

基于模糊 C-均值聚类 and 萤火虫算法的高光谱图像波段选择

张周^{1,2}, 孙旭^{2*}, 刘荣¹, 高连如²

¹东华理工大学测绘工程学院, 江西 南昌 330013;

²中国科学院空天信息创新研究院计算光学成像技术重点实验室, 北京 100094

摘要 传统基于聚类的波段选择方法多属于硬聚类,在对波段进行划分时不够精确。针对该问题,提出一种基于模糊 C-均值聚类(FCM)的无监督波段选择方法,通过引入萤火虫算法(FA)得到 FCM-FA,利用 FA 的全局搜索特性,修正 FCM 在特定情况下获得局部最优解的问题。在两个公开高光谱数据集上进行分类实验,结果表明:在 136 组实验中,所提 FCM-FA 所选波段在 55.9% 的情况下能够达到使用全波段的分类精度;在 77.9% 情况下能够达到最优分类精度;FA 的引入有效地提升了 FCM 的效果,总体精度最大提高了 3.12 个百分点,Kappa 最大提高了 4.26 个百分点。经验证,FCM-FA 能够在大幅减少数据量的同时保留原数据的主要信息,可进一步进行推广和研究。

关键词 遥感; 波段选择; 高光谱图像; 模糊 C-均值聚类; 萤火虫算法

中图分类号 TP751.1 文献标志码 A

DOI: 10.3788/LOP221136

Band Selection of Hyperspectral Images Based on Fuzzy C-Means Clustering and Firefly Algorithm

Zhang Zhou^{1,2}, Sun Xu^{2*}, Liu Rong¹, Gao Lianru²

¹Faculty of Geomatics, East China University of Technology, Nanchang 330013, Jiangxi, China;

²Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

Abstract Traditional clustering-based band selection methods mostly belong to hard clustering, which are not accurate enough to divide the bands. To solve this problem, this paper proposes an unsupervised band selection method based on fuzzy C-means clustering (FCM). By introducing the firefly algorithm (FA), an FCM-FA is obtained, after which the global search feature of FA is used to solve the problem for which the FCM only obtains a locally optimal solution in certain circumstances. Classification experiments on two public hyperspectral datasets show that the proposed FCM-FA achieves the classification accuracy for all bands in 55.9% of 136 experiments, the optimal classification accuracy is achieved in 77.9% of the cases, the introduction of FA effectively improves the effect of FCM, with the overall accuracy increasing by 3.12 percentage points, and Kappa is increased by 4.26 percentage points at most. Hence, our results verify that FCM-FA can significantly reduce the amount of data while retaining the main information of original data that can be further promoted and studied.

Key words remote sensing; band selection; hyperspectral image; fuzzy C-means clustering; firefly algorithm

1 引言

与多光谱图像相比,高光谱图像(HSI)具有更高的光谱分辨率^[1],能够获取更多通道的光谱影像,同一位置的光谱值可以形成近似连续的光谱曲线,增强对

地物的判别能力。但对于更大的影像数据,在分析处理时会产生较高的计算复杂度,波段之间的高度相关性也会产生“Hughes”现象^[2],因此有必要在数据使用前对数据进行降维处理。

目前,降维的方式主要有两种:一种是特征提取

收稿日期: 2022-03-28; 修回日期: 2022-04-15; 录用日期: 2022-06-22; 网络首发日期: 2022-06-29

基金项目: 国家自然科学基金委员会国际(地区)合作与交流项目(62161160336)

通信作者: *sunxu@aircas.ac.cn

(FE);一种是波段选择(BS),又称特征选择。特征提取方法将原始数据通过计算从高维数据转换为低维数据,如主成分分析(PCA)^[3]、独立成分分析(ICA)^[4]和离散小波变换(DWT)^[5]等,但这种方法改变了原始数据的物理信息,然而这些信息在霾检测^[6]、目标检测^[7]、异常检测^[8]等应用中发挥着重要作用。波段选择方法则从所有波段中按一定条件选择一定数量的波段作为子集,这些子集能够代表原始数据,相对于特征提取方法,保留了原始数据各波段数值的物理意义,能够满足特定的应用需求。

波段选择算法可以按照方法原理分为基于稀疏^[9-10]、基于嵌入式学习^[11-12]、基于搜索^[13-15]、基于排序^[16]、基于聚类^[17]及基于混合^[18]的方法。基于聚类的方法将每个波段视为一个对象,通过聚类将所有波段分为若干个簇,使得簇心与簇心之间的距离最大化,簇内对象之间的距离最小化,再从每个簇中选取一个代表该簇的波段形成子集,能够有效减小提取的波段之间的相似性,是高光谱图像降维的主要方法之一。最早利用聚类思想进行波段选择的方法是 Martinez-Uso 等^[19]在 2007 年提出的利用互信息的 Ward 联动策略(WaLuMI)和利用发散的 Ward 联动策略(WaLuDI),它们是基于 Ward's linkage 的分层聚类算法。2009 年, Qian 等^[20]提出一种无监督方法,它是亲和传播(AP)算法。之后 Su 等^[21]又在 AP 的基础上提出一种基于样本的半监督的自适应 AP,而这种方法在聚类过程中需要标签样本数据,但在实际应用中,获取地面真实标签是困难的,因此无监督方法可行性更大。Yang 等^[22]在 K-均值聚类基础上提出代表性波段挖掘方法, Yuan 等^[23]将 K-均值聚类算法扩展为利用上下文分析的双聚类方法。这两种基于 K-均值聚类的方法属于硬聚类,即它们将一个波段归属到特定的一个类别,而不是像软聚类一样给出一个波段归属于每个类别的概率值。然而,由于光谱波段本身的复杂性,一个波段可能同时属于不同簇,硬聚类在进行对象划分时不够精确,而计算对象对每一类的隶属度的模糊聚类方法能够很好地解决该问题。

本文提出的 FCM-FA 在模糊 C-均值聚类(FCM)^[24]的基础上,通过加入萤火虫算法(FA)^[25]对目标函数进行优化来选择波段,是一种无监督波段选择方法。FCM 属于软聚类方法,允许样本以不同概率同时属于多个簇,但随机初始化过程中容易获得一个局部最优解。FA 是一种全局搜索算法,能够利用萤火虫在全局的移动找到全局最优解。实验结果表明,选择特定波段的 FCM 能够获得优秀的分类精度,而加入 FA 之后,FCM 的效果有明显提高。

2 所提算法内容

2.1 模糊 C-均值聚类

FCM 算法通过计算出每个对象在不同子集中的

隶属度,将对象划分到隶属度值最大的对应簇中,每个隶属度值的确定可以定义为最优化问题:

$$\min_{u_{ij}} J = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m d_{ij}^2 \quad (1)$$

$$\text{s. t. } \begin{cases} u_{ij} \in [0, 1] \\ \sum_{j=1}^C u_{ij} = 1, \quad \forall i = 1, 2, \dots, N \\ d_{ij} = \|\mathbf{c}_j - \mathbf{x}_i\| \end{cases} \quad (2)$$

