

鲁棒多特征谱聚类的高光谱影像波段选择

孙伟伟¹, 杨刚¹, 彭江涛², 孟祥超³

1. 宁波大学 地理与空间信息技术系, 宁波 315211;

2. 湖北大学 数学与统计学学院, 湖北省应用数学重点实验室, 武汉 430062;

3. 宁波大学 信息科学与工程学院, 宁波 315211

摘要: 传统谱聚类的高光谱影像波段选择模型中, 采用的波段相似矩阵受到噪声或异常值的影响且仅能表征波段的单一相似特征, 导致波段子集的选取结果受到限制。本文从波段选择的目的出发, 提出鲁棒多特征谱聚类方法, 整合多个特征的波段相似矩阵来形成综合相似矩阵以解决上述问题。该方法假设4种相似性度量包括光谱信息散度、光谱角度距离、波段相关性和拉普拉斯图谱能够共同揭示波段聚类的内在结构特征, 通过构建低秩稀疏矩阵分解模型来表征单一相似矩阵与综合相似矩阵的内在关系。进一步, 采用增强拉格朗日乘子算法来优化求解综合相似矩阵, 利用常规谱聚类方法来聚合所有波段至不同的类别, 并选取代表性波段。采用两个常用的高光谱影像数据, 对比5种常用的波段选择方法来进行实验验证。实验结果表明, 鲁棒多特征谱聚类方法优于改进稀疏子空间聚类、常规谱聚类方法和其他主流波段选择方法, 而且计算效率较高。

关键词: 遥感, 高光谱遥感, 降维, 波段选择, 分类, 鲁棒多特征谱聚类

引用格式: 孙伟伟, 杨刚, 彭江涛, 孟祥超. 2022. 鲁棒多特征谱聚类的高光谱影像波段选择. 遥感学报, 26(2): 397–405

Sun W W, Yang G, Peng J T and Meng X C. 2022. Robust multi-feature spectral clustering for hyperspectral band selection. National Remote Sensing Bulletin, 26(2): 397–405 [DOI:10.11834/jrs.20209165]

1 引言

以高分五号为代表的高光谱遥感是一种先进的对地观测技术, 利用较窄的光谱波段来获取地物地表的连续光谱响应信息, 有利于区分具有较小光谱响应差异的地表地物 (Tong 等, 2014; 童庆禧 等, 2016; Zhong 等, 2018)。然而, 影像的光谱维数很高且波段相关性很强, 存在明显的信息冗余和计算量大问题, 给数据处理和应用造成较大困难 (Zhang 等, 2011; 张兵, 2016; 张良培和李家艺, 2016; 张良培 等, 2018)。波段选择能够选取代表性的波段子集, 解决上述问题并最大程度地保持原始影像的应用效果 (Sun 等, 2015b, 2016, 2017a; 孙伟伟 等, 2018)。

目前国内外学者提出多种策略来选取合适的波段子集, 如排序策略、搜索策略、稀疏策略和

聚类策略等 (Sun 和 Du, 2018)。基于排序的方法通过比较每个波段的重要性来选取若干波段, 如最大方差主成分分析法 MVPCA (Maximum Variance Principal Component Analysis) (Chang 等, 1999) 和限制波段选择法 (Chang 和 Wang, 2006)。然而该方法忽略不同波段的差异性, 容易选取相关性较强的波段。基于搜索的方法转换波段选择为准则函数的优化求解问题, 如线性预测 LP (Linear Predication) (Du 和 Yang, 2008) 和多目标优化搜索 (Pan 等, 2019)。搜索策略相比排序策略能够选择更好的波段子集, 然而该方法大多需要较高的计算量。基于稀疏的方法构建高光谱波段的稀疏表达模型 (孙伟伟 等, 2017), 通过求解稀疏限制的优化问题来选择代表性波段, 如稀疏非负矩阵分解 (Sun 等, 2015a)、差异性加权稀疏表达模型 (Sun 等, 2016) 和图正则的鲁棒主成分分析模

收稿日期: 2019-05-25; 预印本: 2019-09-07

基金项目: 国家自然科学基金 (编号: 42122009, 42171351, 41971296, 61871177); 浙江省自然科学基金 (编号: LR1901D0001); 湖北省自然科学基金 (编号: 2021CFA087); 宁波市科技计划项目 (编号: 2021S089, 2021Z107); 浙江省省属高校基本科研业务费专项资金资助 (编号: SJLZ2022002); 浙江省自然资源厅科技项目 (编号: 2021-30, 2021-31)

第一作者简介: 孙伟伟, 1985年生, 教授, 研究方向为多源遥感协同智能解译与海岸带健康监测。E-mail: sunweiwei@nbu.edu.cn

通信作者简介: 彭江涛, 1982年生, 教授, 研究方向为高光谱遥感解译与机器学习。E-mail: pengjt1982@126.com

型 (Sun 和 Du, 2018)。稀疏系数估计对目标函数的收敛较为敏感, 波段选择的结果变化较大。基于聚类的方法将所有波段聚合至不同的类别或子空间中, 选择离聚类中心最近的波段, 如双聚类方法 (Yuan 等, 2016) 和最优聚类框架 (Wang 等, 2018) 等。

