

基于多分类器集成的“北京一号” 小卫星遥感影像分类研究

单丹丹^{①,②}, 杜培军^{①,②}, 夏俊士^{①,②}

(① 国土环境与灾害监测国家测绘局重点实验室, 徐州 221116;

② 中国矿业大学测绘与空间信息工程研究所, 徐州 221116)

摘要: 将多分类器集合应用于“北京一号”小卫星多光谱遥感数据土地覆盖分类, 首先构建分类器集合, 应用最小距离分类、最大似然分类、支持向量机(SVM)、BP神经网络、RBF神经网络和决策树等进行土地覆盖分类, 然后利用 Bagging、Boosting、投票法、证据理论和模糊积分法等分类器集成方法, 得到综合不同分类器输出的最终分类结果。试验表明, 多分类器集成能够有效提高“北京一号”小卫星土地覆盖分类的精度, 具有广泛的应用前景。

关键词: “北京一号”小卫星; 多分类器集成; 土地覆盖分类

doi: 10.3969/j.issn.1000-3177.2011.02.013

中图分类号: TP79 **文献标识码:** A **文章编号:** 1000-3177(2011)114-0069-10

Classification of Remote Sensing Image of Beijing-1 Micro-satellite Based on Multiple Classifier Combination

SHAN Dan-dan^{①,②}, DU Pei-jun^{①,②}, XIA Jun-shi^{①,②}

(① Key Laboratory for Land Environment and Disaster of State Bureau of Surveying and Mapping (SBSM),
China University of Mining and Technology, Xuzhou 221116;

② Institute of Surveying and Spatial Information Engineering, China University of Mining and Technology,
Xuzhou 221116)

Abstract: Using the multi-spectral remote sensing image of Beijing-1 micro-satellite in Xuzhou city as the data source, it is shown that the multiple classifier combination can improve the classification accuracy effectively. Firstly, the different member classifiers, including Minimum Distance Classifier, Maximum Likelihood Classifier, Support Vector Machine, BP neural network, RBF neural network and decision tree, are used to land use/land cover classification. Then different multiple classifier integration strategies, including bagging, boosting, majority voting, Dempster-Shafer theory and fuzzy integral are used, to combine the results from the different classification method. Experiments show that the combination of multiple classifiers not only has the potential to improve the accuracy, but also has broad application prospects.

Key words: Beijing-1 small satellite; multiple classifier combination; land cover classification

1 引言

遥感应用对影像分类和信息提取精度的要求越来越高, 提高遥感信息处理和解释精度的需求也越来

越迫切。多分类器集成作为模式识别新技术, 是应用于遥感影像分类的一个热门方向, 具有重要的应用前景, 在进一步提高遥感分类精度的同时, 还可以推动模式识别与机器学习理论方法在遥感领域的应用。

收稿日期: 2010-04-07 修订日期: 2010-06-17

基金项目: 国家高技术研究发展计划(863计划)专项课题(2007AA12Z162)、国家自然科学基金(40871195)和北京宇视蓝图信息技术有限公司北京一号小卫星应用开放基金联合资助。

作者简介: 单丹丹(1986~), 研究方向: 资源环境遥感。

E-mail: shanddan005@163.com

近年来,国内外很多学者对遥感领域多分类集成的应用开展了许多研究工作。Briem 等系统研究了用于多源遥感数据的多分类器问题,针对多源遥感数据和地理数据分类的要求,提出了一些适用的单独和多分类器系统,重点将基于 Bagging、Boosting 和统计一致性理论的多分类器应用于多源遥感数据,结果表明多分类器的总体精度、生产者精度和用户精度都优于单一分类器^[1-2]。Brian 等将多分类器组合应用于土地覆盖分类制图,采用乘积规则和简单非参数分类器两种分类器合并方法,表明乘积准则合并分类器后的性能与非约性堆栈回归法相近,精度比约束堆栈回归方法的效果更好^[3]。Tobias 等提出一种混合分类技术,将非监督分类和监督分类方法结合,采用主成分集作为输入并利用非监督算法 ISODATA 进行聚类,然后采用超类进行类别合并,取得了较好的分类效果^[4]。目前一些国外卫星获取的遥感影像包括中等分辨率的 Landsat TM/ETM+、ASTER 和高分分辨率的 SPOT5、QuickBird、IKONOS 等都得到了较多应用^[5],相对而言,国产卫星遥感影像如 CBERS、“北京一号”小卫星、环境与灾害小卫星等的信息处理方法和应用还有待进一步推动。本文旨在针对“北京一号”小卫星 32m 多光谱数据和 4m 全色波段数据配置的特点,进行多分类器集成的应用,首先利用 Bagging 和 Boosting 方法对单个基分类器进行增强,然后采用投票法、证据理论、模糊积分方法等集成方法参与遥感分类,结果表明分类器集成的分类精度优于单分类器,能够有效提高“北京一号”小卫星遥感信息提取的精度和可靠性。

2 研究区和数据

2.1 “北京一号”小卫星数据及其应用

“北京一号”小卫星于 2005 年 10 月 27 日成功发射。轨道高度 686km,预计在轨寿命 5 年以上,是“国际灾害监测星座”(DMC, Disaster Monitoring Constellation)成员,又称“中国 DMC+ 4”卫星^[6]。“北京一号”小卫星影像已应用于国土资源调查、环境保护、灾害监测、“奥运科技行动计划”等。“北京一号”小卫星参数如表 1 所示。

