

文章编号: 1001-3806(2022)06-0808-09

低秩稀疏和改进 SAM 的高光谱图像误标签检测

刘 焯¹, 渠慎明^{1,2*}

(1. 河南大学 软件学院, 开封 475001; 2. 河南大学 智能网络系统研究所, 开封 475001)

摘要: 为了解决基于监督学习的高光谱图像分类算法训练样本中存在的噪声标签会降低后续的分类精度的问题, 采用了一种基于低秩稀疏表示和改进光谱角制图(SAM)的高光谱图像误标签检测算法。首先对高光谱图像中信号子空间进行预测, 根据预测到的子空间对原始高光谱图像重构并去噪; 然后通过基于归一化的光谱角制图算法来获取每一类样本间的距离信息, 得到每类样本间的光谱相似度, 并利用密度峰值聚类算法得到每个训练样本的局部密度; 最后采用基于局部密度的决策函数对噪声标签进行检测, 使用支持向量机在两个真实数据集上验证。结果表明, 该算法比先进的层次结构的高光谱图像误标签检测算法提高了 1.91% 的总体精度。这一结果对高光谱图像分类是有帮助的。

关键词: 图像处理; 低秩稀疏表示; 归一化光谱角制图; 密度峰值聚类算法; 噪声标签检测

中图分类号: TP751 **文献标志码:** A **doi:** 10.7510/jgjs.issn.1001-3806.2022.06.016

False label detection in hyperspectral image based on low rank sparse and improved SAM

LIU Xuan¹, QU Shenming^{1,2}

(1. School of Software, Henan University, Kaifeng 475001, China; 2. Institute of Intelligence Networks System, Henan University, Kaifeng 475001, China)

Abstract: In order to solve the problem that reduction of the subsequent classification accuracy in the hyperspectral image classification algorithm based on supervised learning due to the presence of noise labels in the training samples, a false label detection algorithm based on low rank sparse representation and improved spectral angle mapping (SAM) was adopted. Firstly, the signal subspace of hyperspectral image was predicted, and the original hyperspectral image was reconstructed and denoised according to the predicted subspace. Next, the normalized spectral angle mapping algorithm was used to obtain the distance information between each class of samples, and the spectral similarity between each class of samples was obtained. Then, the density peak clustering algorithm was used to get the local density of each training sample. Support vector machine was used to verify the results on two real datasets. The experimental results show that the overall accuracy is improved by 1.91% compared with the advanced hierarchical structure of hyperspectral image false label detection algorithm. This result is helpful for hyperspectral image classification.

Key words: image processing; low rank sparse representation; normalized spectral angle mapping; density peak clustering algorithm; noise label detection

引 言

高光谱图像是由成像光谱仪接收的数十上百个波段所反射回来的地物的光谱特性组成。高光谱图像由

基金项目: 河南省科技发展计划资助项目(212102210538); 河南大学研究生教育创新与质量提升计划资助项目(SYL20040121)

作者简介: 刘 焯(1994-), 男, 硕士研究生, 主要研究方向为高光谱图像处理。

* 通讯联系人。E-mail: qsm@vip.henu.edu.cn

收稿日期: 2021-07-16; 收到修改稿日期: 2022-03-10

两个空间维和一个光谱维构成, 光谱维中的光谱向量代表了高光谱图像中相应像素独特的光谱特征。由于光谱特征在特征识别方面的优势, 目前高光谱图像处理技术已经被广泛应用到各种场景中^[1-2], 例如精准农业^[3]、海洋监测^[4]以及城乡规划^[5]等。在这些应用场景中, 高光谱图像分类起到了重要作用。近年来, 一些空谱联合分类算法被用来提升分类精度^[6-8]。这些方法用于学习训练样本标签是可行的, 然而在实际应用中并非如此。

有监督的高光谱图像分类算法要求样本是标记完成的, 但是手动标记过程非常困难, 仅凭视觉解释的训

练样本并不可靠。具体来说,引入误标签的原因有如下几点:(1)全球定位系统会对目标对象的空间位置产生不准确的估计,导致很难确定高光谱像素的精确位置;(2)对于一些场景,比如海洋和湿地,这样的场景人类无法到达,在这种情况下,基于人类视觉解读的训练样本标签不可避免会产生噪声;(3)当标记一个包含许多不规则形状土地覆盖物的场景时,人工贴标签的过程中会产生错误。

为了解决训练样本的误标签问题,对计算机视觉领域进行了深入的研究。LU 等人^[9]提出一种基于曼哈顿距离优化的学习模型来检测弱噪声标签。FOODY 等人^[10]发现,噪声标签会影响基于支持向量机的机载制图分类。虽然许多研究已经解决了计算机视觉领域的噪声标签问题,但由于高光谱图像的高维和非线性结构,这些方法不能直接扩展到高光谱图像误标签分类中。最近几年,关于带有噪声标签的高光谱图像分类算法得到了关注。KANG 等人^[11]首次提出了基于光谱检测和边缘保持滤波的噪声标签检测和校正方法。TU 等人^[12]通过融合光谱角度和局部离群值因子来检测高光谱图像中的噪声标签,实验结果表明,该算法能有效地检测出有噪声的标签。密度峰值(density peak, DP)聚类算法作为一种鲁棒的聚类算法首次在科学杂志上被提出^[13]。TU 等人^[14]首次利用 DP 聚类算法来检测高光谱图像训练样本中的误标签,基于 DP 聚类的高光谱图像误标签检测算法在检测过程中没有考虑相邻光谱像素之间的空间相关性。为了解决这一问题,TU 等人^[15]提出一种新的基于空间 DP 聚类(k -spatial density peak, K-SDP)的噪声标签检测算法,该算法通过加入中心样本的邻域样本来进一步检测中心样本的异常程度。然而,参考文献[14]和参考文献[15]中没有考虑原始高光谱图像中存在稀疏噪声的问题。参考文献[16]中提出一种基于核熵分量分析(kernel entropy component analysis, KECA)的噪声标签检测方法,但是,该算法在检测过程中没有考虑到训练样本的上下文信息。多种基于约束能量最小化(constrained energy minimum, CEM)算法已被广泛应用于高光谱图像处理中。ZOU 等人^[17]提出一种用于高光谱图像目标检测的二次约束能量最小化检测器。此外,ZHANG 等人^[18]提出一种混合稀疏性和 CEM 的检测器,以提高目标检测的性能。CEM 也有效地应用到了高光谱图像误标签检测上。TU 等人^[19]提出了一种层次约束能量最小值(hierarchical constrained energy minimum, HCEM)方法来检测经过监督任务训练的原