式中: N 为聚类对象的总数; C 为聚类的子集数; u_{ij} 为第 i 个对象 \mathbf{x} 对于第 j 个类的隶属度,由实际物理意义可知,一个对象对于所有类的隶属度之和为 1; d_{ij} 为第 i 个对象 \mathbf{x} 与第 j 个聚类中心 \mathbf{c} 之间的欧氏距离; m 为 $[1, \infty)$ 范围的控制隶属度的加权指数,通常为 2。

对于式(1),可以采用拉格朗日乘子法进行求解,结果为

$$\bar{J} = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m d_{ij}^2 + \sum_{i=1}^N \lambda_i \left(\sum_{j=1}^C u_{ij} - 1 \right) \quad (3)$$

\bar{J} 对 u_{ij} 求偏导数,并令其等于 0,根据 u_{ij} 的约束条件可以推导出隶属度计算公式:

$$u_{ij} = \left[\sum_{s=1}^C \left(\frac{\|\mathbf{c}_j - \mathbf{x}_i\|}{\|\mathbf{c}_s - \mathbf{x}_i\|} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (4)$$

将式(4)代入式(3)中, \bar{J} 对 \mathbf{c}_j 求偏导数,并令其等于 0,推导可得聚类中心计算公式:

$$\mathbf{c}_j = \frac{\sum_{i=1}^N u_{ij}^m \mathbf{x}_i}{\sum_{i=1}^N u_{ij}^m} \quad (5)$$

由式(4)可知,在计算隶属度时,需要已知聚类中心坐标,而通过式(5)计算 \mathbf{c}_j 时,又需要先计算 u_{ij} ,因此在计算时,通常随机初始化一个满足 u_{ij} 约束条件的隶属度矩阵 \mathbf{U} ,再计算出聚类中心矩阵 \mathbf{C} ,由此通过迭代进行计算,直到达到迭代次数或满足收敛后停止迭代,收敛条件为 $J^{(t+1)} - J^{(t)} < \epsilon$ 或 $\max \|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \epsilon$, ϵ 为误差阈值,通常为 0.0001。

由 FCM 的聚类过程可以看出,聚类结果一定程度上依赖于隶属度矩阵 \mathbf{U} 的初始化效果,不合适的初始化结果会带来不良的聚类结果,且由于拉格朗日乘子的特点,往往获得的结果是一个局部最优解。

2.2 萤火虫算法

群智能算法是一种基于生物群体社会行为的智能优化算法,也是基于概率型的全局优化算法,能够有效得到全局最优解,与传统梯度方法和传统演化算法相比,具有鲁棒性强、自组织能力好、易于实现等优点^[26]。Yang^[25]根据萤火虫的生物学特性提出了一种全局搜索算法,即萤火虫算法,该算法在效率和成功率方面均优于粒子群算法和遗传算法。

根据萤火虫发光程度的强弱,评价萤火虫所在位置的优劣,亮度高的萤火虫所在位置更优,具有更大的吸

引力,亮度低的萤火虫向亮度高的萤火虫移动,亮度相同的萤火虫则随机移动。亮度和吸引力是FA的两个关键元素。从数学的角度可以对FA进行以下描述。

\mathbf{z}_k 表示第 k 只萤火虫的位置, $r_{kl} = \|\mathbf{z}_k - \mathbf{z}_l\|$ 表示第 k 只萤火虫和第 l 只萤火虫之间的欧氏距离, 萤火虫 k 看到萤火虫 l 的亮度表示为

$$I_{kl} = I_0 \exp(-\gamma r_{kl}^2), \quad (6)$$

式中: I_0 为萤火虫 l 自身最大亮度, 由被优化的目标函数决定; γ 为 $[0, \infty)$ 范围的吸光系数。萤火虫 k 看到的最大亮度的萤火虫 l^* 为

$$l^* = \arg \max(I_{kl}). \quad (7)$$

萤火虫 l 对萤火虫 k 的吸引力定义为

$$\beta(r_{kl}) = \beta_0 \exp(-\gamma r_{kl}^2), \quad (8)$$

式中: β_0 为 $r=0$ 时的最大吸引力。

根据吸引力和第 l^* 只萤火虫的位置, 更新第 k 只萤火虫的位置, 可以表示为

$$\mathbf{z}_k^{(t+1)} = \mathbf{z}_k^{(t)} + \beta(r_{kl^*})(\mathbf{z}_{l^*} - \mathbf{z}_k^{(t)}) + \alpha(N_{\text{rand}} - \frac{1}{2})\mathbf{I}, \quad (9)$$

式中: α 为步长因子, 是一个 $[0, 1]$ 范围的常数; N_{rand} 为 $[0, 1]$ 范围的随机数; \mathbf{I} 是和 \mathbf{z}_k 一样的列向量, 且该列向量所有元素值为 1。

FA 的搜索过程可以简单描述为: 初始化所有萤火虫位置后, 根据目标函数和式(6)计算萤火虫亮度; 根据式(7)确定所看到的最高亮度萤火虫并利用式(8)

计算吸引力; 根据式(9)更新萤火虫位置, 根据新的位置重新计算亮度和吸引力并再次移动, 直到满足最大移动次数或目标函数收敛时停止迭代计算, 得到每只萤火虫的最终位置, 选出全局最优解。由上述可知, FA 可以提高获得全局最优解的概率, 能够很好地弥补FCM的缺点, 因此可以利用FA对式(1)进行优化。

2.3 基于模糊C-均值聚类的萤火虫算法

将HSI看作数据集 X , 每一个波段为聚类对象, 所选择的波段数量为聚类数。与文献[14]不同的是, 这里没有将波段的索引值作为每一只萤火虫的位置坐标, 而是直接将聚类中心矩阵 C 作为一只萤火虫的位置, 矩阵 C 中的一列为一个聚类中心的坐标, 其维度为单波段的像素数量, 具体数值表示为高光谱图像的灰度值, 将灰度值视为坐标值, 通过距离度量, 能够更加直接地体现出波段之间的相关性。并且, 由于式(6)没有参与式(9)的计算, 因此可以直接将目标函数值作为判断依据, 目标函数值大的萤火虫向目标函数值小的萤火虫移动, 通过萤火虫的移动代替式(5)的聚类中心计算, 具体过程如图1所示。本文将FA融入到FCM的计算过程中, 针对HSI数据邻近波段数据具有相关性的特点, 在利用数据聚类原理去除冗余信息的同时, 针对HSI数据量大、易在聚类过程中陷入局部最优的特点, 又利用萤火虫算法的随机搜索特性增大跳出局部最优解的概率, 最终形成了完整的波段选择算法。