谱聚类 SC (Spectral Clustering) 是一种经典的非监督聚类方法, 广泛应用于高维数据的非监督聚类且效果较好 (von Luxburg, 2007)。谱聚类能够利用不同波段的相似性度量来构建连通图谱 (或相似矩阵), 实现波段在图谱上的非监督聚类并选取波段子集。采用的相似性度量如光谱角度距离 SAD (Spectral Angle Distance)、相关系数 CC (Correlation Coefficient) 和光谱信息散度 SID (Spectral Information Divergence)。近年来, 随着稀疏理论引入至高光谱遥感领域, 波段的稀疏系数也被用来构建相似矩阵, 其中稀疏系数代表波段之间的重构概率 (Sun 等, 2015b, 2016)。然而, 上述相似性度量假设高光谱遥感成像服从高斯或线性分布, 在表征波段的相似程度时, 相似矩阵容易受到光谱噪声或异常值影响, 从而降低波段子集的选取效果 (Xia 等, 2014; Sun 等, 2018)。同时, 上述相似性度量仅能描述波段在某一角度的相似程度, 并不能够完全表征波段的全部相似特征 (Sun 等, 2015b)。以相关系数 CC 为例, 相似矩阵仅代表波段之间的相关强弱, 无法体现波段之间的信息可区分性。波段选择目的是寻找信息量较大、相关性较低且差异度较高的代表性波段。因此, 常规谱聚类方法构建的相似矩阵并不能够代表波段选择所需的特征信息, 无法得到满足分类应用需求的波段子集。

针对上述问题, 本文提出鲁棒多特征谱聚类方法 RMSC (Robust Multi-feature Spectral Clustering) 来改进传统 SC 方法, 选取更优代表性的波段子集。相比传统 SC 方法, RMSC 认为每一特征的相似矩阵仅能反映波段聚类的部分结构信息且受到噪声的负面影响。RMSC 通过整合单个相似特征矩阵来得到涵盖多个特征相似信息的综合相似矩阵, 实现不同角度的相似信息的汇总表达, 提升相似矩阵的信息总量并降低噪声的影响, 改善波段子集的选取结果。假设每一特征 (如信息熵、波段相关性和波段差异性) 构建的波段相似矩阵具有低阶的结构特征 (Xia 等, 2014), RMSC 能够分解单一

特征相似矩阵为综合相似矩阵和噪声矩阵, 并将综合相似矩阵的估计问题转换为低秩稀疏矩阵分解问题, 采用增强拉格朗日乘子 ALM (Augmented Lagrangian Multiplier) (Lin 等, 2015) 来进行优化求解。进一步, RMSC 利用谱聚类方法来聚合所有波段至不同的类别中, 从各个类别选取代表性波段。本文采用 Indian Pines 和 PaviaU 两个常用的高光谱影像数据集, 对比 5 种主流的波段选择方法来验证方法的效果。

2 鲁棒多特征谱聚类的波段选择

2.1 高光谱波段的单一特征相似矩阵

假设高光谱数据的所有波段构成二维矩阵 $Y = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbf{R}^{M \times N}$, 其中 M 和 N 分别为像素个数和波段数量, $\mathbf{y}_i$ 为第 i 个波段对应的向量; 选取得到的波段子集为 $\Phi = Y(:, \kappa) \in \mathbf{R}^{M \times k}$, 其中 κ 为所选波段的索引集合, k 为所选的波段子集大小。考虑到波段子集的较高信息量、较低相关性和较大差异性要求, 本文采用光谱信息散度 SID、相关系数 CC、拉普拉斯图谱 LG (Laplacian Graph) (Sun 和 Du, 2018) 和光谱角度距离 SAD 来构建多个特征的相似矩阵。SID 表征不同波段的相对信息量差异, 构建波段信息量特征的相似矩阵。CC 表征波段的相关性, 本文利用相关系数平方和来构建描述相关性特征的相似矩阵。LG 考虑波段的内在聚类结构, 利用波段的邻域结构, 构建邻域差异性特征的相似矩阵。SAD 考虑波段的光谱角度差异, 构建反映波段向量形状差异特性的相似矩阵。上述 4 个特征度量, 每一个特征的相似矩阵都能够涵盖部分的波段聚类的相似性结构信息, 都为非负相似矩阵。

2.2 综合相似矩阵的估计模型

假设上述 4 个特征的波段相似矩阵都能从某一方面来表征波段的聚类结构, RMSC 通过构建波段综合矩阵来汇总 4 个特征矩阵的波段相似信息, 更好揭示所有波段的内在聚类结构特征。相应地, 每个特征的波段相似矩阵 $S^{(i)}$ 可以分解为两部分,

$$S^{(i)} = B + E^{(i)}, \text{ s.t., } S^{(i)} \geq 0, \text{ rank}(B) = r, \|E^{(i)}\|_0 \leq K, i = 1, 2, 3, 4 \quad (1)$$

式中, $S^{(i)} \in \mathbf{R}^{N \times N}$ 分别为光谱信息散度 SID、相关系数 CC、拉普拉斯图谱 LG 和光谱角度距离 SAD

构建得到的波段相似矩阵, $B \in \mathbf{R}^{N \times N}$ 为待求的综合相似矩阵, $E^{(i)} \in \mathbf{R}^{N \times N}$ 为每个特征的误差矩阵, 代表其噪声或粗差的影响。约束项 $\text{rank}(B) = r$ 限制矩阵 B 为低秩。波段的子空间聚类结构, 导致位于相同子空间 (或类别) 的波段相似系数较大, 不同类别的波段相似系数较小, 从而使得矩阵 B 具有低秩结构。每一特征的相似矩阵都能代表大部分的波段聚类结构信息, 因此 $\|E^{(i)}\|_0 \leq K$ 限定每一相似矩阵与综合相似矩阵的差异较小。利用式 (1) 的低秩和稀疏分解, 能够分离出单个相似矩阵的噪声信息, 降低噪声影响并汇总 4 个特征矩阵的波段聚类结构信息。

