

引用格式: CHEN Jing, ZHANG Zhenxing. Greedy Unsupervised Hyperspectral Image Band Selection Method Based on Variable Precision Rough Set[J]. Acta Photonica Sinica, 2021, 50(2):0210004

陈婧,张振兴.基于变精度粗糙集的贪婪无监督高光谱图像波段选择方法[J].光子学报,2021,50(2):0210004

基于变精度粗糙集的贪婪无监督高光谱图像 波段选择方法

陈婧,张振兴

(鲁东大学信息与电气工程学院,山东烟台 264000)

摘 要:由于高光谱图像分类中现有的无监督波段选择方法无法计算出波段之间的相似性,以及在选
择过程中存在的高维度特性,本文提出了一种基于变精度粗糙集的贪婪无监督高光谱波段选择方法。
首先利用变精度粗糙集定义了一种新的依赖度量,使得它对变精度粗糙集中误分类参数变得不敏感,
从而充分利用波段之间的相似性。其次,提出了一种新的判别准则,找出未选择和已选择波段子集中
具有较高和较低的相似性值的波段。然后,采用一阶增量搜索法,逐条选择出所需的信息波段,从而避
免大量信息的产生,减少计算复杂度。最后通过使用三个高光谱数据集将所提出的波段选择技术与五
种最新技术进行比较。结果显示提出的方法对所有数据集均具有较好的分类精度,且在 50% 标记像素
条件下平均分类精度相对于所有像素点平均分类精度分别仅下降了 1.9%,3.1% 以及 4.1%。所提方法
能够保证良好的分类性能与数据集泛化能力,并且对参数具有鲁棒性。

关键词:高光谱图像;波段选择;变精度粗糙集;图像分类

中图分类号:TP391

文献标识码:A

doi:10.3788/gzxb20215002.0210004

Greedy Unsupervised Hyperspectral Image Band Selection Method Based on Variable Precision Rough Set

CHEN Jing, ZHANG Zhenxing

(College of Information and Electrical Engineering, Ludong University, Yantai, Shandong 264000, China)

Abstract: Because the existing unsupervised band selection methods in hyperspectral image classification could not calculate the similarity between bands and the high-dimensional characteristics in the selection process, a greedy unsupervised hyperspectral band selection method based on variable precision rough set was proposed. Firstly, a new dependency measure was defined by using the variable precision rough set, which made it insensitive to the misclassification parameters of the variable precision rough set, so as to make full use of the similarity between wavebands. Secondly, a new criterion was proposed to find out the bands with higher and lower similarity values in the unselected and selected segment subsets. Then, the first order incremental search method was used to select the required information band one by one, so as to avoid the generation of a large amount of information and reduce the computational complexity. Finally, three hyperspectral datasets were used to compare the proposed band selection technique with the five latest techniques. The results show that the proposed method has good classification accuracy for all datasets, and the average classification accuracy is only 1.9%, 3.1% and 4.1% lower than the average classification accuracy of all pixels under the condition of 50% marked pixels, which proves that the proposed method

基金项目:国家自然科学基金(No.61304052),烟台市重点研发计划项目(No.2015ZH058)

第一作者:陈婧(1982-),女,讲师,博士,主要研究方向为模式识别,图像处理。Email: chjiloko@126.com

收稿日期:2020-08-14;录用日期:2020-09-18

<http://www.photon.ac.cn>

can guarantee good classification performance and generalization ability of data set, and has robustness to parameters.

Key words: Hyperspectral image; Band selection; Variable precision rough set; Image classification

OCIS Codes: 100.4145; 110.4234; 100.4994; 100.3020

0 引言

高光谱图像(Hyperspectral Images, HSI)作为一个丰富的光谱信息来源,可以精确地描述物体,因此在海洋勘探、军事目标探测、林业、水文等各个领域均得到了广泛的应用^[1-2]。尽管如此,高光谱数据的高维性和标记样本的稀缺性,使得高光谱数据的分类成为一个具有挑战性的课题,吸引着众多学者进行探索^[3-4]。

对于高光谱图像分类来说,标记样本可用性数量以及限制证明了非监督波段选择技术比监督波段选择技术更实用^[5]。众多的无监督波段选择技术分为四种:基于排名的、基于聚类的、基于进化的和基于贪婪的^[6-10]。基于排名的方法通过使用不同的统计标准为每个波段分配一个排名,然后选择所需的排名靠前的波段数量来表示高光谱图像,该方法忽略了所选波段之间的相关性。基于聚类的方法可以根据不同的相似性度量将相似的波段组合成一个簇,并从每个簇中选择最具代表性的波段,但是该方法不一定能够选择信息量更大的波段。而进化算法通常陷入局部最优,从非常巨大的搜索空间中寻找最优解,使得计算复杂度较大。基于贪婪的方法虽然能够有效实现最优波段选择,但是依旧存在波段之间相关性不强,无法充分挖掘高光谱数据,且在计算复杂度上并无较大优势。

粗糙集理论(Rough Set Theory, RST)是一种处理数据不确定性、模糊性和不完全性的范式。它是在数据挖掘和知识发现时代具有重要应用价值的一种数据分析工具,为有监督的和无监督的特征选择提供了依赖性度量的概念^[11-12]。尽管基于RST的无监督依赖性度量在不同的应用领域得到了成功的应用,但它们可能不是度量高光谱波段之间相似性的合适度量。高光谱图像波段值是连续的,并且RST的基本概念是基于等价类的生成,等价类只能从离散数据中获得。为了使用RST进行HSI波段选择,连续的带值需要转换为离散值,在离散化之后,为了以无监督的方式计算波段之间的相似性,有的文献提出了相对独立性作为相似性度量,或者是利用粗糙集理论的下近似来计算两者之间的依赖性属性^[13-14]。但是上述方法限制性很强,而且带值的离散化是以无监督的方式进行的,有可能会计算的两个波段之间的依赖度为零。因此,上述方法提出的依赖度无法准确衡量波段之间的相似性。

针对上述问题,提出了一种基于变精度粗糙集的贪婪无监督高光谱波段选择方法。通过三个实验,利用支持向量机对五种最新的选段技术在平均总分类精度、平均kappa精度和标准差方面进行了比较,证明了该选段方法的有效性。

1 变精度粗糙集

变精度粗糙集(Variable Precision Rough Set, VPRS)是Ziarko提出的粗糙集理论的扩展^[15-16]。该方法提供了用于定义RST的上下近似的子集算子。它根据数据的统计趋势来评估数据模式,而不是功能性趋势。变精度粗糙集的主要目的是允许分类具有受控的不确定性程度或错误分类。

设 X 和 Y 是 U 的非空子集, X 相对于 Y 的错误分类率用 $e(X, Y)$ 表示,计算为

$$e(X, Y) = \begin{cases} 1 - \frac{|X \cap Y|}{|X|}, & |X| > 0 \\ 0, & |X| = 0 \end{cases} \quad (1)$$

设 $\beta(0 \leq \beta < 0.5)$ 为某一允许误差水平, $X \subseteq U$ 。 X 的 B^β -上近似和 B^β -下近似分别由 $\overline{B}^\beta X$ 和 $\underline{B}^\beta X$ 表示,定义为

$$\begin{aligned} \underline{B}^\beta X &= \cup \{ [X_i]_B | e([x_i]_B, X) \leq \beta \} \\ \overline{B}^\beta X &= \cup \{ [X_i]_B | e([x_i]_B, X) < 1 - \beta \} \end{aligned} \quad (2)$$