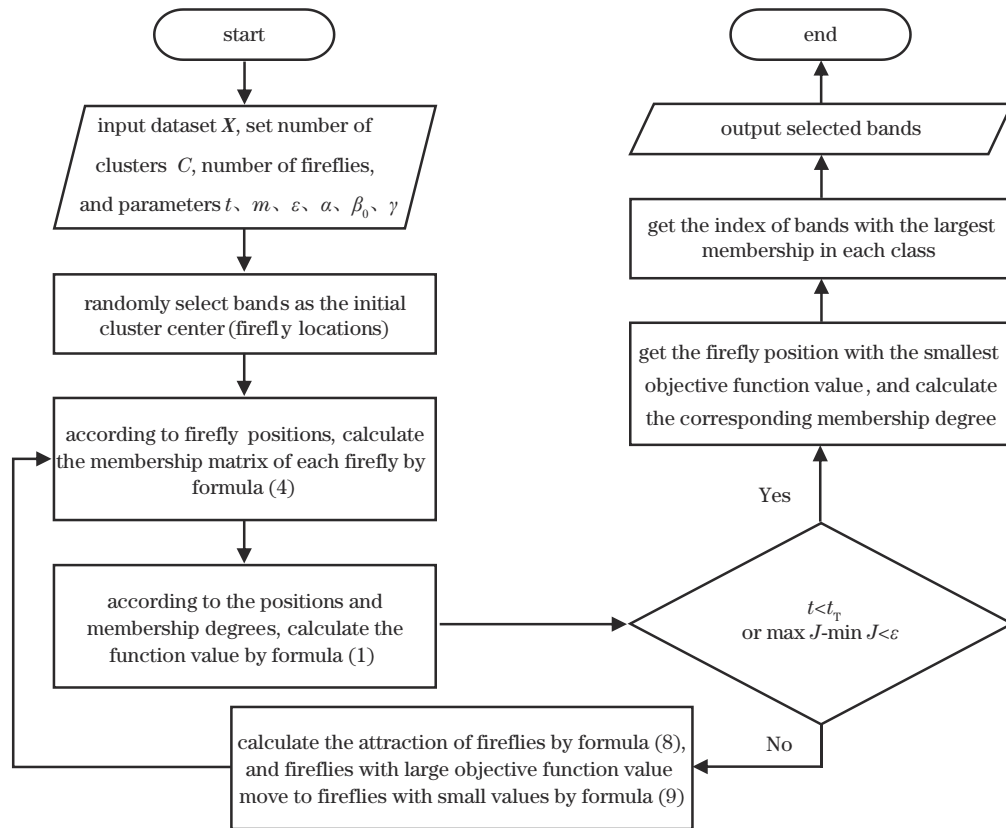


图1 FCM-FA 波段选择流程

Fig. 1 FCM-FA band selection process

3 实验与分析

在两个公开 HSI 数据集上进行分类实验对比,来证明所提方法的优越性。首先介绍两个数据集的基本信息;其次介绍对比方法和参数设置及分类实验所选取的分类器和评价标准;然后对分类结果进行分析,并对所选波段形成的光谱特征曲线与全波段特征曲线进行对比;接着比较不同方法的计算时间;最后分析萤火虫数量对所选波段质量和计算时间的影响。

3.1 实验数据

Indian Pines 数据集:该数据集通过 AVIRIS 传感器在 1992 年对位于印第安纳州西北部的印第安农林试验场进行采集而获得,由 145×145 个像素和 224 个光谱反射波段组成,波长范围为 $0.4 \sim 2.5 \mu\text{m}$,空间分辨率为 20 m,光谱分辨率为 10 nm。去除覆盖吸水区域的波段(第 104~108、150~163、220 波段),将波段的数量减少到 200 个。实际地物可分为 16 类。

PaviaU 数据集:该数据集通过 ROSIS 传感器于 2003 年在意大利北部帕维亚城进行拍摄而获得,由 610×340 个像素和 115 个光谱反射波段组成,波长范围为 $0.43 \sim 0.86 \mu\text{m}$,空间分辨率为 1.3 m。去除 12 个受噪声影响的波段,剩余 103 个波段。实际地物可分为 9 类。

3.2 对比方法及参数设置

选取基于稀疏的方法:改进的稀疏子空间聚类

(ISSC)^[9]和基于稀疏的波段选择(SpaBS)^[10]。选取基于排序的方法:基于排序的高效图卷积自表示(EGCSR-R)^[27]。选取基于聚类的方法:基于聚类的高效图卷积自表示(EGCSR-C)^[27]。选取基于搜索的方法:基于正交投影的波段选择法(OPBS)^[15]。对这些方法得到的结果与FCM和所有波段的分类结果进行对比。ISSC的正则化参数设置为0.001;SpaBS的稀疏度为0.05;EGCSR-R和EGCSR-C的邻近值设置为10;FCM的迭代次数为100,误差阈值 ϵ 为0.0001,加权指数 m 为2;FCM-FA的迭代次数、阈值和加权指数与FCM相同,使用10只萤火虫进行搜索,步长因子 α 为0.5, β_0 为1,吸光系数 γ 为 10^{-12} 。

3.3 不同波段数量分类实验

实验选取支持向量机(SVM)和K最邻近(KNN)作为分类器。SVM分类器选取径向基函数(RBF)作为核函数,惩罚系数和核函数系数通过交叉验证确定,误差精度为0.001。KNN邻近数量为3,距离采用欧氏距离。使用每种地物20%的像素作为训练样本。使用总体分类精度(OA)和Kappa系数对分类结果进行评估。Indian Pines数据集中选择波段数量为4~40个,步长为4;PaviaU数据集中选择3~30个波段,步长为3。对于所有实验,均采用10次独立实验的平均结果作为最终结果。

图2展示了在Indian Pines数据集上使用不同方法选择波段子集,作为分类器的输入数据后得到的分类

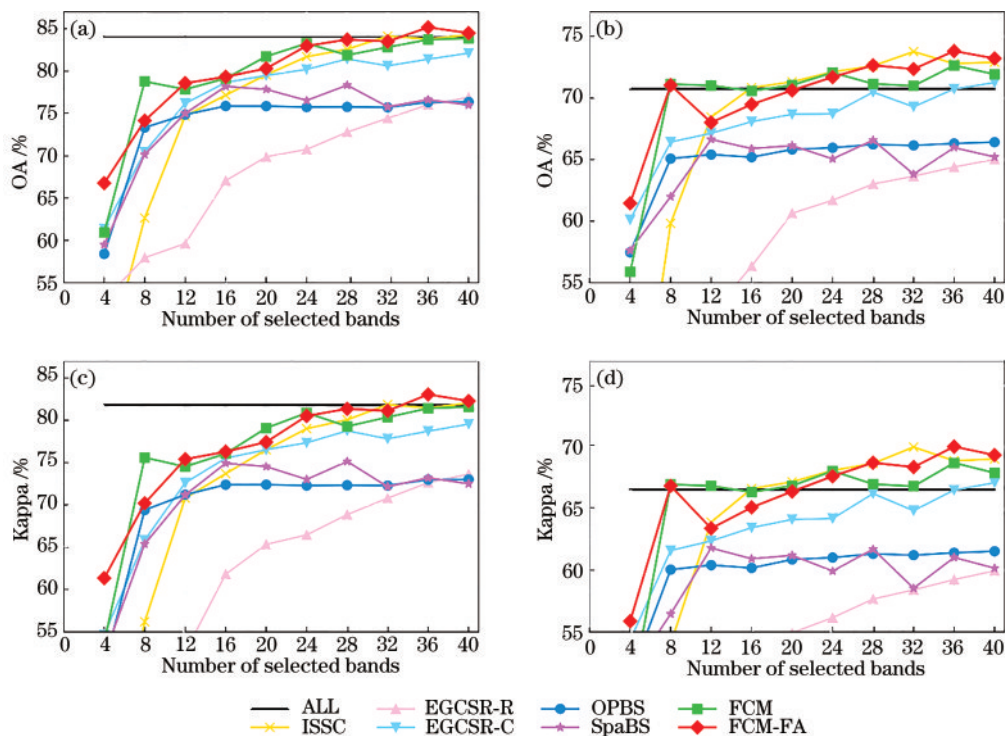


图2 Indian Pines数据集上不同方法的分类精度。(a)SVM分类器的OA;(b)KNN分类器的OA;(c)SVM分类器的Kappa系数;(d)KNN分类器的Kappa系数

