



多视图主动学习的多样性样本选择方法研究

陈立伟, 房赫, 朱海峰

引用本文:

陈立伟, 房赫, 朱海峰. 多视图主动学习的多样性样本选择方法研究[J]. 智能系统学报, 2021, 16(6): 1007–1014.

CHEN Liwei, FANG He, ZHU Haifeng. Diversity sample selection method of multiview active learning classification[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(6): 1007–1014.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202007037>

您可能感兴趣的其他文章

面向不平衡数据的融合谱聚类的自适应过采样法

Spectral clustering–fused adaptive synthetic oversampling approach for imbalanced data processing

智能系统学报. 2020, 15(4): 732–739 <https://dx.doi.org/10.11992/tis.201909062>

一种自训练框架下的三优选半监督回归算法

Three–optimal semi–supervised regression algorithm under self–training framework

智能系统学报. 2020, 15(3): 568–577 <https://dx.doi.org/10.11992/tis.201905033>

基于PageRank的主动学习算法

Active learning through PageRank

智能系统学报. 2019, 14(3): 551–559 <https://dx.doi.org/10.11992/tis.201804052>

SUCE:基于聚类集成的半监督二分类方法

SUCE: semi–supervised binary classification based on clustering ensemble

智能系统学报. 2018, 13(6): 974–980 <https://dx.doi.org/10.11992/tis.201711027>

动态数据约简的神经网络分类器训练方法研究

Reducing training times in neural network classifiers by using dynamic data reduction

智能系统学报. 2017, 12(2): 258–265 <https://dx.doi.org/10.11992/tis.201605031>

在线学习的大规模网络流量分类研究

Large–scale network traffic classification based on online learning

智能系统学报. 2016, 11(3): 318–327 <https://dx.doi.org/10.3969/j.issn.1673–4785.201603033>



关注微信公众号，获取更多资讯信息

DOI: 10.11992/tis.202007037

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210922.1140.002.html>

多视图主动学习的多样性样本选择方法研究

陈立伟, 房赫, 朱海峰

(哈尔滨工程大学 信息与通信工程学院, 黑龙江 哈尔滨 150001)

摘要: 为了去除高光谱图像多视图主动学习分类中的所选样本的冗余, 降低人工标记成本, 本文提出了两种用于多视图主动学习分类中的多样性样本选择方法。将高光谱图像进行超像素分割, 将所选样本中属于不同的超像素的样本加入训练集, 其余样本加入候选集; 比较各视图对样本的预测标签, 将所选样本中预测标签不完全相同的样本加入训练集, 其余样本加入候选集。本文分别用这两种方法对传统多视图主动学习的样本选择方法进行改进, 并用两组高光谱图像数据进行实验。实验结果表明: 使用这两种方法改进后, 所得分类精度不变, 使用的训练样本数量大幅减少。

关键词: 高光谱图像分类; 多视图主动学习; 多样性; 样本选择; 超像素; 训练样本数量; 预测标签; 分类精度
中图分类号: TP753 **文献标志码:** A **文章编号:** 1673-4785(2021)06-1007-08

中文引用格式: 陈立伟, 房赫, 朱海峰. 多视图主动学习的多样性样本选择方法研究 [J]. 智能系统学报, 2021, 16(6): 1007-1014.

英文引用格式: CHEN Liwei, FANG He, ZHU Haifeng. Diversity sample selection method of multiview active learning classification[J]. CAAI transactions on intelligent systems, 2021, 16(6): 1007-1014.

Diversity sample selection method of multiview active learning classification

CHEN Liwei, FANG He, ZHU Haifeng

(College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: To remove the redundancy of selected samples in the multiview active learning classification of hyperspectral images and reduce the cost of manual marking, this paper proposes two methods for the selection of diverse samples in the multiview active learning classification. First, hyperspectral images are divided into superpixel segments, then samples belonging to different superpixel segments are added to the training set, and the remaining samples are put back into the candidate set. Second, the prediction labels of the samples from each view are compared, then the samples with different prediction labels are added into the training set, and the remaining samples are put back into the candidate set. In this study, the two methods are used to improve the sample selection method in the traditional multiview active learning classification, and experiments are conducted in two groups of hyperspectral image data. The results show that the accuracy of classification is unchanged, yet the number of training samples is greatly reduced after using the two methods.

Keywords: hyperspectral image classification; multiview active learning; diversity; sample selection; superpixel; number of training samples; prediction labels; accuracy of classification

随着遥感技术的迅速发展, 高光谱图像 (hyperspectral image, HSI) 在土地覆盖物分类中得到了广泛的应用^[1-3]。训练一个 HSI 分类器, 通常需要大量的标记样本, 而标记样本的采集过程既昂

贵又费时^[4-5]。主动学习 (active learning, AL) 方法可以有效解决 HSI 标记样本少的问题^[6-7]。在 AL 方法中, 多视图主动学习 (multiview active learning, MVAL) 方法可以从多个视图中提取互补信息, 大大减少训练样本的数量^[8-10]。

学者们对 MVAL 的样本选择方法展开了广泛研究: 文献 [11] 提出了自适应最大不一致 (ad-

收稿日期: 2020-07-23. 网络出版日期: 2021-09-23.

