

遥感影像多分类器集成的关键技术与系统实现

夏俊士, 杜培军, 张 伟

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 为克服单一分类器在遥感影像分类精度和效率方面的限制, 有必要构建多分类器系统, 集不同分类器的优点, 获得比单一分类器更高的精度。针对遥感影像的特点和分类的需求, 在遥感影像多分类器集成系统需求分析和系统设计的基础上, 运用 IDL 语言在 ENVI 遥感影像处理平台下实现系统开发。遥感影像多分类器集成系统的主要功能包括遥感影像文件处理、特征选择与提取、分类预处理、分类、多种模式的多分类器集成 (固定组合模式、用户自定义模式、向导模式和推荐模式) 等。通过分类实例对系统应用进行介绍, 表明本系统能够有效地提高遥感影像分类精度。

关键词 多分类器集成; 差异性测度; 遥感影像; 交互式数据语言

中图分类号 TP79

文献标识码 A

doi 10.3981/j.issn.1000-7857.2011.21.002

Key Techniques and Implementation of Multiple Classifier System for Remote Sensing Images

XIA Junshi, DU Peijun, ZHANG Wei

Key Laboratory for Land Environment and Disaster Monitoring of SBSM; China University of Mining and Technology, Xuzhou 221116, Jiangsu Province, China

Abstract In order to overcome the limitation of the accuracy and efficiency of a single classifier, Remote Sensing Multiple Classifier System (RSMCS) is proposed. The system is able to combine the advantages of different classifiers and in the meantime acquires higher accuracy than that for a single classifier. According to the characteristics of remote sensing images and the requirement of classifications, based on the requirement analysis and design for the system, the system is developed by using IDL language on the ENVI remote sensing information processing platform. The major functions of the system include remote sensing image file processing module, image feature selection and extraction module, remote sensing image classification pre-processing module, and remote sensing multiple classifier ensemble module (fixation combination style, user-defined style, wizard style, and recommend style). Furthermore, two experiments are presented to illustrate that the system could effectively improve the accuracy of remote sensing image classification.

Keywords multiple classifier ensemble; diversity measure; remote sensing image; interactive data language

0 引言

近年来, 随着遥感技术的不断发展, 遥感影像的空间、光谱和时间分辨率越来越高, 对遥感影像处理如分类、变化检测等都提出了新的挑战。遥感影像常用的遥感分类器包括最大似然分类器、最小距离分类器、面向对象分类器、人工免疫系统分类器、支持向量机分类器和神经网络分类器等^[1-4]。但

是, 每一种遥感分类器都有其自身的优势和局限性, 其精度和适用范围也有一定限度^[5]。Gincinto 和 Roli 对各种分类器在遥感数据分类中的应用进行了比较研究, 得出“没有哪一个分类器是万能的”的结论^[6]。除了发展更先进的分类器之外, 进行多分类器集成有望提高分类精度, 由于不同分类器能够提供关于待分类模式的互补信息, 因此组合分类器作

收稿日期: 2010-12-15; 修回日期: 2011-07-04

基金项目: 国家自然科学基金项目(40871195); 高等学校博士学科点专项科研基金课题(20070290516); 教育部留学回国人员科研启动基金项目

作者简介: 夏俊士, 博士研究生, 研究方向为遥感影像处理及模式识别、资源环境遥感, 电子信箱: xiajunshi@126.com; 杜培军 (通信作者), 教授, 研究方向为资源环境遥感及遥感信息智能化处理, 电子信箱: dupjrs@cumt.edu.cn

为一种有效的方法,能够获取比任何单一分类器更好的分类精度。

多分类器集成学习的效果不仅取决于组合规则,还取决于分类器成员选择的好坏,一般认为选择差异性较大的分类器参与集成效果较好。集成规则主要有基于算术运算的组合规则、投票法、根据后验概率的线性加权法、贝叶斯平均、证据推理法、模糊推理法、统计意见一致法、分类器动态选择法以及将分类结果作为新输入特征的神经网络组合方法、多级分类方法等^[7-18]。除组合单分类器输出的方法外,基于训练样本的组合方式(Adaboost 和 Bagging 算法^[19-20])和基于特征选择组合的方式(Random Forest 和 Rotation Forest)也受到了研究人员的重视^[21-22]。差异性指标主要有交互信息、Q 统计量、相关系数、双错测度、不一致测度、带有权重的错误和正确结果等^[8,23]。