给定系统 (U, A, V, f) ,设 C 和 D 为 A 的任意子集, β 是误差因子, D 对 C 的依赖度定义为

$$\gamma_C^\beta(D) = \frac{\sum_{x \in U/b} |C^\beta(X)|}{|U|} \quad (3)$$

2 本文方法

在HSI分类中,通过去除冗余波段来选择波段子集,解决了维数灾难问题,降低了计算复杂度^[17-19]。因此,波段选择方法对HSI的有效分析具有重要意义。此外,在许多HSI应用中,经常存在很少甚至没有标记样本可用。由于监督方法选择的波段直接取决于标记样本的质量和数量,因此非监督波段选择方法比监督波段选择方法更具实用性。在数据挖掘和知识发现时代,粗糙集理论已经成为处理模糊、不确定和不一致信息的有效工具。在这一部分中,提出了一种新的基于变精度粗糙集的无监督波段选择方法。

设 $U = \{x_1, x_2, \dots, x_n\}$ 是包含于HSI, H 和 $A = \{b_1, b_2, \dots, b_m\}$ 是包含 H 的所有波段的非空属性集, H 中的信息可以存储在二维矩阵中,如 $\{w_{ij} | i = 1, 2, \dots, n; j = 1, 2, \dots, m\}$ 。矩阵的第 i 行表示 H 的像素,以 $(w_{i1}, w_{i2}, \dots, w_{im})$ 的 x_i 形式表示,其中 w_{ij} 是 b_j 波段的 x_i 值,即 $w_{ij} = V_{b_j}(x_i)$ 。因此,根据信息系统的定义,HSI 可以表示为 (U, A, V, f) 。

为了使用RST进行HSI波段选择,必须将连续波段值转换为离散值。对于离散化,可以使用等宽度间隔的分箱、基于均值和标准差的离散化、粗糙化等。本文采用无监督等宽度区间组合方法进行离散化。在等宽区间分箱离散化方法中,首先计算属性的最小值和最大值,然后将该范围划分为用户定义的等宽离散间隔数。

离散化后,为了从 A 中选择一个子集的信息波段,定义了一个度量任意两个波段之间冗余度的相似度量。本文定义了一种基于变精度粗糙集的相似度量方法。由于VPRS对用于定义下近似的子集算子提供了松弛,因此它降低了在波段对之间产生零依赖的概率。

虽然与RST相比VPRS提供了更好的依赖性度量,但它依赖于 β 参数。设 $\gamma_{b_j}^\beta(b_k)$ 为 b_k 对 b_j 的依赖度,允许 β 误差。当 β 较大时,可能存在多个分区 $x (x \in U/b_k)$, 其中容许误差 $e([x_i]b_j, x)$ 不为 β 。这意味着 U 中的 x_i 对象可能属于 $\sum_{x \in U/b_k} b_j^\beta(x)$, 即使它的等价类 $[x_i]b_j$ 的元素分布在 U/b_k 的多个分区中。这可能会导致 X 较低的下近似值。为了减少这种问题的影响,使依赖度量不依赖于 β 参数,在 U/b_k 中,如果对 $e([x_i]b_j, x)$ β 满足两个等价类,则在 $\sum_{x \in U/b_k} b_j^\beta(x)$ 中不包含对象 x_i 。

给定一个信息系统 (U, A, V, f) , 并设 b_j 和 b_k 为 A 中的波段, β 为误差因子。 b_k 对 b_j 的依赖度定义为

$$r_{b_j}^\beta(b_k) = \frac{|\sum_{x \in U/b_k} b_j^\beta(X)|}{|U|} \quad (4)$$

其中,如果对于超过2个 X 在 $r_{b_j}^\beta(b_k)$ 满足 $e([x_i]b_j, X) \leq \beta$, 则 x_i 不存在 $\sum_{x \in U/b_k} b_j^\beta(X)$ 。

从上述定义可以看出, $r_{b_j}^\beta(b_k)$ 和 $r_{b_k}^\beta(b_j)$ 的值并不总是相同的。为了计算 A 中任何一对波段之间的相似度量,本文定义了一个相似度量,其平均依赖度为

$$\Omega(b_j, b_k) = \frac{r_{b_j}^\beta(b_k) + r_{b_k}^\beta(b_j)}{2} \quad (5)$$

使用式(5)形成一个对称矩阵 Ω , 其中包含 A 中每对波段之间的相似度量。在我们提出的工作中,基于生成的 Ω 矩阵,从 A 中选择一个子集 Y , 其中仅包含信息波段。最初 $Y = \text{空值}$, $S = A$ 。我们的方法首先找到一个波段对 (b_p, b_q) , $p \neq q$ 在 Ω 中具有最高相似值。然后选择 b_p 或 b_q 作为第一信息波段,即如果

$$(b_p, b_q) = \arg \max_{\substack{b_i \in S, b_j \in S \\ i \neq j}} \Omega(b_i, b_j) \quad (6)$$

则选择 b_p 为信息波段。选择第一个波段 b_p 后, Y 和 S 将分别更新 $Y = \{b_p\}$ 和 $S = \{S - b_p\}$ 。对于剩余波段的选择,使用一个简单的增量搜索,称为一阶增量搜索,一次选择一个波段。在一阶增量搜索中,假设已选择 $(d-1)$ 波段。根据一个准则,从 S 中的剩余波段中选择最佳第 d 波段。在这项工作中,为了在每一次序搜索中选择一个最佳波段,我们定义了一个新的函数 $\text{rdnt}(b_j)$, $b_j \in S$ 为

$$\text{rdnt}(b_j) = \max_{\substack{b_k \in S \\ j \neq k}} \{\Omega(b_k, b_j)\} - \max_{b_i \in Y} \{\Omega(b_i, b_j)\} \quad (7)$$

$\text{rdnt}(b_j)$ 有两项。第一项计算 b_j 与 S 中其余波段之间的最大相似度值。第二项计算 b_j 与 Y 中所有已被选择为信息波段的所有波段之间的最大相似度值。第一项的较高值表示除了 b_j 之外, S 中至少存在一个与 b_j 非常相似的波段。第二项的较高值表示 Y 中有一个与未选择的 b_j 非常相似的波段。对于未选择的波段 b_j , 第一项的较高值和第二项的较低值意味着, 如果我们将 b_j 包含在 Y 中, 它将是 Y 中的非冗余波段。因此, 在本工作中, 对于每个顺序搜索, S 中产生最大 $\text{rdnt}()$ 值的波段包含在 Y 中。在更详细的信息中, 设 $(d-1)$ 波段已经在 Y 中。利用一阶搜索, 选择最佳第 d 波段 $b_k \in S$

$$b_k = \arg \max_{b_i \in S} \{ \text{rdnt}(b_i) \} \quad (8)$$

选择 b_k 后, 选定和未选定的波段子集将更新为 $Y = \{Y \cup b_k\}$ 和 $S = \{S - b_k\}$ 。搜索过程将重复, 直到选择所需的波段数。算法 1 给出了所提出的波段选择方法的步骤。