表 1 “北京一号”小卫星波段参数

波段	波长(μm)	分辨率(m)
Pan	0.5~0.8	4
Band 1(绿)	0.52~0.62	32
Band 2(红)	0.63~0.69	32
Band 3(近红外)	0.76~0.90	32

2.2 研究区和数据预处理

研究选择徐州市某一区域进行。范围为一个大小为 220×220 像素的矩形典型区域,既包括徐州中心城区,也包括部分城市边缘区和郊区,既有以城市为主的建筑物,又有耕地、公共绿地、林地和水体,能比较典型地反映徐州市区的情况。图 1 为研究区遥感图像。

本文中使用的数据是“北京一号”2007 年 2 月 2 日的 32m 多光谱遥感影像和 2007 年 9 月 10 日获取的 4m 高空间分辨率的全色影像。



(a)多光谱影像
(R:Band 3,G:Band 2,B:Band 1)



(b)全色影像

图 1 研究区多光谱和全色影像

为了充分利用高分辨率影像中包含的细节信息,选取平均值、变异和对比度等描述纹理特征,与原始全色数据构成高分辨率数据集,用于分类和后续处理。

3 多分类器集成

3.1 基于 Bagging 和 Boosting 算法的多分类器集成

3.1.1 Bagging 算法

Bagging 算法是通过随机有放回的选取训练样本,改变训练样本集合,对同一基分类器的多个分类结果进行组合,得到一个性能改进的组合分类器。其基本思想是:从原始训练数据集中,分别独立随机地采用放回采样的方式,抽取 k 个数据($k <$

= T , T 代表数据集中样本的个数) 形成自助数据集。抽样过程独立进行许多次, 直到产生很多个独立的自助数据集。然后, 每一个自助数据集都被独立地用于训练一个“分量分类器”(component classifier)。对于 Bagging 算法, 组合多个分类器能通过减小误差方差从而减小期望误差值, 越多的分类器参与, 误差方差就越小^[7]。

3.1.2 Boosting 算法

Boosting 算法也称为增强法或推进法, 它的目标是提高任何给定的学习算法的分类准确率。Boosting 方法可以处理带有权值的数据, 该方法通过增加分类错误的样本的权值, 使学习算法将精力集中于特定的实例样本上。Boosting 算法有许多不同的变形, 其中最为流行的就是 AdaBoost 算法^[8]。在 AdaBoost 算法中, 每一个训练样本都被赋予一个权重, 表明它被某个分量分类器选入训练样本的概率。如果某个样本点已经被准确地分类, 那么在构造下一个训练样本时, 它被选中的权重就被降低; 相反, 如果某个样本点没有正确分类, 那么它的权重就得到提高。

3.2 基于分类器组合的多分类集成

当有多个分类器时, 可以采用对硬分类器或软分类器输出进行相应组合处理, 实现分类器的集成。对输出为类别代码的分类器, 可以采用抽象级的多分类器集成方法; 对输出为类别概率的分类器, 可以采用测量级的多分类器集成方法。

3.2.1 投票法/加权投票法

投票法是最常用的抽象级分类器集成方法。抽象级的分类器集成过程就是采用合理的方法解决分类器的分类不一致问题, 最大化利用互补信息, 从而获得正确的遥感分类结果。多数投票规则将多数分类器一致分类的类别判为待分类像元的类别。当多个类别获得的投票数目相同时, 往往随机选择其中一个类别作为最终的结果。此外也可以对各个不同的结果按照分类好坏设置不同的权重, 进行加权投票。

3.2.2 证据理论

证据理论也称为 Dempster-Shafer 理论。与 Bayesian 理论相比, 证据理论将概率论中的单点赋值扩展到集合赋值, 能更容易地处理未知因素引起的不确定性^[9]。Dempster-Shafer 理论通过辨别框架、信任函数、似然函数和概率分配函数进行知识的表达和处理。

多分类器集成中, 每一个分类器的结果可以作为

一个证据。概率分配函数可以采用分类器的相应类别的分类精度, 如, 对每一个像元, 某分类器将其分到 i 类时, 基本概率分配如下: $m(C_i) = P_i$, $m(\Theta) = 1 - P_i$, 其中 P_i 为该分类器的第 i 类别的分类精度。证据合成完成以后, 采用证据值最大的类别作为最后的决策结果。

由于 Dempster-Shafer 证据理论在合成高度冲突的证据时, 合成结果有悖常理^[9], 孙权等提出了一种新的证据理论合成公式, 并用几个假定的证据数据验证了该公式, 结果表明该方法在处理证据高度冲突的情况时优于一般的证据合成方法。由于遥感分类中对于待识别像元, 分类器之间识别结果不同时, 往往会产生高度冲突的证据, 因此将该新的合成理论应用于分类器集成, 称为改进的证据理论方法^[10]。

3.2.3 模糊积分

模糊积分法是一种有效的信息融合方法。相对于证据理论中各个证据要求相互独立, 模糊积分则不要求相互独立, 同时考虑到了分类器的可靠性。Sugeno 积分是一种常用的模糊积分方法^[11-12]。

3.2.4 分类器差异性测量

一个好的多分类器系统不仅取决于组合规则, 而且取决于分类器成员选择的好坏^[13]。在组合策略一定的情况下, 除了分类器各自分类精度影响外, 多分类器组合结果的好坏在一定程度上与各个分类器分类输出之间的差异性程度有关^[14]。通常认为, 组合多个完全一致的分类器(输入输出均完全一样)是不会对性能有任何帮助的, 而差异性较大的分类器进行组合可以充分发挥分类器之间的互补性, 具有提高分类精度的潜力。测量分类器之间的差异性可以帮助选择合适的组合分类器成员, 从而可以提高组合分类器分类精度。