面对遥感影像日益需求的高精度分类要求,遥感影像多分类器集成系统 (Remote Sensing Multiple Classifier System, RSMCS)将模式识别中的多分类器集成引入到遥感影像分类领域,通过对遥感影像的单分类器的选择与组合,获得比任何单一分类器更高的精度。

1 多分类器集成系统

本系统采用 ENVI+IDL 的系统二次开发策略,这样既可以保留一些 ENVI 现有的功能,同时利用交互式数据语言 (Interactive Data Language, IDL)编写多分类器系统的核心功能模块,并将两者有效地结合在一起,构成遥感影像多分类器集成系统。

系统设计成典型的从遥感影像输入、预处理、特征选择与提取、分类、多分类器组合、分类器差异性测度计算、显示输出到精度评价的遥感影像处理系统。基于以上分析,将 RSMCS 划分为 5 大子模块。

(1) 遥感影像文件模块:主要包括打开影像、矢量文件,保存影像、矢量文件以及对感兴趣区的一系列等文件基本操作。

(2) 遥感影像预处理模块:主要包括辐射定标、影像裁剪、影像旋转、影像叠加、影像掩模、格式转换等预处理工具,其

中格式转换工具主要是对 ENVI 和开源数据挖掘软件 Weka 之间的数据进行转换,以及对最小距离、光谱角制图、光谱信息散度等的规则文件进行概率换算。

(3) 遥感影像的特征选择与提取模块:主要包括主成分变换、独立成分变换、最大噪声分离、植被指数提取、缨帽变换和纹理特征提取等。

(4) 遥感影像单分类器分类模块:主要是利用 ENVI 和 Weka 中的分类器,在影像预处理模块已设置了文件格式转换程序,可以有效地将 ENVI 遥感影像互操作和 Weka 的分类功能结合起来,主要的分类方法有:最小距离、最大似然、光谱角制图、光谱信息散度、神经网络、支持向量机、随机森林、决策树等。

(5) 遥感影像多分类器集成模块:主要包括固定组合模式、用户自定义模式、向导模式和推荐组合模式,界面如图 1 所示。固定多分类器组合选择 ENVI 自带的 6 种分类器:最大似然、最小距离、光谱角制图、光谱信息散度、神经网络和支持向量机为单分类器,选择不同的单分类器组合及组合方法进行集成,如选择最大似然、最小距离、光谱角制图的单分类器组合,组合方法选择投票法。系统要求输入为待分类图像和分类样本,输出为各单分类器及组合分类结果。用户自定义模式主要分为多分类器组合和分类器差异性测度计算,选择相应的组合和差异性测度方法、单分类器个数,输入分类文件即可输出组合分类结果和分类器差异性测度指标值。向导模式分多分类器组合模式和基于差异性测度的向导模式,前者选择两个或两个以上的单分类器(同固定组合的 6 种单分类器)之后,再选择组合方法;后者是选择两个或两个以上的单分类器之后,根据差异性度量指标,优化单分类器组合方案,最后根据推荐方案选择组合方法。推荐组合模式是根据经验分别对高光谱和多光谱数据进行推荐固定分类器组合选择。由于以上 4 种模式的核心是多分类器组合和分类器差异性测度,而这也是建立多分类器系统的关键技术,因此下文对多分类器组合方法和差异性测度两项关键技术进行探讨。