复杂性分析: 对于具有 n 个像素, 每个像素用 m 个波段表示的 HSI, 由该算法计算的一对波段之间的相似度值需要 $O(n)$ 比较。因此, 要计算 $m \times m$ 大小的相似矩阵 Ω , 需要 $O(nm^2)$ 比较, 其中 $n \gg m$ 。生成 Ω 后, 所提出的方法选择第 i 波段所需的比较次数为 $m(m-i+1)$ 。因此, 从 HSI 个波段中选择 k 个波段所需的比较次数为 $O(km^2)$ 。结果表明, 该算法的计算复杂度为 $O(nm^2) + O(km^2)$, 相当于 $O(nm^2)$ 。值得注意的是, HSI 的像素 n 的数目远远大于波段 m 的数目。

算法 1: 无监督 HSI 波段选择算法

输入: 一个 HSI 和 β 参数值

输出: HSI 波段子集

1) 初始化 $A = \{b_1, b_2, \dots, b_m\}$, $S = A$ 和 $Y = \{\emptyset\}$; 2) 使用式(5)计算 A 中每对波段间的相似度值 $\Omega(b_i, b_j)$; 3) 使用式(6)选择具有最高相似值的波段 $b_p \in S$; 4) 更新 $Y = \{Y \cup b_p\}$ 和 $S = \{S - b_p\}$; 5) 对于每个 $b_i \in S$, 使用式(7)计算 $\text{rdnt}(b_i)$; 6) 使用式(8), 选择与 Y 和 S 中的波段具有低冗余度和高冗余度的 $b_k \in S$; 7) 更新 $Y = \{Y \cup b_k\}$ 和 $S = \{S - b_k\}$; 8) 重复搜索直到选择所需的波段数。

3 实验与结果

3.1 数据集描述

使用三个 HSI 数据集, 即 Botswana、KSC 和 Indian Pines (IP), 评估提出的波段选择方法的有效性。

Botswana 数据集: 2001 年 5 月 1 日, 博茨瓦纳奥卡万戈三角洲地区的传感器捕捉到了博茨瓦纳的第一个 HSI 数据集。数据在 400~2 500 nm 的光谱范围内捕获, 光谱分辨率为 10 nm。它由 1 476 × 256 像素和 242 个波段组成, 空间分辨率为 30 m。德克萨斯大学空间研究中心对数据进行了预处理。去除未校准、低信噪比和吸水波段后, 仅使用 145 个波段进行分析。它获得了反映洪水对研究区植被影响的 14 个不同的地面真值类。图 1 显示了博茨瓦纳高光谱图像的假彩色组成和参考图。

KSC 数据集: 1996 年 3 月 23 日, 佛罗里达州肯尼迪航天中心上空的 AVIRIS 传感器捕捉到第二个 HSI 数据集 KSC。这些数据是从大约 20 公里的高度捕获的, 空间分辨率为 18 m/像素。它由 512 × 614 个像素和 224 个波长为 10 nm、中心波长为 0.4~2.5 μm 的光谱波段组成。除去吸水率和低信噪比波段后, 224 个波段中只有 176 个保留下来进行分

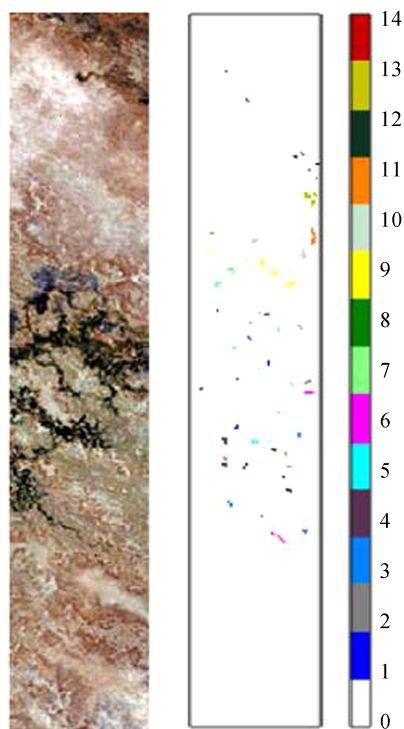


图 1 高光谱 Botswana 图像及其参考地图

Fig.1 Hyperspectral Botswana image and its reference map

析。它获得了13个不同的地面真值类,代表了不同的土地覆盖类型。图2显示了高光谱KSC图像的假彩色组成和参考图。

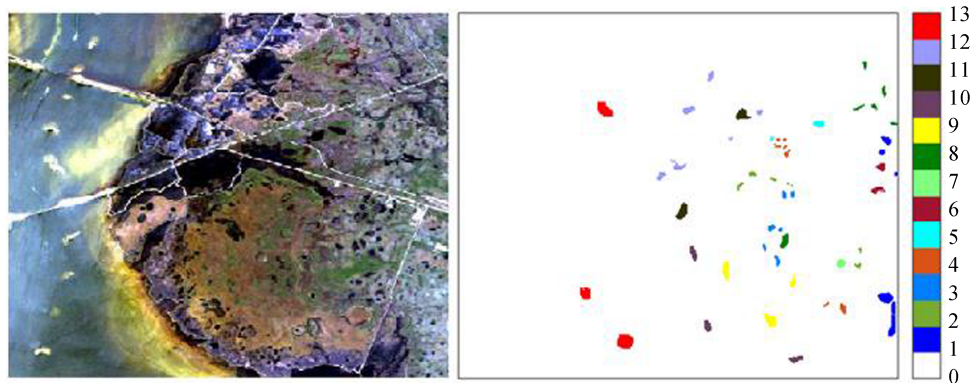


图2 高光谱KSC图像及其参考地图
Fig.2 Hyperspectral KSC image and its reference map

Indian Pines数据集:第三个HSI数据集是由阿维里斯传感器在1992年在印第安纳州西北部的农业区。数据在400~2500 nm的光谱范围内捕获,光谱分辨率为10 nm。它由145×145像素和220个光谱波段组成,空间分辨率为20 m/像素。去除吸水率和低信噪比波段后,220个波段中只有185个保留下来进行分析。它获得了16个不同的基本真值类。图3显示了高光谱印度松树图像的假彩色组成和参考图。

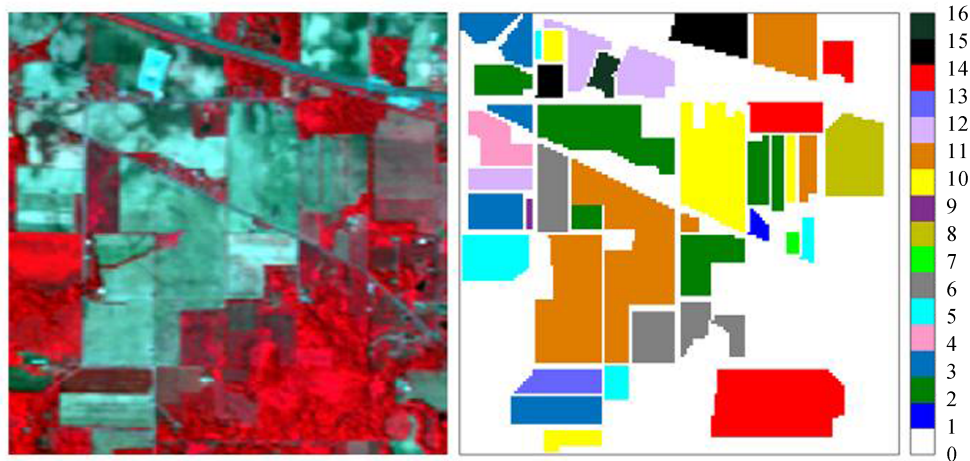


图3 高光谱Indian Pines图像及其参考地图
Fig.3 Hyperspectral Indian Pine image and its reference map

