引文格式：杨小波,苗李莉. 利用空间邻域信息和多分类器集成分类高光谱影像[J]. 地理空间信息, 2025, 23(5): 71-75. doi:10. 3969/j.issn. 1672-4623.2025.05.015

利用空间邻域信息和多分类器集成分类高光谱影像

杨小波[1]，苗李莉[2*]
（1. 四川省公路规划勘察设计研究院有限公司，四川 成都 610041；2. 国家基础地理信息中心，北京 100830）

摘　要：目前半监督高光谱影像分类方法中，针对未考虑标记样本邻域信息、迭代训练中样本量的研究较少，因此提出了一种基于空间邻域信息和多分类器集成的半监督高光谱影像分类方法；并分别基于Salinas Valley AVIRIS和Pavia ROSIS高光谱影像数据集对该分类方法进行了实验分析。结果表明：①初始模型训练时通过对标记样本进行邻域生长，保证初始模型训练的可靠性；②通过SVM和MLC的预测结果综合评估非标记样本，可实现精准标记；③对于模型训练样本的选择，通过抽取部分样本对分类模型进行多次投票试验，取票数多的作为最终分类结果，保证了分类结果的准确性。该方法能很好地改善较少样本情况下图像分类模型的泛化能力，提高分类精度。

关键词：半监督分类；高光谱影像；邻域信息；多分类器集成；SVM分类
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2025）05-0071-05

Hyperspectral Image Classification Based on Spatial Neighborhood Information and Multi-classifier Integration

YANG Xiaobo[1], MIAO Lili[2]
(1. Sichuan Highway Planning, Survey, Design and Research Institute Co., Ltd., Chengdu 610041, China;
2. National Geomatics Center of China, Beijing 100830, China)

Abstract: We proposed a semi-supervised hyperspectral image classification method based on spatial neighborhood information and multi-classifier integration, in order to address the lack of consideration for labeled sample neighborhood information and the sample size in model iterative training. Then, we conducted the experimental comparative analysis on this method based on two hyperspectral image datasets, Salinas Valley AVIRIS and Pavia ROSIS. The results indicate that ①during the initial model training, neighborhood growth is performed on labeled samples to ensure the reliability of initial model training. ②By using SVM and MLC prediction results to comprehensively evaluate unlabeled samples, accurate labeling can be achieved. ③For the selection of model training samples, multiple voting experiments are conducted on the classification model by extracting some samples, and the one with the most votes is taken as the final classification result, ensuring the accuracy of classification results. The method proposed in this article can effectively improve the generalization ability of image classification model with fewer samples, and improve classification accuracy.

Key words: semi-supervised classification, hyperspectral image, neighborhood information, multi-classifier integration, SVM classification

高光谱遥感影像是成像光谱仪在电磁波谱的紫外、可见光、近红外和中红外区域获得的大量光谱连续且分辨率较高的图像数据[1]。随着高光谱遥感技术的迅猛发展，其应用需求不断增加，如土地利用变化评估、大气成分监测等[2-4]，使得高光谱遥感影像分类成为行业研究重点。传统分类方法主要分为人工神经网络、SVM等监督分类[5]和K-means聚类[6]、模糊聚类[7]等非监督分类，监督分类方法需要大量有标签样本对网络进行训练，将耗费大量的人力物力，成本高昂[8]；半监督分类方法最早由Shahshahani等[9]展开研究，基于少量标记样本和大量未标记样本，采用不同的学习方式提升分类器性能，以获取较好的分类效果。在目前的高光谱影像半监督分类方法中[10-13]，多数仅利用少量标记样本初次训练模型，未考虑标记样本的邻域信息；且将新选择样本全部参与模型训练，对训练样本量的研究较少。鉴于此，本文提出了一种基于空间邻域信息和多分类器集成的半监督高光谱影像分类方法；并通过Salinas Valley AVIRIS和Pavia ROSIS两组高光谱影像数据集对该方法进行了实验分析。