基金项目: 国家自然科学基金项目 (61675051).

通信作者: 朱海峰. E-mail: zhuhaifeng@hrbeu.edu.cn.

aptive maximum disagreement, AMD) 的样本选择方法, 该方法利用各分类器对样本预测结果的不一致性选择样本; 文献 [12] 提出了加权投票熵 (weighted voting entropy, WVE) 的样本选择方法, 由于不同视图对于不同类别的区分能力不同, 该方法通过各视图的权重体现了不同视图对样本的辨别能力的差异; 文献 [13] 提出了一种 IEUE 样本选择方法, 该方法综合考虑了视图内和视图间的不确定性。这些方法均取得了良好的分类结果, 但是这些方法只考虑了样本的不确定性, 没有考虑样本的多样性, 导致所选样本中存在冗余。并且, 随着迭代次数的增加, 不同视图训练的分类器会趋同, 具有相同的分歧程度的样本越来越多, 样本冗余问题会进一步加剧^[11-14]。

目前, 学者们对 MVAL 中所选样本的多样性研究较少。文献 [15] 提出了一种用于 MVAL 的基于聚类的多样性样本选择方法, 该方法采用局部聚类密度度量方法对 HSI 样本进行聚类, 采用光谱角距离作为聚类的距离准则。然而, 该方法在聚类过程中只使用了光谱信息, 没有考虑样本的空间信息, 由此选出的相似样本会产生同谱异物的问题^[16-17]。

本文提出了一种基于超像素分割^[18]的 MVAL 多样性样本选择方法。HSI 的超像素分割方法同时基于样本的光谱特性和空间特性, 可以有效地避免样本选择过程中的同谱异物问题。

在 MVAL 中, 多个视图训练的多个分类器, 彼此独立并互相补充, 共同对样本选择过程和得到最终分类结果起作用^[19]。分类器对不同样本的预测结果的一致性直接与样本间的相似性有关, 因此本文又提出了一种基于多视图预测标签一致性的样本选择方法。

1 算法框架

1.1 MVAL 的基本流程

基于 MVAL 的 HSI 分类的基本流程是: 首先将 HSI 的全部已标记样本分为训练集和候选集, 然后对 HSI 采用某种视图生成方法得到多个视图。每个视图分别训练一个分类器, 并使用每个分类器对全部样本进行预测, 得到其预测的结果和精度。根据预测结果, 使用 MVAL 样本选择策略从候选集中选出信息量大的样本。依次迭代, 直到满足停止条件。根据各个分类器的预测结果得到最终的分类结果和分类精度^[20-21]。停止条件一般为达到了最大迭代次数或分类精度达到某个值。基于 MVAL 的 HSI 分类的基本流程图如

图 1 所示。

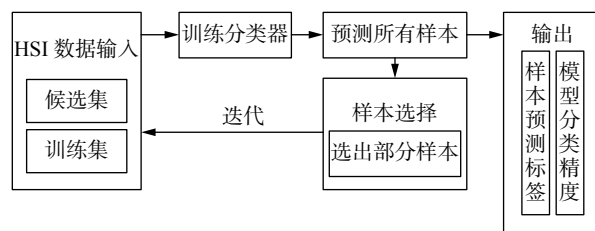


图 1 基于 MVAL 的 HSI 分类的基本流程

Fig. 1 Basic flow chart of HSI classification based on MVAL

1.2 视图生成方法

本文使用了 3D-Gabor 滤波视图生成方法, 通过使用不同频率和方向的 3D-Gabor 滤波器将原始 HSI 转换成多个具有不同频谱空间特征的数据集^[22]。利用这些数据集提供的频谱空间信息, 将 HSI 的光谱和空间信息结合起来。3D-Gabor 内核为

$$G_{\omega, \varphi, \theta}(x, y, \lambda) = g(x, y, \lambda) \exp[j(x\omega_x + y\omega_y + \lambda\omega_\lambda)]$$

式中: ω 为波向量的中心频率; φ 为波向量和光谱维度的夹角; ω_x 、 ω_y 、 ω_λ 分别是样本的特征向量在高光谱图像的横纵坐标轴 x 、 y 和光谱方向 λ 上的投影; θ 是样本的特征向量在地面 x 、 y 上的投影与 x 轴的夹角; $g(x, y, \lambda)$ 是在 (x, y, λ) 域的三维高斯包络线。频率和方向的参数设置为: $\omega \in \left\{ \frac{1}{4}, \frac{1}{8}, \frac{1}{12}, \frac{1}{16}, \frac{1}{20} \right\}$, $\{\varphi, \theta\} = \left\{ 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4} \right\}$ 。当 $\varphi = 0$ 时, 波向量与不同的 θ 矢量方向相同, 共 13 个方向。HSI 经 3D-Gabor 滤波后, 得到 65 个不同频率和方向的 Gabor 立方体, 得到的 Gabor 立方体与原始 HSI 大小相同。