3.2 实验设置

为了评估所提出的方法的有效性,通过使用上述三个真实HSI数据集,将其与四种无监督和一种有监督的最新波段选择方法进行了比较。为此,第一种用于比较的无监督方法是基于奇异值分解的方法(Singular Value Decomposition, SVD)。该方法通过逼近HSI的主成分来选择信息波段子集。第二种方法是用于无监督波段选择的信息散度测度(ID)^[20]。该方法通过计算相对熵来度量波段子集的概率密度函数和高斯概率密度函数之间的信息散度。其他两种用于比较的无监督最新方法,使用信息度量生成一个不同矩阵,以存储每对HSI波段之间的不同值。一种方法利用K-L散度(WaLuDi),另一种方法利用互信息计算两个波段之间的差异值(WaLuMi)^[21]。在计算不同矩阵后,两种方法都遵循连锁聚类策略对相似波段进行聚类,然后从每个聚类中选出信息量最大的波段。上述方法都已经证明可以有效的处理波段选择问题,且在一定程度上代表了目前研究的前沿创新方法,具有较高的对比价值。为了进一步验证所提出的方法,还将其与基于有监督RST的最新方法(RRS)进行了比较。RRS方法根据相关性和显著性准则,利用RST来识别HSI中信息量最大的波段。一阶增量搜索用于选择信息波段。

在进行实验之前,第3.1节中描述的数据集的所有波段都在0到1之间被规范化。由于RST是在离散数据上工作的,需要对连续的波段值进行离散化处理。因此使用等宽度间隔的分格,将分格大小调整为100。对于所有三个数据波段,VPRS的误差参数 β 固定为0.1。所有这些方法的性能都是通过选择不同数量的波段来进行比较的,这些波段是范围内5的倍数。使用一对多支持向量机分类器分析了不同波段选择方法选择的相同波段数量的有效性。每个二值支持向量机的核函数都是用径向基函数实现的。采用十次交叉验证方法,得到了支持向量机参数 $\{\sigma, C\}$ 的最优值。在实验中,50%的可用标记样本被随机选择来训练支持向量机分类器,剩余的标记样本被用来评估不同波段选择方法选择波段的有效性。

3.3 结果

第一个实验是利用三个HSI数据集,即 Botswana、KSC 和 IP 数据集。比较所提出的波段选择方法与其他最新方法的性能。图4为运行十次得到的在不同方法以及相同数量的波段条件下的平均总分类准确度。从图中可以看出,与四种无监督波段选择方法相比,即SVD、ID、WaLuDi和WaLuMI,所提出的方法为所有数据集提供了非常一致的结果。例如,SVD方法为Botswana和IP数据集提供了更好的结果,但它完全没有为KSC数据集提供可接受的结果。ID方法为Botswana数据集提供了很好的结果,但对于KSC和IP数据集则效果较差。而另一方面,所提出的无监督方法为所有三个数据集均提供了较好的结果,说明了提出的方法具有较好的泛化能力。

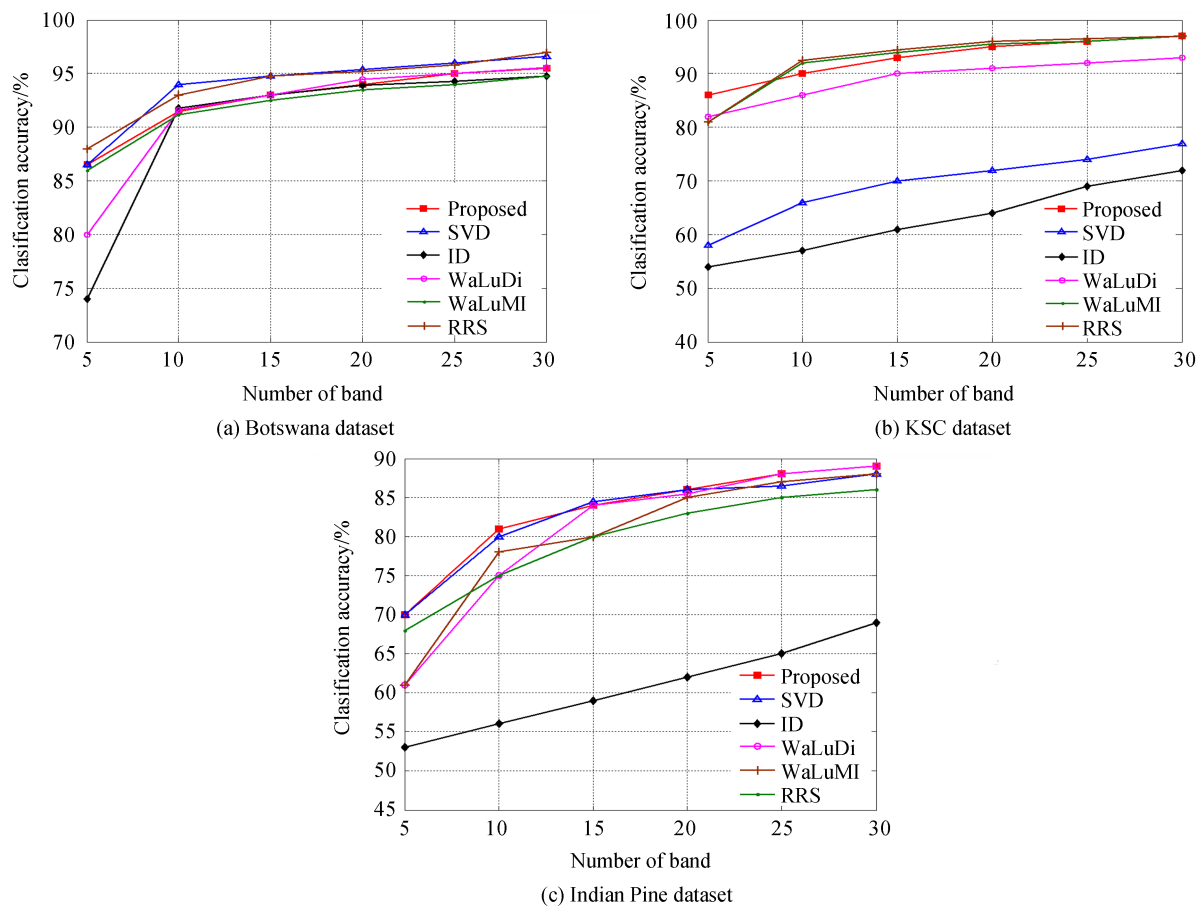


图4 平均总分类准确度对比

Fig.4 Comparison of average total classification accuracy

为了进一步评估所提出的基于无监督VPRS的波段选择方法的有效性,还将其与现有的有监督RRS方法进行比较。从图4可以看出,当选择相同数量的波段时,所提出的无监督方法获得了与现有的有监督RRS可比结果。对于定量分析,表1~3分别显示了Botswana、KSC和IP数据集在10次运行中获得的平均总体分类准确度(Overall Accuracy, OA)、平均kappa准确度(Kappa Accuracy, KA)和标准偏差(Standard Deviation, STD)。从这些表中可以清楚地看到,由所提出的无监督方法选择的波段总是提供比由最新的无监督SVD、