Fig. 2 Classification accuracy of different methods on Indian Pines dataset. (a) OA of the SVM classifier; (b) OA of the KNN classifier; (c) Kappa coefficient of the SVM classifier; (d) Kappa coefficient of the KNN classifier

精度。测试了 7 种不同波段选择方法和 2 种分类器(即 SVM 和 KNN),展示了两种分类精度衡量指标(即总体分类精度 OA 和 Kappa 系数)。SVM 分类器在利用 FCM-FA 选择的 8、20、24 个波段上的分类精度仅次于在 FCM 选择的波段上的分类精度,在使用 32 个波段的分类精度略低于 ISSC 外,其他地方分类精度均为第一,且在使用 28 个波段时的分类精度能够和使用所有波段的分类精度相近,并在 36 和 40 个波段的分类精度明显超过了所有波段的分类精度;ISSC 和 FCM 在 40 个波段也能够达到所有波段的分类精度。KNN 分类器上,虽然 FCM-FA 在使用 12 个波段时分类精度有一个下降,但在 20 个波段时与 FCM 和 ISSC 的精度就没有太大区别,并达到了所有波段的分类精度,且随着波段数的增加逐渐优于所有波段的分类精度,24 个波段之后明显优于 FCM 的分类精度,在波段子集数量为 28、36、40 时精度为第一。从图 2 也可以看出:OPBS 在从 4 个波段到 8 个波段时精度能够有一个较大的提升,但之后便趋于平稳,没有明显提升;SpaBS 精度不太稳定;EGCSR-C 虽然随着波段数量的增加能够不断提升精度,但与 FCM-FA 的精度有明显差别;EGCSR-R 分类精度同样一直在

提升,但仍然很差。

图 3 展示了在 PaviaU 数据集上在两个分类器上 7 种方法的分类精度。在 SVM 分类器上,仅在使用 3 个和 12 个波段数量上,FCM-FA 分类精度仅略低于 FCM,其余波段数量上的分类精度均为第一,且在使用 18 个波段时就已经达到使用所有波段的分类精度,并一直保持,在波段数为 27 个处达到最大分类精度,为 94.22%,Kappa 系数为 92.33%。在 KNN 分类器上,FCM-FA 仅在使用 3 个波段处的分类精度略低于 FCM,在其他情况下分类精度均为第一,在波段子集数量为 15 时接近使用所有波段进行分类的分类精度,OA、Kappa 均优于所有波段的分类精度。FA 对 FCM 的优化有明显效果,使用 21 个波段时 OA 最高提高了 3.12 个百分点、Kappa 系数最高提高了 4.26 个百分点。在 Indian Pines 数据集上表现优秀的 ISSC 的分类精度在 PaviaU 数据集有所下降,整体趋势低于 FCM-FA、FCM、OPBS、EGCSR-C 方法;OPBS 在 PaviaU 数据集上的表现好于在 Indian Pines 数据集上的表现,在波段数量较多时分类精度与 FCM 相近;EGCSR-C 虽然大多数时候有较好的分类精度,但是有较大起伏;SpaBS 和 EGCSR-R 的分类效果依然很差。

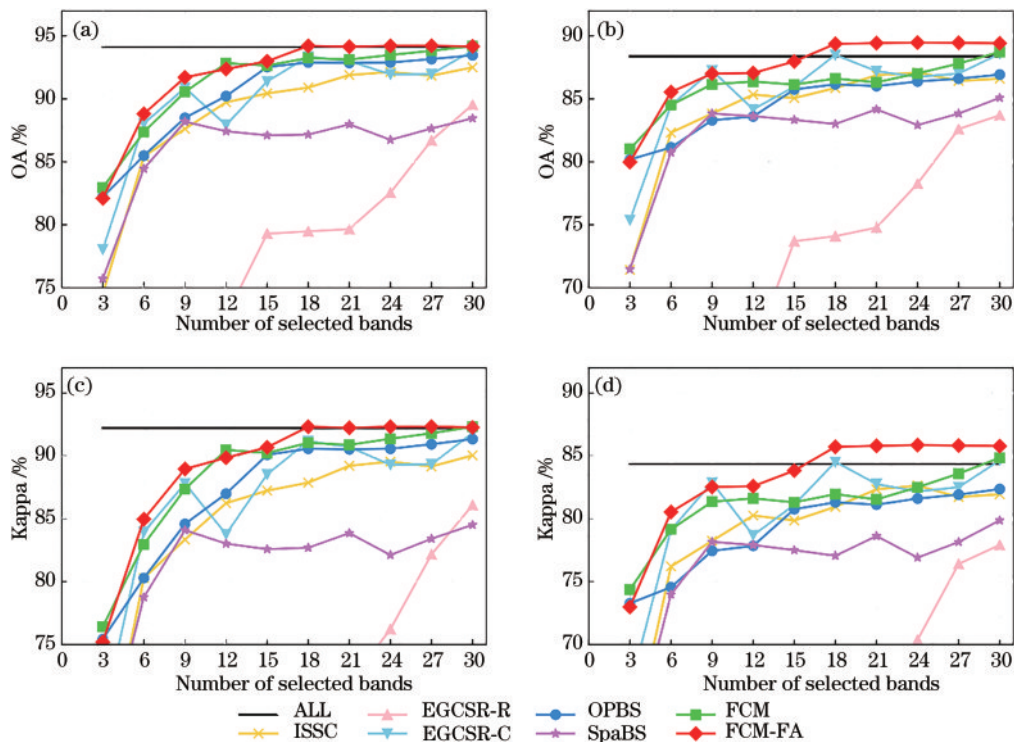


图 3 PaviaU 数据集上不同方法的分类精度。(a)SVM 分类器的 OA;(b)KNN 分类器的 OA;(c)SVM 分类器的 Kappa 系数;(d)KNN 分类器的 Kappa 系数

Fig. 3 Classification accuracy of different methods on PaviaU dataset. (a) OA of SVM classifier; (b) OA of KNN classifier; (c) Kappa coefficient of SVM classifier; (d) Kappa coefficient of KNN classifier

在两个分类器上利用平均分类精度(AA)对所选的不同波段数量和不同训练样本率的分类精度进行评价,AA 数值大小与 OA 近似,这里不再进行图片展示,

精度趋势可以参考 OA 图像。

由上述分析可得,在大多数情况下 FCM-FA 能够获得最好的分类精度,少数不是最好的情况下也能够

排在第二、三名,且与第一名的精度差别不大。FCM-FA 选取少数波段就能够达到所有波段的分类效果,且在 KNN 分类器上,在两个数据集中选取一定数量的波段后的分类精度明显优于使用所有波段后的分类精度,有效地达到了降维的目的。