本文采用的差异性测量方法有双错误测量(Double Fault)^[15]、Kappa 统计^[14]、WCEC^[16]。由于这些指标都是一对一的差异性测量, 所以它们是通过分类器两两测量差异性后取平均得到的。

4 实验与分析

4.1 多光谱数据分类的试验

4.1.1 单分类器实验

采用的单分类器有最大似然分类器、SVM 分类器、BP 神经网络分类器、RBF 神经网络分类器、决策树分类, 分类结果图和分类精度表如图 2 和表 2。

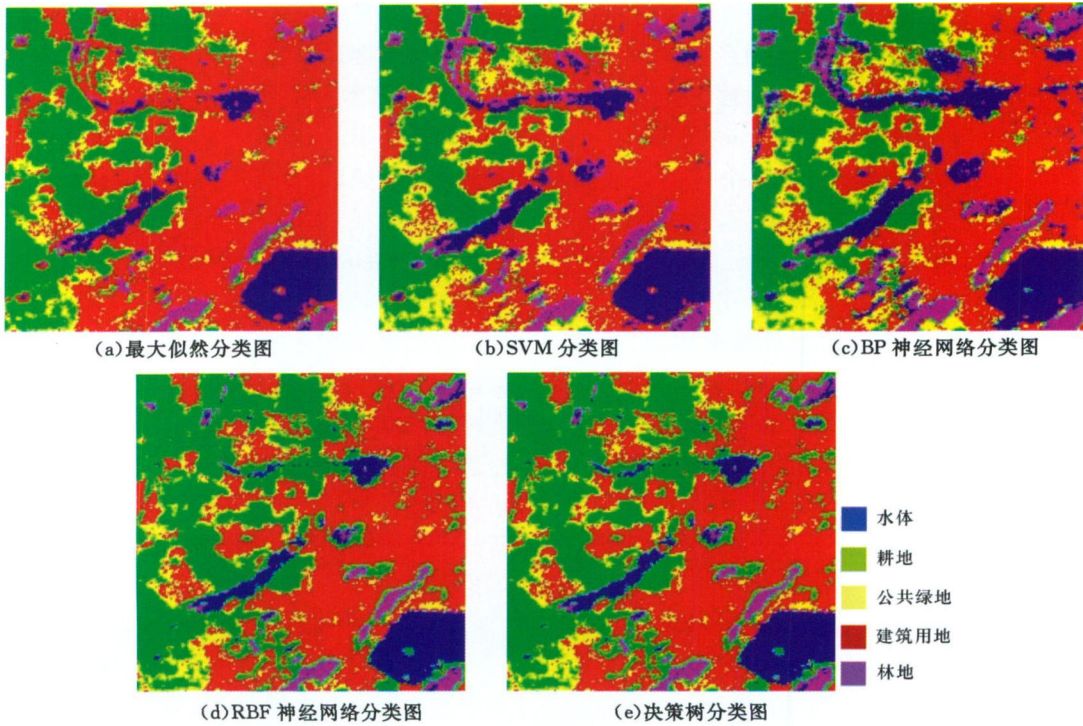


图 2 分类结果图

表 2 多光谱数据单分类器分类精度表

	水体	耕地	建筑用地	林地	公共绿地	总体精度	Kappa 系数
最大似然分类器	76.47	99.05	89.46	96.43	53.62	90.2423	0.8495
SVM 分类器	86.27	99.05	79.92	87.86	76.81	86.1821	0.7964
BP 神经网络分类器	91.18	98.33	81.05	100.00	79.71	88.1467	0.8244
RBF 神经网络分类器	86.27	98.81	87.33	92.14	55.07	89.3910	0.8370
决策树分类器	78.43	97.14	84.57	98.57	69.57	88.2122	0.8230

从各个类别的分类精度可以看出,没有一个分类器可以对所有的类别都取得最高的分类精度。实验中,最大似然分类器在耕地和建筑用地上取得了最高的精度;SVM 分类器在耕地上取得了和最大似然分类器一样的分类精度,但是总体精度不如最大似然分类器;BP 神经网络分类器在水体、林地和公共绿地上取得最高的精度;RBF 神经网络和决策树分类器虽然没有在单个类别上取得较高的分类精度,但是它们

的总体精度要较 BP 神经网络高一些。因此,各个分类器之间的性能是可以互补的,采用合理的方式组合多个单分类器来提高分类精度是有潜力的。

4.1.2 基于 Boosting 和 Bagging 算法的试验

基于 Bagging 的实验中,对 BP 神经网络分类器、RBF 神经网络分类器、决策树分类器分别基于 Bagging 算法进行了多分类器集成的实验, Bagging 分别运用不同采样率进行分类,结果如表 3 所示。

表 3 基于训练样本的多分类集成分类精度表

	BP 神经网络分类器	RBF 神经网络分类器	决策树
基分类器	88.1467	89.3910	88.2122
25% 样本 Bagging	88.6051	88.4741	81.7944
50% 样本 Bagging	88.6228	89.5874	81.7289
75% 样本 Bagging	88.2954	89.7184	80.3536
100% 样本 Bagging	88.8847	90.0458	81.4669
Boosting	90.2423	90.1768	88.2122