表1 Botswana数据集分类性能对比
Table 1 Comparison of classification performance of Botswana dataset

Number of bands	Proposed method($\beta=0.1$)			SVD			ID			WaLuDi			WaLuMI			RRS		
	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD
5	86.64	0.855	0.577	86.21	0.851	0.488	73.64	0.714	0.329	79.76	0.781	0.861	86.12	0.85	0.425	88.77	0.878	0.481
10	90.73	0.9	0.268	94	0.935	0.293	91.14	0.904	0.192	90.85	0.901	0.407	90.88	0.901	0.32	92.89	0.923	0.227
15	93.31	0.928	0.275	94.64	0.942	0.329	93.2	0.926	0.341	93.15	0.926	0.276	92.19	0.915	0.336	94.67	0.929	0.284
20	94.08	0.936	0.262	95.3	0.949	0.37	94.23	0.938	0.251	94.64	0.942	0.316	93.46	0.929	0.287	95.12	0.936	0.249
25	95.51	0.951	0.322	95.83	0.955	0.247	94.63	0.942	0.249	95.39	0.95	0.281	94.32	0.938	0.355	95.78	0.943	0.226
30	96.02	0.957	0.343	96.72	0.964	0.373	94.97	0.946	0.191	96.18	0.959	0.187	95.02	0.946	0.37	96.95	0.954	0.252

表2 KSC数据集分类性能对比
Table 2 Comparison of classification performance of KSC dataset

Number of bands	Proposed method($\beta=0.1$)			SVD			ID			WaLuDi			WaLuMI			RRS		
	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD
5	84.91	0.831	0.241	58.25	0.53	0.958	52.98	0.465	0.383	81.96	0.798	0.505	81.38	0.791	0.216	80.77	0.785	0.322
10	88.97	0.877	0.205	66.08	0.618	0.541	57.25	0.515	0.459	85.84	0.842	0.295	90.04	0.889	0.442	90.31	0.892	0.172
15	91.74	0.908	0.198	69.45	0.657	0.498	61.26	0.561	0.42	89.54	0.883	0.359	92.49	0.916	0.307	92.76	0.919	0.269
20	92.99	0.922	0.275	70.94	0.674	0.389	63.67	0.589	0.466	90.2	0.891	0.287	93.32	0.926	0.183	93.48	0.927	0.273
25	94.12	0.934	0.14	72.66	0.693	0.287	68.57	0.646	0.341	90.97	0.899	0.439	93.9	0.932	0.253	94.11	0.934	0.306
30	94.44	0.938	0.099	76.54	0.737	0.505	71.17	0.676	0.414	92.2	0.911	0.316	94.51	0.939	0.245	94.44	0.938	0.308

表3 IP数据集分类性能对比
Table 3 Comparison of classification performance of Indian pine dataset

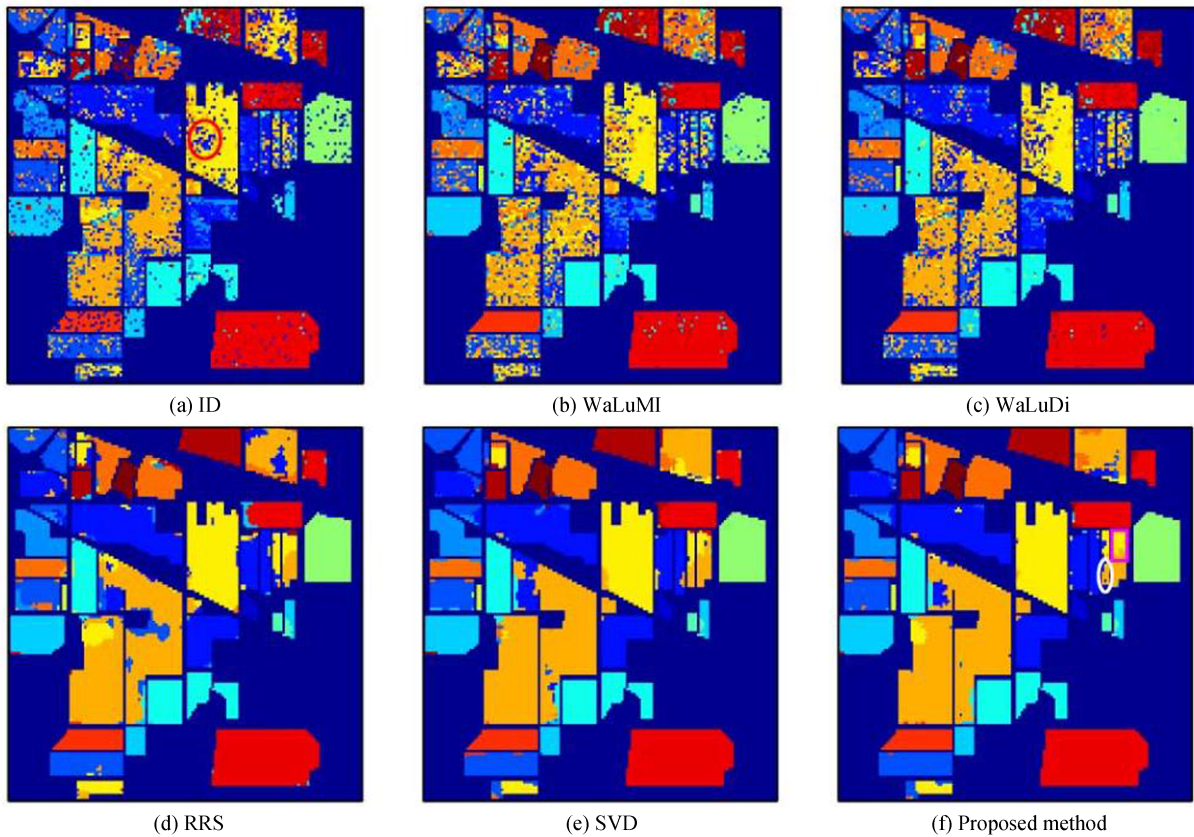
Number of bands	Proposed method($\beta=0.1$)			SVD			ID			WaLuDi			WaLuMI			RRS		
	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD	OA	KA	STD
5	69.61	0.649	0.272	69.88	0.65	0.517	52.46	0.443	0.416	60.59	0.535	0.356	60.9	0.541	0.341	67.03	0.616	0.496
10	80.78	0.78	0.183	80.16	0.773	0.2	55.87	0.483	0.651	75.15	0.714	0.216	77.85	0.745	0.221	75.19	0.714	0.236
15	83.92	0.816	0.237	84.62	0.824	0.222	58.71	0.518	0.584	84.06	0.818	0.232	80.24	0.773	0.17	80	0.771	0.256
20	86.38	0.844	0.242	86.24	0.843	0.219	62.31	0.557	0.465	85.78	0.837	0.223	85.39	0.833	0.206	83.17	0.807	0.408
25	87.42	0.856	0.192	86.92	0.851	0.193	64.93	0.593	0.381	87.75	0.86	0.206	86.44	0.845	0.173	84.83	0.826	0.229
30	88.7	0.871	0.252	87.93	0.862	0.305	69.24	0.647	0.205	88.73	0.871	0.168	87.69	0.859	0.259	86.65	0.847	0.275