始训练集的错误标记样本,该方法可以准确地去除原始训练集的噪声标签,有效地提高监督分类任务的性能。但是,该算法的一个缺点是使用原始的光谱角制图算法(spectral angle mapping, SAM)来衡量光谱向量的相似度。原始的 SAM 是一种全局性的描述指标,当部分波段属性值有变化、或全部波段属性值具有不同的变化值时,往往导致光谱角余弦的失真。

为了解决参考文献[14]~参考文献[16]和参考文献[19]中所出现的问题,本文作者提出基于低秩稀疏和改进光谱角制图的密度峰值聚类算法(low rank sparse-normalized spectral angular mapping density peak clustering, LRS-NSAMDP)。相比于 DP 聚类算法^[14]和 K-SDP^[15]算法,本算法的改进是去除原始高光谱图像中的稀疏噪声,提取高光谱图像中的低秩成分,降低每一类样本中的加权平均局部密度,从而减少了光谱向量中的误标签数目,提高了分类精度。相比于基于层次约束能量最小值的高光谱图像误标签分类算法^[19],本算法对原始的 SAM 算法进行改进,将光谱向量在波段上的属性值除以该光谱向量的模进行归一化,相比于 SAM 算法降低了同类像元之间的光谱角,使同类像元更加接近,从而更容易检测出训练样本中的像元之间差异较大的误标签。通过以上两个改进,相比于其它先进的遥感图像误标签分类算法,提升了总体精度(overall accuracy, OA)、平均精度(average accuracy, AA)和 kappa 系数。

1 相关技术

1.1 信号子空间估计

一幅原始高光谱图像 $\mathbf{Y} = [y_1, y_2, \dots, y_Q]$, Q 代表每一波段的像素数。由于高光谱图像相邻波段之间的高相关性,根据线性回归理论和最小二乘法理论^[20],假设 \mathbf{z}_i 为传感器在第 i 波段读取的相关系数向量,所以有:

$$\mathbf{z}_i = \mathbf{Z}_{\hat{\theta}_i} \boldsymbol{\beta}_i + \boldsymbol{\xi}_i \quad (1)$$

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{R}_{\hat{\theta}_i, \hat{\theta}_i} - \mathbf{R}_{\hat{\theta}_i, i} \mathbf{R}_{i, \hat{\theta}_i}' / \mathbf{R}_{i, i}) \hat{\mathbf{R}}_{\hat{\theta}_i, i} \quad (2)$$

式中, $\mathbf{Z}_{\hat{\theta}_i}$ 表示去除第 i 波段后的相关系数矩阵, $\boldsymbol{\beta}_i$ 表示第 i 波段的回归向量, $\boldsymbol{\xi}_i$ 代表稀疏噪声向量, $\hat{\mathbf{R}} = (\mathbf{Y}\mathbf{Y}^T)$ 为自相关矩阵, \mathbf{R}' 为自相关矩阵 $\hat{\mathbf{R}}$ 的逆, $\mathbf{R}_{\hat{\theta}_i, \hat{\theta}_i}'$ 表示删除第 i 行和第 i 列的自相关矩阵的逆;上标 $\hat{\cdot}$ 表示估计值。

接下来根据预测到的稀疏噪声向量 $\hat{\xi}_i$ 来估计信号子空间。计算原始高光谱图像 Y 自相关矩阵 \hat{R}_Y ,同理计算稀疏噪声向量 ξ_i 的自相关矩阵 \hat{R}_{ξ} ,计算真实信号原始光谱向量 x 的自相关矩阵 \hat{R}_x ,并计算其特征向量 E ,将高光谱图像所在的空间分解为 k 维子空间 E_k 和 E_γ , E_γ 表示子空间 E_k 的正交补空间。设 U_k 为子空间 E_k 的投影矩阵, U_γ 为子空间 E_γ 的投影矩阵, $\hat{x}_k = U_k y$,为观测到的光谱向量 y 在子空间 E_k 上的投影,在此将 \hat{x}_k 和原始光谱向量 x 之间的最小均方误差作为估计子空间的准则^[21],可得:

$$K = (\hat{k}, \hat{\pi}) = \underset{k, \pi}{\operatorname{argmin}} \{ \operatorname{tr}(U_R \hat{R}_Y) + 2\operatorname{tr}(U_k \hat{R}_{\xi}) \} \quad (3)$$

式中, K 为设置项, \hat{k} 为 $\delta = \operatorname{tr}(U_R \hat{R}_Y) + 2\operatorname{tr}(U_k \hat{R}_{\xi})$ 取负值的个数, $\hat{\pi}$ 为置换项, π 为波段 $1 \sim i$ 的矩阵排列。将 δ 取负值时所对应的特征向量 E 中的子集作为预测到的子空间 \hat{S} ,具体来说,子空间 \hat{S} 可以根据 \hat{k} 和 $\hat{\pi}$ 所对应的特征向量进行检索得到^[21]。