研究表明, 核范数和 L_1 范数能够分别很好地近似表达低秩和 L_0 范数约束条件 (Sun 等, 2017b)。因此, 综合相似矩阵 B 的求解转换为一个凸优化问题:

$$\min_{B, E^{(i)}} \|B\|_* + \lambda \|E^{(i)}\|_1, \text{ s.t., } S^{(i)} = B + E^{(i)}, \quad (2)$$

$$B \geq 0, i = 1, 2, 3, 4$$

式中, $\|\cdot\|_*$ 为核范数, λ 为正则化参数来平衡两个约束项。本文采用 ALM 算法来优化求解问题式 (2), 引入辅助变量 $Q \in \mathbf{R}^{N \times N}$, 可以转化为

$$\min_{B, Q, E^{(i)}} \|Q\|_* + \lambda \|E^{(i)}\|_1, \text{ s.t., } B = Q, S^{(i)} = B + E^{(i)}, B \geq 0, i = 1, 2, 3, 4 \quad (3)$$

目标函数式 (3) 的拉格朗日方程可以展开为

$$\Psi(B, Q, E^{(i)}) = \|Q\|_* + \lambda \|E^{(i)}\|_1 + \frac{\mu}{2} \|B - Q\|_F^2$$

$$\sum_{i=1}^4 \text{tr}(\Lambda^{(i)T}(B + E^{(i)} - S^{(i)})) + \text{tr}(\Delta^T(B - Q)) +$$

$$\frac{\mu}{2} \sum_{i=1}^4 \|B + E^{(i)} - S^{(i)}\|_F^2, i = 1, 2, 3, 4 \quad (4)$$

式中, $\Lambda^{(i)}$ 和 Δ 是拉格朗日乘子, μ 是惩罚参数, $\text{tr}(\cdot)$ 是矩阵的迹。在 $t+1$ 次循环中, 当固定其他变量时, 变量 Q 的更新可以转换为

$$\min_Q \|Q\|_* + \text{tr}(\Delta^T(B - Q)) + \frac{\mu}{2} \|B - Q\|_F^2 + \frac{\|\Delta\|_F^2}{2\mu} - \frac{\|\Delta\|_F^2}{2\mu} \quad (5)$$

进一步, 式 (5) 可以等价转换为

$$\min_Q \|Q\|_* + \frac{\mu}{2} \left\| B - Q + \frac{\Delta}{\mu} \right\|_F^2 - \frac{\|\Delta\|_F^2}{2\mu} \quad (6)$$

考虑到第 3 项 $\frac{\|\Delta\|_F^2}{2\mu}$ 与变量 Q 的更新与优化求解无

关, 式 (6) 可以简化为

$$\min_Q \|Q\|_* + \frac{\mu}{2} \left\| B - Q + \frac{\Delta}{\mu} \right\|_F^2 \quad (7)$$

类似地, 固定其他变量, 求解式 (8) 来更新变量 $E^{(i)}$

$$\min_{E^{(i)}} \lambda \|E^{(i)}\|_1 + \frac{\mu}{2} \sum_{i=1}^4 \left\| E^{(i)} - \left(S^{(i)} - B - \frac{\Lambda^{(i)}}{\mu} \right) \right\|_F^2 \quad (8)$$

$$i = 1, 2, 3, 4$$

接下来, 固定其他变量, 求解式 (9) 来更新变量 B ,

$$\min_B \sum_{i=1}^4 \left\| B + E^{(i)} - S^{(i)} + \frac{\Lambda^{(i)}}{\mu} \right\|_F^2 + \left\| B - Q + \frac{\Delta}{\mu} \right\|_F^2, \text{ s.t., } B \geq 0, i = 1, 2, 3, 4 \quad (9)$$

最后, 固定其他变量, 来依次更新拉格朗日乘子 $\Delta_{(t+1)}$ 和 $\Lambda_{(t+1)}^{(i)}$,

$$\Delta_{(t+1)} = \Delta_{(t)} + \mu(B - Q) \quad (10)$$

$$\Lambda_{(t+1)}^{(i)} = \Lambda_{(t)}^{(i)} + \mu(B + E^{(i)} - S^{(i)}), i = 1, 2, 3, 4 \quad (11)$$

重复上述迭代过程式 (7) — (11), 直至满足收敛条件 $\|B + E^{(i)} - S^{(i)}\|_\infty \leq \varepsilon$ 和 $\|B - Q\|_\infty \leq \varepsilon$, 最终得到最优的波段综合相似矩阵 $\hat{B} = B_{(t+1)}$ 。

2.3 利用综合相似矩阵选取波段子集

利用优化估计的综合相似矩阵 $\hat{B}$, 采用谱聚类方法来选取最优波段子集。首先, 利用 $\hat{B}$ 来计算归一化拉普拉斯矩阵。其次, 通过本征值分解来得到矩阵 U , 其每一列为拉普拉斯矩阵的前 k 个最小的非零本征值对应的本征向量。接下来, 将矩阵 U 的每一列进行归一化处理, 采用 K-均值聚类方法来聚类至 k 个不同的类别中。最后, 选择每一类中距离中心最近的行向量对应的波段来构成波段子集。RMSC 选取 k 个波段的技术流程包括:

(1) 转换高光谱影像数据立方体为二维的波段矩阵 Y , 利用 SID, CC, SAD 和 LG 度量, 分别构建 4 个特征的波段相似矩阵;

(2) 构建多特征谱聚类模型 (2), 将每个特征的相似矩阵分解为低秩矩阵和稀疏误差矩阵项;

(3) 采用 ALM 算法来依次求解式 (7) — (11), 得到最优的波段综合相似矩阵 $\hat{B}$;

(4) 采用谱聚类方法来聚类综合相似矩阵 $\hat{B}$, 划分所有波段到 k 个不同的类别中;