比较采用不同基分类器的 Bagging 和 Boosting 算法的实验结果,可以看出:

(1)采用 Bagging 算法和 Boosting 算法, BP 神经网络和 RBF 神经网络分类器都取得了高于基分类器的分类精度。其中基于 BP 神经网络分类器集成的分类精度提高最多,同时取得了所有实验中的最高分类精度。

(2)实验中,采用 Boosting 算法构造的分类器集成的分类精度高于采用 Bagging 算法构造的分类器集成的分类精度。

(3)基于 Bagging 算法,不同的采样样本率可以得到不同的分类精度, BP 神经网络和 RBF 神经网络构建的分类器集成的分类精度提高的程度高于决策树的分类精度。随着采样样本率的提高,精度略有下降或者波动但总体呈上升趋势,原因之一是迭代次数不够多。决策树的分类算法效果欠佳,出现这种情况可能是组合的分类器对数据产生了过拟合。

4.1.3 分类器差异性测量

对于多光谱数据的单分类结果,同一组数据中不同分类器分类采用相同的分类样本和分类特征,分类器间并不是独立的,具有一定相关性。采用 Double Fault、WCEC、Kappa 值分别对每组实验中成员分类器之间的差异性进行测量,综合几种差异性测量结果,从 5 个单分类器中选择 3 个差异性最大的分类器构成多分类器集合。参与实验的分类器集合包含的分类器有:①最大似然分类器,②SVM 分类器,③BP 神经网络分类器,④RBF 神经网络分类器,⑤决策树分类器。根据差异性测量指标的定义,在选择时应该选择 Kappa 系数小、Double Fault

指数小和 WECE 指标大的组合,综合表 4 中各指标值及各分类器分类精度,选择最优分类器集合由最大似然分类器、RBF 神经网络分类器和决策树分类器构成。

表 4 多样性指标计算结果

分类器组合	Kappa	Double Fault	WECE
1, 2, 3	0.658274	0.0829513	0.499454
1, 2, 4	0.671245	0.0803318	0.528051
1, 2, 5	0.631806	0.0796769	0.520410
1, 3, 4	0.671332	0.0757477	0.558393
1, 3, 5	0.628061	0.0746562	0.559921
1, 4, 5	0.673201	0.0757477	0.562104
2, 3, 4	0.700793	0.0892818	0.492032
2, 3, 5	0.709784	0.0932111	0.444008
2, 4, 5	0.656533	0.0842611	0.518227
3, 4, 5	0.669597	0.0807684	0.549225

4.1.4 基于分类器组合的分类器集成

根据差异性测量的结果,选择最大似然分类器、RBF 神经网络分类器和决策树分类器的组合来比较不同的多分类器集成策略。表 5 为多分类器组合的分类精度表,图 3 为组合分类器集成分类的结果图。