然后, 采用文献 [13] 中提出的 FR 准则衡量视图的充分性。从得到的全部视图中选出充分性最大的 5 个视图作为 MVAL 的多视图。FR 准则为

$$\varphi_{FR} = \frac{1}{B} \frac{2}{r(r-1)} \sum_b \sum_{i,j \in \{1,2,\dots,r\}, i \neq j} \sum_{x \in D_s} \frac{(\mu_i - \mu_j)(\mu_i - \mu_j)^T}{(\mu_i - x)(\mu_i - x)^T + (\mu_j - x)(\mu_j - x)^T}$$

式中: D_s 为初始标记训练样本; r 为地物类别数量; $(\mu_i - \mu_j)(\mu_i - \mu_j)^T$ 为第 i 、 j 类的均值类间散射矩阵; $(\mu_i - x)(\mu_i - x)^T + (\mu_j - x)(\mu_j - x)^T$ 为第 i 、 j 类的方差类内分散矩阵。

1.3 传统的 AMD 查询策略

传统样本选择方法通过比较各分类器对样本的不同预测结果个数, 衡量样本的不确定性, 并从候选集中选出不同预测结果个数最多的样本进行查询。该方法称为自适应最大不一致策略 (adaptive maximum disagreement, AMD)。具体表达式为

$$Q_{AMD} = \arg \max_{x_i \in D_C} |L_k(x_i)|_{\text{count}}, \quad k = 1, 2, \dots, K$$

式中: $L_k(x_i)$ 为第 k 个分类器对样本 x_i 的分类结果, 共有 K 个分类器; D_C 为候选集; $|\cdot|_{\text{count}}$ 为其中不同元素的个数。

2 MVAL 样本选择方法的多样性选择策略

针对 MVAL 样本选择过程中存在冗余样本的问题, 本文提出了基于超像素分割和基于预测标签一致性的两种 MVAL 多样性样本选择策略。在 MVAL 的传统 AMD 样本选择方法后, 使用提出的多样性方法对所选样本进行进一步筛选, 减少训练样本个数, 从而降低人工标记成本。

2.1 基于超像素分割的多样性选择策略

传统的超像素分割方法有 Meanshift、简单线性迭代聚类 (simple linear iterative clustering, SLIC)、归一化分割 (normalized cut)、基于熵率 (entropy rate) 等。其中, SLIC 算法能够根据 HSI 的同质性和非均匀性特点进行超像素分割, 使不同超像素的空间相干性大大降低。同时只需设置一个预分割超像素数量参数即可在运行速度、紧凑整齐度等方面有一定优势^[23]。因此, 本文采用 SLIC 超像素算法^[24]用于 HSI 的 MVAL 多样性样本选择过程。

利用 SLIC 超像素方法对 MVAL 样本选择过程进行多样性改进的步骤为: 首先对整个 HSI 进行 SLIC 超像素分割, 分割后的 HSI 数据如图 2。设共得到 k 个超像素, 每个样本 x_i 均对应一个超像素标签 L_i , 同一超像素的样本的超像素标签相同。然后使用一种样本选择方法 (如 AMD 策略) 选出不确定性最大的 m 个样本, 记为 $X_{AMD} = [x_1 \ x_2 \ \dots \ x_m]$ 。再从这 m 个样本找出属于相同超像素的样本。接下来对其中属于同一个超像素的样本, 随机选出其中一个作为代表, 剩下的样本放回候选集中。令选出的样本必须来自不同的超像素, 从而实现样本选择过程的多样性改进。