项目来源：四川省交通运输科技资助项目（2021-ZL-07）。
第一作者简介：杨小波（1976—），本科，高级工程师，从事高速勘测与管理工作，E-mail：34661785@qq.com。
通信作者：苗李莉（1978—），博士，高级工程师，从事遥感影像智能化信息提取相关工作，E-mail：mll0826@163.com。

1 分类算法

为提升高光谱影像分类精度，本文主要从影像数据标定样本选择、预测以及标记非标记样本等方面进行优化。

1.1 邻域生长

地理学第一定律指出，地理事物或属性在空间分布上互为相关，因此样本邻域信息在半监督分类中得到了广泛应用[5,14]。本文采用八邻域方法，以标记样本点为中心，对其进行八邻域生长。其关键因素在于阈值的设定，本文选择标记样本点八邻域内未标记样本点与该标记样本点像元值之差的绝对值中值作为阈值，进行标记样本点邻域生长。另外，考虑到高光谱成像中的混合像元问题，邻域生长时加入了边缘限制，利用Canny算子[15]使结果不包含地物的边缘点。高光谱影像包括几十甚至几百个波段，且各波段穿透能力不同，每个波段都会在不同程度上反映地物对象的不同光谱信息，因此利用不同波段对标记样本点进行邻域生长将得到不同的结果。为使同一地物标记样本点邻域生长结果的光谱信息符合真实地物情况，选择近似真彩色3波段的光谱影像作为邻域生长和提取边缘影像，能有效减少边缘混合像元混入初始训练样本，改善数据处理的冗余问题。

1.2 非标记样本选择

为降低标记样本量对高光谱影像分类精度的影响，不增加前期人工样本标记工作，本文将合适的非标记样本加入标记样本中，逐步完成对分类模型的训练，达到提高分类精度的目的。因此，利用邻域生长算法标记样本的邻域信息，并对非标记样本进行选择标记，成为增加标记样本数量、提升分类模型分类性能的有效途径。选择非标记样本时，首先确定一定半径的结构元素，将标记样本半径内的非标记样本作为候选样本；然后对候选样本进行预测，结果与标记样本相同的点将加入标记样本，形成新标记样本。由于本文采用两种分类器组合的方式对高光谱影像进行分类训练，因此将两种分类器训练标记一致的样本作为新标记样本。

1.3 多分类器集成训练

本文采用多分类器集成的方式训练样本。以不同分类器确定样本时采用的不同假设理论为根据，更全面地训练样本，能弥补单一分类器的局限，减少样本对单一分类器的依赖[16]。SVM最早由Vapnik[17]提出，由于已有标记样本数量有限，仅采用SVM对未标记样本进行训练，将使得训练结果与真实地物情况差异较大，分类器难以正确预测非标记样本，若将错误结果加入新的训练样本中，将直接影响下次分类器的训练，造成错误累积。MLC是一种稳定性和鲁棒性较好的分类方法[18]。本文将SVM与MLC相结合进行半监督高光谱影像分类，首先从标记样本中随机抽取部分样本作为训练样本，分别采用SVM和MLC对高光谱影像进行分类；然后将标记样本邻域内的非标记样本作为候选标记样本，并比较两种分类器的结果，结果一致且与标记样本类别相同时则将该样本点加入标记样本；最后在新的标记样本集合中随机抽取部分样本作为新训练样本，重复上述步骤，直至达到预设条件为止。