表 5 基于多分类器组合的多分类集成分类精度表

分类器	分类精度	Kappa 系数
投票法	90.3078	0.8513
证据理论	90.4388	0.8532
模糊积分	90.2423	0.8499

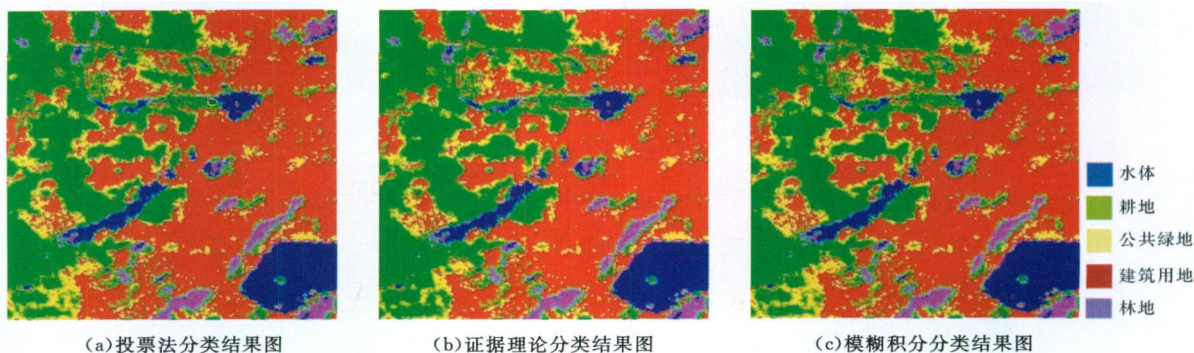


图 3 组合分类器集成分类结果图

可以看出:分类器的集成能有效提高分类器的分类精度,是一种很有潜力的方法,能够得到相当或优于任何单个分类器高的分类精度,其中基于证据

理论的多分类器集成方法要比其他两个的分类精度都高。

4.2 对融合数据的多分类器集成

本文选用 IHS 变换整合, Brovey 融合, Gram-Schmidt 融合和 PCA 融合方法进行比较^[17], 选取融

合效果最好的对徐州市土地利用情况进行分析, 融合结果如图 4 所示。

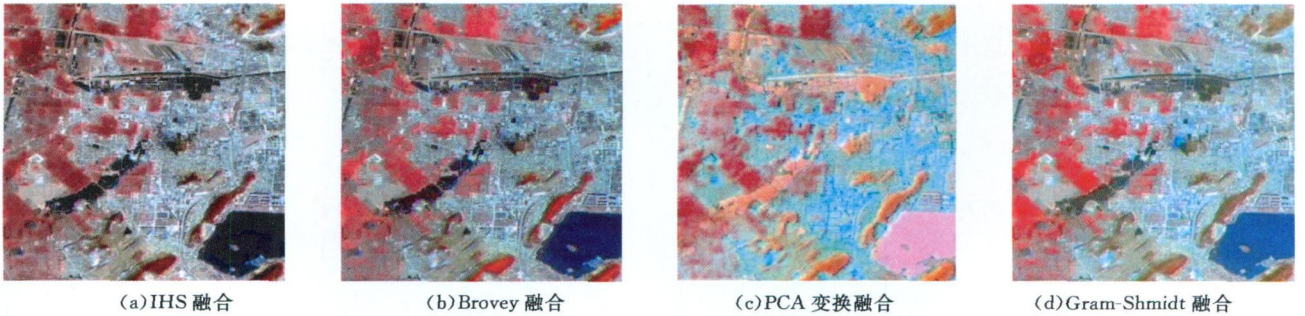


图 4 不同方法融合结果影像

为了客观地评价融合结果, 从主观定性和客观定量两方面进行评价。定量评价融合效果的统计参数较多, 选取常用的信息熵、平均梯度、相关系数及偏差指数这 4 个参数评价各种融合算法在光谱保真度、高频信息融入度和影像清晰度等方面的能力, 如表 6。

表 6 融合效果评价参数

	平均梯度	信息熵	偏差指数	相关系数
原始多光谱影像	0.3104	4.9138		
IHS	10.0988	6.7517	0.4222	0.6246
PCA	1.3892	4.8111	0.7357	0.1033
Brovey	1.9635	5.1423	0.6812	0.5208
Gram-Schmidt	1.6211	4.9523	0.1814	0.1039

综合目视效果和定性评价, 在“北京一号”小卫星影像的融合中, IHS 融合影像不仅清晰度较好, 信息熵最大, 图像信息量最丰富, 相关系数较大, 偏差指数较小。而且 IHS 融合能够较好地提高多光谱影像的空间分辨率, 同时很好地保持多光谱影像的光谱信息, 光谱畸变较小, 在本实验区域具有较好的融合效果。

4.2.1 融合后影像的单分类器实验

为了便于对比分析, 融合后的影像也选择最大似然分类器、SVM 分类器、BP 神经网络分类器、RBF 神经网络分类器和决策树分类 5 个单分类器, 其结果如图 5 和表 7。

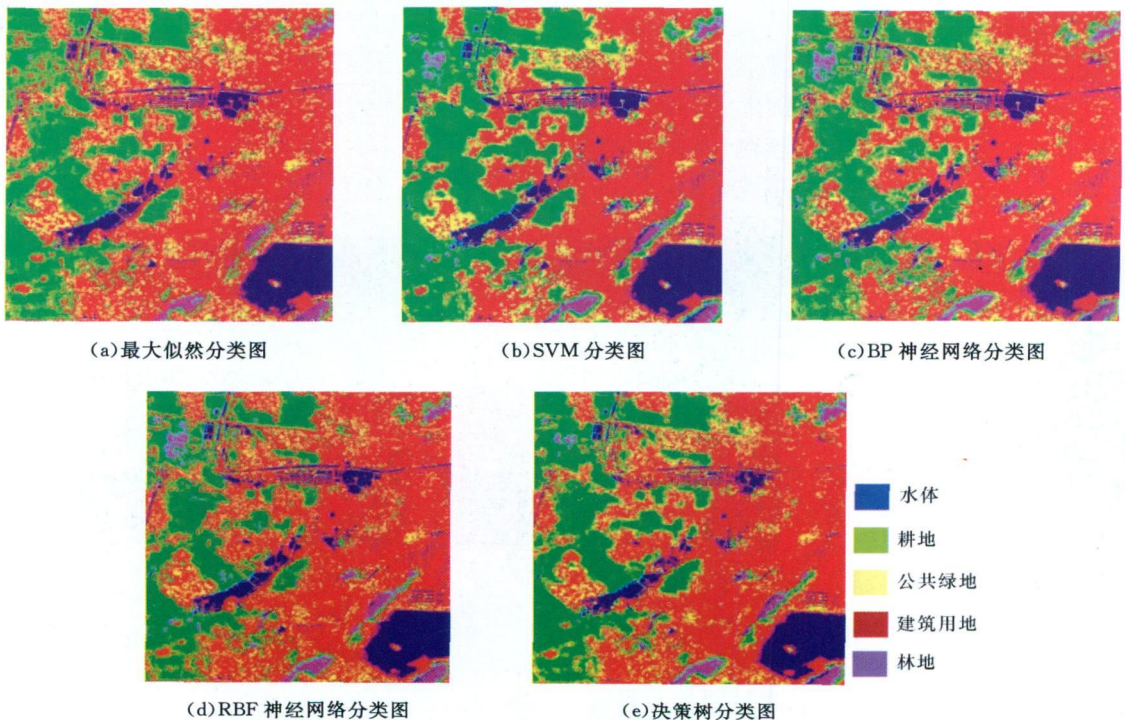


图 5 融合后图像单分类器分类结果图

表7 融合后图像单分类器分类精度表

	水体	耕地	建筑用地	林地	公共绿地	分类精度	Kappa 系数
最大似然分类器	96.43	92.37	91.72	90.18	81.16	92.0019	0.8925
支持向量机分类器	98.17	87.90	92.08	94.11	92.90	92.2201	0.8964
BP 神经网络分类器	97.37	79.66	94.00	100.00	74.84	88.0635	0.8430
RBF 神经网络分类器	97.87	79.15	92.62	94.41	56.52	85.5429	0.8094
决策树分类器	96.83	90.42	93.76	93.81	83.35	92.2685	0.8966