1.2 光谱角制图

SAM是KRUSE等人在1993年提出的^[22],把图像中的每一个像元的光谱视为一个高维向量,通过计算两向量之间的夹角来度量光谱间的相似性,夹角越小,两光谱越相似,属于同类地物的可能性越大,因而可根据光谱角的大小来辨别未知数据的类别。分类时,通过计算未知数据与已知数据间的光谱角,并把未知数据的类别归为最小光谱角对应的类别中,如下式所示:

$$\cos \alpha = \frac{A \cdot B}{|A| |B|} = \frac{\sum_{i=1}^L A_i B_i}{\sqrt{\sum_{i=1}^N A_i A_i} \sqrt{\sum_{i=1}^N B_i B_i}} \quad (4)$$

式中, L 为波段数, A 和 B 分别表示两个光谱向量在 L 个波段上的属性值, α 为光谱角。夹角越小,余弦值较大;相反夹角大,相应的余弦值就较小。

2 基本低秩稀疏和改进SAM的新方法

图1是所提出的LRS-NSAMDP的流程图。主要分为5个步骤:(1)基于低秩稀疏表示的高光谱图像低秩特征提取;(2)计算各个类中训练样本间的距离;(3)训练样本局部密度的计算;(4)检测误标签训练样本;(5)支持向量机分类。

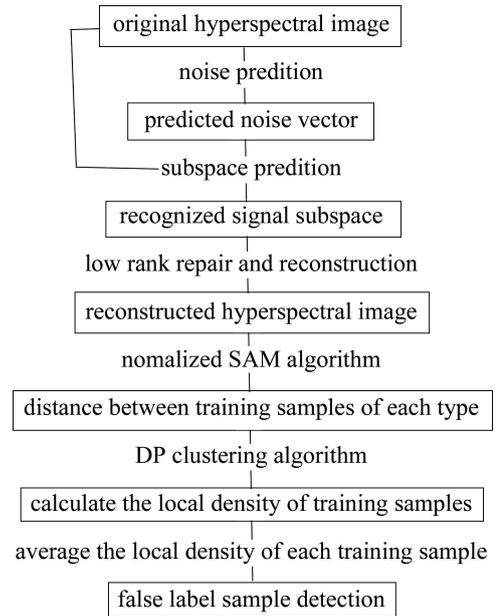


Fig. 1 Flow chart of LRS-NSAMDP algorithm

2.1 基于低秩稀疏表示的高光谱图像低秩特征提取

根据第1.1节中得到原始高光谱图像 Y 的信号子空间 \hat{S} ,该问题可表述为:

$$z = \min_z \frac{1}{2} \| M(I \otimes \hat{S})z - y \|^2 + \delta \phi(z) \quad (5)$$

式中, \otimes 表示克罗内克乘积, $\| \cdot \|$ 表示矩阵的范数, I 表示单位矩阵, Z 表示相关系数矩阵, z 为向量化 Z , M 为掩模, $\phi(z)$ 表示正则化函数, δ 是正则化系数, y 为向量化图像 Y 。设掩模 M_p 作用在未观测到的像素 p 所对应的高光谱图像 y_p ,所以有:

$$y_p = M_p \hat{S} z_p + n_p \quad (6)$$

式中, n_p 表示噪声部分,根据参考文献[21]可得 \hat{z}_p ,计算 $\hat{y}_p = \hat{S} \hat{z}_p$ 得到完全观测到的高光谱图像 \hat{Y} 。在恢复未观测到的分量后,需要去除原始噪声。所以有:

$$Z = \min_Z \frac{1}{2} \| \hat{S} Z - \hat{Y} \|^2 + \delta \phi(Z) \quad (7)$$

对于(7)式的优化问题,在此利用3维块匹配滤波算法进行去噪,得到相关系数矩阵 Z ,所以重构的高光谱图像 X 可由 $\hat{X} = \hat{S} Z$ 得到。

2.2 计算各类中训练样本间的距离

根据得到的重构后的高光谱图像 \hat{X} 的训练样本 \hat{T} , m 表示训练样本 \hat{T} 中类的个数, \hat{T}_j 表示第 j 类中的

n 个训练样本,在此提出基于归一化的光谱角计算方法来度量光谱相似度。

$$\begin{cases} \hat{\mathbf{g}}_a = \frac{\hat{\mathbf{T}}_{ja}}{\sum_{a=1}^n \sqrt{(\hat{\mathbf{T}}_{ja})^2}} \\ \hat{\mathbf{g}}_b = \frac{\hat{\mathbf{T}}_{jb}}{\sum_{b=1}^n \sqrt{(\hat{\mathbf{T}}_{jb})^2}} \end{cases} \quad (8)$$

$$d_{ab} = \arccos \left[\frac{\hat{\mathbf{g}}_a^T \hat{\mathbf{g}}_b}{\|\hat{\mathbf{g}}_a\| \|\hat{\mathbf{g}}_b\|} \right] \quad (9)$$

式中, $\hat{\mathbf{g}}_a$ 和 $\hat{\mathbf{g}}_b$ 分别表示归一化后的第 a 个和第 b 个样本的光谱向量, d_{ab} 表示两样本之间的距离。通过计算第 j 类中第 l 个像素和类中其它像素之间的距离,第 l 个像素的距离数组可以构造如下:

$$\mathbf{d}_l = [d_{l,1}, d_{l,2}, \dots, d_{l,l_j}]^T \quad (10)$$

式中, l_j 代表第 j 类中的所有像素数,每一类样本的距离矩阵 $\mathbf{D}_j = \{d_1, d_2, \dots, d_j\}$ 。

2.3 计算训练样本的局部密度

为了计算训练样本间的局部密度,定义截止距离 d_c , 计算方式如下:

$$d_c = S(t), \left(t = \left\langle \frac{N_j \cdot (N_j - 1)}{100} \cdot \theta \right\rangle \right) \quad (11)$$