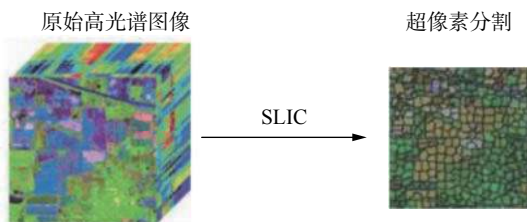


图 2 高光谱图像的 SLIC 超像素分割图

Fig. 2 SLIC superpixel segmentation schematic of hyperspectral images

得到超像素标签的表达式为

$$L_i = L(x_i), \quad L_i \in [1, k]$$

提出的基于超像素的多样性样本选择方法的定义公式为

$$X_{SLIC} = \text{unique}[L(x_1) \ L(x_2) \ \dots \ L(x_m)]$$

式中 X_{SLIC} 为 X_{AMD} 经过提出的基于超像素的多样性样本选择方法后得到的最终所选样本。unique 函数的具体过程如图 3。

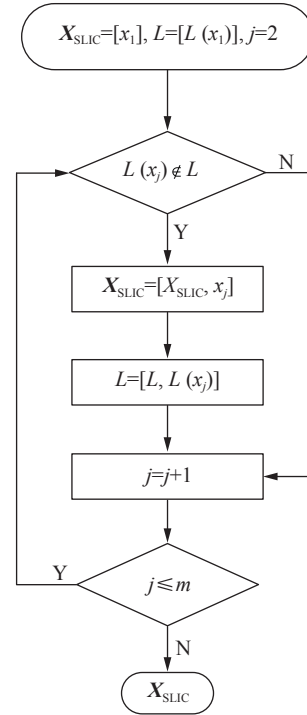


图 3 unique 函数的流程

Fig. 3 Flow of unique function

此方法先对原始 HSI 进行超像素分割, 该过程同时考虑了样本的空间相邻性和光谱相似性。将 HIS 中光谱相似且空间相邻的样本判别为相似样本, 放入一个超像素区域中。通过超像素分割方法得到相似样本可以有效地避免同谱异物问题, 从而更好地进行多样性样本选择。

2.2 基于预测标签一致性的多样性选择策略

在 HSI 分类的 MVAL 方法中, 每个分类器均对样本进行预测。对于两个特别相似的样本, 各分类器对它们的分类结果大概率相同。基于 MVAL 方法特有的多视图特点, 以及每个视图分别训练分类器的特点, 提出一种基于预测标签一致性的去冗余算法。该算法通过比较分类器对不同样本预测结果的一致性, 找出所选样本中的相似样本, 并去掉其中冗余样本。此方法的定义为

$$D = \sum_{k=1}^K |L_k(x_i), L_k(x_j)|_{\text{equal}}$$

$$|a, b|_{\text{equal}} = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases}$$

式中: $x_i, x_j \in U_C$, 且 $x_i \neq x_j$; D 为样本 x_i 和 x_j 的相似程

度, D 越大, x_i 和 x_j 越相似。

基于预测标签一致性的多样性样本选择方法如图4所示。首先使用一种样本选择方法(如AMD策略)选出一批信息量大的样本 $\{X_1, X_2, \dots, X_8\}$, 然后对从中找出预测标签完全相同的样本。从图4中可以看出, 样本 X_1 与样本 X_3 的预测标签完全相同, 均为 $[3, 4, 4, 6, 7]$; 样本 X_5 与样本 X_6 的预测标签完全相同, 均为 $[6, 6, 7, 12, 7]$ 。从样本 X_1 和样本 X_3 中随

机选择一个作为代表加入训练集, 如样本 X_1 , 其余样本放回候选集; 再从样本 X_5 和样本 X_6 中随机选择一个作为代表加入训练集, 如样本 X_5 , 其余样本放回候选集, 最终选择的训练样本为 $\{X_1, X_2, X_4, X_5, X_7, X_8\}$ 。从图4中可以看出, 训练样本数由8个减少为6个, 通过基于预测标签一致性的去冗余方法可以实现去除多余的训练样本的目的, 减少人工标记成本。

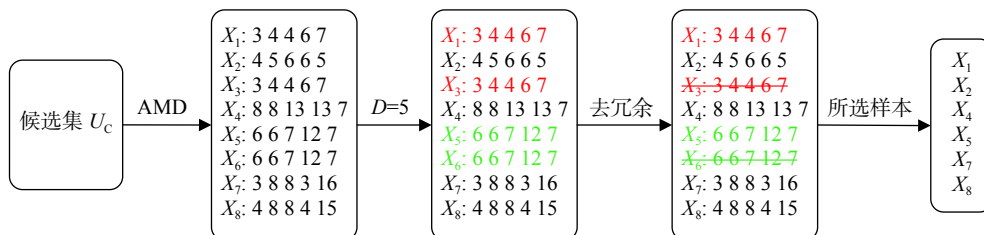


图4 基于预测标签一致性的多样性样本选择方法示意

Fig. 4 Graphical representation of a diversity sampling method based on predictive label consistency

3 实验结果与讨论

3.1 实验数据集

为了验证本文算法的有效性, 采用了2个常用的高光谱图像进行仿真实验, 分别是 Indian Pines 数据集和 Salinas 数据集。Indian Pines 数据集包括16个类别, 共有21025个样本, 其中有真实标签的样本为10249个; Salinas 数据集也包括16个类别, 共有111104个样本, 其中有真实标签的样本为54129个。