从融合后图像的单分类器的分类结果和各地物的分类精度可以看出:对于最大似然分类器、支持向量机分类器和决策树分类器,融合后影像的总分类精度都比多光谱数据的分类精度有显著提高,除了BP神经网络和RBF神经网络分类器。而且各地物类别的分类精度都有所提高,除了耕地略有下降。这说明图像融合技术是解决多源遥感数据综合、提高遥感图像空间分辨率的有效方法,对于土地的覆

盖分类具有可靠性。

4.2.2 基于 Boosting 和 Bagging 算法的试验

类似多光谱数据, BP 神经网络分类器、RBF 神经网络分类器、决策树分类器分别基于 Bagging 算法进行了多分类器集成的实验。3 种不同的分类器分别基于 4 个不同的训练样本采样率进行了 4 次实验。分类精度结果如表 8 所示。

表8 基于训练样本的多分类集成分类精度表

	BP 神经网络分类器	RBF 神经网络分类器	决策树
基分类器	88.0635	85.5429	92.2685
25% 样本 Bagging	90.6689	84.7067	91.6142
50% 样本 Bagging	90.0267	83.4828	91.4687
75% 样本 Bagging	89.5540	83.4707	91.8686
100% 样本 Bagging	88.8997	84.4644	91.9292
Boosting	89.2874	86.1003	92.0625

通过 Bagging 和 Boosting 实验结果可以看出:

(1) 采用 Bagging 和 Boosting 算法, BP 神经网络取得了比基分类器高的分类精度,但是对于 Bagging 算法,随着采样率的提高精度略有下降。而对于 RBF 神经网络分类器, Bagging 算法效果欠佳, Boosting 算法的分类精度比基分类器有所提高。

(2) 与多光谱数据结果略有不同,对于 BP 神经网络,采用 Boosting 算法的精度并不优于 Bagging 算法,而对于 RBF 神经网络, Boosting 算法优于 Bagging 算法。

(3) 基于 Bagging 算法,采用不同的采样样本率时,可以得到不同的分类精度,而且随着采样率的提高,精度也略有下降或者波动。但是使用融合数据后,决策树分类器的分类效果较多光谱数据有所提高。

4.2.3 多分类器集成实验与分析

对于融合后的图像,同样采用 Double Fault、WCEC、Kappa 值分别对每组实验中成员分类器之间的差异性进行测量如表 9,综合几种差异性测量结果,从 5 个单分类器中选择最大似然分类器、SVM 分类器、决策树分类器构成多分类器集合。表

10 和图 6 分别为多分类器组合的分类精度表和分类结果图。

表9 融合后图像各种不同分类器组合时差异性测量指标值

分类器组合	Kappa	Double Fault	WCEC
1、2、3	0.517633	0.0225591	0.849028
1、2、4	0.485681	0.0217425	0.851580
1、2、5	0.231596	0.00842138	0.928393
1、3、4	0.601791	0.0261828	0.832491
1、3、5	0.287385	0.0107181	0.919002
1、4、5	0.300511	0.0115858	0.914153
2、3、4	0.546813	0.0233247	0.844128
2、3、5	0.308060	0.0108712	0.918032
2、4、5	0.258052	0.00918695	0.924718
3、4、5	0.299124	0.0108202	0.918032

表10 基于多分类器组合的多分类集成分类精度表

分类器	分类精度	Kappa 系数
投票法	92.9593	0.9059
证据理论	92.9593	0.9059
模糊积分	93.1774	0.9087

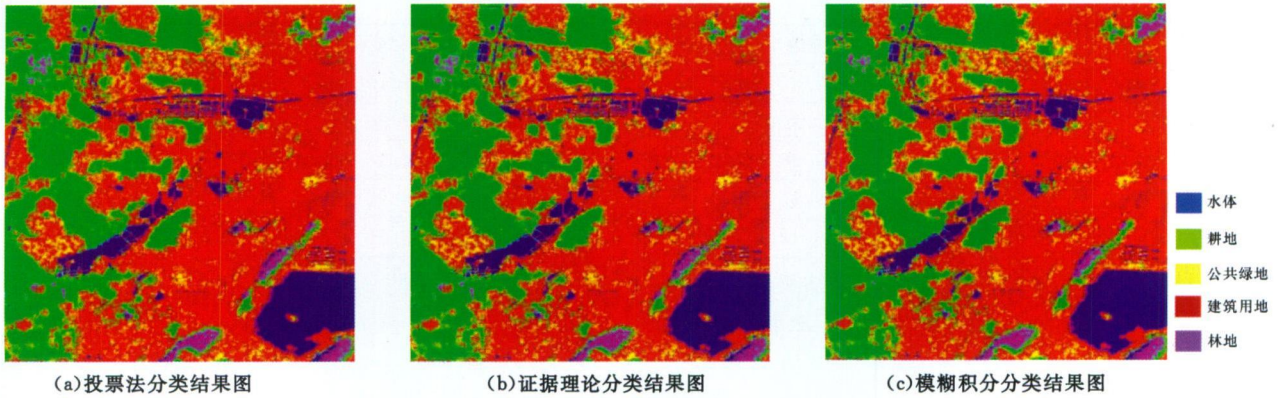


图 6 融合数据多分类器集成分类结果图

通过实验可以看出: 与多光谱数据相似, 分类器的集成实验能有效提高分类器的分类精度, 另外与投票法和证据理论比较, 模糊积分的多分类器集成方法要比其他两个的分类精度都高。