式中, $S(t)$ 为将 \mathbf{D}_j 的上三角矩阵中的非零元素从最小到最大排序得到的矩阵, N_j 为第 j 重样本总数, θ 为随机参数, $\langle \cdot \rangle$ 为四舍五入运算。根据得到的 d_c 矩阵计算每一类的局部密度 $\rho = \sum \exp[-(\mathbf{D}_j/d_c)^2]$ 。

2.4 检测误标签训练样本

根据每一类中每一个训练样本的局部密度,误样本可以通过线性阈值决策函数计算得到:

$$\mathbf{W} = \begin{cases} \mathbf{H}_{ja}, (\rho_{ja} \geq \lambda \bar{\rho}_j) \\ \Phi, (\text{otherwise}) \end{cases} \quad (12)$$

式中, \mathbf{W} 为检测完成后并删除误样本后的训练集, \mathbf{H}_{ja} 为检测为正确标签的第 j 类中的第 a 个样本, ρ_{ja} 为第 j 类第 a 个样本的局部密度, $\bar{\rho}_j$ 为第 j 类样本的平均局部密度, λ 为随机参数,。

3 结果与分析

3.1 实验数据集

Kennedy Space Center (KSC) 数据集是由 AVIRIS

高光谱仪于 1996 年在佛罗里达州肯尼迪太空中心采集的 512 像素×614 像素大小的高光谱图像,包含 224 个波段,经过噪声去除后还剩下 176 个波段,空间分辨率是 18m,有 13 个地物类别,总样本大小为 5211。

University of Pavia (PaviaU) 数据集是由 ROSIS 高光谱仪在帕维亚大学上空采集的 610 像素×340 像素大小的高光谱图像,共包含 9 类地物 115 个波段,去掉含噪声波段后,其余 103 个波段作为实验数据集,总体样本大小为 42776。本算法的实验以及对比算法实验的运行环境为 12G 内存,英特尔酷睿 i5 2.2GHz 的 CPU,开发环境为 MATLAB R2018a。图 2 和图 3 分别给出了两种数据集的假彩色图像、地物真值图和每一类物种。

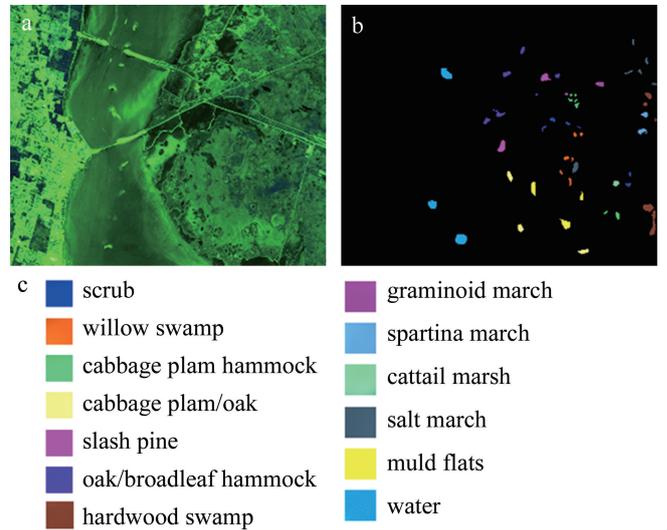


Fig. 2 KSC dataset

a—false color image b—ground object truth map c—name of each species

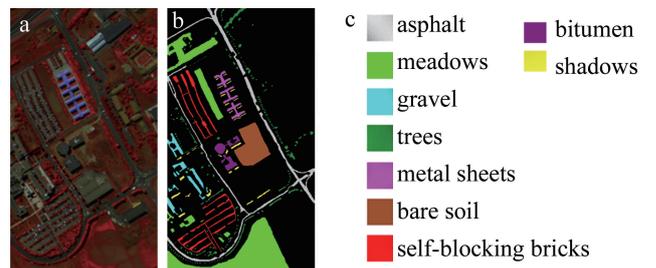


Fig. 3 PaviaU dataset

a—false color image b—ground object truth map c—name of each species

3.2 实验参数分析

本算法提出的两个参数分别为:计算局部密度的随机参数 θ 和检测误标签训练样本的随机参数 λ ,图 4 和图 5 中分别展示了在 KSC 和 PaviaU 两个数据集上的不同参数下对 OA 的影响。为了证明算法的有效性,后续实验使用广泛应用在高光谱图像分类算法中的支持向量机 (support vector machine, SVM) 作为分类

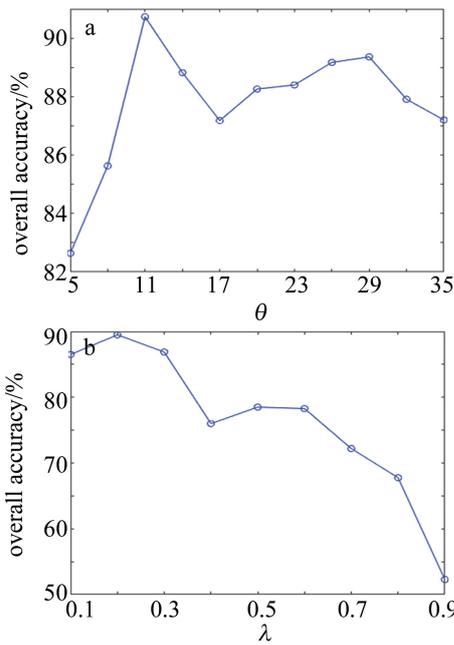


Fig. 4 On KSC dataset, the parameter θ and λ coefficient of different local densities λ impact on OA