(5) 从各个聚类中寻找靠近中心最近的波段

来选取得到最终的波段集 $\Phi = Y(:, \kappa)$ 。

3 实验与分析

3.1 实验数据

Indian Pines 数据由美国普渡大学的遥感应用实验室提供，于 1992-06-12 通过美国 JPL 成像光谱仪采集得到。影像的空间分辨率为 20 m，光谱区间为 200—2400 nm，预处理后的波段数为 200 个。图 1 为美国印第安纳州西拉法叶地区西部 8 km 的一小块区域，影像大小为 145×145 像素，包含 16 类地物，其真实样本信息见表 1。

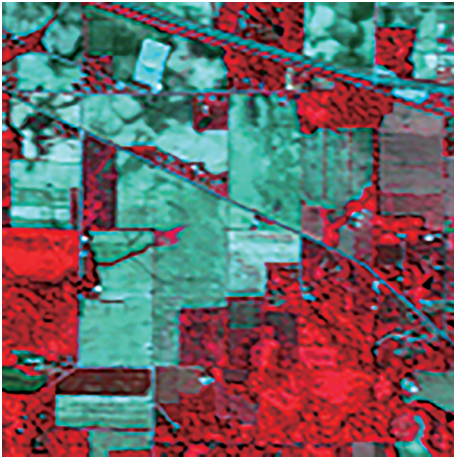


图 1 Indian Pines 影像
Fig. 1 The Indian Pines image

表 1 Indian Pines 数据的地物真实样本信息
Table 1 Ground truth of main classes on Indian Pines

标签	类别	样本数量
1	苜蓿	46
2	免耕玉米地	1428
3	少耕玉米地	830
4	玉米	237
5	草地/牧场	483
6	草地/树木	730
7	修建的草地	28
8	甘草—堆料	478
9	燕麦	20
10	免耕大豆地	972
11	少耕大豆地	2455
12	大豆	593
13	小麦	205
14	树木	1265
15	建筑	386
16	石钢塔	93
总计		10249

PaviaU 数据由自西班牙巴斯克大学智能课题组提供，由 ROSIS 成像光谱仪采集得到，共 103 个波段，空间分辨率为 1.3 m，光谱区间为 430—860 nm。图 2 为帕维亚大学区域的部分影像，大小为 610×340 像素，包含 9 类地物，其真实样本信息见表 2。



图 2 PaviaU 影像
Fig. 2 The PaviaU image

表 2 PaviaU 数据的地物真实样本信息
Table 2 Ground truth of main classes on PaviaU

标签	类别	样本数量
1	柏油路	6631
2	草地	18649
3	沙砾	2099
4	树木	3064
5	金属板	1345
6	裸土	5029
7	沥青屋顶	1330
8	地砖	3682
9	阴影	947
总计		42776

3.2 实验分析

本节采用 Indian Pines 和 PaviaU 数据集，设计分类实验来验证 RMSC 方法。对比的主流波段选择方法包括聚类策略方法 WaluDI (Martínez-Usómartinez-Uso 等, 2007)、排序策略方法快速密度峰值聚类方法 FDPC (Fast Density-Peak based Clustering) (Jia 等, 2016)、正交投影方法 OPBS (Orthogonal Projections based Band Selection) (Zhang 等, 2018) 和稀疏策略方法 ISSC。同时，对比传统谱聚类方法 SC-SID，其采用的波段相似特征

性度量光谱信息散度SID。实验分类器采用支持向量机SVM (Support Vector Machine), 定量评价指标为总体分类精度OCA (Overall Classification Accuracy)。SVM的核函数采用径向基函数, 采用交叉验证方法来确定方差和惩罚因子的参数取值 (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/> [2019-05-25])。对每个数据集进行10次进行独立试验, 得到下述平均结果。

(1) 不同波段子集大小的分类精度对比。Indian Pines 和 PaviaU 数据集中波段数的选择区间为5—60, 步长为5, 随机选取10%作为训练样本, 其余作为测试样本。Indian Pines 和 PaviaU 数据集中, ISSC的正则化因子分别取值为0.1和0.001; RMSC的正则化因子分别取值为0.005和0.001。

图4可以看出, OPBS的OCA曲线低于其他方法。RMSC的OCA在6种方法中表现最优, 高于SC-SID, 这表明综合相似矩阵相比单一相似矩阵用于选择波段子集的显著优势。RMSC的OCA略高于ISSC。这是因为RMSC采用综合相似矩阵, 降低噪声负面影响并提升波段相似矩阵的信息量, 从而提升波段选取效果。图3和图5为各种方法在波段数量为30时得到的分类结果图。RMSC波段子集的分类精度高于其他5种方法, 与图4的结果

保持一致。因此, 在不同波段数量的条件下, RMSC的总体分类精度OCA表现最优。

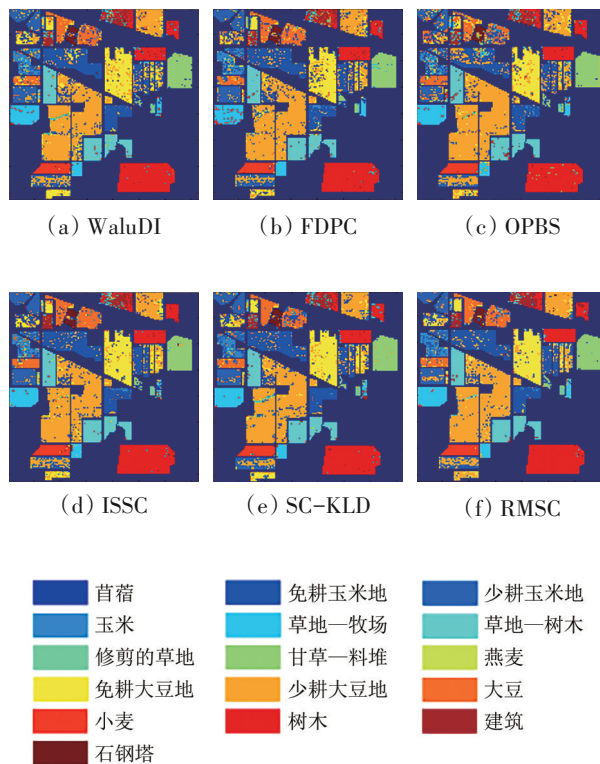


图3 不同方法在 Indian Pines 数据的 SVM 分类图
Fig. 3 Different classification maps of Indian Pines

