

联合空间信息的高光谱遥感协同表示动态集成分类算法

虞瑶¹, 苏红军², 陶旻¹

1. 江苏省基础地理信息中心, 南京 210013;

2. 河海大学 地球科学与工程学院, 南京 211100

摘要: 近年来, 集成学习成为高光谱遥感影像分类的研究热点, 尤其是动态集成算法根据测试样本的特征自适应地选择最佳分类器, 其分类性能显著提升。然而现有的动态集成方法仅考虑测试样本与验证样本的光谱信息, 忽略了高度规则化的高光谱遥感影像包含的丰富空间信息。为进一步提升高光谱遥感影像动态集成算法分类的准确性和可靠性, 提出了联合空间信息的可变K邻域动态集成算法 VKS (Variable K-neighborhood and Spatial Information) 和联合自适应邻域空间信息的可变K邻域动态集成算法 VKSA (Variable K-neighborhood with Shape-Adaptive)。两种算法第一阶段综合考虑分类器精度与相似度自适应地改变测试样本的K邻域, 第二阶段分别设计固定窗口和自适应窗口的嵌入方式增加地物的局部空间近邻关系, 充分利用高光谱遥感影像地物复杂的空间形态结构信息。实验部分采用3组通用的高光谱遥感影像数据对所提出算法的性能进行综合评价。结果表明相比于传统的动态集成算法, 本文提出的联合空间信息的动态集成模型能显著提升分类精度, 其中基于自适应窗口方式的VKSA算法明显优于基于固定窗口的VKS算法。

关键词: 高光谱遥感, 动态集成, 自适应邻域, 协同表示, 影像分类

中图分类号: TP701/P2

引用格式: 虞瑶, 苏红军, 陶旻. 2024. 联合空间信息的高光谱遥感协同表示动态集成分类算法. 遥感学报, 28(1): 187-202

Yu Y, Su H J and Tao Y. 2024. Dynamic selection algorithm for collaborative representation of hyperspectral remote sensing based on joint spatial information. National Remote Sensing Bulletin, 28(1): 187-202 [DOI: 10.11834/jrs.20221704]

1 引言

高光谱遥感技术利用连续的电磁波波段获取地物目标的相关理化信息, 从而实现对地物目标的精细探测与解译 (Bioucas-Dias 等, 2013; 张良培和张立福, 2005)。利用高光谱遥感数据对地物目标进行精细分类, 一直以来是高光谱遥感领域的热点方向。然而, 高光谱遥感影像分类面临诸多挑战 (Chang, 2003; 童庆禧 等, 2016; 苏红军, 2022)。一方面, 高光谱遥感影像的高维特征造成数据量大和冗余度高。另一方面, 在监督分类领域, 可利用的标签样本数量有限且质量好坏不均衡。当训练样本数量一定时, 分类精度随着

波段数目的增加往往会出现先升后降的现象, 即“Hughes”现象。针对高光谱遥感影像分类面临的小样本、分类器局限性等难题, 亟需发展新型的高光谱遥感影像分类方法。

集成学习并非特指某种分类算法, 而是集成多个基分类器共同决策的机器学习方法。通过集成多个结构较为简单的基分类器, 可表现出优于单个基分类器的分类效果 (杜培军 等, 2016)。集成学习包括静态分类器集成和动态分类器集成 DCS (Dynamic Classifier Selection) (Benediktsson 等, 2007; Woloszynski 和 Kurzynski, 2011)。现有的多分类器集成大部分为静态集成, 静态集成方式对测试样本进行分类时选择完全相同的分类器

收稿日期: 2021-11-06; 预印本: 2022-07-05

基金项目: 国家自然科学基金(编号:42122008, 41871220), 江苏省自然资源科技项目(编号:2021041)

第一作者简介: 虞瑶, 研究方向为高光谱遥感、资源环境遥感。E-mail: yuyaoyao_yy@163.com

通信作者简介: 苏红军, 研究方向为高光谱遥感、资源环境遥感。E-mail: hjsu@hhu.edu.cn

组合方案；然而，该集成方式忽略了分类器对每个测试样本的分类性能具有差异性。动态集成方式则根据测试样本的特征自适应地选择最佳分类器，从而实现更优的预测结果（Didaci 和 Giacinto, 2004；Kumar 等, 2002）。

动态集成方式的关键问题在于找到与测试样本相似的样本集，即测试样本的能力区域 RoC (Region of Competence) (Kuncheva, 2000)。目前，常见的能力区域生成方式主要是基于 KNN (K-Nearest Neighbor) 准则的方法 (García 等, 2018)。该方法主要选用某个相似性距离准则在验证集中寻找与测试样本相似的验证样本，将搜索到的验证样本集作为能力区域。确定测试样本的能力区域之后，采用某种评判标准评价基分类器的分类性能就显得更为关键，由此产生了多种基于 KNN 准则的动态集成算法。Woods 等 (1997) 首次提出利用局部精度估计的方法评估分类器的性能进行多分类器的动态集成。考虑到邻域样本重要性不同，Smits 等 (2002) 提出了基于距离加权方式改进的局部精度估计方法。Giacinto 和 Roli (2001) 通过分析分类器在验证集和测试样本中的输出情况，设计了基于多分类器行为 MCB (Multiple Classifier Behavior) 的多分类器动态集成算法，剔除能力区域内的“假邻居”。

从多个基分类器中选出一个局部性能最优的基分类器对测试样本分类，使得集成后“过泛化”风险增大。因此 Albert Ko 等 (2008) 提出基于 Oracle 性能上限的动态集成方法。Oracle 的概念认为只要参与集成的分类器中有一个可以将测试样本分类正确则集成系统就可以将该样本正确分类。由于 Oracle 性能上限过于理想很难达到，Albert Ko 等 (2008) 将 Oracle 概念用于局部，研究了 4 种基于 Oracle 的多分类器动态选择算法均取得较好性能。但基于 Oracle 的动态集成方法构建 K 邻域时仅考虑分类器的分类精度，而基于分类器行为的动态集成算法优先考虑分类器的输出情况，再考虑分类器的分类精度，使得能力区域的设置受分类器行为相似度影响较大。因此本文首次采用同时限制分类器的分类精度和相似度的方式自适应地改变测试样本的 K 邻域，使得能力区域的设置更加合理有效。

基于 KNN 准则构建 K 邻域时仅仅利用了光谱信息，而研究表明联合空间信息可以有效地改善算法的分类效果 (Li 等, 2019；Hou 等, 2016)。

为解决高光谱遥感影像分类动态集成模型中空间信息利用不足的问题，Damodaran 等 (2015) 结合马尔可夫随机场模型和动态集成选择算法应用于高光谱遥感影像的分类问题中，大大改善了分类效果。苏红军和刘浩 (2017) 提出一种综合利用空间和光谱信息确定测试样本类别的动态分类器选择算法，但该算法在地表较为复杂多变的影像上分类效果并不理想。

基于此，本文研究联合空间信息的动态集成算法从而提升动态集成效果，首先提出的联合空间信息的可变 K 邻域动态集成算法 VKS (Variable K-neighborhood and Spatial Information)，第一阶段生成测试样本的 K 邻域时，通过同时限制分类器的分类精度和相似度的行为，找到测试样本的最佳的 K 邻域并形成能力区域，从而将光谱判别性能好的测试样本优先分类，并设置为标记点。第二阶段对光谱判别性能较差的测试样本而言，若不存在满足条件的 K 邻域，则统计测试样本邻域内标记点所占的比例；若标记点所占比率超过一定的阈值，则考虑利用空间邻域样本的标签信息进行预测。只有满足一定条件下，该算法才考虑采用光谱和空间信息结合的方式。为进一步提升空间信息的利用效率，本文设计了自适应窗口挖掘更深层次的空间信息，提出了联合自适应邻域空间信息的可变 K 邻域动态集成算法 VKSA (Variable K-neighborhood with Shape-Adaptive)。

实验部分使用 Purdue Campus、Indian Pines 和 Salinas 这 3 组高光谱遥感影像数据对所提出方法的有效性进行了实验验证。

2 动态集成算法

动态集成学习分类算法的核心思想是每个测试样本分布在样本特征空间的不同区域，而每个基分类器对测试样本的分类性能存在差异。通过分析比较各个基分类器在测试样本分布区域周围的分类表现，自适应地选择最佳基分类器，最终共同决策判定测试样本的地物类别。

动态集成学习分类算法的核心目的是分类测试样本时最大概率地选择那些可以将其正确的基分类器。该技术考虑到每个基分类器的错误信息分布是不同的，即每个基分类器都有特长，尽可能选择那些在测试样本的特征区域附近有特长的基分类器对其进行分类 (Cruz 等, 2018；Damodaran

和Nidamanuri, 2014)。

2.1 基于KNN准则的动态集成方法

依据KNN准则在验证集中寻找测试样本的K个邻域样本，以此形成测试样本的能力区域。分析比较各个分类器在测试样本能力区域中的分类表现，按照某种组合策略将基分类器的输出结果进行集成（Dos Santos等，2008；Cruz等，2015）。

该算法的主要思想是：利用KNN准则寻找测试样本在验证集中的能力区域，计算测试样本能力区域内各基分类器的分类精度，选择能力区域内分类精度最高的基分类器对该测试样本进行分类，该分类器的输出结果作为测试样本的最终分类结果。该算法同时提出了两种计算局部精度的方法：总体局部精度OLA（Overall Local Accuracy）和局部类精度LCA（Local Class Accuracy）。OLA方法计算能力区域内验证样本被正确分类的百分比，选取百分值最高的分类器的输出结果作为该测试样本的最终标签值；LCA算法则计算能力区域内各类别的验证样本被分类器正确分类的百分比，选取百分值最高的类别标签作为该测试样本的标签值。基于局部精度的多分类器动态选择算法主要步骤如下：

- (1) 获取所有基分类器的分类输出结果；
- (2) 在验证集中搜索测试样本的能力区域；
- (3) 计算每个基分类器在能力区域内的分类精度；
- (4) 选择能力区域内分类精度最高的基分类器输出结果作为测试样本的最终标签。