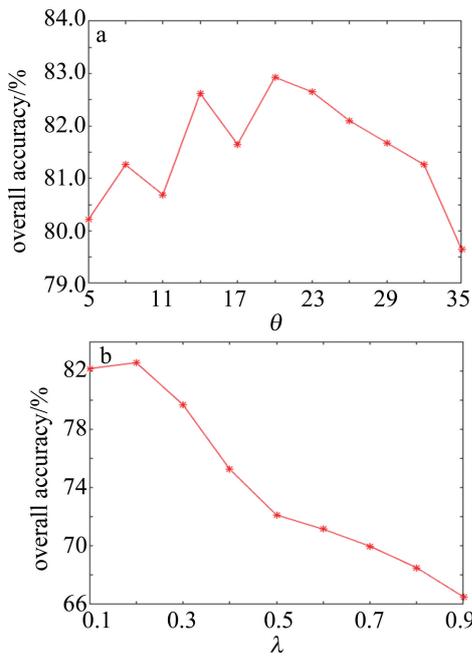


Fig. 5 On PaviaU dataset, the parameter θ and λ coefficient of different local densities λ impact on OA

器,具体使用 LIBSVM 工具箱中的分类器,SVM 的参数采用交叉验证的方式来确定。对于 KSC 数据集,对每一类随机选取 25 个真实样本和 5 个不确定标记样本,对于 PaviaU 数据集,对每一类随机选取 50 个真实样本和 10 个不确定标记样本。

根据图 4 和图 5 可以看出,随机参数 θ 在两个数据集上的波动范围相较于系数 λ 较小,比如在 PaviaU 数据集上,最大的局部密度只比最小的局部密度高 2%左右,而在 KSC 数据集上最大的局部密度比最小的局部密度要高 15%以上,因此决定本算法的精度值主要是随机参数 θ 。从图 4 和图 5 还可以看出,在 KSC 数据集上,当取 $\theta=11, \lambda=0.2$ 时,可以取得局部最优的 OA 值;PaviaU 数据集上,当取 $\theta=20, \lambda=0.2$ 时,可以取得局部最优的 OA 值。因此对于一个新的数据集,建议取 $\theta=20, \lambda=0.2$ 作为提出算法的参数设置。

3.3 算法创新性实验验证

为了验证本算法的有效性,在此计算成功检测出的误样本数做为判别标准。具体验证方式为:根据 DP 聚类算法计算每一类样本的局部密度,得到聚类中心标签集合,根据(12)式统计 $\rho_{ia} < \lambda \bar{\rho}_j$ 的数目 r_1, r_1 越小表示误标签越少,从而判别为正确标签的训练样本数目就越多,后续的分类精度就越高。

表 1 中,T(true)表示真实标签,U(uncertain)表示不确定标签。从表中可以看出,DP 聚类算法^[14]检测到的误标签个数要大于 K-SDP 算法^[15]和 LRS-NSAM-DP 算法,删除掉误标签样本后的正确训练样本 $W = \{W_1, W_2, \dots, W_m\}$ 越少,后续的分类精度就越低。K-SDP^[15]算法通过加入中心样本的邻域样本来进一步检测中心样本的异常程度,相比于 DP 聚类算法^[14]考虑了相邻光谱像素之间的空间相关性,提升了后续分类精度。LRS-NSAMDP 算法提取了原始高光谱图像的低秩成分,去除了原始图像中的稀疏噪声,使用基于归一化的光谱角制图算法来计算光谱相似度,减少了类样本之间像素的距离,增大了截止距离,根据(12)式推导可知,降低了阈值 $\lambda \bar{\rho}_j$,从而 r_1 的值会变小,误

Table 1 Umber of false labels in each class detected by different detection algorithms under different uncertain samples

		number of true labels and uncertain labels in each training sample				
		25T+3U	25T+5U	25T+7U	25T+9U	25+11U
		number of false labels detected in each class				
different false label detection algorithms	DP ^[14]	1.3	0.6	2	2	1
	K-SDP ^[15]	1	1	1	1	1
	LRS-NSAMDP	0	0.6	0.3	1.3	0.6

标签数目减少,相比于其它两种误标签检测算法,判别为正确标签的训练样本数目就多,后续的分类精度就越高。

为了验证改进光谱角制图算法的优越性,在此将多种距离度量算法应用在本算法当中,比如,欧几里得距离 (Euclidean distance, ED)^[23]、光谱信息散度 (spectral information divergence, SID)^[24]、相关系数 (correlation coefficient, CC)^[25] 以及 SAM^[22]。根据表 2 可以明显看出,本文中提出的归一化光谱角相似度算法取得了较优的分类精度,因此后续的对比如实验采用本文中提出的归一化光谱角相似度算法来度量距离。

Table 2 Classification performance of KSC dataset under the false labels detected by different distance measurement algorithms

classification accuracy	ED ^[23]	SID ^[24]	CC ^[25]	SAM ^[22]	LRS-NSAMDP
OA/%	88.70	87.63	89.17	89.46	89.74
AA/%	83.95	83.05	85.21	84.65	85.38
kappa	0.8739	0.8620	0.8790	0.8824	0.8854

3.4 不同的误标签检测算法性能比较

将本算法和先进的误标签检测算法进行对比,具体包括 SVM 算法^[26]、DP 聚类算法^[14]、K-SDP 算法^[15]、KECA 算法^[16]和 HCEM 算法^[19]。本算法的实验参数采用第 3.2 节中给出的参数,为了保持对比算法在最优的条件下进行对比,所有参数采用文献中给出的默认参数。在 KSC 数据集上,实验采用每一类 25 个正确样本加 5 个不确定样本、25 个正确样本加 15 个不确定样本。在 PaviaU 数据集上,实验采用每一类 50 个正确样本加 10 个不确定样本、50 个正确样本加