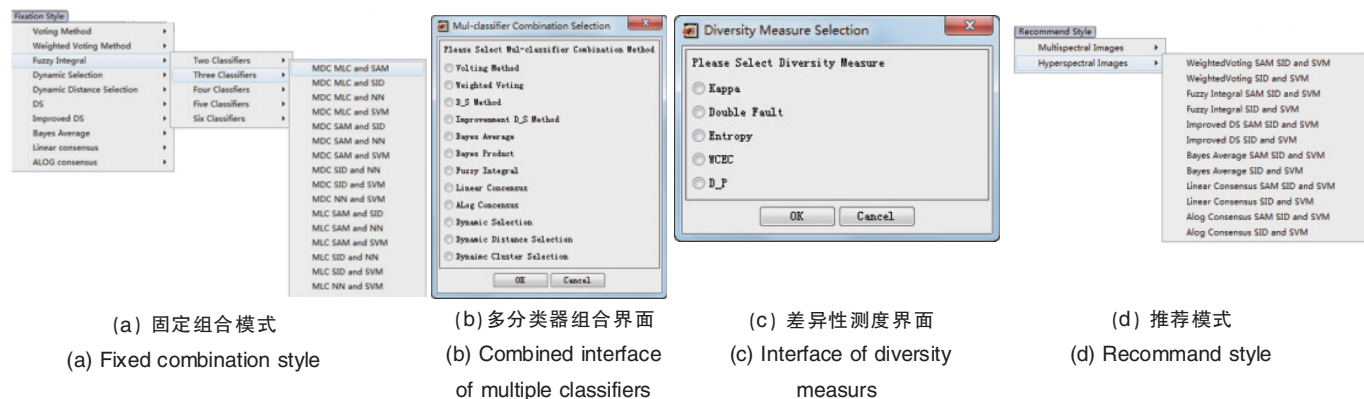


图 1 多分类器集成模块界面
Fig. 1 Interface of multiple classifiers ensemble

2 关键技术

2.1 多分类器组合方法

本系统实现的多分类器组合方法主要有投票法/加权投票法、模糊积分法、贝叶斯平均法、证据理论法、动态分类器选择方法等。

投票法是将大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,随机选择其中一个类别作为最终结果。对各个不同的结果按照分类精度设置不同的权重,进行加权投票。意见一致性理论通过组合多个单分类器的类别概率,根据隶属度最大原则进行类别判断。通常使用线性意见一致性模型和对数意见一致性模型。具体原理见参考文献[11]—[12]。贝叶斯平均法是针对软分类器的分类器集成方法,平均各个分类器的后验概率估计,然后根据平均的后验概率大小,确定待分类像元的类别。模糊积分是一种有效的信息融合方法,它要求证据相互独立,同时考虑到了分类器的可靠性^[13]。证据理论也称为 Dempster-Shafer 理论,其主要优点是将概率论中的单点赋值扩展到集合赋值,因此能更容易地处理未知因素引起的不确定性。DS 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^[15]。对于遥感分类中的待识别像元,当分类器之间识别结果不同时,往往会产生高度冲突的证据。由于证据理论在合成高度冲突的证据时,合成结果有悖常理。因此,本研究中将该新的合成理论应用于遥感图像的分类器集成中,称为改进的证据理论方法。改进的证据理论计算公式和原理见文献[16]。

分类器的动态选择方法有普通分类器选择方法、基于距离权值的分类器选择。普通动态分类器选择方法是对不同的输入样本选择最可能对其进行正确分类的分量分类器,并将该分类器的输出结果作为最后的决策^[17]。基于距离权值的分类器选择是对每一个待分类的像元,计算所有训练样本到该像元的距离,选择 k 个最邻近的像元(实验中 $k=15$),并根据距离计算每一个邻近像元的权值^[18]。

d_k, d_t, d_l 分别代表按距离递增排序后,第 k 个样本、第 t 个样本、第 l 个样本到待分类像元的距离。第 t 个最邻近像元的权值计算如下:

$$W_t = \begin{cases} \frac{d_k - d_l}{d_k - d_t} & d_k \neq d_l \\ 1 & \text{其他} \end{cases} \quad (1)$$

第 j 个分类器的局部分类精度 L 为

$$L = \frac{1}{k} \sum_{\theta(X_t) = C_j(X_t)} W(t) \quad (2)$$

其中, $\theta(X_t)$ 为第 t 个邻近样本的类别值, $C_j(X_t)$ 为第 j 个分类器判别的第 t 个邻近样本的类别值。选择局部精度 (L) 最大的分类器的分类结果作为最终的分类结果。