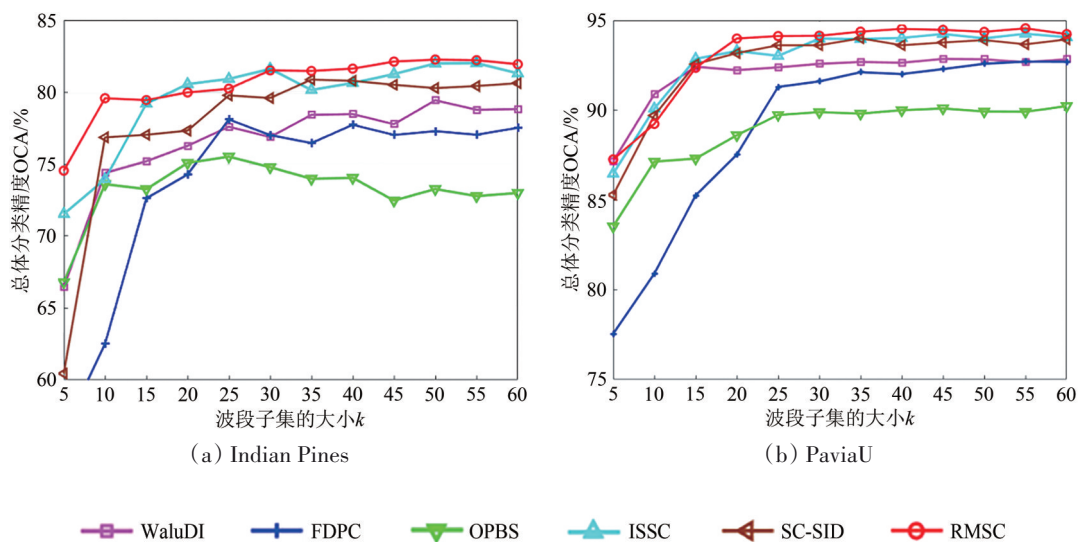


图4 不同波段子集下的各种方法的OCA曲线对比

Fig. 4 The comparison of different OCA curves with different number of selected bands

(2) 不同训练样本大小下的分类结果对比。实验中的采样比率区间为 $[0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5]$, 波段子集大小 k 都设置为30。图6为RMSC和其他几种方法的总体分类精度OCA曲线图, 总体结果与上述实验保

持一致。随着训练样本的采样百分比增加, 各种方法的总体分类精度OCA逐渐提升。OPBS的OCA最低; FDPC和WaluDI的OCA曲线高于OPBS, 但明显低于ISSC、SC-SID和RMSC。RMSC的OCA在所有方法中表现最优, 优于SC-SID和ISSC。