20 个不确定样本。限于篇幅,图 6 和图 7 中分别展示了在 KSC 数据集上 25 个正确样本加 5 个不确定样本和 PaviaU 数据集上 50 个正确样本加 10 个不确定样本下的不同误标签检测算法随机一次地物分类图。表 3 和表 4 中分别展示了不同误标签检测算法在 KSC 和

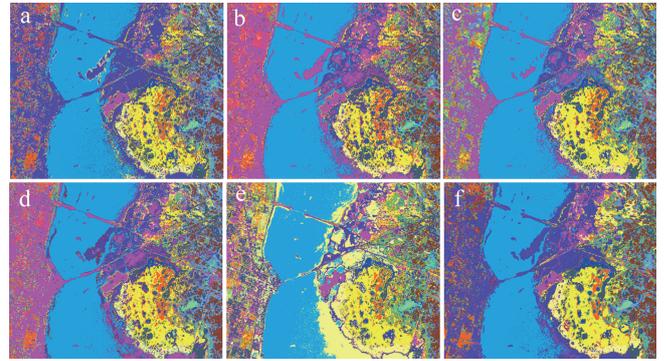


Fig. 6 Feature classification map (25T+5U) obtained by different algorithms in KSC dataset
 a—SVM, OA: 85.20% b—DP, OA: 87.04% c—K-SDP, OA: 86.41% d—KECA, OA: 86.72% e—HCEM, OA: 87.90% f—LRS-NSAMDP, OA: 88.51%



Fig. 7 Feature classification map (50T+10U) obtained by different algorithms in PaviaU dataset
 a—SVM, OA: 75.63% b—DP, OA: 79.01% c—K-SDP, OA: 80.44% d—KECA, OA: 81.43% e—HCEM, OA: 82.72% f—LRS-NSAMDP, OA: 83.28%

Table 3 Classification accuracy under different false label algorithms on KSC dataset

class	number of true samples and uncertain samples											
	25T+5U						25T+15U					
	SVM	DP	K-SDP	KECA	HCEM	LRS-NSAMDP	SVM	DP	K-SDP	KECA	HCEM	LRS-NSAMDP
scrub	93.53	95.30	95.04	93.29	92.87	96.44	96.11	95.73	95.30	93.38	87.79	96.83
will-S	79.86	83.77	81.30	85.60	89.90	88.49	77.02	74.53	76.62	79.93	86.39	88.27
cabb-H	91.84	84.55	83.43	82.76	83.67	78.18	86.19	84.34	85.04	81.06	84.16	70.89
cabb-O	58.72	64.64	64.31	62.60	61.99	65.89	56.02	57.00	56.64	60.02	54.48	57.89
slash-P	51.88	61.92	62.86	62.45	63.28	58.04	57.14	57.82	57.34	56.72	75.00	51.57
broad	51.72	57.08	59.27	55.13	54.59	75.09	36.10	48.74	53.65	57.81	51.60	67.20
hardwood	56.83	70.00	64.99	66.96	65.14	84.38	53.19	65.48	66.03	60.36	64.65	76.47
graminoid	82.58	78.09	77.49	80.34	78.24	89.57	52.58	69.17	73.83	80.82	89.49	77.63
spartina	89.27	84.96	86.89	89.07	92.67	92.52	88.55	82.63	87.10	86.75	87.52	88.94
cattail	80.85	92.55	88.93	94.48	100.00	90.96	83.21	87.27	88.86	94.13	91.72	89.93
salt	86.72	87.33	91.52	92.60	94.29	94.89	97.78	90.71	90.95	90.83	90.59	95.49
muld	83.62	90.24	94.40	93.48	84.92	91.45	82.76	87.10	85.11	92.81	84.88	93.05
water	99.18	96.62	99.50	99.60	100.00	98.79	100.00	98.65	99.0	98.07	96.33	98.87
OA/%	83.57	85.47	86.19	86.67	87.52	89.43	79.90	82.60	84.97	85.19	85.23	86.49
AA/%	77.43	80.54	80.76	81.42	81.66	84.98	74.36	76.86	79.35	79.44	80.35	81.00
kappa	0.8171	0.8381	0.8460	0.8514	0.8607	0.8821	0.7766	0.8062	0.8325	0.8349	0.8345	0.8493

Table 4 Classification accuracy of PaviaU dataset with different false label algorithms

class	number of true samples and uncertain samples											
	50T+10U						50T+20U					
	SVM	DP	K-SDP	KECA	HCEM	LRS-NSAMDP	SVM	DP	K-SDP	KECA	HCEM	LRS-NSAMDP
asphalt	89.43	95.72	97.09	93.83	80.57	96.25	88.69	89.63	92.52	95.94	95.82	96.51
meadows	93.05	93.80	94.03	94.66	95.72	97.43	92.46	92.64	93.69	92.78	93.71	92.91
gravel	57.31	58.11	60.86	60.75	57.14	61.38	56.47	56.90	57.10	65.44	58.16	67.37
trees	71.22	75.79	80.51	73.04	88.67	86.84	67.07	70.52	72.58	81.67	78.59	89.60
M-sheets	85.51	94.23	95.41	86.01	99.10	98.35	84.50	88.67	89.40	81.22	86.05	81.48
B-soil	57.34	56.00	59.96	66.78	44.17	47.28	48.12	49.53	56.25	50.86	60.03	60.64
bitumen	48.44	43.99	52.73	51.71	56.16	58.94	46.09	40.77	48.89	67.86	70.43	68.52
self-Bricks	77.10	74.95	80.00	75.89	81.17	83.28	71.47	73.33	76.47	59.66	56.37	66.26
shadows	83.26	74.66	95.89	88.34	83.01	93.82	87.30	77.30	92.22	70.53	69.26	70.19
OA/%	76.73	78.91	79.55	81.11	82.25	83.49	72.46	75.27	76.32	76.53	78.12	81.42
AA/%	73.84	73.93	79.50	76.77	76.19	80.40	71.96	73.70	75.93	74.00	74.27	77.05
kappa	0.7135	0.7187	0.7542	0.7577	0.7601	0.7783	0.6649	0.6772	0.6973	0.6982	0.7158	0.7551