4.3 融合前后分类结果对比

4.3.1 单分类器分类结果对比

表 11 多光谱与融合后数据的单分类器结果比较

分类器	影像	水体	耕地	建筑用地	林地	公用绿地	总体精度	Kappa 系数
最大似然分类器	多光谱	76.47	99.05	89.46	96.43	53.62	90.2423	0.8495
	IHS 融合	96.43	92.37	91.72	90.18	81.16	92.0019	0.8925
SVM 分类器	多光谱	86.27	99.05	79.92	87.86	76.81	86.1821	0.7964
	IHS 融合	98.17	87.90	92.08	94.11	92.90	92.2201	0.8964
BP 神经网络分类器	多光谱	91.18	98.33	81.05	100.0	79.71	88.1467	0.8244
	IHS 融合	97.37	79.66	94.00	100.0	74.84	88.0635	0.8430
RBF 神经网络分类器	多光谱	86.27	98.81	87.33	92.14	55.07	89.3910	0.8370
	IHS 融合	97.87	79.15	92.62	94.41	56.52	85.5429	0.8094
决策树分类器	多光谱	78.43	97.14	84.57	98.57	69.57	88.2122	0.8230
	IHS 融合	96.83	90.42	93.76	93.81	83.35	92.2685	0.8966

4.3.2 基于 Bagging 和 Boosting 算法的结果对比

综合表 3 和表 8 得到了表 12, 从表中可以看出: 对于 BP 神经网络, 多光谱数据和融合后数据在 Bagging 和 Boosting 算法都取得了比基分类器高的分类精度, 虽然融合后数据的分类精度要高, 但是对于 Bagging 算法随着采样率的提高精度略有下降。对于 RBF 神经网络分类器, 融合后数据不如多光谱数据效果好。对于决策树分类, 融合后数据比多光谱数据普遍分类精度要高, 但是这两种方法都没有比基分类器的精度有所提高, 因此 Bagging 和 Boosting 算法不适应决策树。

为了方便进行多光谱和 IHS 融合后数据的对比, 综合表 2 和表 7 得到了表 11, 从表中可以看出: 各个地物中除了耕地外, 融合后数据的精度都较多光谱有所增加; 对于总体精度, 除了 BP 神经网络和 RBF 神经网络外, 融合后数据的分类精度都有显著提高。

综上所述, 基于“北京一号”小卫星分析城市变化时, 可以根据需要选择融合后数据基于 BP 神经网络分类器 25% 样本的 Bagging 算法, 多光谱数据基于 BP 神经网络分类器和 RBF 神经网络分类器的 Boosting 算法。

4.3.4 基于多分类器组合的结果分析

综合表 4 和表 9 得到了表 13, 从表中可以看出: 3 种方法中融合后影像的分类精度都较多光谱数据要高, 模糊积分相对提高的较多。因此在以后的分类器组合中, 如果选用合理的分类器, 再运用融合技术, 对于分类精度的提高将有很好的作用。

表 12 多光谱与融合后数据的 Bagging 和 Boosting 算法结果比较

	影像	BP 神经网络分类器	RBF 神经网络分类器	决策树
基分类器	多光谱	88.1467	89.3910	88.2122
	IHS 融合	88.0635	85.5429	92.2685
25% 样本 Bagging	多光谱	88.6051	88.4741	81.7944
	IHS 融合	90.6689	84.7067	91.6142
50% 样本 Bagging	多光谱	88.6228	89.5874	81.7289
	IHS 融合	90.0267	83.4828	91.4687
75% 样本 Bagging	多光谱	88.2954	89.7184	80.3536
	IHS 融合	89.5540	83.4707	91.8686
100% 样本 Bagging	多光谱	88.8847	90.0458	81.4669
	IHS 融合	88.8997	84.4644	91.9292
Boosting	多光谱	90.2423	90.1768	88.2122
	IHS 融合	89.2874	86.1003	92.0625

表 13 多光谱与融合后数据的多分类器组合分类精度比较

分类器	影像	分类精度	Kappa 系数
投票法	多光谱	90.3078	0.8513
	IHS 融合	92.9593	0.9059
证据理论	多光谱	90.4388	0.8532
	IHS 融合	92.9593	0.9059
模糊积分	多光谱	90.2423	0.8499
	IHS 融合	93.1774	0.9087

5 结论与展望

随着不同领域应用对遥感分类精度的要求越来越高,研究人员提出了各种用来提高分类精度的新方法。多分类器集成是一个具有广泛应用前景的方法。本研究以“北京一号”小卫星多光谱遥感图像和多光谱与全色融合图像为数据源,进行多分类器集成的试验和分析。得出以下结论:

(1) 采用 Bagging 算法和 Boosting 算法的多分

类器集成中,综合比较 3 种分类器, BP 神经网络分类器取得了较好的效果,对于“北京一号”小卫星的影像分析很有潜力。而且 Boosting 策略对错误样本更加敏感,对分类精度的提高略优于 Bagging 方法。