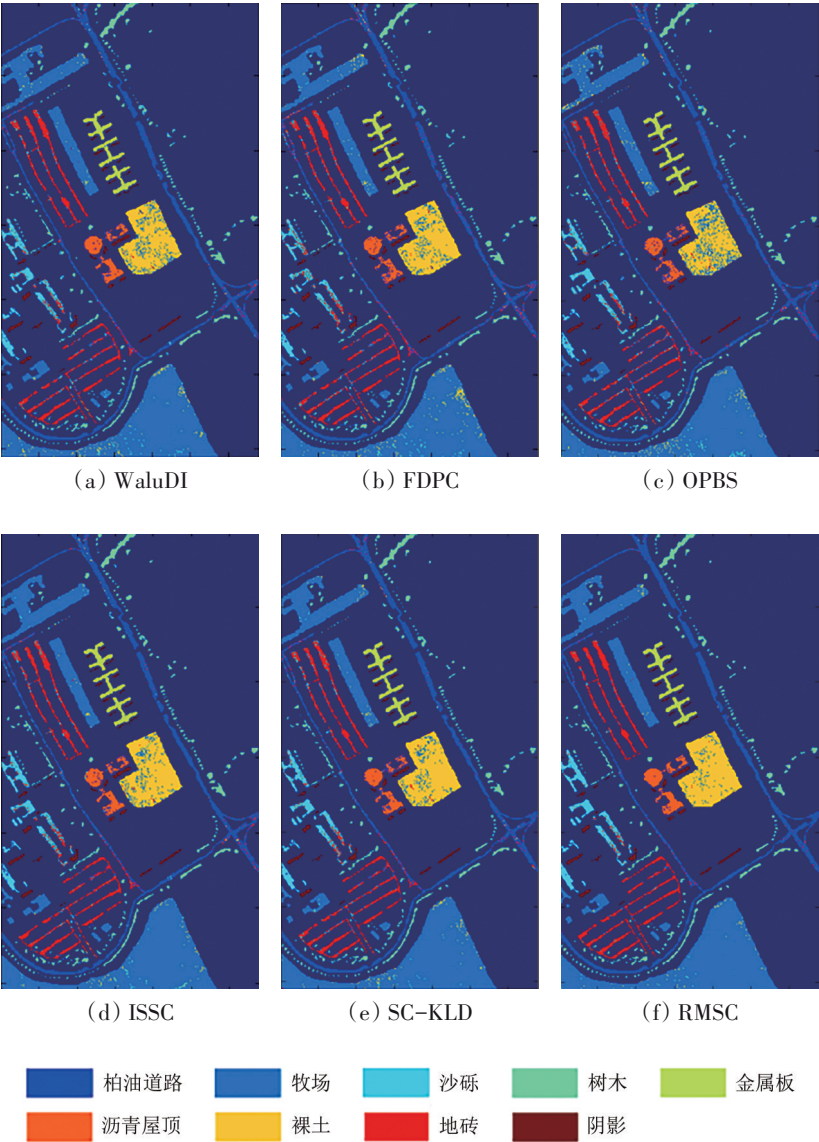


图5 不同方法在PaviaU数据的SVM分类图
Fig. 5 Different classification maps of PaviaU

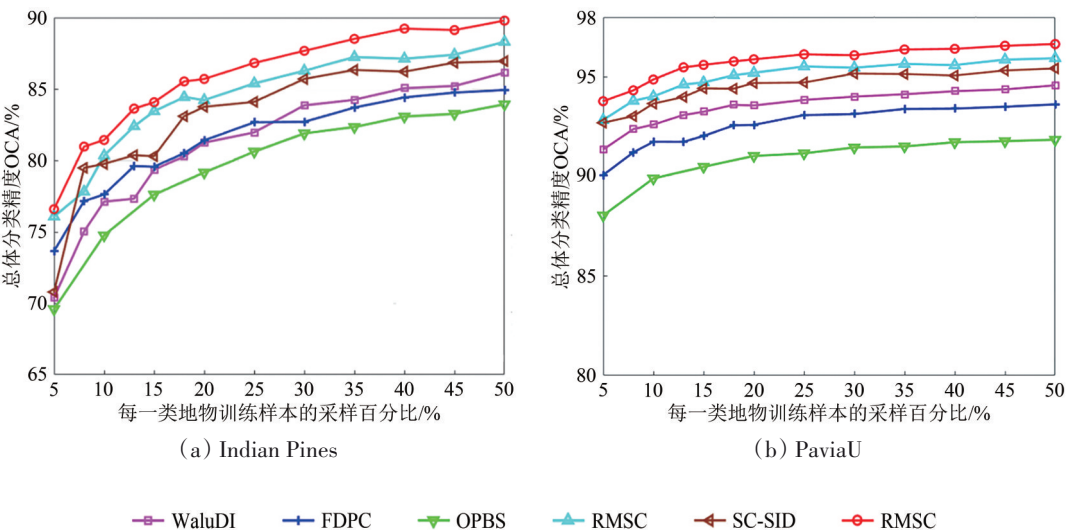


图6 不同训练样本大小下的各种波段选择方法的OCA曲线
Fig. 6 OCA curves of different methods with different size of training samples per class

(3) 不同方法的计算效率对比。表3为不同波段子集大小条件下各种方法的计算时间，两个数据的波段子集大小的取值范围均为10—50。实验的计算平台为戴尔Win10台式计算机，硬件配置为Inter Xeon Gold 6130 CPU 2.10 GHz和64 GB内存，代码的执行环境为MATLAB 2014a。各种方法的计算时间随着波段数量增加逐渐上升。结合实验(1)

和(2)，可以看出，OPBS方法的波段选择效果较差，但计算速度较快。WaluDI的计算效率最低，选取的波段子集却优于FDPC和OPBS。FDPC和OPBS的计算效率相比最高，但选取的波段子集结果较差。ISSC、SC-SID和RMSC的计算效率较高，尤其ISSC和SC-SID，但RMSC波段子集的分类精度最高。

表3 不同波段选择方法的计算时间对比
Table 3 Lists of computational time from different methods on two datasets

数据	<i>k</i>	波段选择的计算时间/s						数据	<i>k</i>	波段选择的计算时间/s					
		WaluDI	FDPC	OPBS	ISSC	SC-SID	RMSC			WaluDI	FDPC	OPBS	ISSC	SC-SID	RMSC
Indian Pines	10	542.1	3.27	0.11	0.83	0.39	10.62	PaviaU	10	1760	5.91	0.87	0.47	0.83	3.30
	20	543.0	3.33	0.20	1.18	0.71	10.92		20	1760	6.04	1.03	0.62	0.94	3.91
	30	544.3	3.41	0.28	1.66	0.92	11.93		30	1760	6.09	1.37	1.04	1.02	4.52
	40	544.7	3.59	0.37	2.05	1.14	12.74		40	1760	6.21	2.14	1.21	1.25	4.64
	50	545.3	3.72	0.41	2.60	1.30	13.11		50	1760	6.73	2.31	1.61	1.51	4.82

(4) 参数λ对RMSC波段子集的影响分析。实验中，Indian Pines和PaviaU数据的波段子集大小*k*设置为30，参数λ的选择区间为[0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10]。表4列出Indian Pines和PaviaU数据中不同λ对应的总体分类精度

OCA。可以看出，随着λ的逐渐增加，RMSC波段子集的OCA呈现波动但总体较为稳定。因此，参数λ的选择对RMSC的影响有限，可以选择较小的λ来确保波段子集的较高分类精度。

表4 不同正则化参数λ下的RMSC总体分类精度
Table 4 The OCA of RMSC with different choices of regularized parameter λ

数据集	OCA/%								
	λ=0.001	λ=0.005	λ=0.01	λ=0.05	λ=0.1	λ=0.5	λ=1	λ=5	λ=10
Indian Pines	81.89	82.12	82.47	82.19	81.99	82.07	91.98	82.07	82.28
PaviaU	94.76	95.12	95.04	94.81	94.57	95.02	94.78	94.98	95.22