2 实验结果与分析

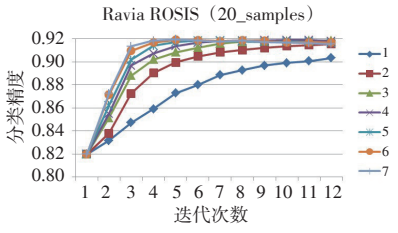
为验证本文方法，分别选用Salinas Valley AVIRIS和Pavia ROSIS高光谱影像数据集进行实验分析。Pavia ROSIS原始影像包括115个波段，波长范围为0.43～0.86 μm，剔除受噪音影响严重的波段后采用103个波段；影像大小为610像素×340像素，分辨率为1.3 m。Salinas Valley AVIRIS原始影像包括224个波段，剔除受水汽吸收影响严重的波段后采用204个波段；影像大小为512像素×217像素，分辨率为3.7 m。

2.1 结构元素半径的确定

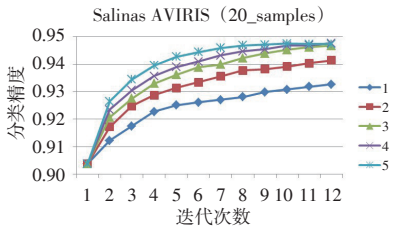
为实现最佳邻域生长效果，确定形态学结构元素半径是其关键，决定了被选入非标记样本的数量，影响着整个实验的运算效率。通过大量试验发现，Pavia ROSIS和Salinas Valley AVIRIS数据的最优半径分别为初始标记样本数10个和20个。Pavia ROSIS和Salinas Valley AVIRIS数据结构元素半径选择对分类结果的影响见图1，可以看出，Pavia ROSIS数据将结构元素的半径设置为6时，分类精度达到要求的运行效率最高，而Salinas Valley AVIRIS数据将结构元素的半径设置为3时，分类精度即可满足要求。

2.2 抽取样本量的确定

控制训练样本量不仅可减少数据冗余、提高运算效率，而且可减少错误样本对分类模型的影响。本文对Pavia ROSIS、Salinas Valley AVIRIS数据分别采用半径为6和3的结构元素进行训练，确定最佳抽取样本量。不同抽取样本量对分类精度的影响见图2，可以看出，随着样本抽取量的增加，影像分类精度曲线平稳性持续变好，但样本量增加到一定程度后，再增加样本量对提升精度没有明显效果；将抽取样本量设置



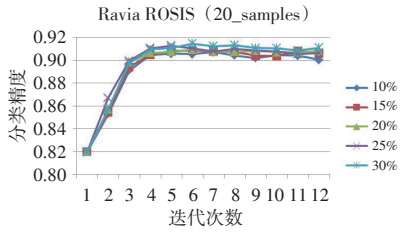
a Pavia ROSIS



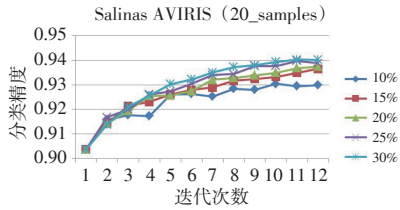
b Salinas Valley AVIRIS

图1　不同结构元素半径的实验结果

为25%和30%时，分类精度没有明显差异。在不影响实验精度的条件下，对于两组数据，本文选择在每次迭代继承过程中抽取标记样本的25%作为模型训练样本集，对模型进行训练，能获得最佳效率。



a Pavia ROSIS

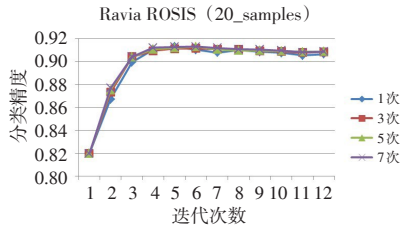


b Salinas Valley AVIRIS

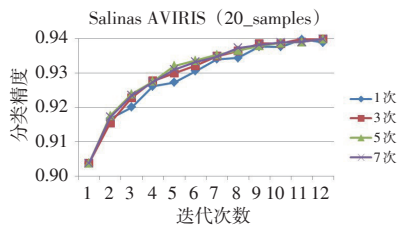
图2　不同抽取样本量的实验结果

2.3　投票次数的确定

投票次数对分类精度的影响见图3，可以看出，随着初始标记样本数量的增加，投票次数对Pavia ROSIS数据分类精度的影响减少；由于Salinas Valley AVIRIS数据同类地物的分布比较集中，除20个初始标记样本抽取1次条件下的分类结果精度滞后于其他组外，投票次数对分类结果的影响均不明显，因此各组实验中均选用抽取3次的分类精度再迭代，可实现最佳运行效率，且与更多次抽取的分类精度无明显差距。