测试样本表示为 $\mathbf{y}$ ，测试样本的能力区域表示为 $\theta = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K\}$ ， K 表示能力区域内验证样本的数目。 $\mathbf{x}_k$ 表示属于能力区域内的验证样本，分类器集合 $\mathbf{C} = \{c_1, c_2, \dots, c_M\}$ ， c_i 表示某个分类器。 M 表示基分类器的数目。 $P(w_l | \mathbf{y}, c_i)$ 表示分类器 c_i 对测试样本分类得到的后验概率。分类器 c_i 的输出结果表示为 $\omega_{l,i}$ 。

$$\delta_i = \frac{1}{K} \sum_{k=1}^K P(\omega_l | \mathbf{x}_k \in \omega_{l,i}, c_i) \quad (1)$$

OLA算法和LCA算法未考虑验证集中验证样本的重要性，Smits等基于OLA和LCA两种算法提出改进的MLA（Modified Local Accuracy）算法。MLA算法的主要思想是：对每个测试样本 $\mathbf{x}_i$ ，寻找测试样本于验证集中的能力区域，即选择 K 个邻

域样本构成能力区域。之后根据距离计算每个邻域样本的权值。 d_k 表示第 k 个样本到测试样本的距离。邻域样本权值的计算公式如式(2)所示，MLA的表达式如(3)所示：

$$W_k = \frac{1}{d_k} \quad (2)$$

$$\delta_i = \sum_{k=1}^K P(\omega_l | \mathbf{x}_k \in \omega_{l,i}, c_i) W_k \quad (3)$$

2.2 基于分类器行为的动态集成方法

为了消除“假邻居”对分类器性能分析产生的影响，Giacinto和Roli在2001年提出了基于分类器行为的动态集成算法（Multiple Classifier Behavior, MCB）（Giacinto和Roli，2001）。该算法利用多分类器行为方式能在测试样本的邻域集中辨识和剔除“假邻居”样本。若多个分类器判定测试样本与某些验证样本不属于同一类别，那么这些验证样本与测试样本很可能不属于同类。测试样本的真实标签不可知，但测试样本的输出结果在一定程度上可以反映测试样本的真实标签。MCB = $\{c_1(\mathbf{x}), \dots, c_M(\mathbf{x})\}$ 表示为 M 个分类器在样本 $\mathbf{x}$ 上的行为。设

$$B_i(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} 1, & c_i(\mathbf{x}_1) = c_i(\mathbf{x}_2) \\ 0, & c_i(\mathbf{x}_1) \neq c_i(\mathbf{x}_2) \end{cases} \quad (4)$$

定义两个样本 $\mathbf{x}_1, \mathbf{x}_2$ 的基于MCB之间的相似度为

$$S(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{M} \sum_{i=1}^M B_i(\mathbf{x}_1, \mathbf{x}_2) \quad (5)$$

假设两个样本 $\mathbf{x}_1, \mathbf{x}_2$ 多分类器行为的相似度取值区间为 $[0, 1]$ 。若相似度 $S(\mathbf{x}_1, \mathbf{x}_2)$ 数值为1，表明所有分类器都判定 $\mathbf{x}_1, \mathbf{x}_2$ 两个样本属于同一类别，若相似度 $S(\mathbf{x}_1, \mathbf{x}_2)$ 数值为0，表明所有分类器都判定 $\mathbf{x}_1, \mathbf{x}_2$ 两个样本属于不同类别。利用该相似度准则可以有效地剔除“假邻居”样本。假设 $\mathbf{x}_k$ 表示测试样本 $\mathbf{y}$ 的第 k 个邻域样本，计算测试样本的输出结果 $\tilde{\mathbf{y}}$ 和第 k 个邻域样本 $\mathbf{x}_k$ 的相似度 $S(\tilde{\mathbf{y}}, \mathbf{x}_k)$ ，相似度大于设定阈值的邻域样本可作为有效的近邻样本，从而生成有效邻域 $\mathbf{N}(\mathbf{y})$ 。

$$\mathbf{N}(\mathbf{y}) = \{\mathbf{x} | S(\tilde{\mathbf{y}}, \mathbf{x}) \geq \text{相似度阈值}\} \quad (6)$$

3 提出的VKS与VKSA算法

3.1 VKS动态集成算法

高光谱遥感影像具有图谱合一的特性，表明

高光谱遥感影像不仅具有精细的光谱信息, 还包含丰富的地物空间分布信息 (Peng 等, 2021; Tang 等, 2021; He 等, 2018)。遥感影像空间相邻像元通常是高度相关的, 表明相邻像元很大概率上由同一地物构成, 即相邻像元很大概率上属于同一类别。然而现有动态集成算法大多计算测试样本和验证样本光谱曲线的相似性, 但鲜有学者利用空间邻域信息研究动态集成算法。而联合空间信息是实现高光谱遥感影像分类性能提升的重要途径, 因此本文首先提出联合空间信息的可变 K 邻域动态集成算法 VKS 算法。该算法的改进主要体现在: (1) 同时考虑分类器的分类精度和相似度自适应地改变测试样本的 K 邻域; (2) 综合利用光谱信息和空间信息, 若判别性能较差的测试样本未找到合适的 K 邻域时, 则考虑利用局部空间近邻样本的标签信息。该算法只有邻域样本符合一定的条件才利用空间信息, 其原因在于尽管高光谱遥感影像上地物分布在空间上具有连续性, 但这种连续并不是绝对的。

假设验证集 $V = \{(x_1, l_1), (x_2, l_2), \dots, (x_N, l_N)\}$, 测试样本表示为 y , 分类器集合 $C = \{c_1, c_2, c_i, \dots, c_M\}$, 测试样本在验证集中的 K 邻域 $\theta = \{x_1, x_2, \dots, x_K\}$, 基分类器的能力估计值 $\delta_i, i = \{1, 2, \dots, M\}$, 基分类器集成 C 对测试样本的分类结果 $\tilde{y} = \{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_M\}$, 基分类器选择切空间协同表示分类器 TCRC (Tangent Space Collaborative Representation Classifier) (Su 等, 2016)。VKS 算法的示意图如图 1 所示。该算法的具体步骤为

- (1) 输入训练集 X , 验证集 V , 测试样本 y , 设置邻域 K 值, 比例阈值 threshold ;
- (2) 利用 Bagging 方式, 生成 M 个 TCRC 分类器;
- (3) 在验证集 V 中找到测试样本 y 的 K 邻域 $\theta = \{x_1, x_2, \dots, x_K\}$ 作为能力区域;
- (4) 对各分类器意见一致的测试样本直接赋予分类器的输出结果, 并在分类结果中进行标记, 成为标记点;
- (5) 同时考虑分类器的分类精度和相似度。要求局部分类总体精度高于 $\delta_i > 0.9$ 和分类器行为相似度高于 $S_i > 0.7$ 。若没有满足条件的 K 值, 去除 K 邻域内最不相似的样本, 邻域大小变为 $\theta' = \{x_1, x_2, \dots, x_{K-1}\}$, K 变为 $K-1$;

(6) 若存在满足条件的 K 邻域, 选择 K 邻域内精度最高的分类器, 对测试样本 y 赋予标签值。处理后, 一部分测试样本被分类, 称为标记点。另一部分测试样本未找到满足条件的分类器, 仍未被分类;

(7) 对于这部分未分类的测试样本 y , 统计 3×3 窗口邻域中标记点的比例;

(8) 如果标记点超过设定的阈值, 则利用空间邻域信息进行测试样本 y 类标签的计算;

(9) 若没有满足条件的阈值, 则使用多数投票法。

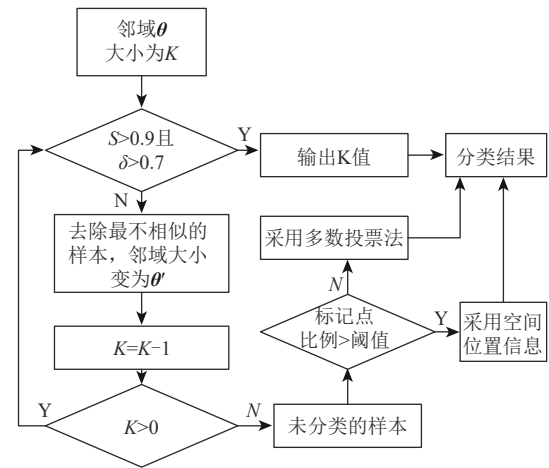


图1 联合空间信息的动态集成的模型示意图

Fig. 1 Illustration of the dynamic selection model of joint spatial information

VKS 算法主要分为两个阶段, 第一阶段生成测试样本 y 的 K 邻域时, 通过同时限制分类器的分类精度和相似度, 找到最佳的 K 邻域形成能力区域, 若存在满足条件的 K 邻域 θ , 则使用在该邻域内局部总体分类精度最高的分类器对测试样本 y 赋予标签。若没有满足条件的 K 邻域 θ , 则去除能力区域内最不相似的邻域样本, 样本邻域变为 θ' 。通过第一阶段的动态集成, 一些光谱判别性能较好的测试样本被赋予了标签值, 称这部分测试样本为标记点, 剩下一部分光谱判别性能较差的测试样本未被分类。对于未分类的测试样本, 则利用测试样本 y 的空间邻域信息。统计测试样本邻域内标记点所占的比例, 如果标记点占比超过设定的阈值, 则利用空间邻域信息进行类标签的计算, 即将该邻域内大部分标记点的标签值赋给测试样本。若不满足阈值条件, 则使用多数投票法的方式。VKS 算法通过同时限制分类器精度和分类器

行为, 将光谱判别性能好的测试样本优先分类; 而对光谱判别性能较差的测试样本则通过增加局部空间的近邻关系, 可以提升分类结果的准确度和鲁棒性。

3.2 VKSA 动态集成算法

高光谱遥感影像分类领域中采用邻域约束 (常采用固定窗口的方式) 引入空间信息是一种常见的手段 (Peng 等, 2019; Li 和 Du, 2014)。然而, 影像空间分辨率较低和地表覆盖错综复杂, 地物在遥感影像上往往交错分布, 且呈现出大小不一, 形状各异的空间特征, 故采用单一形状、固定大小的窗口将不能真实地反映地物复杂多变的形态特征。