2.2 差异性测度

本系统采用 Kappa 统计值、双错测度、熵值、错误和正确结果的权重化等计算分类器的差异性。Kappa 统计值、双错测

度、熵值等差异性测度见文献[23]—[25]。

当一个分类器对某一像元的分类正确,而后续的分类器对该像元的分类也正确,对最终分类器集成结果的改善效果将越来越小,因此,多样性与精度指标能够降低后续分类正确的分类器的贡献,能在分类器精度和多样性之间取得一个折衷。当多样性与精度指标取最大值时对应的分类器集合作为最优分类器成员。

具体计算步骤如下。

(1) 设置 IDL 临时变量等于 0,将所有单个分类器按照总体精度从大到小排序。

(2) 对每一个测试样本,当某分类器分类正确时,对第一个正确分类的分类器,IDL 临时变量 T 增加值为该分类器对该类别的分类精度(生产者精度),对第 M 个正确分类的分类器,IDL 临时变量 T 的增加值为该分类器对该类别的分类精度的 $(1/2)^{M-1}$;

(3) 计算多样性与精度指标值,具体公式如下:

$$D_t = T/S \quad (3)$$

其中, S 为测试样本的数目。

当分类器的数目大于 2 时,计算两两之间的差异性测度,然后取平均。

3 实例分析

本文利用高分辨率 QuickBird 和多光谱 Landsat ETM⁺遥感影像进行分类器集成测试研究。以下所有涉及到的单分类器,均在 ENVI、Weka 中完成。在 QuickBird 实验中,投票法、加权投票法、证据理论、改进的证据理论、模糊积分和动态选择、基于聚类的动态选择组合方法实验中,参与的单分类器包括最大似然分类器、支持向量机分类器、BP 神经网络分类器、RBF 神经网络分类器、简化模糊 H-ARTMAP 神经网络分类器;贝叶斯平均、对数意见一致性、线性意见一致性组合方法中,参与的单分类器有支持向量机分类器、BP 神经网络分类器、RBF 神经网络分类器、马氏距离分类器、决策树分类器。单分类器和分类器组合结果的精度及 Kappa 系数如表 1 所示。部分单分类器和多分类器组合结果见图 2。

ETM⁺遥感图像的多分类器集成实验中,RBF 神经网络分类器、决策树分类器、马氏距离分类器、最大似然分类器、最小距离分类器 5 个分类器被用来进行分类器集成。分类器组合方法采用投票法、加权投票法、证据理论、模糊积分进行组合。单分类器和多分类器集成精度见表 2。部分单分类器和多分类器组合结果见图 3。

通过比较 QuickBird 和 ETM⁺的单分类器和多分类器组合精度,可以发现: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,能够得到比任何单个分类器高的分类精度。不过实验中的各种分类器组合方法,运算时间除动态分类器选择方法时间稍长外,其余方法运算时间都很短;动态分类器选择方法与样本的数量关系是正比关系,当样本数量越大,运算时间越长。

参考文献 (References)