3.4 不同训练样本分类实验和地物光谱特征曲线

为了验证不同训练样本比例与波段选择结果之间的关系,应用不同波段选择算法提取相同数量的波段,再利用波段选择结果在不同分类器和不同训练样本比例的条件进行分类,训练样本选取 10%~80%,步长为 10%。由 3.3 节实验结果可知:SVM 分类器得到的分类精度比 KNN 分类器高;在 SVM 分类器上,Indian Pines 数据集上在选择 28 个波段和 PaviaU 数据集上在选择 18 个波段时的分类精度与使用所有波段的分类

精度相近。因此使用上述波段数量进行分类,分类器设置和评估方法与 3.3 节相同。

图 4 展示了在 Indian Pines 数据集上不同波段选择方法选择 28 个波段作为两个分类器的输入数据,并使用不同比例的训练样本后的分类精度。两个分类器上,FCM-FA 的两种精度评价指标均为第一,并且在 KNN 分类器上的分类精度能够超过使用所有波段的分类精度,OA 最多提高 3.63 个百分点、Kappa 最多提高 4.17 个百分点。ISSC 和 FCM 在 SVM 分类器上的分类精度几乎一致。KNN 分类器上,ISSC 和 FCM 的分类精度同样高于使用所有波段的分类精度;EGCSR-C 的分类精度与使用所有波段的精度相近,当训练样本超过 50% 时,EGCSR-C 的分类精度能够略高于使用所有波段的分类精度。

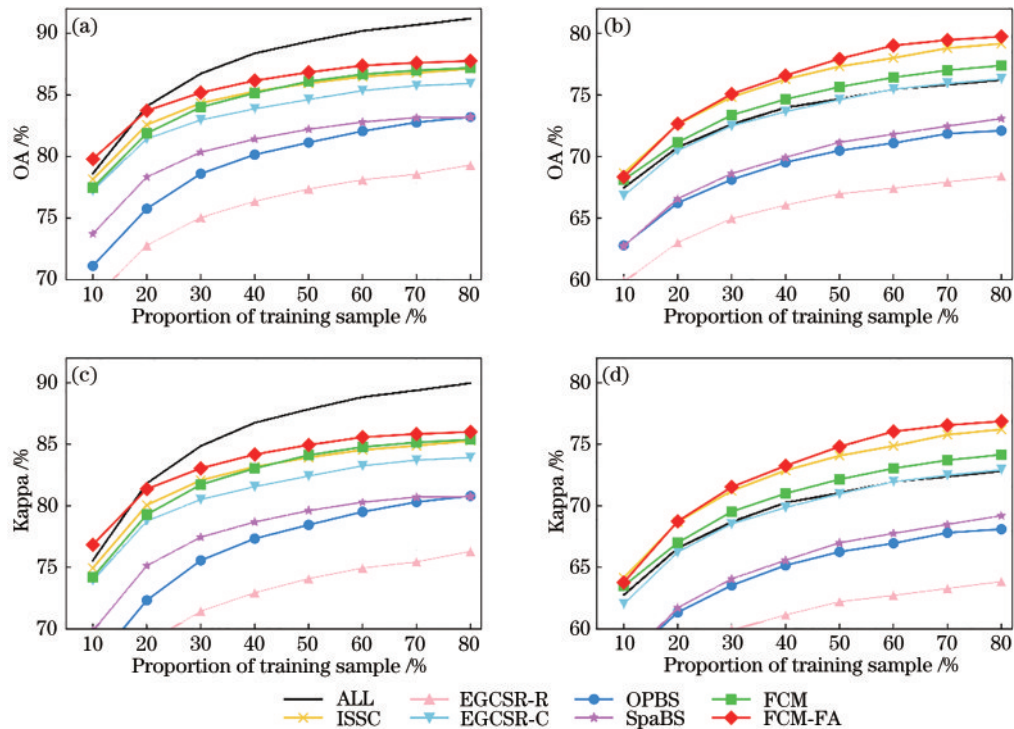


图 4 Indian Pines 数据集上不同方法在所选择的 28 个波段下使用不同比例的训练样本的分类精度。(a)SVM 分类器的 OA;

(b)KNN 分类器的 OA;(c)SVM 分类器的 Kappa 系数;(d)KNN 分类器的 Kappa 系数

Fig. 4 Classification accuracy of different methods in 28 selected bands on Indian Pines dataset using different proportions of training samples. (a) OA of SVM classifier; (b) OA of KNN classifier; (c) Kappa coefficient of SVM classifier; (d) Kappa coefficient of KNN classifier

图 5 展示了在 PaviaU 数据集上选取 18 个波段和使用不同训练样本时方法的分类精度。在 SVM 分类器上,选择特定波段子集的 FCM-FA 的分类精度与使用全波段的分类精度几乎相同,除 10% 和 30% 的训练样本外,在其余样本率情况下选择特定波段子集的 FCM-FA 所能达到的分类精度均高于使用全波段的分类精度,且优于其他方法。KNN 分类器上,EGCSR-C 的分类精度与使用全波段的分类精度相似,但低于 FCM-FA;FCM-FA 相较于使用所有波段,

OA 最少提高了 0.62 个百分点、Kappa 最少提高了 0.85 个百分点,OA 最多提高了 0.99 个百分点、Kappa 最多提高了 1.37 个百分点;EGCSR-R 方法由于分类精度过低,在图 5 的取值范围内显示不出;SpaBS 的 Kappa 值过低,不能显示。

由图 4 和图 5 可得,随着训练样本数量的增加,分类精度也随之增加,但基本不改变波段选择方法的排名情况,所提波段选择算法对待分类图像的训练样本数量具有较好的鲁棒性。

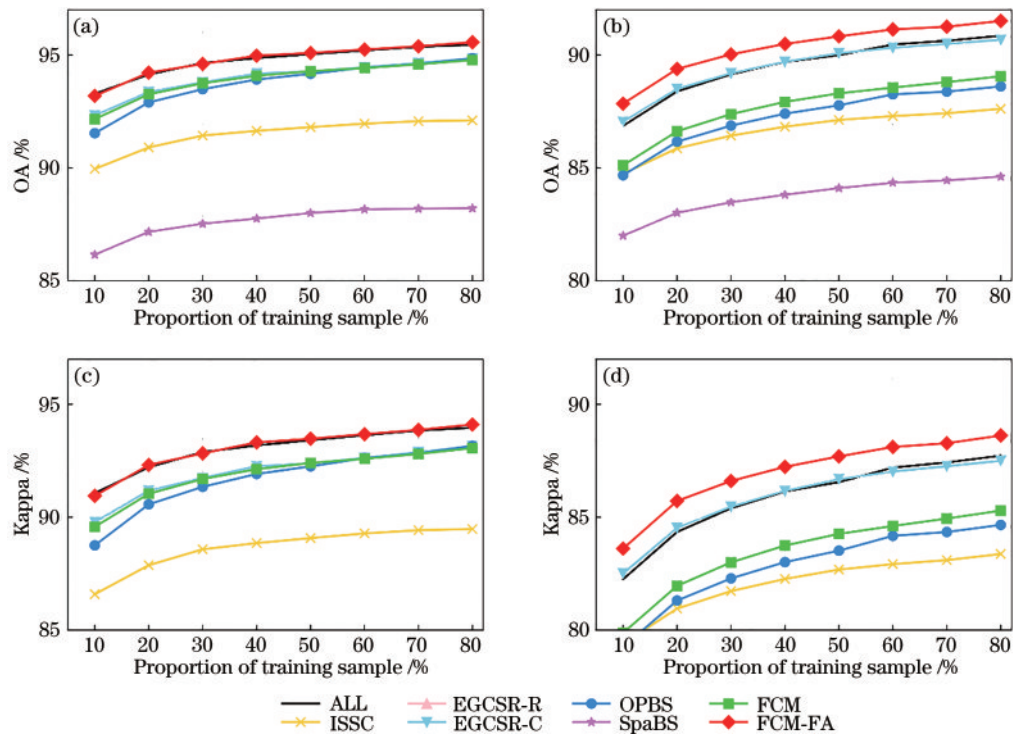


图5 PaviaU数据集上不同方法所选择的18个波段使用不同比例的训练样本的分类精度。(a)SVM分类器的OA;(b)KNN分类器的OA;(c)SVM分类器的Kappa系数;(d)KNN分类器的Kappa系数