PaviaU 数据集上随机运行 10 次后求平均值的分类精度表格。

根据图 6 可知,本文中提出的 LRS-NSAMDP 算法和真实地物分类图更相似,证明了相比于 SVM、DP、K-SDP、KECA 和 HCEM 算法,本文中提出的算法可以有效地去除带有误标签的训练样本。由表 3 可以看出,当每一类训练样本中包含 5 个不确定样本时,本文中提出的算法的 OA 值要比 SVM 的 OA 值高 5.86%,此外,和两个高光谱图像误标签检测算法 DP 聚类和 K-SDP 聚类相比,本算法对原始高光谱图像提取了低秩成分,提高了原始高光谱的质量。根据第 3.3 节可知,本算法相比于 DP 聚类和 K-SDP 聚类可以有效减少训练样本中的误标签,在 KSC 数据集上的不同误标签样本下,都提升了分类精度。相比于 KECA 算法,本算法使用改进的光谱角制图算法充分获取每一类训练样本间的上下文信息,在每一类包含 5 个不确定的训练样本上 OA 提升 2.76%。相比于使用未改进光谱角制图的 HCEM 算法,本算法克服了原始光谱角余弦的失真问题,能够抑制误差,在每一类包含 15 个不确定的训练样本上 OA 提升 1.26%。同时根据图 7 和表 4 可以得出同样的结论,例如,当每一类训练样本中包含 10 个不确定样本时,本算法相比于 SVM、DP、K-SDP、KECA 和 HCEM 算法,OA 分别提高了 6.76%,4.58%,3.94%,2.38%,1.24%,在两种数据集上充分证明了本算法的有效性。

图 8 中给出了不同训练集下使用不同误标签检测算法的总体精度柱状图。其中包括本算法和 5 种不同对比算法进行 10 次重复实验后获得的 OA 平均值。

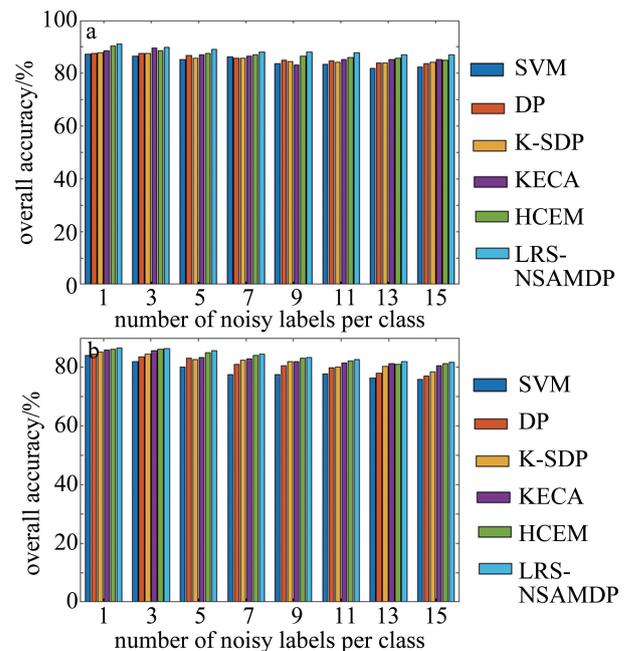


Fig. 8 OA obtained by using different false label detection algorithms in different training sets
a—KSC b—PaviaU

可以看出,本算法相比于 SVM、DP、K-SDP、KECA 和 HCEM 算法,两个数据集上都提高了 OA 值,证明了在误标签的检测过程中,提出的算法相比于对比算法更具有鲁棒性。

3.5 提出的误标签检测算法的误差分析

表 5 中给出本方法在两个数据集上所有检测失败的不确定样本数,检测失败的不确定样本造成检测算法的误差。其中,6×13 表示在训练样本中所有的不确定样本数目,6 表示每一类中的不确定样本数目,13 表示类的数目(实验数据重复 5 次求得平均值),其它类

Table 5 Detection performance of false labels for the proposed method on two datasets

dataset		KSC				PaviaU	
total uncertain samples		6×13	9×13	12×13	10×9	15×9	20×9
undetected false samples	DP	7.4	6.8	5.7	21.7	21.1	20.7
	K-SDP	6.1	5.4	4.8	20.4	18.2	17.8
	LRS-NSAMDP	5.6	4.4	3.8	17.6	15.8	14.1

推。首先明显看出,本算法所有检测失败的不确定样本相比于 DP 算法和 K-SDP 算法检测失败的不确定样本数目少,证明本算法检测性能优越。仍然会出现检测误差的原因主要还是算法本身造成的。(12)式采用简单的线性阈值决策函数,该函数可能会导致无法准确度量、检测、去除临界值附近的待检测样本,造成系统误差。此外,本算法采用改进 SAM 算法来衡量光谱相似度和利用光谱信息,在检测过程中并没有利用到遥感图像的空间信息,这也会造成系统误差,可以采取自适应调节的软阈值决策函数、联合样本之间的空间上下文信息等来减少系统误差。另一方面,测量仪器、设备装置和环境会导致随机误差,可以通过增加求平均值的次数以及使用最小二乘法求得最优值来减少随机误差。但是随机误差仍是不可避免的。