4 结 论

本文提出鲁棒多特征谱聚类RMSC模型来克服传统谱聚类方法中相似特征矩阵的不足，提升传统SC用于高光谱波段选择的相似特征信息量，改善波段选择结果。RMSC认为波段选择的目的是选取信息量大、相关性低且差异性强的波段子集。该方法整合光谱信息散度、相关系数、光谱角度距离和拉普拉斯图谱4个度量项的相似矩阵，转换波段综合相似矩阵的求解问题为单一相似矩阵的稀疏低秩分解模型；利用ALM方法来优化求解目标函数，采用谱聚类方法来选取最佳波段子集。基于Indian Pines和PaviaU数据来设计4组实验，验证RMSC方法的效果。实验结果表明，在不同的波段子集大小和不同地物的训练样本采样比例条

件下，RMSC波段子集的OCA明显优于其他5种主流方法WaluDI、FDPC、OPBS、ISSC和SC-SID，而且计算效率较高。同时，正则化参数λ对RMSC波段子集的分类效果影响较小，较小的λ能够带来较高分类精度。后续工作将对更多的波段选择方法，采用更多的数据集包括国产GF-5卫星高光谱影像来进一步验证本文提出的方法。

参考文献(References)

Chang C I, Du Q, Sun T L and Althouse M L G. 1999. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 37(6): 2631-2641 [DOI: 10.1109/36.803411]
Chang C I and Wang S. 2006. Constrained band selection for hyper-

- spectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6): 1575-1585 [DOI: 10.1109/TGRS.2006.864389]
- Du Q and Yang H. 2008. Similarity-based unsupervised band selection for hyperspectral image analysis. *IEEE Geoscience and Remote Sensing Letters*, 5(4): 564-568 [DOI: 10.1109/LGRS.2008.2000619]
- Jia S, Tang G, Zhu J S and Li Q Q. 2016. A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1): 88-102 [DOI: 10.1109/TGRS.2015.2450759]
- Lin Z C, Chen M M, Wu L Q and Ma Y. 2015. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-rank Matrices. UILU-ENG-09-2215. Coordinated Science Laboratory, University of Illinois at Urbana-Champaign
- Martínez-Usó, Martínez-Usó A, Pla F, Sotoca J M and García-Sevilla P. 2007. Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12): 4158-4171 [DOI: 10.1109/TGRS.2007.904951]
- Pan B, Shi Z Z and Xu X. 2019. Analysis for the weakly Pareto optimum in multiobjective-based hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6): 3729-3740 [DOI: 10.1109/TGRS.2018.2886853]
- Sun W W and Du Q. 2018. Graph-regularized fast and robust principal component analysis for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 56(6): 3185-3195 [DOI: 10.1109/TGRS.2018.2794443]
- Sun W W, Jiang M and Li W Y. 2017. Band selection using sparse self-representation for hyperspectral imagery. *Geomatics and Information Science of Wuhan University*, 42(4): 441-448 (孙伟伟, 蒋曼, 李巍岳. 2017. 利用稀疏自表达实现高光谱影像波段选择. *武汉大学学报(信息科学版)*, 42(4): 441-448) [DOI: 10.13203/j.whugis.20150052]
- Sun W W, Li W Y, Li J and Lai Y M. 2015a. Band selection using sparse nonnegative matrix factorization with the thresholded Earth's mover distance for hyperspectral imagery classification. *Earth Science Informatics*, 8(4): 907-918 [DOI: 10.1007/s12145-014-0201-3]
- Sun W W, Tian L, Xu Y, Zhang D F and Du Q. 2017a. Fast and robust self-representation method for hyperspectral band selection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11): 5087-5098 [DOI: 10.1109/JSTARS.2017.2737400]
- Sun W W, Yang G, Du B, Zhang F F and Zhang L P. 2017b. A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7): 4032-4046 [DOI: 10.1109/TGRS.2017.2686842]
- Sun W W, Zhang D F, Yang G and Li W Y. 2018. Band selection for hyperspectral imagery based on weighted probabilistic archetypal analysis. *Journal of Remote Sensing*, 22(1): 110-118 (孙伟伟, 张殿发, 杨刚, 李巍岳. 2018. 加权概率原型分析的高光谱影像波段选择. *遥感学报*, 22(1): 110-118) [DOI: 10.11834/jrs.20186446]
- Sun W W, Zhang L P, Du B, Li W Y and Lai Y M. 2015b. Band selection using improved sparse subspace clustering for hyperspectral imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6): 2784-2797 [DOI: 10.1109/JSTARS.2015.2417156]
- Sun W W, Zhang L P, Zhang L F and Lai Y M. 2016. A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9): 4374-4388 [DOI: 10.1109/JSTARS.2016.2539981]
- Tong Q X, Xue Y Q and Zhang L F. 2014. Progress in hyperspectral remote sensing science and technology in China over the past three decades. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1): 70-91 [DOI: 10.1109/JSTARS.2013.2267204]
- Tong Q X, Zhang B and Zhang L F. 2016. Current progress of hyperspectral remote sensing in China. *Journal of Remote Sensing*, 20(5): 689-707 (童庆禧, 张兵, 张立福. 2016. 中国高光谱遥感的前沿进展. *遥感学报*, 20(5): 689-707) [DOI: 10.11834/jrs.20166264]
- von Luxburg U. 2007. A tutorial on spectral clustering. *Statistics and Computing*, 17(4): 395-416 [DOI: 10.1007/s11222-007-9033-z]
- Wang Q, Zhang F H and Li X L. 2018. Optimal clustering framework for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10): 5910-5922 [DOI: 10.1109/TGRS.2018.2828161]
- Xia R K, Pan Y, Du L and Yin J. 2014. Robust multi-view spectral clustering via low-rank and sparse decomposition//*Proceedings of the 28th AAAI Conference on Artificial Intelligence*. Québec, Canada: AAAI: 2149-2155
- Yuan Y, Lin J Z and Wang Q. 2016. Dual-clustering-based hyperspectral band selection by contextual analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3): 1431-1445 [DOI: 10.1109/TGRS.2015.2480866]
- Zhang B. 2016. Advancement of hyperspectral image processing and information extraction. *Journal of Remote Sensing*, 20(5): 1062-1090 (张兵. 2016. 高光谱图像处理与信息提取前沿. *遥感学报*, 20(5): 1062-1090) [DOI: 10.11834/jrs.20166179]
- Zhang B, Li S S, Jia X P, Gao L R and Peng M. 2011. Adaptive Markov random field approach for classification of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 8(5): 973-977 [DOI: 10.1109/LGRS.2011.2145353]
- Zhang L P and Li J Y. 2016. Development and prospect of sparse representation-based hyperspectral image processing and analysis. *Journal of Remote Sensing*, 20(5): 1091-1101 (张良培, 李家艺. 2016. 高光谱图像稀疏信息处理综述与展望. *遥感学报*, 20(5): 1091-1101) [DOI: 10.11834/jrs.20166050]
- Zhang L P, Liu R and Du B. 2018. Hyperspectral remote sensing im-