a Pavia ROSIS

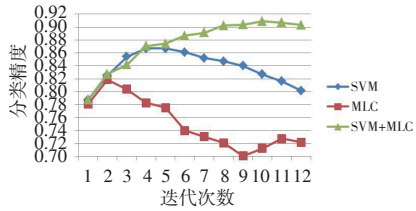


b Salinas Valley AVIRIS

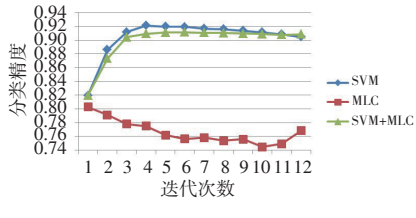
图3　不同投票次数的实验结果

2.4　分类器融合对实验结果的影响

由各参数实验结果可知，实验抽取标记样本的25%，Pavia ROSIS、Salinas Valley AVIRIS数据的结构元素半径分别为6、3，采用SVM分类、MLC分类、SVM与MLC相结合分类方法分别进行3次投票实验。在每类10、20个初始标记样本的条件下利用3种方法对Pavia ROSIS数据进行分类的精度见图4，可以看出：①在每类初始标记样本与迭代次数一定的情况下，仅单一分类方法的精度随迭代次数的增加而下降；②MLC分类对Pavia ROSIS数据适用性较差，在20个初始标记样本条件下采用SVM分类的部分结果优于SVM与MLC相结合分类的结果；③总体来看，采用SVM与MLC相结合分类得到的精度随迭代次数的增加而上升，在升高到一定程度后，达到较稳定的状态，未有明显下降趋势。



a 10个初始标记样本



b 20个初始标记样本

图4　Pavia ROSIS的分类实验结果

为了更直观地表现3种方法的分类效果，分别在

每类10、20个初始标记样本条件下循环迭代12次，每次抽取样本的25%，结构元素半径为6，Pavia ROSIS数据的分类结果见图5，可以看出，在少量样本的条件下，MLC与SVM相结合分类的结果明显优于MLC分类、SVM分类。



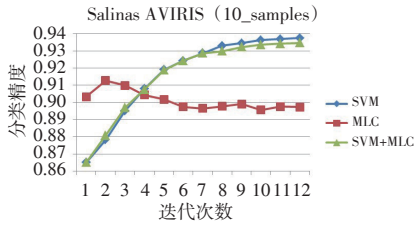
a 10个样本MLC分类 b 10个样本SVM分类 c 10个样本SVM+MLC分类



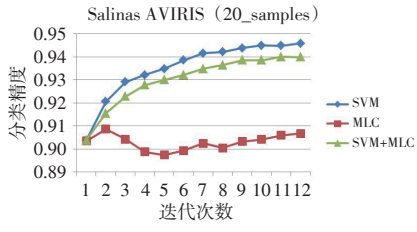
d 20个样本MLC分类 e 20个样本SVM分类 f 20个样本SVM+MLC分类

图5 Pavia ROSIS数据不同分类方法的分类结果

在每类10、20个初始标记样本条件下利用3种方法对Salinas Valley AVIRIS数据进行分类，精度见图6，相较于Pavia ROSIS数据结论发现，在初始样本较多的情况下，SVM分类精度略优于SVM与MLC相结合分类。为了更直观地表现3种方法的分类效果，分别在每类10、20个初始标记样本条件下循环迭代12次，每次抽取样本的25%，结构元素半径为3，