ID、WaLuDi和WaLuMI方法选择的波段更好或可比较的分类结果,以及与最新的有监督RRS方法的对比结果。通过分析各个方法的原理可知,SVD在小尺寸物体分类中提供的信息较少。而当数据服从非高斯分布时,ID方法效果较差。WaLuDi和WaLuMI都使用聚类方法将相似的波段分组到一个簇中,然后从每个簇中选择信息波段。因此,如果聚类方法不能将信息波段划分成不同的聚类,那么这些方法就无法选择信息量最大的聚类波段。另一方面,提出的方法与聚类结果无关,它是基于一个新的依赖值来选择信息波段,该依赖值是根据选择的和未选择的波段计算出来的。为了更加直观地显示各个方法对于分类精度的影响,图5为IP数据集在选择20波段数条件下的各方法的分类结果。从图中可以直观地看出提出方法的分类效果更加细节化,进一步证明了提出方法的有效性。

第二个实验是为了进一步比较所提出的方法和不同的最新方法,通过计算 z 分数进行统计显著性检验,它描述了由两种不同算法获得的两个分类结果之间的差异的显著性,表示为

$$z = (\mu_1 - \mu_2) / |\sigma_1^2 - \sigma_2^2| \quad (9)$$

式中, μ_1 和 μ_2 分别是由算法1和2获得的kappa系数平均值, σ_1^2 和 σ_2^2 是相应的方差。如果 $|z| > 1.96$,则假设两种算法的结果在5%显著性水平上具有统计显著性。表4显示了提出方法与用于比较的五种最新方法之间

图5 20波段数IP数据集分类对比结果($\beta=0.1$)Fig.5 Classification and comparison results of 20 band IP data sets ($\beta = 0.1$)

的 $|z|$ 大小。从表中可以看出,对于所有三个数据集,所提出的方法获得的 $|z|$ 分数总是大于1.96。这表明所提出的方法提供的结果具有统计意义。

表4 获得的 $|z|$ 分数对比Table 4 Comparison of $|z|$ scores obtained

Datasets	SVD	ID	WaLuDi	WaLuMI	RRS
Botswana	3 290.1	1 228.4	175.44	4 675.3	437.5
KSC	6 425.8	12 547.8	2 445.65	131.5	9.67
IP	2 265.6	135 636.3	136.05	19 661.1	19 500

进一步通过计算时间对比各个方法的计算成本,表5为各个数据集上波段数选择为10时的计算时间,从表5中可以看出,提出方法计算时间与其他方法相比并未增大,甚至相比于WaLuDi, WaLuMI, ID有所减少,计算时间仅次于SVD方法,由于篇幅所限,其他波段条件下计算时间结果与表5类似,进一步说明了提出方法在计算量上并无增加,具备较好的实用性。

表5 计算时间对比(单位:秒)

Table 5 Comparison of calculation time(unit: s)

Datasets	SVD	ID	WaLuDi	WaLuMI	RRS	VPRS($\beta=0.1$)
Botswana	357.9	656.2	647.8	700.4	514.9	515.8
KSC	316.6	542.3	540.7	585.9	433.6	430.7
IP	21.8	36.9	36.8	39.8	29.5	29.3

第三个实验证明了该方法对 β 参数的鲁棒性。传统VPRS方法强烈依赖于误差参数 β 。为了减少 β 参数对依赖计算的影响,本文对依赖定义进行了改进。图6和图7为传统VPRS方法与所提出的方法对在0.1

步长范围内改变 β 值的分类结果。从这些图中可以看出,与传统的VPRS依赖定义相比,提出的方法对 β 参数的依赖性较小。特别是对于Botswana和IP数据集,可以看到,用所提出的标准代替传统的依赖性标准,分类结果随 β 值变化的变化显著最小化。结果表明,该方法相对于传统VPRS来说,对参数 β 具有较强的鲁棒性。

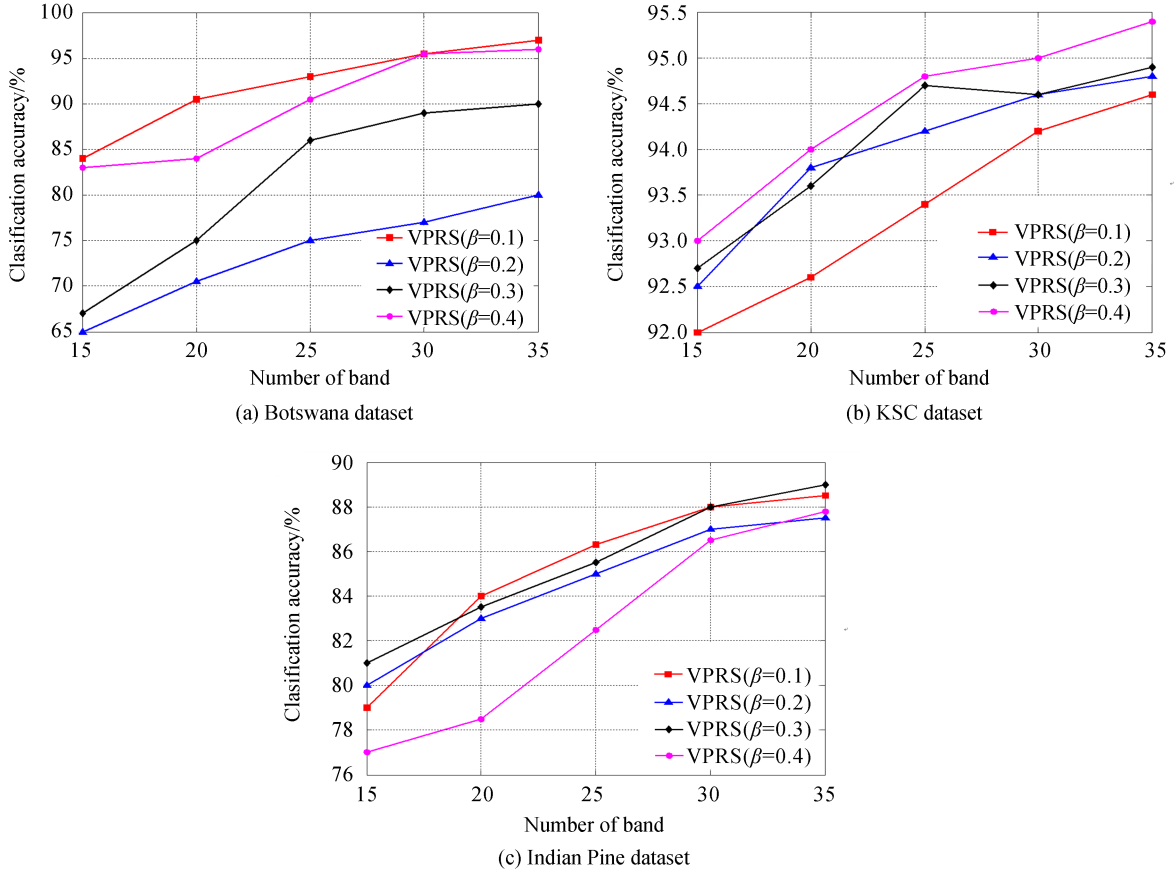
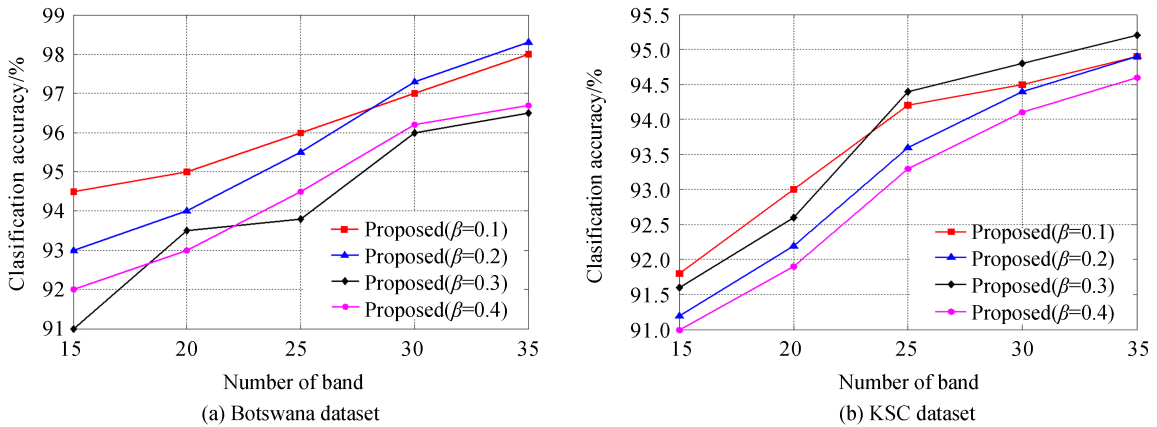