3.2 分类性能评价指标

本文使用综合精度(OA)、平均精度(AA)、Kappa系数(Kappa)对两组对比实验结果进行定量比较。

OA方法通过混淆矩阵判别总体分类精度, 具体表达式为

$$OA = \frac{1}{N} \sum_{i=1}^n m_i$$

式中: N 为全部样本数; n 为类别总数; m_i 为将样本正确分类到第 i 类的数量。

AA表示在每个类别中分类正确的样本所占的比重。如果各类别样本数量相同, 则平均分类精度AA与总体分类精度OA相同。AA的表达式为

$$AA = \frac{1}{N} \sum_{i=1}^n CA$$

式中: $CA = \frac{m_i}{N_i}$ 表示各类别的分类精度; N_i 为第 i 类样本总数。

Kappa系数用来表示分类图与真实图像的一

致性。Kappa系数的具体表达式为

$$Kappa = \frac{N \times \sum_{i=1}^n m_i - \sum_{i=1}^n N_i \times m_i}{N^2 - \sum_{i=1}^n N_i \times m_i}$$

3.3 实验结果对比

本节主要验证本文提出的两种多样性样本选择方法的效果, 并将提出的基于超像素分割的多样性选择方法和基于预测标签一致的去冗余算法分别简称为方法A和方法B。采用选取的2组HSI数据进行3组实验。第1组: 使用传统AMD策略进行样本选择, 记为AMD。第2组: 先使用AMD策略进行样本选择, 再将所选样本使用方法A去除冗余, 最后将去冗余后的所选样本加入训练集, 进行MVAL, 记为AMD+A。第3组: 将方法A换成方法B, 其他操作与第2组实验相同, 记为AMD+B。将第2、3组实验结果与第1组实验结果进行对比, 验证这2种方法的有效性。

以上所有实验均使用MLR分类器^[25], 采用3D-Gabor滤波视图生成方法, 视图数量为5。3组数据使用相同频率和方向的3D-Gabor滤波器, 滤波后均得到65个Gabor立方体, 再通过FR准则选出充分性前5的立方体作为MVAL的5个视图。实验前, 分别在Indian Pines数据集和Salinas数据集中有真实标签的样本中选出一部分样本作为初始训练样本。具体方法为: 从每个类别中随机选出5个样本作为初始训练样本, 共80个样本。其余的有真实标签的样本为候选样

本, 测试样本为有真实标签的全部样本。在 AMD 样本选择方法中, 每次迭代最多选择 15 个样本, 共迭代 20 次。在方法 A 中, 超像素的边长为 4×4 , 超像素个数大约为总样本数除以 16, 超像素中光谱和空间的权重因子为 0.5。

为了进一步测试方法的效果, 本文对比了实验中每 5 次迭代对应的分类精度以及各组实验的耗时情况。Indian Pines 数据集的 AMD 和 AMD+A 实验迭代过程中的分类结果如表 1 所示。Sali-

nas 数据集的 AMD 和 AMD+A 实验迭代过程中的分类结果如表 2 所示。

从表 1、2 可以看出: 无论迭代次数是多少, 通过基于超像素分割的多样性选择方法改进后得到的实验精度基本不变, 而训练样本数量均有不同程度的减少。

两个数据集在 AMD 和 AMD+A 实验中得到的最终分类结果如表 3 所示, 分类结果对比图如图 5、6 所示。

表 1 Indian Pines 数据集的 AMD+A 与 AMD 的实验结果

Table 1 Experimental results of AMD+A and AMD in the Indian Pines dataset

序号	实验	迭代次数	OA	Kappa	AA	训练样本数
1	AMD	5	0.8371	0.8157	0.8916	142
	AMD+A		0.8365	0.8149	0.8912	131
2	AMD	10	0.8958	0.8817	0.9245	168
	AMD+A		0.8950	0.8809	0.9248	157
3	AMD	15	0.9463	0.9388	0.9473	214
	AMD+A		0.9467	0.9393	0.9498	199
4	AMD	20	0.9678	0.9633	0.9654	263
	AMD+A		0.9686	0.9643	0.9693	236

表 2 Salinas 数据集的 AMD+A 与 AMD 的实验结果

Table 2 Experimental results of AMD+A and AMD in the Salinas dataset

序号	实验	迭代次数	OA	Kappa	AA	训练样本数
1	AMD	5	0.9382	0.9312	0.9651	155
	AMD+A		0.9382	0.9312	0.9651	154
2	AMD	10	0.9619	0.9576	0.9773	230
	AMD+A		0.9544	0.9493	0.9739	215
3	AMD	15	0.9744	0.9715	0.9837	305
	AMD+A		0.9746	0.9717	0.9852	272
4	AMD	20	0.9818	0.9798	0.9907	380
	AMD+A		0.9795	0.9771	0.9896	315