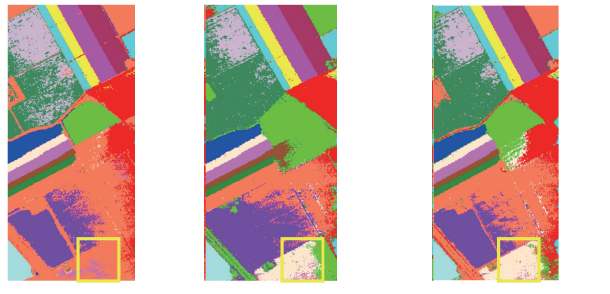
a 10个初始标记样本



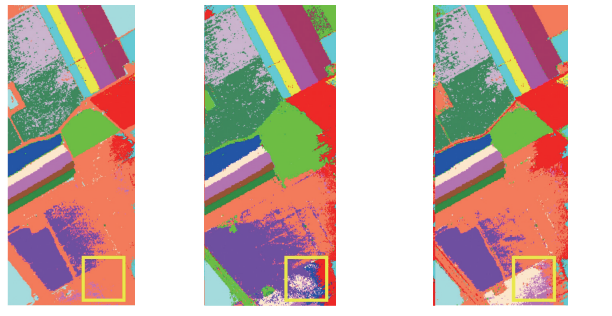
b 20个初始标记样本

图6 Salinas Valley AVIRIS数据的分类实验结果

Salinas Valley AVIRIS数据的分类结果见图7，可以看出，MLC分类效果明显差于其他两种方法；虽然各组实验中SVM分类精度明显优于MLC分类、稍优于SVM与MLC相结合分类，但其获得的3组结果的差异性明显大于其他两种方法。



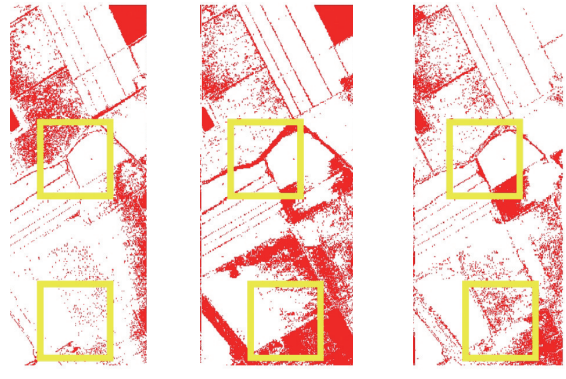
a 10个样本MLC分类 b 10个样本SVM分类 c 10个样本SVM+MLC分类



d 20个样本MLC分类 e 20个样本SVM分类 f 20个样本SVM+MLC分类

图7 Salinas Valley AVIRIS数据不同分类方法的分类结果

为了更好地衡量3组影像分类结果的差异性，分别在每类10、20个初始标记样本条件下绘制3组影像分类结果的差异图（图8），将3幅影像中分类结果不同的点占总影像的百分比定义为图像分类的相异程度，可以看出，SVM分类获得的3组影像的相异性（26.27%）明显高于MLC分类（16.07%）和SVM与MLC相结合分类（15.23%）；相较于SVM分类法，采用SVM与MLC相结合分类法获得的边缘信息有较大改善。



a MLC相异图 b SVM相异图 c SVM+MLC相异图

图8 Salinas Valley AVIRIS数据不同半监督分类方法的分类相异图

3 结 语

针对高光谱影像半监督分类过程中，对模型迭代训练中样本量利用不足的问题，本文提出了一种基于空间邻域信息和多分类器集成的半监督高光谱影像分类方法。

1）在初始模型训练时，通过对标记样本进行邻域生长，可保证初始模型训练的可靠性，同时避免了初始标记样本过少导致的分类模型精度下降。

2）通过SVM和MLC预测的结果对非标记样本进行综合评估，实现了精准标记。

3）对于模型训练样本的选择，本文抽取部分样本对分类模型进行多次投票试验，取票数最多的作为最终分类结果，从而保证分类结果的准确性。

在高光谱影像半监督分类中，本文方法可获得较高的分类精度，采用SVM与MLC相结合分类法能有效改善单独采用SVM分类法边缘数据分类不稳定的问题；但本文方法未对模型训练样本质量以及结构元素半径对分类精度的影响进行深入探究，因此今后将在如何选择高质量的训练样本以及如何确定不同数据约束半径大小方面继续深入研究。