- age processing by using quantum optimization algorithm. *Geomatics and Information Science of Wuhan University*, 43(12): 1811-1818 (张良培, 刘蓉, 杜博. 2018. 使用量子优化算法进行高光谱遥感影像处理综述. *武汉大学学报(信息科学版)*, 43(12): 1811-1818) [DOI: 10.13203/j.whugis.20180231]
- Zhang W Q, Li X Y, Dou Y X and Zhao L Y. 2018. A geometry-based band selection approach for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8): 4318-4333 [DOI: 10.1109/TGRS.2018.2811046]
- Zhong Y F, Wang X Y, Xu Y, Wang S Y, Jia T Y, Hu X, Zhao J, Wei L F and Zhang L P. 2018. Mini-UAV-borne hyperspectral remote sensing: from observation and processing to applications. *IEEE Geoscience and Remote Sensing Magazine*, 6(4): 46-62 [DOI: 10.1109/MGRS.2018.2867592]

Robust multi-feature spectral clustering for hyperspectral band selection

SUN Weiwei¹, YANG Gang¹, PENG Jiangtao², MENG Xiangchao³

1. *Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China;*

2. *Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China;*

3. *Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China*

Abstract: The Hughes problem together with strong intra-band correlations and massive data seriously hinders hyperspectral processing and further applications. Dimensionality reduction using band selection can be used to conquer the abovementioned problems and guarantee the application performance of hyperspectral data. In particular, spectral clustering is a typical method for high-dimensional hyperspectral data. This method finds clusters of all hyperspectral bands on the connected graph and selects the representatives. Unfortunately, the regular similarity measures are negatively affected by outliers or noise of hyperspectral data in measuring the similarity of different bands. They could also only represent one feature of band similarity and have respective limitations. Accordingly, the obtained similarity matrix could not represent the full information of band selection required and could not guarantee obtaining aimed bands from spectral clustering. Therefore, we propose a Robust Multifeature Spectral Clustering (RMSC) method to solve the two problems mentioned above and enhance the performance of hyperspectral band selection from spectral clustering.

The RMSC combines multiple features of similarity measures for pairwise bands, namely, information entropy, band correlation, and band dissimilarity, to construct the integrated similarity matrix. It utilizes spectral information divergence to quantify the information entropy between pairwise bands. The coefficient correlation is utilized to measure the band correlations and construct the similarity matrix of band correlations. The Laplacian graph is also adopted to construct a similarity matrix and show the dissimilarity between different bands considering the inner clustering structure of all bands. The spectral angle distance matrix is constructed as well to reflect the similarity from the aspects of overall differences. The RMSC regards that each similarity matrix of all four features reflect the underlying true clustering information of all bands and has low-rank property. It formulates the estimation of combined dissimilarity matrix into a low-rank and sparse decomposition problem and utilizes the augmented Lagrangian multiplier to solve it. Thereafter, it implements the regular spectral clustering on the integrated similarity matrix and selects the representative bands from each cluster.

Two hyperspectral datasets are used to design four groups of experiments and testify the performance of RMSC. Five state-of-the-art methods, namely, WaluDI, fast density-peak-based clustering, orthogonal projections based band selection, Improved Sparse Spectral Clustering (ISSC) and SC-SID, and support vector machine, are used to quantify the classification accuracy. Experimental results show that RSMC outperforms the five other band selection methods in overall classification accuracy with shorter computational time. The regularization parameter is insensitive to RMSC, and a small candidate could produce high classification accuracy.

RMSC is better in selecting representative bands than current spectral clustering such as ISSC. It can also be a good choice in hyperspectral dimensionality reduction.

Key words: remote sensing, hyperspectral remote sensing, dimensionality reduction, band selection, spectral clustering, robust multi-feature spectral clustering

Supported by National Natural Science Foundation of China (No. 42122009, 42171351, 41971296, 61871177); Natural Science Foundation of Zhejiang Province, China (No. LR19D010001); Natural Science Foundation of Hubei Province, China (No. 2021CFA087); Public Projects of Ningbo City (No. 2021S089, 2021Z107); Fundamental Research Funds for the Provincial Universities of Zhejiang (No. SJLZ2022002); Research Project of Zhejiang Provincial Department of Natural Resources (No. 2021-30, 2021-31)