- [1] Chang C I. Spectral information divergence for hyperspectral image analysis [C]//Proceeding of IEEE International Geoscience and Remote Sensing Symposium. Hamburg: University of Hamburg, 1999: 509–511.
- [2] Kumar N. Multispectral image analysis using the object-oriented paradigm [M]. New York: CRC Press, 2006: 1–9.
- [3] Pal M, Mather P M. Support vector machines for classification in remote sensing[J]. *International Journal of Remote Sensing*, 2005, 26(5): 1007–1011.
- [4] 钟燕飞, 张良培, 龚健雅, 等. 基于人工免疫系统的遥感图像分类[J]. 遥感学报, 2005, 9(4): 374–380.
Zhong Yanfei, Zhang Liangpei, Gong Jianya, et al. *Journal of Remote Sensing*, 2005, 9(4): 374–380.
- [5] Benediktsson J A, Swain P H, Ersoy O K. Neural network approaches versus statistical methods in classification of multisource remote sensing data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1990, 28(4): 540–552.
- [6] Gincinto G, Roli F. Ensembles of neural networks for soft classification of remote sensing images [C]// Proceedings of the European Symposium on Intelligent Techniques. Bari, Italy, March 20–22, 1997: 166–170.
- [7] Al-Ani A, Deriche M. A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence [J]. *Journal of Artificial Intelligence Research*, 2002, 17: 333–361.
- [8] Kuncheva L I. Combining pattern classifiers—methods and algorithms[M]. New Jersey: John Wiley and Sons, 2004.
- [9] 柏延臣, 王劲峰. 结合多分类器的遥感数据专题分类方法研究 [J]. 遥感学报, 2005, 9(5): 555–563.
Bo Yanchen, Wang Jinfeng. *Journal of Remote Sensing*, 2005, 9(5): 555–563.
- [10] Doan H T, Foody G M. Increasing soft classification accuracy through the use of an ensemble of classifiers [J]. *International Journal of Remote Sensing*, 2007, 28(20): 4606–4623.
- [11] 赵书河, 李培军, 冯学智. 遥感影像决策级融合方法实验研究 [J]. 测绘科学技术学报, 2007, 24(4): 247–250.
Zhao Shuhe, Li Peijun, Feng Xuechi. *Journal of Zhengzhou Institute of Surveying and Mapping*, 2007, 24(4): 247–250.
- [12] Benediktsson J A, Sveinsson J R. Hybrid consensus theoretic classification[J]. *IEEE Transaction on Geoscience and Remote Sensing*, 1997, 35(4): 833–842.
- [13] 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.
- [14] 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, 32(12): 2028–2030.
Lin Jian, Bao Guangshu, Wang Runsheng, et al. *Acta Electronica Sinica*, 2004, 32(12): 2028–2030.
- [15] 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.
Liu Chunping, Dai Jinfang, Zhong Wen, et al. *Pattern Recognition and Artificial Intelligence*, 2003, 16(2): 213–218.
- [16] 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, 28(8): 117–119.
Sun Quan, Ye Xiuqing, Gu Weikang. *Acta Electronica Sinica*, 2000, 28(8): 117–119.
- [17] Paul C S. Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection[J]. *IEEE Transaction on Geoscience and Remote Sensing*, 2002, 40(4): 801–813.
- [18] Canuto A, Abreu M, Oliveira L, et al. Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for ensembles [J]. *Pattern Recognition Letters*, 2007, 28(4): 472–486.
- [19] Breiman L. Bagging predictors[J]. *Machine Learning*, 1996, 24(1): 123–140.
- [20] Rarsch G, Onoda T, Muller K. Soft margins for AdaBoost [J]. *Machine Learning*, 2001, 42(3): 287–320.
- [21] Waske B, Benediktsson J A, Amason K, et al. Mapping of hyperspectral AVIRIS data using machine-learning algorithms[J]. *Canadian Journal of Remote Sensing*, 2009, 35(1): 106–116.
- [22] Rodriguez J J, Kuncheva L I, Alonso C J. Rotation forest: A new classifier ensemble method [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(10): 1619–1630.
- [23] 李光丽, 杜培军, 王小美, 等. 多源遥感影像土地覆盖分类一致性评价与集成应用[J]. 地理与地理信息科学, 2009, 25(4): 68–71.
Li Guangli, Du Peijun, Wang Xiaomei, et al. *Geography and Geo-Information Science*, 2009, 25(4): 68–71.
- [24] Aksela M, Laaksonen J. Using diversity of errors for selecting members of a committee classifier[J]. *Pattern Recognition*, 2006, 39(4): 608–623.
- [25] Kuncheva L, Whitaker C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy[J]. *Machine Learning*, 2003, 51(2): 181–207.

(责任编辑 代丽)

· 学术动态 ·

“中韩围填海环境影响与管理政策国际研讨会”征文



中国海洋学会和韩国海洋学会将于 2011 年 11 月 6—7 日在厦门召开“中韩围填海环境影响与管理政策国际研讨会,暨第二届中韩海洋科学技术学术会议”。

会议主题:围填海的环境影响与管理政策。内容包括:围填海的历史、现状及需求;典型围填海工程的社会影响、经济影响与环境影响;围填海管理的政策手段、经济手段和技术手段;围填海管理的国际合作对策。

征文截止日期:2011 年 9 月 10 日。

电话:(0532)88967809, 18660292526;传真:(0532)88964779;电子信箱:quanbingo@126.com。