表 3 AMD+A 与 AMD 实验的最终结果

Table 3 Final results of the AMD+A and AMD experiments

数据集	实验	OA	Kappa	AA	训练样本数	耗时/s
Indian Pines	AMD	0.9678	0.9633	0.9654	263	42.16
	AMD+A	0.9686	0.9643	0.9693	236	45.65
Salinas	AMD	0.9818	0.9798	0.9907	380	313.76
	AMD+A	0.9795	0.9771	0.9896	315	338.68

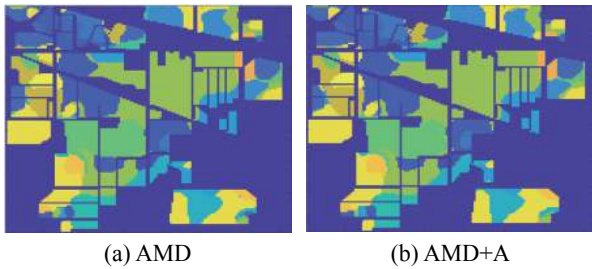


图5 用方法A改进前后的实验结果分析 (Indian Pines)
Fig. 5 Comparison of experimental results before and after improvement with method A (Indian Pines)

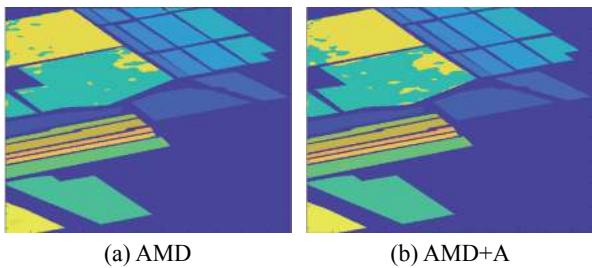


图6 用方法A改进前后的实验结果分析 (Salinas)
Fig. 6 Comparison of experimental results before and after improvement with method A (Salinas)

由AMD和AMD+A的对比实验结果可以看出: 3个HSI数据集在两组实验中的OA、AA、

Kappa值区别不大, 分类结果图也无明显差别, 然而两个数据集在AMD+A实验中用到的训练样本总数比AMD实验中分别减少了10.2%、17.1%, 耗时仅增加了3.49 s、24.92 s。Indian Pines数据集的AMD和AMD+B实验迭代过程中的分类结果如表4所示。Salinas数据集的AMD和AMD+B实验迭代过程中的分类结果如表5所示。

表4 Indian Pines数据集的AMD+B与AMD的实验结果
Table 4 Experimental results of AMD+B and AMD in the Indian Pines dataset

序号	实验	迭代次数	OA	Kappa	AA	训练样本数
1	AMD	5	0.8356	0.8125	0.8807	153
	AMD+B		0.8338	0.8104	0.8737	148
2	AMD	10	0.8944	0.8790	0.9084	179
	AMD+B		0.8939	0.8784	0.9065	172
3	AMD	15	0.9423	0.9342	0.9478	217
	AMD+B		0.9420	0.9339	0.9482	202
4	AMD	20	0.9614	0.9560	0.9638	268
	AMD+B		0.9645	0.9595	0.9628	246

表5 Salinas数据集的AMD+B与AMD的实验结果

Table 5 Experimental results of AMD+B and AMD in the Salinas dataset

序号	实验	迭代次数	OA	Kappa	AA	训练样本数
1	AMD	5	0.8857	0.8734	0.9476	113
	AMD+B		0.8945	0.8831	0.9513	100
2	AMD	10	0.9304	0.9227	0.9530	162
	AMD+B		0.9224	0.9137	0.9478	125
3	AMD	15	0.9497	0.9441	0.9694	182
	AMD+B		0.9449	0.9387	0.9640	138
4	AMD	20	0.9598	0.9553	0.9819	204
	AMD+B		0.9524	0.9470	0.9780	152

从表4、5中可以看出: 无论迭代次数是多少, 通过基于预测标签一致的去冗余算法改进后得到的实验精度基本不变, 而训练样本数量均有

不同程度的减少。两个数据集在AMD和AMD+B实验中得到的最终分类结果如表6所示, 分类结果对比图如图7、8所示。

表6 AMD+B与AMD实验的最终结果

Table 6 Final results of AMD+B and AMD experiments

数据集	实验	OA	Kappa	AA	训练样本数	耗时/s
Indian Pines	AMD	0.9614	0.9560	0.9638	268	41.96
	AMD+B	0.9645	0.9595	0.9628	246	42.84
Salinas	AMD	0.9598	0.9553	0.9819	204	309.34
	AMD+B	0.9524	0.9470	0.9780	152	323.68