因此, 研究自适应邻域构建方法以有效表达错综复杂的地表覆盖情况将进一步提高动态集成的分类效果; 本文进一步提出联合自适应邻域空间信息的可变K邻域动态集成算法。该算法与VKS算法基本思想一致, 唯一的区别在于将固定窗口替换为适应性更强的自适应窗口。

考虑到高光谱遥感影像的高维特点, 基于高光谱遥感影像构建像素的自适应邻域不可避免地导致计算时间消耗大。而高光谱遥感影像第一主成分中保留着大部分有用的光谱信息。因此, 应用高光谱遥感影像的第一主成分构建每个像素的形状自适应SA (Shape Adaptive) 局部邻域。如图2所示, SA邻域为多边形, 其形状和大小由其8个多边形顶点的位置确定。8个方向 θ_k (45° 间隔且 $k = 1, 2, \dots, 8$) 是已知的, 而从中心像素到顶点的8个长度 h_k , $k = 1, 2, \dots, 8$ 需要设置。设置的长度取值范围为 $\{1; 2; 3; 5; 7; 9; 12\}$ 。SA邻域的构建主要包含两个步骤。首先, 采用局部多项式LPA (Local Polynomial Approximation) 滤波核函数 $\{g_{h_k}, \theta_k\}$ 获取8个方向的一组可变尺度估计值和误差。

$$\{X_{h_k, \theta_k}^{PC1}\} = X^{PC1} \otimes \{g_{h_k}, \theta_k\} \quad (7)$$

$$\varepsilon_{h_k, \theta_k} = \left| X_{h_k, \theta_k}^{PC1} - X_{noise_reduced}^{PC1} \right| \quad (8)$$

根据交叉置信区间规则ICI (Intersection of Confidence Interval rule) 比较尺度估计值, 从而确定8个方向上的最佳射线长度。最后连接八条射线

的端点来获得每个中心像素的SA区域。

$$E_{h_k, \theta_k} = [\hat{X}_{h_k, \theta_k}^{PC1} - \varepsilon_{h_k, \theta_k}, \hat{X}_{h_k, \theta_k}^{PC1} + \varepsilon_{h_k, \theta_k}] \quad (9)$$

$$\tilde{h} = \{h_i | E_{h_i, \theta_i} \cap E_{h_j, \theta_j} \neq \emptyset \& E_{h_i, \theta_i} \cap E_{h_j, \theta_j} = \emptyset, \forall h_i \in H\}$$

SA方法能够根据地物空间形态结构自适应地构建局部邻域, 且对带状、圆形和方形等规则结构和不规则多边形结构均具有较强的适应性, 并能够准确地反映地物边缘和细节特征 (Fu 等, 2016)。因此, 采用SA构建自适应邻域, 邻域内部的空间异质性得到了极大地削弱, 邻域的自适应调整使得地物的边缘和细节特征得以很好的保留 (Foi 等, 2007)。图2为SA构建的自适应邻域示意图。

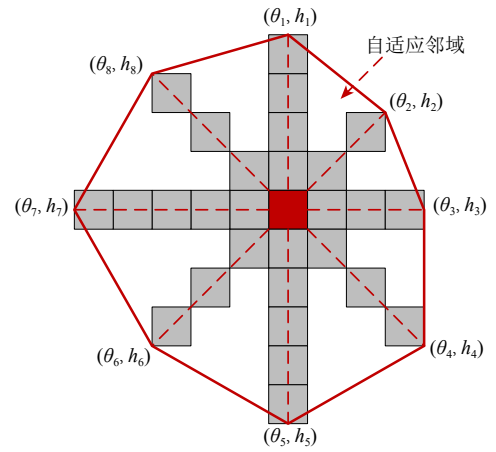


图2 形状自适应构建的自适应邻域示意图

Fig. 2 Illustration of the adaptive neighborhood constructed by shape-adaptive

图3为本文提出的动态集成算法中两种窗口的标记点示意图, 其中图3 (a) 图为固定窗口, 图3 (b) 图为自适应窗口。图3 (a) 固定窗口的大小设置为 3×3 。深灰色表示测试样本, 浅灰色代表标记点, 即第一阶段已经赋予标签的测试样本。统计标记点占窗口所有像元的比例, 若标记点的比例超过设定的阈值, 统计其中每一类别的标记像元的数量。假设第 m 类别的标记像元数目最多, 则测试样本的地物类别选定为第 m 类。图3 (b) 为自适应窗口, 该方式不需要提前设定窗口的大小, 自适应地根据测试样本生成SA邻域。同样地, 统计标记点的占比值, 若占比值超过阈值, 则该区域内存在最多的某类标记点的标签值赋予给测试样本。

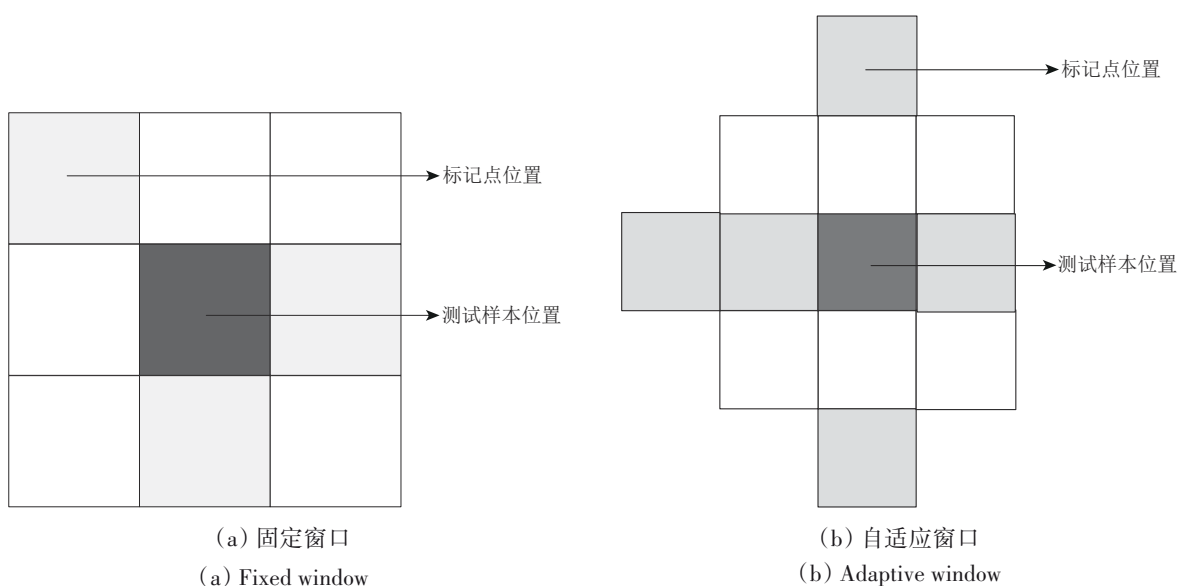


图3 标记点示意图

Fig. 3 Illustration of marking points

4 实验与分析

为验证 VKS 和 VKSA 算法对高光谱遥感影像分类的有效性, 分别采用 Purdue Campus、Indian Pines 和 Salinas 3 组高光谱遥感数据进行验证和分析。

4.1 实验设置

3 组实验分别采用切空间协同表示分类器 TCRC 和基于 Bagging 的 TCRC 算法的多数投票法 MV (Majority Voting), 动态集成方法 OLA、MLA 和 MCB 作为对比算法。其中, TCRC 算法利用测试样本的切平面估计局部流形, 在测试样本的切空间中使用协同表示算法。基于 Bagging 的 TCRC 算法采用 TCRC 为基分类器, 结合 Bagging 的集成方式。根据文献 (Su 等, 2020), 本文中集成次数设置为 20 次, 即构建 20 个 TCRC 的多分类器集合。该算法选择的集成方式为多数投票法, 是一种经典的静态集成方式。动态集成算法利用基于 Bagging 的 TCRC 集成算法提供的 20 个 TCRC 构建的多分类器集合。为方便与动态集成算法对比, 原始的基于 Bagging 的 TCRC 算法记为 MV。根据文献 (Su 等, 2020), 对于 Purdue Campus 数据, 参数 λ 设为 $1\text{E}-9$, η 设为 $1\text{E}-8$, n 设为 8。对于 Indian Pines 数据, 参数 λ 设为 $1\text{E}-9$, 参数 η 设为 $1\text{E}-8$, n 设为 8。Salinas 数据中参数 λ 设置为 $1\text{E}-$

8, 参数 η 设置为 $1\text{E}-8$, n 设置为 8。7 种分类方法重复 10 次实验取平均值, 参数设置采用交叉验证的方法。本文提出 VKS 和 VKSA 两种算法相关参数设置如下, 邻域 K 值的参数范围设置为 $\{5; 10; 15; 20\}$, 阈值的选取范围为 $\{0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9\}$ 。表 1 详细列出了 3 组数据集中 VKS 和 VKSA 两种算法的最优参数。

4.2 Purdue Campus 高光谱影像实验

实验数据一是普渡大学 (Purdue Campus) 西拉法叶分校影像数据, 于 1999 年 9 月 30 日通过机载高光谱制图仪 HYMAP (Hyperspectral Mapper) 系统采集, 在可见光和红外区域 (400—2400 nm) 涵盖 128 条光谱波段。实验中剔除水吸收波段后保留 126 条波段, 空间分辨率为 3.5 m。该实验数据大小为 377 像素 $\times$ 512 像素, 共包含 6 类地物分别为道路、草地、阴影、土壤、树木和建筑物。该高光谱数据假彩色图如图 4 (a) 所示, 样本分布如图 4 (b) 所示。标记样本随机地分为训练样本、验证样本和测试样本。从每类已标注地物类别的样本中随机选取 15 个训练样本作为训练集, 再随机选取 15 个样本共 90 个样本作为验证集, 剩余的标记像元为测试样本, 用于评价算法的分类效果。