(2) 投票法、证据理论法和模糊积分法都在不同程度上提高了分类的精度,而且融合后影像的分类精度要比多光谱数据高。因此分类器集成方法能有效提高遥感图像的分类精度,是一种很有潜力的方法。

(3) 对比融合前后的分类结果,可以看出融合后图像的分类精度在一定程度上要比多光谱图像的分类精度有很大提高。因此将多分类器集成与多分辨率图像融合技术综合,既可以充分利用全色影像和多光谱信息的互补信息,又可以集成不同分类器的优势,对于土地覆盖分类及景观格局分析具有可靠性,也可以有效推进“北京一号”小卫星数据的应用。

参考文献

- [1] 杜培军,袁林山,张华鹏,等. 基于多时相 CBERS 影像分析矿业城市景观格局变化—以徐州市为例[J]. 中国矿业大学学报, 2009, 38(1): 106-113.
- [2] 邬建国. 景观生态学[M]. 北京: 高等教育出版社, 2000.
- [3] 傅伯杰,陈得顶,马克明,等. 景观生态学原理及应用[M]. 北京: 科学出版社, 2001.
- [4] Van Eetvelde, V., Antrop, M.. Analyzing structural and functional changes of traditional landscapes—two examples from Southern France[J]. Landscape and Urban Planning, 2004, 67(1): 79-95.
- [5] 杜培军. 遥感原理与应用[M]. 江苏徐州: 中国矿业大学出版社, 2006: 60-73.
- [6] 李晓琴. 基于遥感的北京山区植被覆盖景观格局动态变化研究[D]. 北京: 中国农业大学, 2003.

- [7] Witten, I. H., Frank, E. . Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations(second edition)[M] . China Machine Press. 2006.
- [8] Freund, Y. , Schapire, R. E. . Experiments with a new boosting algorithm[C] // Proc the Thirteenth International Conference on Machine Learning, Morgan Kaufmann, 1996: 148– 156.
- [9] 刘纯平, 戴锦芳, 钟文, 等. 基于模糊证据理论分类的多源遥感信息融合[J] . 模式识别与人工智能, 2003, 16(2): 213– 218.
- [10] 孙全, 叶秀清, 顾伟康. 一种新的基于证据理论的合成公式[J] . 电子学报, 2000(8): 117– 119.
- [11] Nemmour, H., Chibani, Y. . Multiple support vector machines for land cover change detection: An application for mapping urban extensions[J] . ISPRS Journal of Photogrammetry & Remote Sensing, 2006, 61(2) : 125– 133.
- [12] 林剑, 鲍光淑, 王润生, 等. 基于模糊密度分解的遥感图像光谱和纹理信息的融合[J] . 电子学报, 2004(12) : 2028– 2030.
- [13] Kang, H. J., Doermann, D. . Selection of classifiers for the construction of multiple classifier systems[C] // Proceedings of the Eight International Conference on Document Analysis and Recognition. 2005: 1194– 1198.
- [14] Petrakos, M., Benediktsson, J. A. . The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion[J] . IEEE Transactions on Geoscience and Remote Sensing, 2001, 39(11) : 2539– 2545.
- [15] Giacinto, G., Roli, F. . Design of effective neural network ensembles for image classification purposes[J] . Image and Vision Computing, 2001, 19(9): 699– 707.
- [16] Aksela, M., Laaksonen, J. . Using diversity of errors for selecting members of a committee classifier[J] . Pattern Recognition. 2006, 39(4): 608– 623.
- [17] 赵文吉, 段福州, 刘晓萌, 徐智勇. ENVI 遥感影像处理 专题与实践[M] . 中国环境科学出版社, 2007.

(上接第 35 页)

参考文献

- [1] R. Bamler, P. Hartl. Synthetic aperture radar interferometry[J] . Inverse Problem, 1998: 1– 54.
- [2] R. N. Treuhaft, P. R. Siqueira, . The calculated performance of forest structure and biomass estimates from interferometric radar[J] . Waves in Random Media, 2004(14) : S345– S358.
- [3] S. Oveisgharan, H. A. Zebker. Estimating snow accumulation from InSAR correlation observations[J] . IEEE Transactions on Geoscience and Remote Sensing, 2007(45) : 10– 20.
- [4] R. F. Hanssen. Radar Interferometry, Data Interpretation and Error Analysis[M] . The Netherlands: Kluwer Academic Publishers, 2001.
- [5] H. A. Zebker, K. Chen. Accurate estimation of correlation in InSAR observations[J] . IEEE Geoscience and Remote Sensing Letter, 2005(2) : 124– 127.
- [6] R. Touzi, A. Lopes. Statistics of the Stokes parameters and of the complex coherence parameters in one-look and multi-look speckle fields[J] . IEEE Transactions on Geoscience and Remote Sensing, 1996(34) : 519– 531.
- [7] R. Touzi, et al. . Coherence estimation for SAR imagery[J] . IEEE Transactions on Geoscience and Remote Sensing, 1999(37) : 135– 149.
- [8] R. Abdelfattah. Interferometric SAR coherence magnitude estimation using second kind statistics[J] . IEEE Transactions on Geoscience and Remote Sensing, 2006(44) .
- [9] A. Jarvis, H. I. Reuter, A. Nelson, E. Guevara. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture(CIAT)[EB/OL]. <http://srtm.csi.cgiar.org>. 2008.
- [10] 韦海军, 等. 基于粗精度 DEM 的复杂地形 InSAR 基线估计方法[J] . 国防科技大学学报, 2010(32) : 74– 78.