Fig. 5 Classification accuracy of different methods in 18 selected bands on PaviaU dataset using different proportions of training samples. (a) OA of SVM classifier; (b) OA of KNN classifier; (c) Kappa coefficient of SVM classifier; (d) Kappa coefficient of KNN classifier

地物的光谱特征曲线是判断地物的重要依据,利用FCM-FA所选择的Indian Pines数据集28个波段、PaviaU数据集18个波段和原始数据的全部波段,以波段编号为横坐标,单波段内同类地物像素灰度值的均值为纵坐标,绘制了地物光谱特征图。图6为Indian Pines数据集1~16类的特征图,图7为PaviaU数据集1~9类的特征图,虚线为原始全波段的曲线,实线为FCM-FA选择波段的曲线。由图6和图7可以看出,选择的波段子集能够很好地保留波段全集的地物特征曲线峰值处的波段,很好地反映了地物光谱特征曲线的走势。虽然Indian Pines数据集中地类2、3、4、10、11、12、16的第40波段处,地类7的第30波段处,PaviaU数据集中地类3的第76、85波段,地类4、9的第88波段,地类5的第76波段,地类8的第79、85波段处丢失了特征曲线的极值点,但均不影响特征曲线的整体走势,依然可以代表原始光谱曲线。图7的PaviaU数据集中9种地类的光谱特征曲线有明显差异,能够很好地借助光谱特征对地物进行区分;而图6的Indian Pines数据集中16种地类的特征曲线走势相似,这是因为该数据的大多数地物为农作物,但不同地物之间特征曲线走势仍然具有差别,且部分极值点的大小也会有明显差距。

3.5 复杂度和萤火虫数量对所选波段质量与计算时间的影响

计算复杂度以计算时间进行评估,选择的波段数量同样为Indian Pines数据集28个波段,PaviaU数据集18个波段。表1列出了7种方法在两个数据集上的运行时间。实验基于Python 3.9,使用pycharm2021.3.1在3.70 GHz的Intel i9-10900X CPU上运行,内存大小为64 GB。速度最快的是EGCSR-R,计算时间不足1 s,但需要提醒的是,该方法在分类精度上表现极差,即使拥有较快的运行速度,但该方法仍不推荐使用;ISSC与EGCSR-C在Indian Pines数据集上有相似的计算时间,但在PaviaU数据集上,EGCSR-C拥有更少的计算时间,甚至比在Indian Pines数据集上时间更短;OPBS同样拥有一个较好的计算时间;FCM的计算时间为中等;SpaBS与FCM-FA的计算时间更长;FCM-FA时间最长。与FCM相比,FCM-FA有更长时间的原因:首先,FCM迭代的是一组聚类中心,而FCM-FA中每只萤火虫都是一组聚类中心,迭代过程是计算10只萤火虫的位置;其次,由于FCM计算的是局部最优解,因此在较少的迭代次数内就能够得到结果,而FCM-FA需要更多的迭代次数才能得到全局最优解。

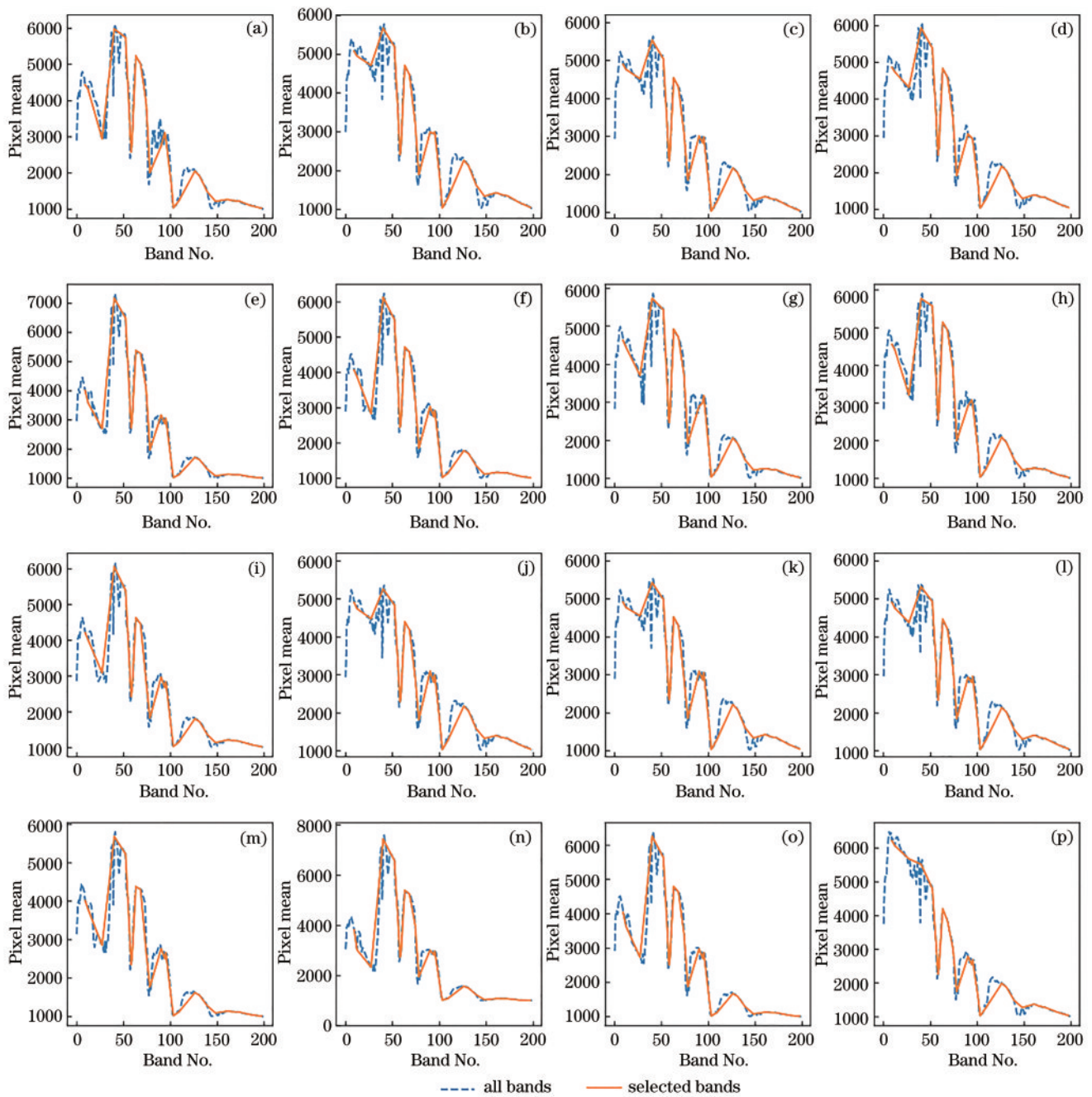


图 6 Indian Pines 数据集中地类光谱特征曲线。(a)~(p)地类 1~16 的特征曲线

Fig. 6 Spectral characteristic curves of land type on Indian Pines dataset. (a)~(p) Characteristic curves of land type 1~16, respectively