4 结 论

针对传统的基于有监督的误标签检测算法检测到的误标签过多而导致后续分类精度下降的问题,提出了一种基于低秩稀疏表示和改进光谱角制图的高光谱图像误标签分类算法。提取原始高光谱图像的低秩成分,使用基于归一化的光谱角制图算法计算光谱相似度。所提出的算法相比于其它误标签检测算法,去除了原始高光谱图像中的混合噪声,降低了阈值 $\lambda \bar{\rho}_j$ 的大小,从而使误标签数目变小。实验结果表明,与其它先进的误标签检测算法相比,本算法提高了分类精度。

参 考 文 献

- [1] ZHANG L, WEI W, ZHANG Y N, *et al.* Cluster sparsity field: An internal hyperspectral imagery prior for reconstruction [J]. *International Journal of Computer Vision*, 2018, 126(8): 797-821.
- [2] LI Sh T, HAO Q N, GAO G H, *et al.* The effect of ground truth on performance evaluation of hyperspectral image classification [J]. *IEEE Transactions on Geoenvironment & Remote Sensing*, 2018, 56(12): 7195-7206.
- [3] PARK B, LU R F. *Hyperspectral imaging technology in food and agriculture* [M]. New York, USA: Springer, 2015: 305-331.
- [4] RUITENBEEK F, DEBBA P, MEER F D, *et al.* Mapping white micas and their absorption wavelengths using hyperspectral band ratios [J]. *Remote Sensing of Environment*, 2006, 102(3/4): 211-222.
- [5] HU Y F, ZHANG Q L, ZHANG Y Z, *et al.* A deep convolution neural network method for land cover mapping: A case study of Qinghai, China [J]. *Remote Sensing*, 2018, 10(12): 2053.
- [6] GUAN Sh H, YANG G, LI H, *et al.* Hyperspectral image classification based on 3-D convolutional recurrent neural network [J]. *Laser Technology*, 2020, 44(4): 485-491 (in Chinese).
- [7] ZHANG X G, GAO Z Y, JIAO L C, *et al.* Multifeature hyperspectral image classification with local and nonlocal spatial information via markov random field in semantic space [J]. *IEEE Transactions on Geoenvironment & Remote Sensing*, 2018, 56(3): 1409-1424.
- [8] FANG L Y, HE N J, LI Sh T, *et al.* A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation [J]. *IEEE Transactions on Geoenvironment & Remote Sensing*, 2018, 56(6): 3534-3546.
- [9] LU Zh W, FU Zh Y, XIANG T, *et al.* Learning from weak and noisy labels for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(3): 486-500.
- [10] FOODY G M. The effect of mis-labeled training data on the accuracy of supervised image classification by SVM [C]// 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). New York, USA: IEEE, 2015: 4987-4990.
- [11] KANG X D, DUAN P H, XIANG X L, *et al.* Detection and correction of mislabeled training samples for hyperspectral image classification [J]. *IEEE Transactions on Geoenvironment & Remote Sensing*, 2018, 56(10): 5673-5686.
- [12] TU B, ZHOU Ch L, KUANG W L, *et al.* Hyperspectral imagery noisy label detection by spectral angle local outlier factor [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(9): 1417-1421.
- [13] ALEX R, LAIO A. Clustering by fast search and find of density peaks [J]. *Science*, 2014, 344(6191): 1492-1496.
- [14] TU B, ZHANG X F, KANG X D, *et al.* Density peak-based noisy label detection for hyperspectral image classification [J]. *IEEE Transactions on Geoenvironment & Remote Sensing*, 2019, 57(3): 1573-1584.
- [15] TU B, ZHANG X F, KANG X D, *et al.* Spatial density peak clustering for hyperspectral image classification with noisy labels [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(7): 5085-5097.
- [16] TU B, ZHANG Ch L, PENG J, *et al.* Kernel entropy component analysis-based robust hyperspectral image supervised classification [J]. *Remote Sensing*, 2019, 11(23): 2823.
- [17] ZOU Zh X, SHI Zh W. Quadratic constrained energy minimization for hyperspectral target detection [C]// 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). New York, USA: IEEE, 2015: 4979-4982.
- [18] ZHANG Y F, XIE B B, SUN J, *et al.* A hybrid sparsity and constrained energy minimization detector for hyperspectral images [C]// 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). New York, USA: IEEE, 2017: 1137-1140.
- [19] TU B, ZHOU Ch L, LIAO X L, *et al.* Hierarchical structure-based noisy labels detection for hyperspectral image classification [J].

- IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 2183-2199.
- [20] KIM S J, KOH K, LUSTIG M, *et al.* An interior-point method for large-scale & 1-regularized least squares[J]. IEEE Journal of Selected Topics in Signal Processing, 2007, 1(4): 606-617.
- [21] BIOUCAS-DIAS J M, NASCIMENTO J M. Hyperspectral subspace identification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(8): 2435-2445.
- [22] KRUSE F A, LEFKOFF A B, BOARDMAN J W, *et al.* The spectral image processing system (SIPS): Software for integrated analysis of AVIRIS data[J]. Remote Sensing of Environment, 1993, 44(2/3): 145-163.
- [23] CUI M S, PRASAD S. Class-dependent sparse representation classifier for robust hyperspectral image classification[J]. IEEE Transactions on Geoscience & Remote Sensing, 2015, 53(5): 2683-2695.
- [24] SU H J, SHENG Y H. Orthogonal projection divergence-based hyperspectral band selection[J]. Spectroscopy and Spectral Analysis, 2011, 31(5): 1309-1313(in Chinese).
- [25] TU B, YANG X C, LI N Y, *et al.* Hyperspectral image classification via superpixel correlation coefficient representation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(11): 4113-4127.
- [26] MELGANI F, BRUZZONE L. Classification of hyperspectral remote sensing images with support vector machines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(8): 1778-1790.