图6 传统VPRS的参数灵敏度对比

Fig.6 Parameter sensitivity comparison of traditional VPRS



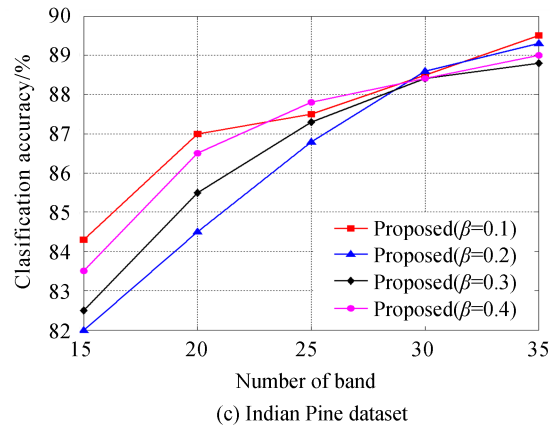


图7 提出方法的参数灵敏度对比

Fig.7 Parameter sensitivity comparison of proposed method

另外,所提出的方法和SVD、ID、WaLuDi和WaLuMI无监督波段选择方法的计算复杂度均为 $O(nm^2)$,其中 n 是用 m 个谱特征表示的样本数,对应于HSI的像素。由于HSI的像素 n 的数目远大于波段 m 的数目,所以所提出的方法的计算时间依赖于 n 的值。高光谱影像分类最主要的问题在于真实数据难以获取且数量有限,选取50%的标记样本是为了验证所提出的无监督波段选择方法的计算精度对于数据样本量大小的鲁棒性,在本实验中,使用包含不同数目样本的三个样本集来测试所提出的方法。第一、第二和第三集合是通过考虑所有像素、仅可用标记像素和仅50%的图像可用标记像素来生成的。图8(a)到8(c)为考虑三个数据集的不同样本数产生的分类结果准确率。从图中可以看出,在考虑少量样本的情况下,由该方法选择的波

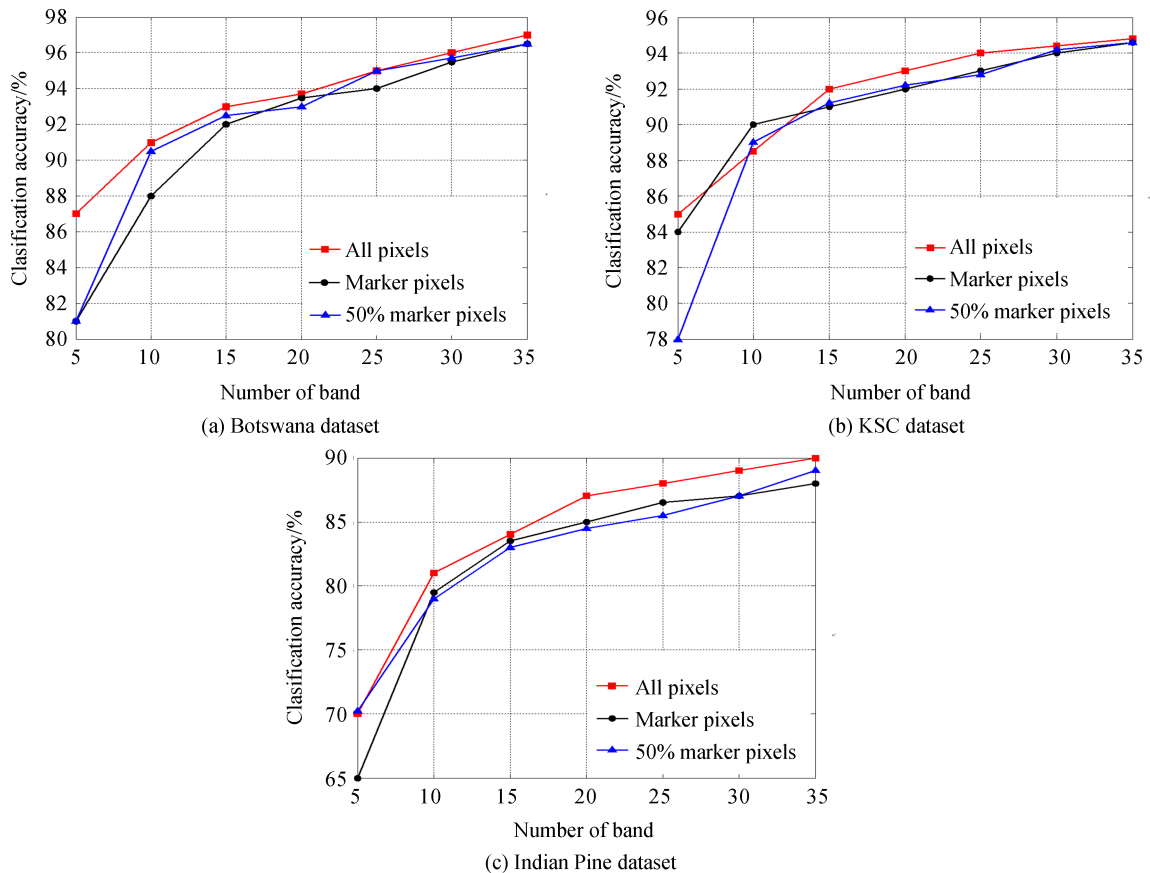


图8 不同样本数产生的分类结果准确率

Fig.8 Accuracy of classification results generated by different sample numbers

段与由图像的所有样本选择的波段具有相似的识别能力。在50%标记像素条件下平均分类精度相对于所有像素点平均分类精度分别仅下降了1.9%,3.1%以及4.1%,进一步说明了提出的方法在能够保证样本数量较少条件下保持较好的分类精度。