在其他参数相同情况下研究不同萤火虫数量对波段选择结果的影响,包括波段的质量和波段选择的时间消耗,波段质量通过分类精度间接评价,分类参数与 3.3 节相同,波段数量与计算复杂度设置相同。根据表 2 可知,在 Indian Pines 和 PaviaU 数据集中,FCM-FA 的计算时间均会随着萤火虫数量的增加而增加。从表 3 中可以看出:在 Indian Pines 数据

集上,除 SVM 分类器上 FCM-FA 对 12 只萤火虫的分类精度比对 10 只萤火虫的分类精度略低外,其余分类精度均会随着萤火虫数量的增加而增加;在 PaviaU 数据集上,FCM-FA 的分类精度并没有随着萤火虫数量的增加而增加,这是因为所选波段下已经是全局最优解,不会再因萤火虫数量的增加而改变。

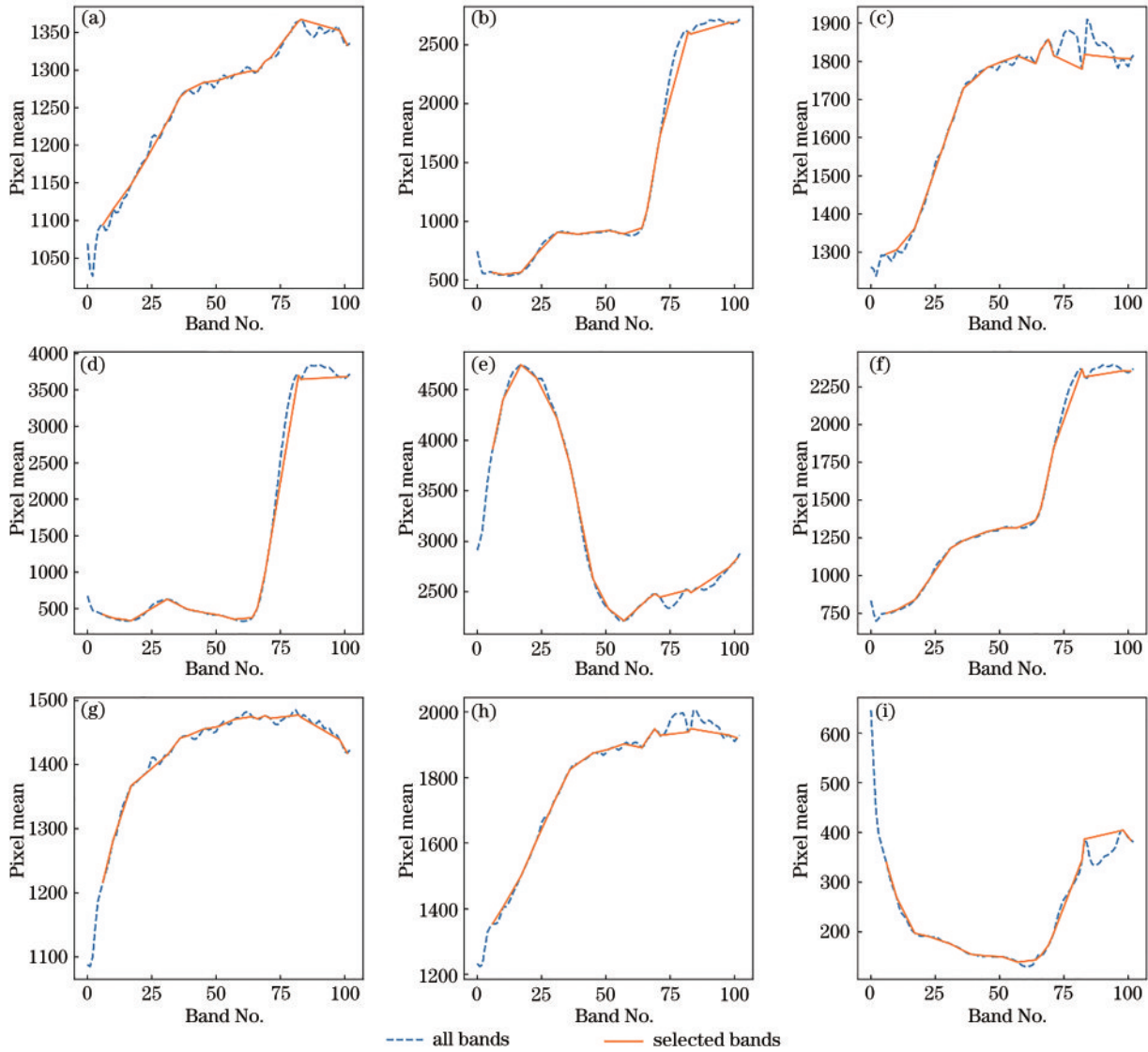


图 7 PaviaU 数据集中地类光谱特征曲线。(a)~(i)地类 1~9 的特征曲线

Fig. 7 Spectral characteristic curves of land type on PaviaU dataset. (a)~(i) Characteristic curves of land type 1~9, respectively

表 1 不同方法的计算时间

Table 1 Calculation time of different methods

unit: s

Dataset	ISSC	EGCSR-R	EGCSR-C	OPBS	SpaBS	FCM	FCM-FA
Indian Pines	9.47	0.12	8.37	4.23	127.13	42.01	163.69
PaviaU	40.63	0.54	2.19	11.22	542.49	258.19	553.92

表 2 FCM-FA 对不同萤火虫数量的计算时间

Table 2 Calculation time of FCM-FA for different fireflies

unit: s

Dataset	$N=5$	$N=8$	$N=10$	$N=12$	$N=15$
Indian Pines	155.19	160.91	163.69	170.61	176.33
PaviaU	523.84	539.57	553.92	563.52	572.99

表 3 萤火虫数量的设置对 FCM-FA 波段选择结果的影响(波段质量通过分类精度间接评价)

Table 3 Effect of the setting of number of fireflies on results of FCM-FA band selection (band quality is indirectly evaluated through classification accuracy)

Dataset	Classifier	Parameter	$N=5$	$N=8$	$N=10$	$N=12$	$N=15$
Indian Pines	SVM	OA	80.368	83.340	83.722	83.568	85.134
		AA	80.216	83.334	83.714	83.536	85.035

表 3 (续)