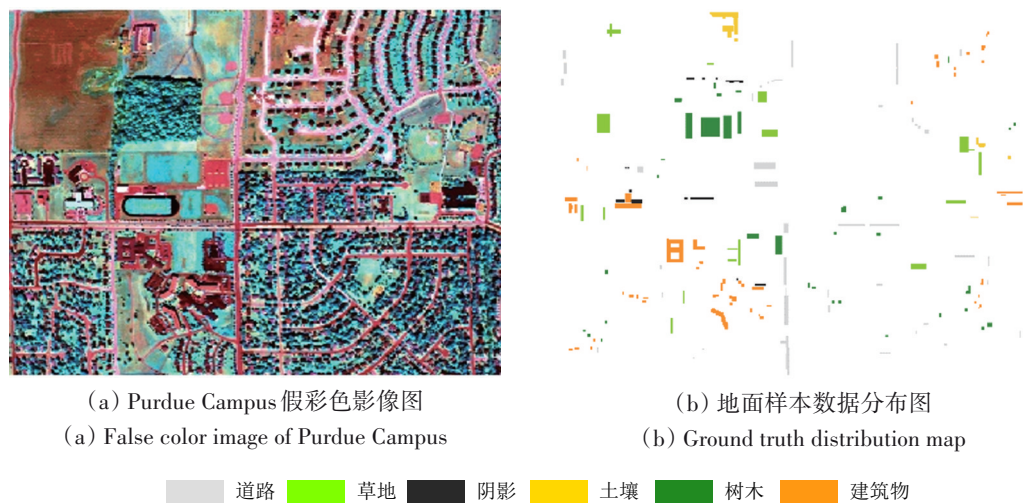


图4 Purdue Campus 假彩色影像图与地面样本数据分布图
Fig. 4 False color image and ground truth distribution map of Purdue Campus

为定量评价本文提出的两种动态集成算法的有效性，选择总体精度 OA（Overall Accuracy）、平均精度 AA（Average Accuracy）和 kappa 系数作为精度定量评价指标。该实验在最优参数环境下进行评价，最优参数如表 1 所示。

表 1 实验设置的最佳参数				
Table 1 The optimal parameters of experimental settings				
数据	参数	Purdue Campus	Indian pines	Salinas
VKS	K	0.5	5	10
	threshold	0.7	0.6	0.1
VKSA	K	5	5	10
	threshold	0.5	0.1	0.1

表 2 记录了 VKS 算法、VKSA 算法和其他对比算法对 HYMAP Purdue Campus 高光谱数据的 VKS 算法、VKSA 算法和其他对比算法具体的分类精度表现。7 种算法（TCRC、MV、OLA、MCB、MLA、VKS 和 VKSA）的总体分类精度分别为 91.04%、92.11%、91.45%、91.38%、91.96%、93.05% 和 93.17%。总体来说，本文提出的两种方法效果最好。就总体分类精度而言，VKSA 算法的分类精度达到 93.17%，比静态集成中多数投票法 MV 高 1.06%，比经典动态集成 OLA 算法、MCB 算法和 MLA 算法分别高了 1.72%、1.79% 和 1.09%，说明动态集成中挖掘局部空间的近邻关系可以有效地提高分类精度和鲁棒性。VKS 算法比静态集成的多数投票法 MV 高了 0.94%，比经典动态集成算法 OLA 算法、MCB 算法和 MLA 算法精度分别高了 1.60%、1.67% 和 0.97%。VKSA 算法的分类表现

略优于 VKS 算法，总体分类精度、平均精度和 kappa 系数分别高了 0.1%、0.4% 和 0.01%。说明动态集成框架中嵌入自适应的局部空间近邻关系能使分类效果更优且鲁棒性更强，在分类性能的提升上具有更大优势。表 2 中 OLA、MCB 和 MLA 3 种动态集成算法的总体分类精度分别低于静态集成的 MV 算法 0.66%、0.73% 和 0.15%，原因可能是训练集规模不足的情况下影响了动态集成的分类效果。图 5 展示了在 HYMAP Purdue Campus 高光谱数据中 7 种算法对于 HYMAP Purdue Campus 高光谱数据的土地覆盖分类效果图，可以发现，本文提出的 VKSA 算法具有最好的分类效果，能明显降低高光谱遥感影像分类的椒盐噪声，分类结果更接近于真实地物分布。VKS 算法的结果明显优于其他 3 种动态集成算法。

4.3 Indian Pines 高光谱影像实验

第二组实验数据采用由机载可见光红外成像光谱仪 AVIRIS（Airborne Visible Infrared Imaging Spectrometer）传感器获取的美国印第安纳州（Indian Pines）实验区的高光谱遥感影像，该数据从可见光到近红外区域（400—2450 nm）覆盖了 220 个波段，空间分辨率约为 20 m。剔除了水吸收波段后，有 200 个波段的影像数据用于实验。该实验区域大小为 145 像素×145 像素，共包括 16 类地物类型。为取得足够的训练样本去除 7 类地物，剩余的 9 类用于实验分析。该区域假彩色影像图如图 6（a）所示，地面样本数据如图 6（b）所示。

表2 Purdue Campus 分类精度和算法运行时间统计

Table 2 Classification accuracy and calculation time statistics of Purdue Campus

类别	训练样本	验证样本	测试样本	TCRC	MV	OLA	MCB	MLA	VKS	VKSA
1	15	15	1272	89.66	90.04	96.56	96.26	96.56	96.60	94.46
2	15	15	1099	84.62	86.04	96.57	95.77	96.49	97.28	98.17
3	15	15	204	91.10	87.93	63.70	59.36	64.76	71.30	74.39
4	15	15	364	86.20	87.18	95.42	94.50	95.42	96.28	95.76
5	15	15	1336	98.24	98.63	98.99	98.30	99.00	99.22	99.47
6	15	15	1270	94.14	96.83	77.07	79.38	79.14	81.73	83.04
OA/%				91.04	92.11	91.45	91.38	91.96	93.05	93.17
AA/%				90.65	91.11	88.05	87.26	88.56	90.40	90.88
Kappa				0.8866	0.9001	0.8921	0.8910	0.8984	0.9121	0.9136
时间/s				—	2.76	33.09	35.42	32.60	1553.33	1365.56

注:加粗字体为某一类地物类型分类精度最高数值。

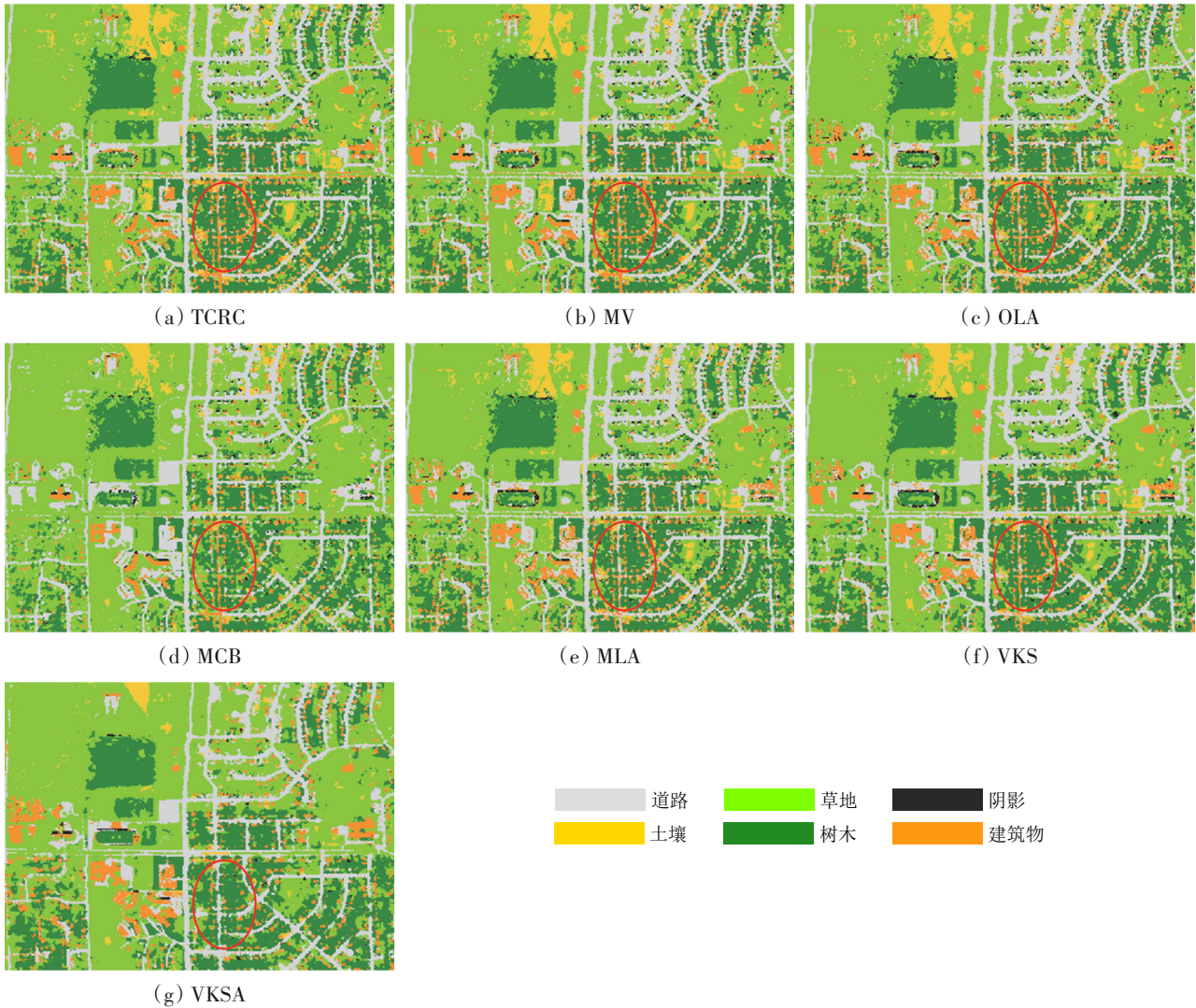


图5 Purdue Campus 数据算法分类土地覆盖图

Fig. 5 Classification maps of land types for different feature extraction algorithms of Purdue Campus

