅助数据的情况下,基于中等分辨率的 MODIS 部分信息就可以获得较高精度的土地覆盖图。

本研究主要依赖于土地覆盖分类总精度,所选择的 3 个光谱波段结合多时相信息作为分类输入参数可以达到总精度的要求,忽略了其他光谱信息。没有使用的光谱波段可能包含了有效区分部分特定类别的信息,如果考虑某一具体类别的精度,这种方法可能不适用。怎样结合多光谱和多时相遥感信息来改善所有类别的分类精度是我们下一步的工作。

参考文献:

- [1] 刘勇洪,牛 铮,王长耀. 基于 MODIS 数据的决策树分类方法研究与应用[J]. 遥感学报,2005,9(4):405-412.
- [2] Hugo Carrão, Paulo Gonçalves, Mário Caetano. Use of Intra - annual Satellite Imagery Time - series for Land Cover Characterization Purpose[J]. EARSeL eProceedings, 2007, 6:1-11.
- [3] Prasad S, Mitchell Schull, Hugh Turrall. Ganges and Indus River Basin Land Use/Land Cover (LULC) and Irrigated Area Mapping Using Continuous Streams of MODIS Data[J]. Remote Sensing of Environment, 2005, 95:317-341.
- [4] Ballantine J A C, Okin G S. Mapping North African Landforms Using Continental Scale Unmixing of MODIS Imagery[J]. Remote Sensing of Environment, 2005, 97:470-483.
- [5] Wessels K J, Fries De R S, Dempewolf J. Mapping Regional Land Cover with MODIS Data for Biological Conservation: Examples from the Greater Yellowstone Ecosystem, USA and Para' State, Brazil [J]. Remote Sensing of Environment, 2004, 92:67-83.
- [6] Matsuoka M, Hayasaka T, Fukushima Y. Land Cover Classification over Yellow River Basin Using Satellite Data[J]. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, 2004(1):231-234.
- [7] Hugo Carrão, Paulo Gonçalves, Mário Caetano. Contribution of Multispectral and Multitemporal Information from MODIS Images to Land Cover Classification[J]. Remote Sensing of Environment, 2008, 112:986-997.
- [8] David Landgrebe. Multispectral Land Sensing: Where from, Where to? [J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(3):414-421.
- [9] Bruzzone L, Roli F, Serpico S B. An Extension to Multiclass Cases of the Jeffries - Matusita Distance[J]. IEEE Transactions on Geoscience and Remote Sensing, 1995(33):1318-1321.
- [10] Cortes C, Vapnik V. Support Vector Networks[J]. Machine Learning, 1995(20):273-297.
- [11] 李 森, 张继贤, 张永红. 支持向量机在 MODIS 影像分类中的方法研究[J]. 测绘与空间地理信息, 2007, 30(3):177-179.
- [12] 刘勇洪, 牛 铮, 徐永明, 等. 基于 MODIS 数据设计的中国土地覆盖分类系统与应用研究[J]. 农业工程学报, 2006, 22(5):99-104.
- [13] 王正兴, 刘 闯, HUETE Alfredo. 植被指数研究进展:从 AVHRR - NDVI 到 MODIS - EVI[J]. 生态学报, 2003, 23(5):979-987.
- [14] Huang C, Davis L S, Townshend J R G. An Assessment of Support Vector Machines for Land Cover Classification[J]. International Journal of Remote Sensing, 2002, 23:725-749.
- [15] Townshend J R, Justice C R, Kalb V. Characterization and Classification of South American Land Cover Types Using Satellite Data [J]. International Journal of Remote Sensing, 1987(8):1189-1207.
- [16] 2007 年中国国土资源公报[EB/OL]. http://www.gov.cn/gzhl/2008-04/17/content_947023.htm

Land Cover Classification in China Based on Chosen Bands of MODIS

ZHAO De - gang^{1,2}, ZHAN Yu - lin³, LIU Xiang⁴, LIU Cheng - lin², ZHUANG Da - fang⁵

(1. Lianyungang Urban Planning & Design Institute Co. , Ltd, Lianyungang 222001, China;

2. College of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, China;

3. LARSIS, Institute of Remote Sensing Applications, CAS, Beijing 100101, China; 4. Beijing Oriental TITAN Technology Co. , Ltd, Beijing 100083, China; 5. Resource and Environmental Science Data Center, CAS, Beijing 100101, China)

Abstract: MODIS data with high spectral and temporal resolutions were used as input parameters for regional land cover classification in China. First, EVI, NDWI and NDSI were calculated as input spectral features on the basis of an annual time series of twelve MODIS 8 - day composite reflectance images (MOD09) acquired during the year of 2007. The three indices were added to the image form a 10 spectral bands image. The authors employed the mean Jeffries - Matusita distance as a statistical separability criterion and classification accuracy of SVM to evaluate the contribution of different bands for land cover classification. Once the aim was achieved, the monthly three largest contribution spectral bands (EVI, B7 and B4) were dealt with. The Principal Component Analysis (PCA) method and its first three principal components were used as input parameters for SVM classification. The result shows that the three largest contribution spectral bands together with temporal information as input parameters can reach certain high classification accuracy (78.04%) at moderate spatial scales without other accessorial data.

Key words: MODIS; J - M distance; Land cover classification; SVM

第一作者简介: 赵德刚(1985 -),男,硕士研究生,主要从事遥感应用方面的研究。

通信作者: 刘成林, E - mail: liuel@ncu.edu.cn.

(责任编辑: 李 瑜)