Dataset	Classifier	Parameter	N=5	N=8	N=10	N=12	N=15
PaviaU	KNN	Kappa	77.514	80.929	81.366	81.193	83.008
		OA	66.557	72.434	72.655	72.686	74.999
		AA	66.079	72.237	72.521	72.603	74.886
		Kappa	61.708	68.491	68.743	68.790	71.400
		OA	94.214	94.214	94.214	94.214	94.214
		AA	94.199	94.199	94.199	94.199	94.199
	SVM	Kappa	92.315	92.315	92.315	92.315	92.315
		OA	89.379	89.379	89.379	89.379	89.379
		AA	89.393	89.393	89.393	89.393	89.393
		Kappa	85.720	85.720	85.720	85.720	85.720
		OA	89.379	89.379	89.379	89.379	89.379
		AA	89.393	89.393	89.393	89.393	89.393

4 结 论

针对高光谱图像数据特征冗余问题,在模糊 C-均值聚类算法的基础上加入全局搜索方法即萤火虫算法,提出了一种波段选择算法,称 FCM-FA。该方法优化了聚类中心的计算方法,能够有效地找到全局最优解。实验结果表明:萤火虫算法的加入有效地提升了 FCM 的分类精度,且 FCM-FA 所选择的波段能够达到甚至超过所有波段的分类精度,达到了降维的目的。FCM-FA 同时利用了聚类方法的去除冗余的优势和搜索方法的求解全局最优解的优势,为以后的算法研究提供了新的思路。未来的研究工作将针对两方面开展:针对选择的波段数量较少时分类精度不佳方面进行改进;针对计算时间较长方面进行优化。

参 考 文 献

- [1] Li J X, Zheng K, Yao J, et al. Deep unsupervised blind hyperspectral and multispectral data fusion[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 6007305.
- [2] Hughes G. On the mean accuracy of statistical pattern recognizers[J]. IEEE Transactions on Information Theory, 1968, 14(1): 55-63.
- [3] Agarwal A, El-Ghazawi T, El-Askary H, et al. Efficient hierarchical-PCA dimension reduction for hyperspectral imagery[C]//2007 IEEE International Symposium on Signal Processing and Information Technology, December 15-18, 2007, Giza, Egypt. New York: IEEE Press, 2008: 353-356.
- [4] Wang J, Chang C I. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(6): 1586-1600.
- [5] Jia S, Qian Y T, Li J M, et al. Feature extraction and selection hybrid algorithm for hyperspectral imagery classification[C]//2010 IEEE International Geoscience and Remote Sensing Symposium, July 25-30, 2010, Honolulu, HI, USA. New York: IEEE Press, 2010: 72-75.
- [6] 陆永帅, 李元祥, 刘波, 等. 基于深度残差网络的高光谱遥感数据霾监测[J]. 光学学报, 2017, 37(11): 1128001. Lu Y S, Li Y X, Liu B, et al. Hyperspectral data haze monitoring based on deep residual network[J]. Acta Optica Sinica, 2017, 37(11): 1128001.
- [7] 吴超, 吴一全. 基于混沌粒子群优化投影寻踪的高光谱图像目标检测[J]. 光学学报, 2011, 31(12): 1211003. Wu C, Wu Y Q. Target detection in hyperspectral image using projection pursuit based on chaotic particle swarm optimization[J]. Acta Optica Sinica, 2011, 31(12): 1211003.
- [8] 王德港, 饶伟强, 孙旭, 等. 结合孪生网络和像素配对的高光谱图像异常检测[J]. 中国图象图形学报, 2021, 26(8): 1860-1870. Wang D G, Rao W Q, Sun X, et al. Siamese network with pixel-pair for hyperspectral image anomaly detection [J]. Journal of Image and Graphics, 2021, 26(8): 1860-1870.
- [9] Sun W W, Zhang L P, Du B, et al. Band selection using improved sparse subspace clustering for hyperspectral imagery classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(6): 2784-2797.
- [10] Li S J, Qi H R. Sparse representation based band selection for hyperspectral images[C]//2011 18th IEEE International Conference on Image Processing, September 11-14, 2011, Brussels, Belgium. New York: IEEE Press, 2011: 2693-2696.
- [11] Cai Y M, Liu X B, Cai Z H. BS-nets: an end-to-end framework for band selection of hyperspectral image[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(3): 1969-1984.
- [12] 关世豪, 杨桃, 卢珊, 等. 基于注意力机制的多目标优化高光谱波段选择[J]. 光学学报, 2020, 40(21): 2128002. Guan S H, Yang G, Lu S, et al. Multi-objective optimization of hyperspectral band selection based on attention mechanism[J]. Acta Optica Sinica, 2020, 40(21): 2128002.
- [13] Su H J, Du Q, Chen G S, et al. Optimized hyperspectral band selection using particle swarm optimization[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(6): 2659-

- 2670.
- [14] Su H J, Yong B, Du Q. Hyperspectral band selection using improved firefly algorithm[J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(1): 68-72.
- [15] 张文强. 基于向量子空间投影的高光谱图像无监督波段选择技术研究[D]. 杭州: 浙江大学, 2020.
Zhang W Q. Research on unsupervised hyperspectral band selection based on vector subspace projection[D]. Hangzhou: Zhejiang University, 2020.
- [16] Chang C I, Du Q, Sun T L, et al. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1999, 37(6): 2631-2641.
- [17] Su H J, Yang H, Du Q, et al. Semisupervised band clustering for dimensionality reduction of hyperspectral imagery[J]. *IEEE Geoscience and Remote Sensing Letters*, 2011, 8(6): 1135-1139.
- [18] Wang Q, Zhang F H, Li X L. Optimal clustering framework for hyperspectral band selection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(10): 5910-5922.
- [19] Martinez-Uso A, Pla F, Sotoca J M, et al. Clustering-based hyperspectral band selection using information measures[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, 45(12): 4158-4171.
- [20] Qian Y, Yao F, Jia S. Band selection for hyperspectral imagery using affinity propagation[J]. *IET Computer Vision*, 2009, 3(4): 213-222.
- [21] Su H J, Sheng Y H, Du P J, et al. Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection[J]. *Applied Optics*, 2012, 51(14): 2656-2663.
- [22] Yang R L, Su L F, Zhao X B, et al. Representative band selection for hyperspectral image classification[J]. *Journal of Visual Communication and Image Representation*, 2017, 48: 396-403.
- [23] Yuan Y, Lin J Z, Wang Q. Dual-clustering-based hyperspectral band selection by contextual analysis[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(3): 1431-1445.
- [24] Bezdek J C, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm[J]. *Computers & Geosciences*, 1984, 10(2/3): 191-203.
- [25] Yang X S. Firefly algorithms for multimodal optimization [M]//Watanabe O, Zeugmann T. Stochastic algorithms: foundations and applications. Lecture notes in computer science. Heidelberg: Springer, 2009, 5792: 169-178.
- [26] 张庆科. 粒子群优化算法及差分进化算法研究[D]. 济南: 山东大学, 2017.
Zhang Q K. Research on the particle swarm optimization and differential evolution algorithms[D]. Jinan: Shandong University, 2017.
- [27] Cai Y M, Zhang Z J, Liu X B, et al. Efficient graph convolutional self-representation for band selection of hyperspectral image[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 4869-4880.