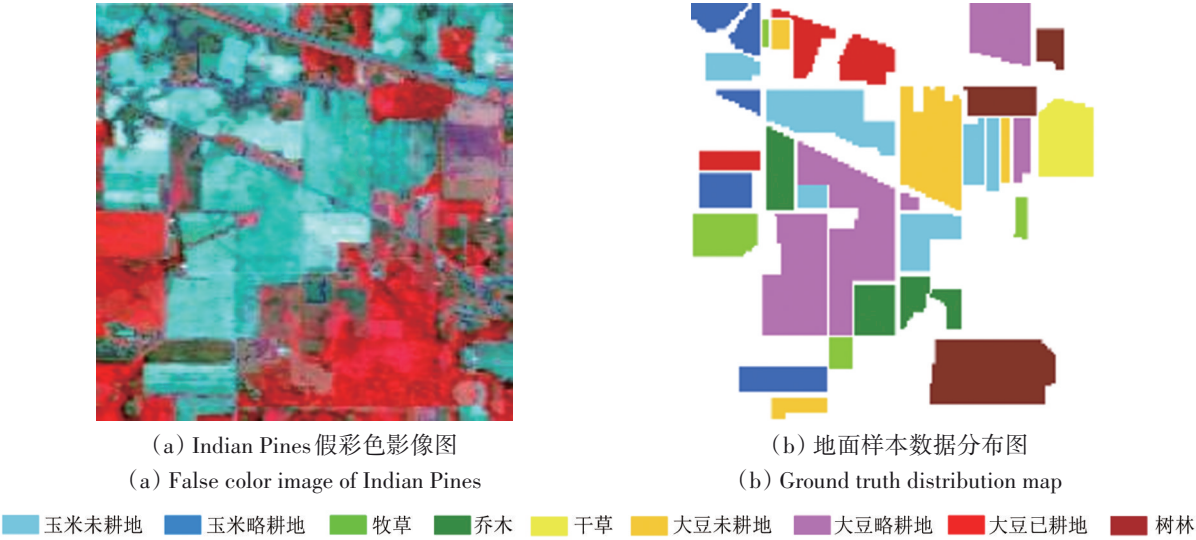


图6 Indian Pines 假彩色影像图与地面样本数据分布图
Fig. 6 False color image and ground truth distribution map of Indian Pines

实验中，地面真实数据随机分为训练样本、验证样本和测试样本。每类别随机选取2%的训练样本作为训练集，随机选取2%的样本作为验证样本，其余标记样本作为测试样本。该实验在最优参数环境下进行评价，最优参数如表1所示。表3中7种算法（TCRC、MV、OLA、MCB、MLA、VKS和VKSA）的总体分类精度分别为80.16%、81.49%、81.11%、80.80%、80.66%、83.07%和84.96%。针对AVIRIS Indian Pines分类场景，本文提出的两种法对比其他方法同样具有更出众的效果。就总体精度而言，VKSA算法精度达到84.96%，比静态集成的多数投票法MV高了3.5%，比经典动态集成算法OLA算法、MCB算法和MLA算法精度分别高了3.8%、4.1%和4.3%。VKS算法

分类精度也有明显的提升，比静态集成的多数投票法MV高了1.58%，比经典动态集成算法OLA算法、MCB算法和MLA算法精度分别高了1.86%、2.27%和2.41%。就平均精度和kappa系数来说，本文提出的VKS和VKSA两种动态集成算法展现了优异的分类性能。就总体精度而言，VKS算法的总体分类精度低于VKSA算法的1.9%。实验说明对Indian Pines数据而言，自适应窗口的设计方式明显优于固定窗口。7种分类算法对Indian Pines高光谱数据进行分类的土地覆盖情况如图7所示，图7（g）为VKSA算法的土地覆盖类型分类图，图中椒盐现象极大地被削弱，充分地反映了VKSA算法中联合自适应空间信息的有效性。

表3 Indian Pines分类精度和算法运行时间统计
Table 3 Classification accuracy and calculation time statistics of Indian Pines

类别	训练样本	验证样本	测试样本	TCRC	MV	OLA	MCB	MLA	VKS	VKSA
1	72	72	1284	76.83	77.59	82.41	83.36	82.35	85.90	89.58
2	42	42	746	81.53	80.68	47.53	47.21	48.02	49.43	54.48
3	25	25	433	97.66	97.89	88.45	87.43	87.39	88.28	91.18
4	37	37	656	94.57	95.32	97.46	96.95	97.01	99.24	99.39
5	24	24	430	99.54	99.82	99.95	99.81	100	100	100
6	49	49	874	79.86	80.95	62.29	57.92	60.64	63.13	63.20
7	123	123	2209	66.28	67.66	86.64	87.01	86.08	91.12	95.36
8	30	30	533	80.75	88.23	52.04	53.02	50.84	49.53	42.33
9	64	64	1137	98.46	98.76	99.67	99.44	99.65	99.65	99.88
OA/%				80.16	81.49	81.11	80.80	80.66	83.07	84.96
AA/%				86.16	87.43	79.61	79.13	79.11	81.35	81.71
Kappa				0.7623	0.7758	0.7748	0.7704	0.7692	0.7977	0.8199
时间/s				—	0.36	147.78	486.37	494.68	508.30	460.53

注：加粗字体为某一地物类型分类精度最高数值。

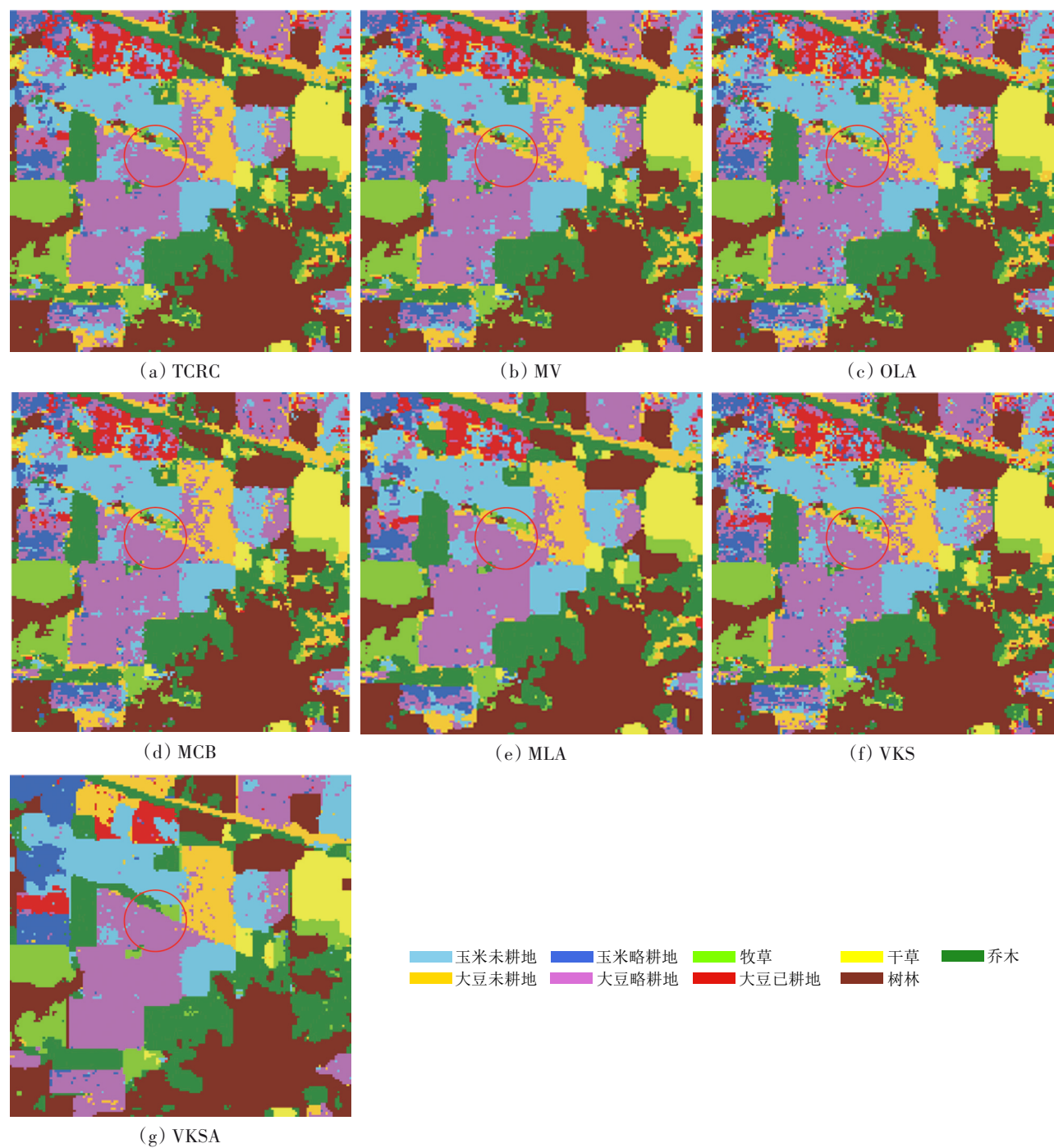


图7 Indian Pines数据算法分类土地覆盖图

Fig. 7 Classification maps of land types for different feature extraction algorithms of Indian Pines

4.4 Salinas 高光谱影像实验

第3组实验数据是美国加利福尼亚州Salinas山谷的高光谱遥感数据，该数据于1998年采集。该数据分辨率较高，达到3.7 m，空间尺寸为512像素×217像素。实验中去除水吸收波段后共保留204个