4 结论

针对高光谱图像分类中现有的无监督波段选择方法无法计算出波段之间的相似性,以及在选择过程中存在的高维度特性,提出了一种基于变精度粗糙集的贪婪无监督高光谱波段选择方法。通过分类实验发现:1)提出的方法与现有的无监督和有监督方法相比,具有较好的数据集泛化性能。且具有良好的分类性能,计算时间也能够保证该方法的实用性能;2)依赖性度量的引入使得提出的方法对误分类参数具有较好的鲁棒性,降低了传统分类方法对参数的依赖,并且所提出的方法获得的 $|z|$ 分数总是大于1.96,证明了结果具有统计意义;3)在50%标记像素条件下平均分类精度相对于所有像素点平均分类精度分别仅下降了1.9%,3.1%以及4.1%,说明提出方法在不影响波段选择结果的情况下,可以通过较小的样本数来进一步降低该方法的计算复杂度,并且还能够保证较高的分类精度。

参考文献

- [1] XU Dongdong, CHENG Deqiang, CHEN Liangliang, et al. Hyperspectral image classification based on hierarchical guided filtering and nearest neighbor regularization subspace [J]. *Acta Photonica Sinica*, 2020, 49(4): 0410004.
徐冬冬,程德强,陈亮亮,等.基于分层引导滤波与最近邻正则化子空间的高光谱图像分类[J].*光子学报*,2020,49(4):0410004.
- [2] LIU Jiamin, ZHANG Limei, SHI Guangyao, et al. Hyperspectral image classification based on sparse property and neighborhood similarity measure [J]. *Acta Photonica Sinica*, 2018, 47(6): 0610001
刘嘉敏,张丽梅,石光耀,等.联合稀疏特性和邻域相似度量的高光谱图像分类[J].*光子学报*,2018,47(6):0610001.
- [3] YAN Jingwen, CHEN Hongda, LIU Lei. Research progress of hyperspectral image classification [J]. *Optical Precision Engineering*, 2019, 27(3): 680-693.
闫敬文,陈宏达,刘蕾.高光谱图像分类的研究进展[J].*光学精密工程*,2019,27(3):680-693.
- [4] CHEN Shanxue, CHEN Wenwen. Combined sparse representation hyperspectral image classification based on secondary dictionary[J]. *Systems Engineering and Electronic Technology*, 2020, 42(3): 550-556.
陈善学,陈雯雯.基于二级字典的联合稀疏表示高光谱图像分类[J].*系统工程与电子技术*,2020,42(3):550-556.
- [5] GAO Wei, PENG Yu. Hyperspectral image classification based on Mahalanobis distance multi-core learning [J]. *Journal of Instrumentation*, 2018, 39(3): 250-257.
高巍,彭宇.基于马氏距离多核学习的高光谱图像分类[J].*仪器仪表学报*,2018,39(3):250-257.
- [6] BARMAN B, PATRA S. A novel technique to detect a suboptimal threshold of neighborhood rough sets for hyperspectral band selection[J]. *Soft Computing*, 2019, 23(8):13709-13719.
- [7] WANG Q, ZHANG F, LI X. Optimal clustering framework for hyperspectral band selection[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2018, 56(10): 5910-5922.
- [8] FAN Y, ZHANG C, LIU Z. Cost-sensitive stacked sparse autoencoder models to detect striped stem borer infestation on rice based on hyperspectral imaging[J]. *Knowledge-based Systems*, 2019, 168(8): 49-58.
- [9] BARMAN B, PATRA S. Empirical study of neighbourhood rough sets based band selection techniques for classification of hyperspectral images[J]. *IET Image Process*, 2019, 13(8): 1266-1279.
- [10] XIE Fuding, LEI Cun, LI Fangfei, et al. Unsupervised band selection based on fuzzy c-means algorithm and artificial bee colony algorithm [J]. *System Science and Mathematics*, 2018, 38(12): 1417-1428.
谢福鼎,雷存,李芳菲,等.基于模糊c均值算法和人工蜂群算法的无监督波段选择[J].*系统科学与数学*,2018,38(12):1417-1428.
- [11] XIE W, LI Y, ZHOU W. Feature extraction of hyperspectral images with a matting model[J]. *International Journal of Remote Sensing*, 2018, 39(5):1510-1527.
- [12] ZHANG Yue, GUAN Yunlan. Unsupervised band selection method of hyperspectral image based on clustering[J]. *Beijing Surveying and Mapping*, 2018, 32(1): 1-4.
张悦,官云兰.基于聚类的高光谱图像非监督波段选择方法[J].*北京测绘*,2018,32(1):1-4.
- [13] WANG Ligu, ZHAO Liang, SHI Yao. Band selection of hyperspectral remote sensing image based on the maximum and minimum distance [J]. *Journal of Intelligent Systems*, 2018, 13(1): 131-137.
王立国,赵亮,石瑶.基于最大最小距离的高光谱遥感图像波段选择[J].*智能系统学报*,2018,13(1):131-137.
- [14] LIU Yao, LI Zinan, WU Tao, et al. Soybean variety recognition algorithm and comprehensive performance evaluation based on hyperspectral image and neighborhood rough set theory [J]. *Soybean Science*, 2018, 37(4): 596-605.

- 刘瑶,李梓楠,吴涛,等.基于高光谱图像和邻域粗糙集理论的大豆品种识别算法及其综合性能评估[J].大豆科学,2018,37(4):596-605.
- [15] LIU Yao. Research on hyperspectral band selection algorithm based on neighborhood rough set [D]. Harbin Engineering University, 2017.
刘瑶.基于邻域粗糙集的高光谱波段选择算法研究[D].哈尔滨工程大学,2017.
- [16] WU Dan, ZHANG Yan. Clustering method of hyperspectral forest biomass based on artificial bee colony optimization algorithm [J]. Laser Journal, 2020, 41(5): 101-104.
吴丹,张艳.基于人工蜂群优化算法的高光谱森林生物量聚类方法[J].激光杂志,2020,41(5):101-104.
- [17] YIN Yuping, WEI Lin, LIU Wanjun. Hyperspectral image classification by integrating cumulative variation ratio and over limit learning machine [J]. Chinese Journal of Image Graphics, 2020, 25(5): 1053-1068.
尹玉萍,魏林,刘万军.融合累积变异比和集成超限学习机的高光谱图像分类[J].中国图象图形学报,2020,25(5):1053-1068.
- [18] LI Jiayin, ZHU Wenquan, MENG Fanyun. Hyperspectral image compression algorithm based on reducing mapping prediction residual error [J]. Computer Engineering and Science, 2020, 42(5): 825-834.
李佳颖,朱文泉,孟繁蕴.基于降低映射预测残差的高光谱图像压缩算法[J].计算机工程与科学,2020,42(5):825-834.
- [19] CUI Binge, ZHONG Liwei, LU Yan. Hyperspectral image classification method based on multi feature image integration [J]. Journal of Shandong University of Science and Technology (Natural Science Edition), 2020, 39(2): 86-94.
崔宾阁,钟利伟,路燕.基于多特征图像集成的高光谱图像分类方法[J].山东科技大学学报(自然科学版),2020,39(2):86-94.
- [20] CHANG C I, WANG S. Constrained band selection for hyperspectral imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(6): 1575-1585.
- [21] 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.