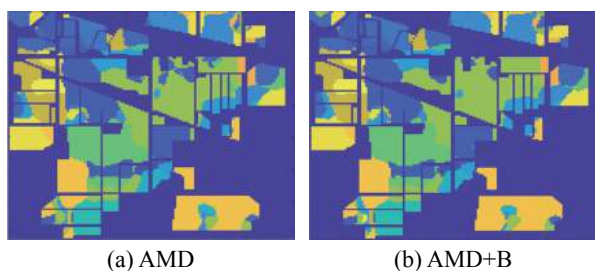


图 7 用方法 B 改进前后的实验结果分析 (Indian Pines)
Fig. 7 Comparison of experimental results before and after improvement with method B (Indian Pines)

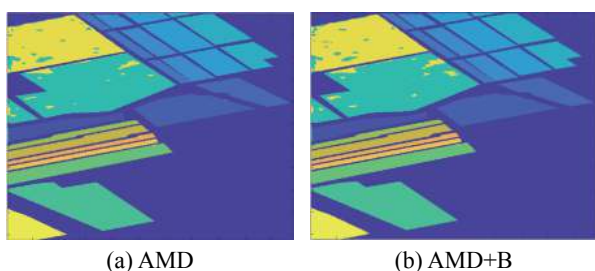


图 8 用方法 B 改进前后的实验结果分析 (Salinas)
Fig. 8 Comparison of experimental results before and after improvement with method B (Salinas)

由 AMD 和 AMD+B 的对比实验结果可以看出: 3 个 HSI 数据集在两组实验中的 OA、AA、Kappa 值区别不大, 分类结果图也无明显差别, 然而两个数据集在 AMD+B 实验中用到的训练样本总数比 AMD 实验中分别减少了 8.2%、25.4%。AMD+B 比 AMD 实验的耗时增加 0.88 s、14.34 s。

通过观察以上实验结果可以看出, AMD+A 方法和 AMD+B 方法相对于传统 AMD 方法具有明显优势。从实验结果来看, 将本文提出的两种样本多样性选择方法用在传统 AMD 样本选择方法后, OA、AA、Kappa 及分类结果图均无明显变化, 训练样本数量均有不同程度的减少。使用这两种改进方法虽然会少量地增加耗时, 但增加的时间成本与节省的人工标记成本相比可以忽略不计。

4 结束语

本文基于 SLIC 超像素分割方法和各视图预测结果的一致性, 提出了 2 种 MVAL 多样性样本选择方法, 有效地解决了传统 MVAL 样本选择过程存在冗余样本的问题。在 2 组 HSI 中进行实验, 验证了这两种方法能够有效地去除传统样本选择过程中的冗余样本, 在分类精度不变的前提下, 减少训练样本总数, 进而减少人工标记成本。

参考文献:

[1] LU Xiaohui, XIA Zhengyan, QU Fangfang, et al. Identification of authenticity, quality and origin of saffron us-

ing hyperspectral imaging and multivariate spectral analysis[J]. Spectroscopy letters, 2020, 53(2): 76–85.

[2] 张晓玲, 张培强, 沈兰荪. 基于信息量失真测度的 VQ 及在高光谱图像无损压缩中的应用 [J]. 遥感学报, 2004, 8(5): 414–418.

ZHANG Xiaoling, ZHANG Peiqiang, SHEN Lansun. A VQ based on information distortion measure and its application to lossless compression of hyperspectral image[J]. Journal of remote sensing, 2004, 8(5): 414–418.

[3] WANG Meng, HUA Xiansheng. Active learning in multimedia annotation and retrieval: a survey[J]. ACM transactions on intelligent systems and technology, 2011, 2(2): 10.

[4] TU Bing, KUANG Wenlan, ZHAO Guangzhe, et al. Hyperspectral image classification by combining local binary pattern and joint sparse representation[J]. International journal of remote sensing, 2019, 40(24): 9484–9500.

[5] YANG Jihai, LI Shijun, XU Wenning. Active learning for visual image classification method based on transfer learning[J]. IEEE access, 2017, 6: 187–198.

[6] LIN Jianzhe, ZHAO Liang, LI Shuying, et al. Active-learning-incorporated deep transfer learning for hyperspectral image classification[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2018, 11(11): 4048–4062.

[7] 赵忠明, 高连如, 陈东, 等. 卫星遥感及图像处理平台发展 [J]. 中国图象图形学报, 2019, 24(12): 2098–2110.

ZHAO Zhongming, GAO Lianru, CHEN Dong, et al. Development of satellite remote sensing and image processing platform[J]. Journal of image and graphics, 2019, 24(12): 2098–2110.

[8] 程圆娥, 周绍光, 袁春琦, 等. 基于主动深度学习的高光谱影像分类 [J]. 计算机工程与应用, 2017, 53(17): 192–196, 248.

CHENG Yuan'e, ZHOU Shaoguang, YUAN Chunqi, et al. Hyperspectral image classification based on active deep learning[J]. Computer engineering and applications, 2017, 53(17): 192–196, 248.

[9] 吴超. 高光谱图像处理若干关键技术研究 [D]. 南京: 南京航空航天大学, 2012.