光谱波段。每类随机选取10个训练样本作为训练集，每类随机选择10个样本作为验证集，其余标记样本作为测试样本。该实验区域大小为512像素×217像素，共包括16类地物类型。该区域假彩色影像如图8(a)所示，地面样本数据如图8(b)所示。

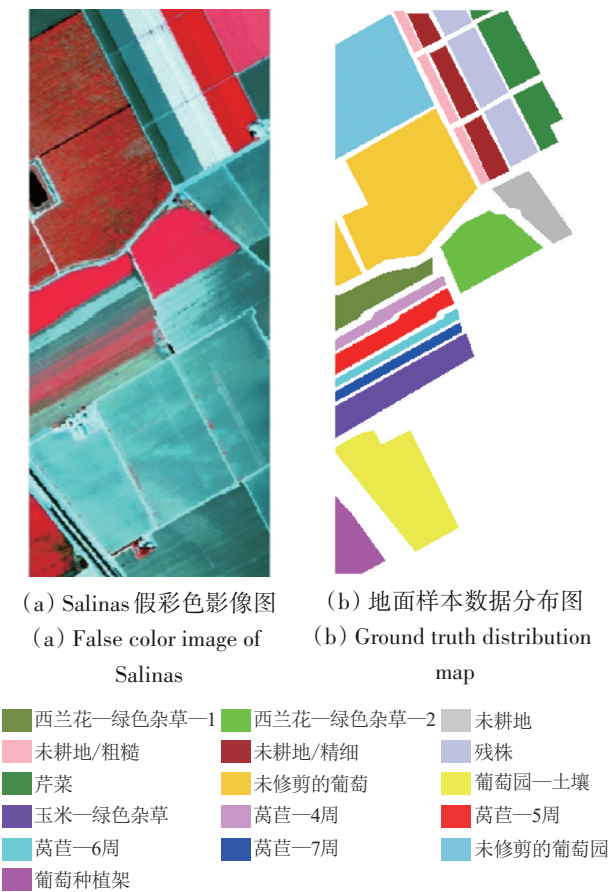


图8 Salinas假彩色影像图与地面样本数据分布图
Fig. 8 False color image and ground truth distribution map of Salinas

该实验在最优参数环境下进行评价，最优参数如表1所示。表4为7种分类算法对于Salinas高光谱数据的分类结果。7种算法（TCRC、MV、OLA、MCB、MLA、VKS和VKSA）的总体分类精度分别为84.51%、85.19%、86.11%、85.61%、86.75%、87.18%和87.76%。类似于前两组实验数据，提出的VKS和VKSA同样具有不错的分类表现。就总体精度而言，VKSA达到87.76%，比多数投票法MV分类精度高2.5%，比动态集成中的OLA算法、MCB算法和MLA算法分别提高了1.6%、2.1%和1%，比VKS算法提高0.58%。对Salinas数据而言，自适应窗口相较于固定窗口在动态集成分类中更能有效地挖掘欧洲航天局局部邻关系。VKS算法要略差于VKSA算法，但与其他动态集成分类算法相比，结果都有不同程度的提升。本文7种分类算法的分类效果图如图8所示，图8（g）为VKSA算法的土地覆盖分类图，分类图定性证实了VKSA算法增强了对空间同质区域的分类效果，提升了地物的可分性。总体而言，对Salinas数据来说，本文提出的联合空间信息的动态集成方法具有更出众分类性能，且相较于固定窗口，自适应窗口可提供更强的优化性能。

表4 Salinas数据各算法的分类精度和算法运行时间统计
Table 4 Classification accuracy and calculation time statistics of Salinas

类别	训练样本	验证样本	测试样本	TCRC	MV	OLA	MCB	MLA	VKS	VKSA
1	10	10	1989	99.91	99.94	99.24	97.65	99.24	99.55	99.80
2	10	10	3706	99.21	99.24	99.97	99.48	99.97	99.46	99.42
3	10	10	1956	97.04	97.62	97.55	93.72	97.80	98.84	98.68
4	10	10	1374	98.19	98.20	93.61	93.77	93.41	97.21	98.51
5	10	10	2658	94.97	94.47	98.40	97.38	98.37	95.96	96.43
6	10	10	3939	99.89	99.96	100	100	100	99.96	99.96
7	10	10	3659	99.79	99.79	99.98	99.88	99.96	99.98	99.76
8	10	10	11251	65.68	68.27	68.73	68.51	71.14	72.81	73.93
9	10	10	6183	94.89	94.83	99.91	99.53	99.91	97.18	98.10
10	10	10	3258	91.00	91.23	87.57	87.91	88.59	91.70	94.00
11	10	10	1048	96.03	96.32	98.09	97.42	98.21	96.38	96.85
12	10	10	1907	95.76	95.36	85.56	91.00	87.98	97.36	97.68
13	10	10	896	76.05	77.49	92.19	91.96	91.47	90.53	91.20
14	10	10	1050	84.10	85.58	93.49	94.11	93.96	87.00	90.81
15	10	10	7248	59.17	60.06	60.35	57.97	60.29	60.52	60.84
16	10	10	1787	100	100	97.48	97.39	97.72	100	100
OA/%				84.51	85.19	86.11	85.61	86.75	87.18	87.76
AA/%				90.73	91.15	92.21	91.73	92.38	92.78	93.50
Kappa				0.8276	0.8352	0.8462	0.8400	85.26	85.77	86.38
时间/s				—	1.63	414.42	455.85	417.06	569.37	596.00

注：加粗字体为某一地物类型分类精度最高数值。

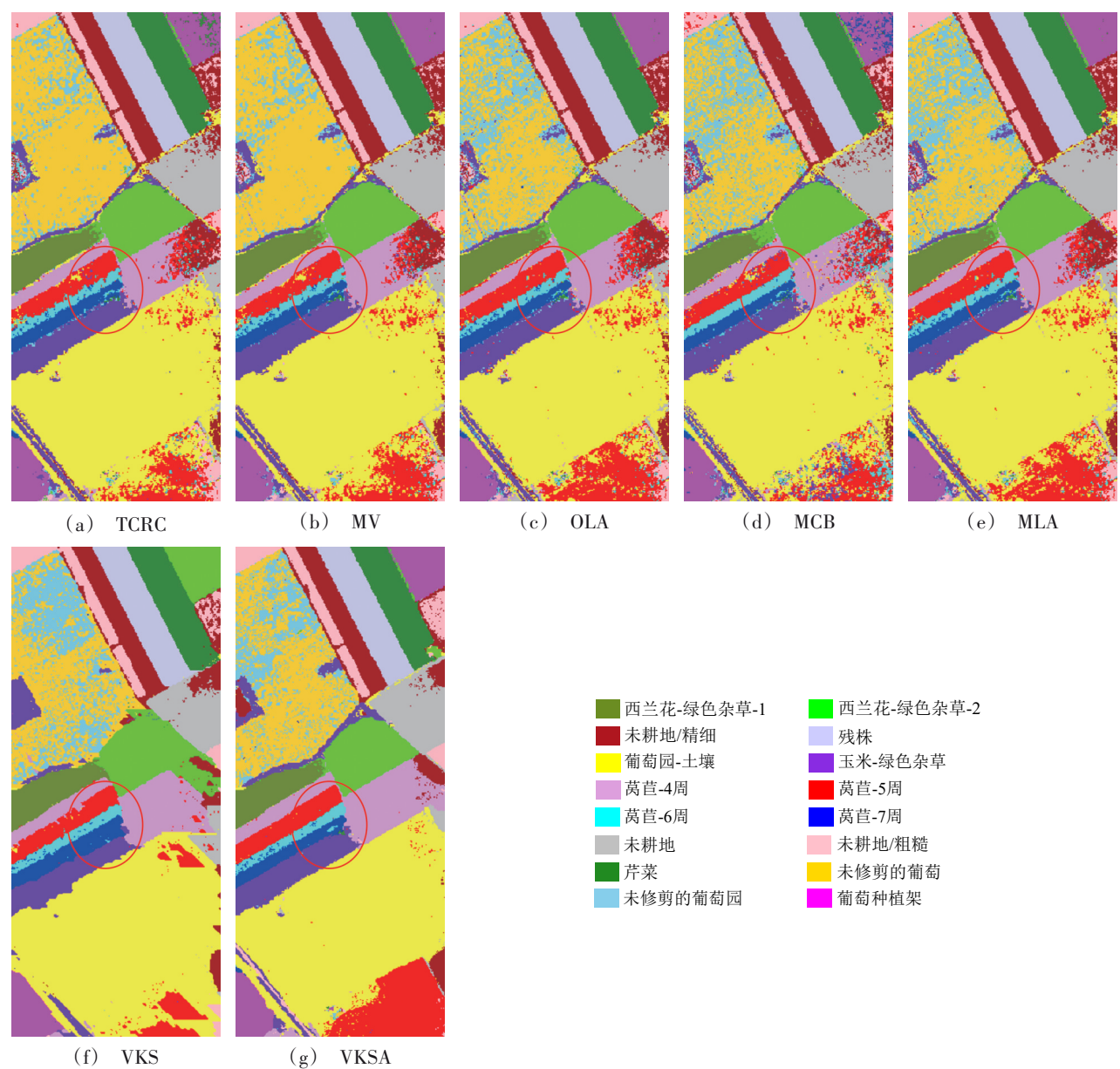


图9 Salinas 数据算法分类土地覆盖图
Fig. 9 Classification maps of land types for different feature extraction algorithms of Salinas

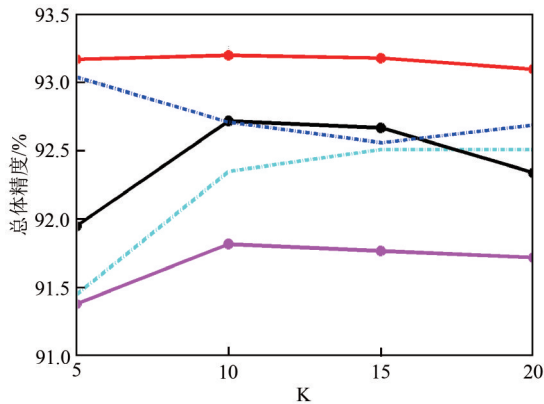
4.5 参数分析

(1) 时间复杂度分析。表2、表3和表4统计了7种算法10次运行的平均时间（仅集成过程，不包括前期分类器生成，分类器的分类过程）。本文对比的5种集成算法中，MV算法运算速率最快，OLA算法次之，MCB算法运算速率最慢。VKSA算法对比算法相比时间消耗明显更多，但与VKS算法时间消耗相差无几。但VKSA算法的分类精度却明显优于VKS算法。综合考虑分类精度和时间复杂度，VKSA算法更优异。