参考文献

[1] 杜培军，夏俊士，薛朝辉，等. 高光谱遥感影像分类研究进展[J]. 遥感学报，2016，20(2):236-256

[2] Chen W，Ouyang S，Yang J，et al. JAGAN: A Framework for Complex Land Cover Classification Using Gaofen-5 AHSI Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，2022，15(6): 1 591-1 603

[3] Elcoroaristizabal S，Amigo J M. Near Infrared Hyperspectral Imaging as a Tool for Quantifying Atmospheric Carbonaceous Aerosol[J]. Microchemical Journal，2021，160:105 619

[4] 刘宇承，罗芳，杜清运，等. 深圳市宝安区土地利用遥感分类方法对比分析[J]. 时空信息学报，2023，30(4):491-499

[5] Tan K，Hu J，Li J，et al. A Novel Semi-supervised Hyperspectral Image Classification Approach Based on Spatial Neighborhood Information and Classifier Combination[J]. ISPRS Journal of Photogrammetry and Remote Sensing，2015，105:19-29

[6] Liu G，Li X，Wang S，et al. An Improved Genetic K-means Clustering Algorithm Based on Image Segmentation[J]. Journal of Naval University of Engineering，2009(1): 447-451

[7] Li J，Gao X，Jiao L. A New Feature Weighted Fuzzy Clustering Algorithm[J]. ACTA Electronica Sinica，2006，34(1):412-420

[8] 杨星，方乐缘，岳俊. 高光谱遥感影像半监督分类研究进展[J]. 遥感学报，2024，28(1):20-41

[9] Shahshahani B M，Landgrebe D A. The Effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon[J]. IEEE Transactions on Geoscience and Remote Sensing，1994，32(5):1 087-1 095

[10] Dopido I，Li J，Plaza A，et al. Semi-supervised Classification of Hyperspectral Data Using Spectral Unmixing Concepts[C]. IEEE Tyrrhenian Workshop on Advances in Radar and Remote Sensing (TyWRRS)，2012

[11] Gu Y，Feng K. L1-graph Semi-supervised Learning for Hyperspectral Image Classification[C]. IEEE International Geoscience and Remote Sensing Symposium，2012

[12] Dópido I，Li J，Marpu P R，et al. Semi-supervised Self-learning for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing，2017，51(7):4 032-4 044

[13] Cavalli R M. Capability of Remote Sensing Images to Distinguish the Urban Surface Materials: a Case Study of Venice City[J]. Remote Sensing，2021，13(19):3 959

[14] Wang L，Hao S，Wang Y，et al. Spatial-spectral Information-based Semi-supervised Classification Algorithm for Hyperspectral Imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，2014，7(8): 3 577-3 585

[15] Ding L，Goshtasby A. On the Canny Edge Detector[J]. Pattern Recognition，2021，34(3):721-725

[16] Yuan P，Chen Y，Jin H，et al. MSVM-kNN: Combining SVM and k-NN for Multi-class Text Classification[C]. IEEE International Workshop on Semantic Computing and Systems，2008

[17] Vapnik V N. An Overview of Statistical Learning Theory[J]. IEEE Transactions on Neural Networks，1999，10(5):988-999

[18] 左晓庆，刘怀鹏. 基于Landsat 8影像的冰面提取光谱模型构建[J]. 遥感信息，2022，37(3):122-128