WU Chao. Research on some key technologies of hyperspectral image processing[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012.

[10] JAMSHIDPOUR N, SAFARI A, HOMAYOUNI S. Multiview active learning optimization based on genetic algorithm and Gaussian mixture models for hyperspectral data[J]. IEEE geoscience and remote sensing letters, 2020, 17(1): 172–176.

[11] DI Wei, CRAWFORD M M. View generation for multiview maximum disagreement based active learning for

- hyperspectral image classification[J]. IEEE transactions on geoscience and remote sensing, 2012, 50(5): 1942–1954.
- [12] DI Wei, CRAWFORD M M. Multi-view adaptive disagreement based active learning for hyperspectral image classification[C]//2010 IEEE International Geoscience and Remote Sensing Symposium. Honolulu, USA, 2010: 1374–1377.
- [13] 甘继生. 基于广义组合核的高光谱图像 GPU 并行分类 [D]. 南京: 南京理工大学, 2018.
GAN Jisheng. Hyperspectral Image GPU parallel classification based on generalized composite cores[D]. Nanjing: Nanjing University of Science and Technology, 2012.
- [14] HU Jie, HE Zhi, LI Jun, et al. 3D-gabor inspired multiview active learning for spectral-spatial hyperspectral image classification[J]. Remote sensing, 2018, 10(7): 1–24.
- [15] XU Xiang, LI Jun, LI Shutao. Multiview intensity-based active learning for hyperspectral image classification [J]. IEEE transactions on geoscience and remote sensing, 2018, 56(2): 669–680.
- [16] 魏飞鸣, 李小文, 顾行发, 等. 基于形状参数的遥感图像“同谱异物”目标区分 [C]//第十四届全国图象图形学术会议论文集. 福州, 中国, 2008: 477–481.
WEI Feiming, LI Xiaowen, GU Xingfa, et al. Shape-based classification of “Spectral Similar” objects in remote sensing image processing[C]//Proceedings of the 14th National Conference on Image Graphics. Fuzhou, China, 2008: 477–481.
- [17] 王立国, 魏芳洁. 结合 APO 算法的高光谱图像波段选择 [J]. 哈尔滨工业大学学报, 2013, 45(9): 100–106.
WANG Liguang, WEI Fangjie. Artificial physics optimization algorithm combined band selection for hyperspectral imagery[J]. Journal of Harbin Institute of Technology, 2013, 45(9): 100–106.
- [18] 王阳. 基于超像素的图像分割算法研究 [D]. 兰州: 兰州理工大学, 2018: 1–57.
WANG Yang. Research on image segmentation algorithm based on superpixel[D]. Lanzhou: Lanzhou University of Technology, 2018: 1–57.
- [19] LIU Peng, HUANG Jiwei, ZHANG Shiwu, et al. Multiview hyperspectral topography of tissue structural and functional characteristics[J]. Journal of biomedical optics, 2016, 21(1): 1–12.
- [20] SUN Shiliang. A survey of multi-view machine learning[J]. Neural computing and applications, 2013, 23(7/8): 2031–2038.
- [21] ZHAO Jing, XIE Xijiong, XU Xin, et al. Multi-view learning overview: recent progress and new challenges [J]. Information fusion, 2017, 38: 43–54.
- [22] HE Lin, LI Jun, PLAZA A, et al. Discriminative low-rank Gabor filtering for spectral-spatial hyperspectral image classification[J]. IEEE transactions on geoscience and remote sensing, 2017, 55(3): 1381–1395.
- [23] 段无悔. 基于超像素的高光谱图像分类算法研究 [D]. 长沙: 湖南大学, 2015: 1–64.
DUAN Wuhui. Researches on hyperspectral image classification algorithms based on superpixel[D]. Changsha: Hunan University, 2015: 1–64.
- [24] PATRA S, BHARDWAJ K, BRUZZONE L. A spectral-spatial multicriteria active learning technique for hyperspectral image classification[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2017, 10(12): 5213–5227.
- [25] BIOUCASDIAS B J, FIGUEIREDO M. Logistic regression via variable splitting and augmented lagrangian tools[R]. Instituto Superior Técnico, TULisbon, 2009.

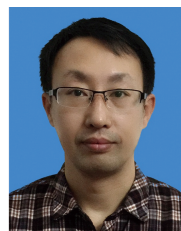
作者简介:



陈立伟, 副教授, 博士, 中国图像图形学会会员, 黑龙江省生物医学工程学会会员, 主要研究方向为人工智能、深度学习、图像的分割、特征提取、分类识别以及理解。主持省市级科研项目 3 项, 参与国家级项目 5 项。获授权发明专利 5 项, 发表学术论文 40 余篇。



房赫, 硕士研究生, 主要研究方向为高光谱图像分类技术。



朱海峰, 讲师, 主要研究方向为人工智能、高光谱图像处理。