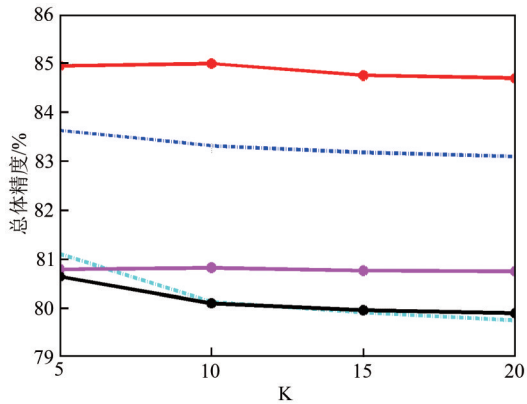
(2) 邻域参数K。图10展示了OLA、MCB、

MLA、VKS和VKSA等5种动态集成算法的总体分类精度值随邻域参数K值变化的趋势。其余参数设置如表1所示。针对HYMAP Purdue Campus 高光谱数据，由图10（a）中可见，OLA、MCB、MLA和VKS等4种动态集成算法的分类精度随K值的变化呈现出不同的变化趋势。但VKSA算法的分类性能明显优于其他4种动态集成算法，且该算法受K值影响较小，说明VKSA算法分类性能高且鲁棒性较强。VKS算法的总体分类精度随着K值的增加呈现先下降随后略有提升的趋势。原因在于Purdue Campus数据为城市区域，地物形状结构复杂多变，

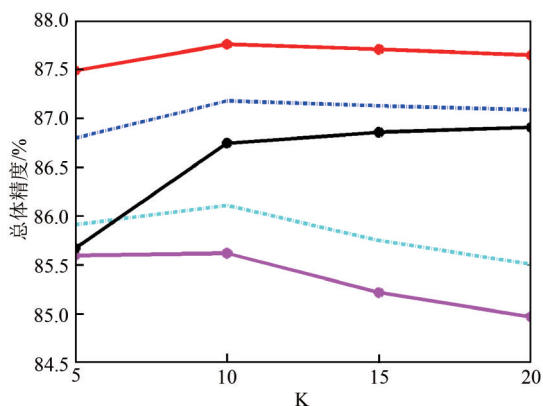
较小的 K 值（即较小的邻域窗口）更能真实地描述复杂的地物形态结构。VKS算法对 K 值较为敏感，当 K 值为15:00，该算法的总体分类精度降到92.5%，甚至低于MLA动态算法的分类精度。由图10（a）中可见，VKS和VKSA两种动态算法的最佳 K 值均为5。



(a) Purdue Campus



(b) Indian Pines



(c) Salinas

— OLA — MCB — MLA
— VKS — VKSA

图10 3组数据算法 K -总体精度图变化曲线图

Fig. 10 Overall accuracy of different algorithms for three data set with varying K

对Indian Pines高光谱数据而言，MLA、MCB、VKS和VKSA 4种动态集成算法相较于OLA动态集成算法受 K 值影响较小。不同于Purdue Campus高光谱城市数据，Indian Pines为形状较为规则统一的高光谱农村数据，所以固定窗口的尺寸大小对VKS动态集成算法影响较小。VKSA算法总体分类精度约为85%，高于VKS算法分类精度约1%，高于其他3种动态集成算法分类精度约4%。由图10（b）中可见， K 值设置为5时，VKS和VKSA两种动态集成算法的分类表现最佳。

针对第3组分类场景Salinas数据，VKSA和VKS两种动态集成算法的分类精度随 K 值的变化呈现出相同的变化趋势。随着 K 值的增加，两种算法的分类精度均先升后降。两种动态集成算法的最佳 K 值均为10。由图10（c）可见，VKSA算法的分类精度明显高于其余4种动态集成分类算法。

（3）阈值参数 threshold。图11对应VKS和VKSA两种动态集成算法的总体分类精度值随阈值的变化趋势。其余参数设置如表1所示。阈值越大，即标记点所占比例越大，代表算法越依赖空间信息。针对第1组Purdue Campus高光谱数据，对于VKS算法而言，最佳阈值参数为0.7。当阈值范围在0.1—0.5区间，VKS算法分类精度保持平稳，之后随着阈值的继续增加，分类精度先上升达到最高点后下降。VKSA算法的最佳阈值为0.5，分类精度随阈值的变化呈现出先上升后下降的趋势。当阈值取值范围于0.2—0.6，VKSA算法分类精度高于VKS算法。其余阈值区间，VKS算法分类表现反而优于VKSA算法。

针对第2组Indian Pines高光谱数据，当阈值位于0.1至0.5取值区间内，VKS算法分类精度保持不变，随后总体精度先上升后下降。该算法的最佳阈值为0.6。对VKSA算法而言，最佳阈值参数为0.1。VKSA算法的总体分类精度随着阈值的增加一直下降。由图11（b）图可见，VKSA算法的总体精度高于VKS算法。

针对第3组Salinas高光谱数据，VKS算法的总体精度先保持不变，之后再下降，最佳阈值设置为0.5。而VKSA算法的分类表现随着阈值增加反而变差，最佳阈值为0.1。由图11（c）中可见，无论 K 值如何变化，VKSA算法的总体精度一直高于VKS算法。

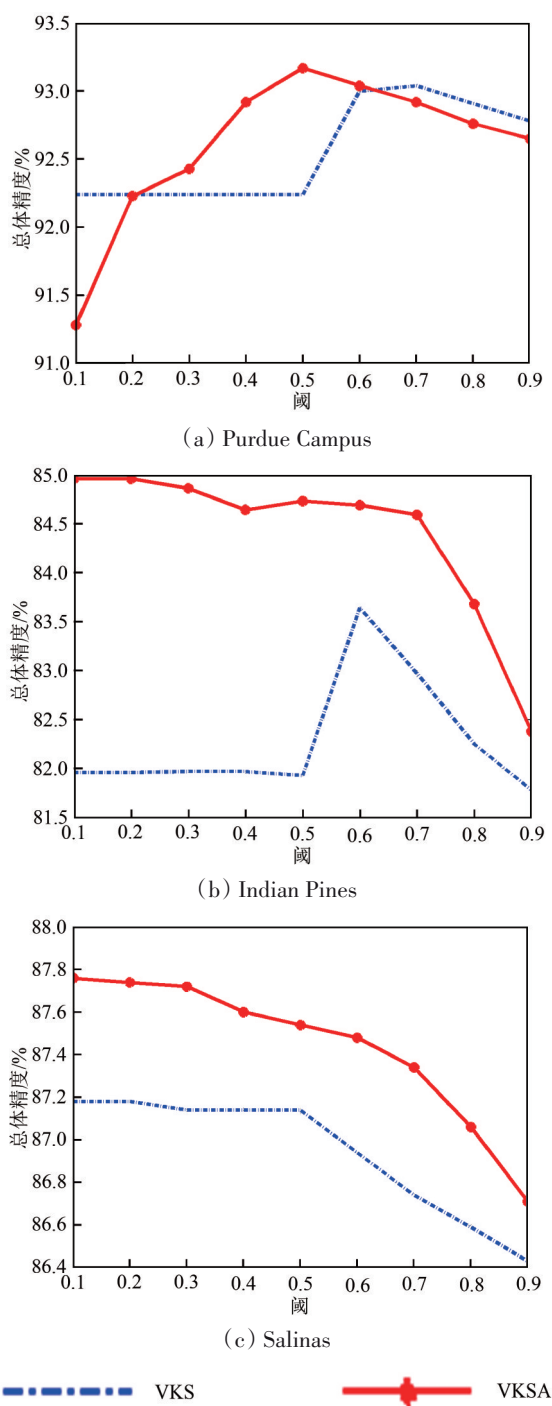


图 11 3组数据算法 threshold-总体精度变化曲线图

Fig. 11 Overall accuracy of different algorithms for three data set with varying threshold

5 结 论

本文提出了两种高光谱遥感影像动态集成分类算法：VKS和VKSA。其中VKS算法综合利用高光谱遥感影像的光谱特性和空间特征，首先通过同时限制分类器的行为和分类器的精度自适应地改变K邻域，若不存在满足条件的K邻域，则运用

固定窗口的方式利用邻域样本的空间信息。考虑到自适应窗口更能精确反映地物的形态特征，进一步设计了基于自适应窗口的VKSA算法。3组高光谱遥感影像实验结果表明与其他几种动态集成方法相比，本文提出的两种联合空间信息的动态集成分类算法能显著提升分类精度，其中与基于固定窗口的VKS算法相比，基于自适应窗口方式的VKSA算法能获得更优的分类效果。

与其他动态集成算法相比，本文提出的两种算法优点如下：

(1) 提出的动态集成算法首次同时考虑分类器的分类精度和相似度自适应地改变测试样本的K邻域，使得设置的能力区域具有更高的可靠性与灵活性。

(2) 提出的动态集成算法增加了地物的局部空间近邻信息。VKS算法和VLSA算法分别设计了固定窗口和自适应窗口的嵌入方式融合了样本的局部空间近邻关系，可充分利用高光谱遥感影像地物复杂的空间形态结构信息。

但本文提出的两种算法涉及的参数比较多，如何考虑不同高光谱影像数据的特点，选择合适的参数组合达到最佳分类效果需要进一步研究。

参考文献(References)

- Benediktsson J A, Chanussot J and Fauvel M. 2007. Multiple classifier systems in remote sensing: from basics to recent developments// 7th International Workshop on Multiple Classifier Systems. Prague: Springer: 501-512 [DOI: 10.1007/978-3-540-72523-7_50]
- Bioucas-Dias J M, Plaza A, Camps-Valls G, Scheunders P, Nasrabadi N and Chanussot J. 2013. Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and Remote Sensing Magazine, 1(2): 6-36 [DOI: 10.1109/mgrs.2013.2244672]
- Chang C I. 2003. Hyperspectral Imaging: Techniques for Spectral Detection and Classification. New York: Plenum Publishing Co
- Cruz R M O, Sabourin R and Cavalcanti G D C. 2015. META-DES. H: a dynamic ensemble selection technique using meta-learning and a dynamic weighting approach//Proceeding of 2015 International Joint Conference on Neural Networks (IJCNN). Killarney: IEEE, 2015: 1-8 [DOI: 10.1109/IJCNN.2015.7280594]
- Cruz R M O, Sabourin R and Cavalcanti G D C. 2018. Dynamic classifier selection: recent advances and perspectives. Information Fusion, 41: 195-216 [DOI: 10.1016/j.inffus.2017.09.010]
- Damodaran B B and Nidamanuri R R. 2014. Dynamic linear classifier system for hyperspectral image classification for land cover mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(6): 2080-2093 [DOI: 10.1109/JSTARS.

- 2013.2294857]
- Damodaran B B, Nidamanuri R R and Tarabalka Y. 2015. Dynamic ensemble selection approach for hyperspectral image classification with joint spectral and spatial information. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6): 2405-2417 [DOI: 10.1109/JSTARS.2015.2407493]
- Didaci L and Giacinto G. 2004. Dynamic classifier selection by adaptive k-nearest-neighbourhood rule//5th International Workshop on Multiple Classifier Systems. Cagliari: Springer: 174-183 [DOI: 10.1007/978-3-540-25966-4_17]
- Dos Santos E M, Sabourin R and Maupin P. 2008. A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. *Pattern recognition*, 41(10): 2993-3009 [DOI: 10.1016/j.patcog.2008.03.027]
- Du P J, Xia J S, Xue Z H, Tan K, Su H J and Bao R. 2016. Review of Hyperspectral remote sensing image classification. *Journal of Remote Sensing*, 20(2): 236-256 (杜培军, 夏俊士, 薛朝辉, 谭琨, 苏红军, 鲍蕊. 2016. 高光谱遥感影像分类研究进展. *遥感学报*, 20(2): 236-256) [DOI: 10.11834/jrs.20165022]
- Foi A, Katkovnik V and Egiazarian K. 2007. Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images. *IEEE Transactions on Image Processing*, 16(5): 1395-1411 [DOI: 10.1109/TIP.2007.891788]
- Fu W, Li S T, Fang L Y, Kang X D and Benediktsson J A. 2016. Hyperspectral image classification via shape-adaptive joint sparse representation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2): 556-567 [DOI: 10.1109/JSTARS.2015.2477364]
- García S, Zhang Z L, Altalhi A, Alshomrani S and Herrera F. 2018. Dynamic ensemble selection for multi-class imbalanced datasets. *Information Sciences*, 445-446: 22-37 [DOI: 10.1016/j.ins.2018.03.002]
- Giacinto G and Roli F. 2001. Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition*, 34(9): 1879-1881 [DOI: 10.1016/S0031-3203(00)00150-3]
- He L, Li J, Liu C Y and Li S T. 2018. Recent advances on spectral-spatial hyperspectral image classification: an overview and new guidelines. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3): 1579-1597 [DOI: 10.1109/TGRS.2017.2765364]
- Hou C Q, Xia Y J, Xu Z R and Sun J. 2016. Learning classifier competence based on graph for dynamic classifier selection//Proceeding of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Changsha: IEEE, 2016: 1164-1168 [DOI: 10.1109/FSKD.2016.7603343]
- Ko A H R, Sabourin R and Britto A S. 2008. From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, 41(5): 1718-1731 [DOI: 10.1016/j.patcog.2007.10.015]
- Kumar S, Ghosh J and Crawford M M. 2002. Hierarchical fusion of multiple classifiers for hyperspectral data analysis. *Pattern Analysis and Applications*, 5(2): 210-220 [DOI: 10.1007/s100440200019]
- Kuncheva L I. 2000. Clustering-and-selection model for classifier combination//KES'2000. Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies. Proceedings (Cat. No.00TH8516). Brighton: IEEE: 185-188 [DOI: 10.1109/KES.2000.885788]
- Li D Y, Wen G H, Li X and Cai X F. 2019. Graph-based dynamic ensemble pruning for facial expression recognition. *Applied Intelligence*, 49(9): 3188-3206 [DOI: 10.1007/s10489-019-01435-2]
- Li W and Du Q. 2014. Joint within-class collaborative representation for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6): 2200-2208 [DOI: 10.1109/JSTARS.2014.2306956]
- Peng J T, Sun W W and Du Q. 2019. Self-paced joint sparse representation for the classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2): 1183-1194 [DOI: 10.1109/TGRS.2018.2865102]
- Peng J T, Zhou Y C, Sun W W, Du Q and Xia L K. 2021. Self-paced nonnegative matrix factorization for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 59(2): 1501-1515 [DOI: 10.1109/TGRS.2020.2996688]
- Smits P C. 2002. Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Transactions on Geoscience and Remote Sensing*, 40(4): 801-813 [DOI: 10.1109/TGRS.2002.1006354]
- Su H J. 2022. Dimensionality reduction for hyperspectral remote sensing: Advances, challenges, and prospects. *National Remote Sensing Bulletin*, 26(8): 1504-1529 (苏红军. 2022. 高光谱遥感影像降维: 进展、挑战与展望. *遥感学报*, 26(8): 1504-1529) [DOI: 10.11834/jrs.20210354]
- Su H J and Liu H. 2017. A novel dynamic classifier selection algorithm using spatial-spectral information for hyperspectral classification. *Remote Sensing for Land and Resources*, 29(2): 15-21 (苏红军, 刘浩. 2017. 一种利用空间和光谱信息的高光谱遥感多分类器动态集成算法. *国土资源遥感*, 29(2): 15-21) [DOI: 10.6046/gtzyyg.2017.02.03]
- Su H J, Yu Y, Du Q and Du P J. 2020. Ensemble learning for hyperspectral image classification using tangent collaborative representation. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6): 3778-3790 [DOI: 10.1109/TGRS.2019.2957135]
- Su H J, Zhao B, Du Q and Sheng Y H. 2016. Tangent distance-based collaborative representation for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 13(9): 1236-1240 [DOI: 10.1109/LGRS.2016.2578038]
- Tang X, Meng F B, Zhang X R, Cheung Y M, Ma J J, Liu F and Jiao L C. 2021. Hyperspectral image classification based on 3-D octave convolution with spatial - spectral attention network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3): 2430-2447 [DOI: 10.1109/TGRS.2020.3005431]
- 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]
- Woloszynski T and Kurzynski M A. 2011. Probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recognition*

- ognition, 44(10-11): 2656-2668 [DOI: 10.1016/j.patcog.2011.03.020]
Woods K, Kegelmeyer W P and Bowyer K. 1997. Combination of multiple classifiers using local accuracy estimates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(4): 405-410
[DOI: 10.1109/34.588027]
Zhang L P and Zhang L F. 2005. Hyperspectral Remote Sensing. Wuhan: Wuhan University Press (张良培, 张立福. 2005. 高光谱遥感. 武汉: 武汉大学出版社)

Dynamic selection algorithm for collaborative representation of hyperspectral remote sensing based on joint spatial information

YU Yao¹, SU Hongjun², TAO Yang¹

1. Basic Geographic Information Center of Jiangsu Province, Nanjing 210013, China;

2. School of Earth Science and Engineering, Hohai University, Nanjing 211100, China

Abstract: Ensemble learning has recently attracted considerable attention for hyperspectral image analysis. This model integrates multiple base classifiers for joint decision making, which is better than using a base classifier. Ensemble learning includes static and dynamic classifier ensembles. In the static ensemble method, the same classifier combination scheme is selected for testing sample. However, this method ignores the difference in classifier performance for each testing sample. Considering the features of testing sample, the best classifier is selected adaptively in dynamic ensemble methods. Therefore, this classifier can generally achieve better performance than static ensemble methods for hyperspectral image classification. However, numerous dynamic ensemble methods only consider the spectral information of the validation and training samples, ignoring the rich spatial information of hyperspectral images.

A Variable K-neighborhood and Spatial information algorithm (VKS) is proposed in this paper to further improve the accuracy and reliability of hyperspectral image classification. Firstly, the VKS algorithm comprehensively considers the accuracy and similarity of the classifier to adaptively adjust the K-neighborhood of the testing sample, increasing the reliability and flexibility of the region of competence setting. Thus, the testing samples with good spectral discrimination performance are preferentially classified. The label information of spatial neighborhood samples is used for predicting the testing samples with poor spectral discrimination performance. A fixed window is designed to provide local spatial information in hyperspectral images. However, fixed windows cannot reveal the complex and changeable morphological characteristics of ground objects. An adaptive window that can effectively reflect complex spatial information is proposed to capture the complex and changeable spatial structure in a hyperspectral image, and a variable K-neighborhood with a shape-adaptive (VKSA) algorithm is further designed.

The Purdue Campus, Indian Pines, and Salinas hyperspectral remote sensing data sets are used to design experiments and verify the performance of the proposed VKS and VKSA algorithms. Four state-of-the-art methods, namely, majority voting, overall local accuracy, modified local accuracy, and multiple classifier behavior, are used to quantify the classification accuracy. Experimental results demonstrate that the VKS and VKSA algorithms outperform static ensemble methods and three classic dynamic ensemble methods in overall classification accuracy. Moreover, the VKSA algorithm with an adaptive window perform better than the VKS algorithm with a fixed window.

Key words: hyperspectral remote sensing, dynamic selection, shape-adaptive neighborhood, collaborative representation, image classification

Supported by National Natural Science Foundation of China (No. 42122008, 41871220)