

# 融合光谱度量标记迁移和 Tri-training 的高光谱遥感图像半监督分类算法

曹峰<sup>1</sup>, 李文涛<sup>1</sup>, 骆剑承<sup>2</sup>, 李德玉<sup>1,3</sup>, 钱宇华<sup>1,3,4</sup>, 白鹤翔<sup>1</sup>, 张超<sup>1</sup>

1. 山西大学计算机与信息技术学院(大数据学院), 山西 太原 030006;
2. 中国科学院空天信息创新研究院遥感科学国家重点实验室, 北京 100101;
3. 山西大学计算智能与中文信息处理教育部重点实验室, 山西 太原 030006;
4. 山西大学大数据与产业研究院, 山西 太原 030006

## 摘要

针对海量的高光谱遥感图像光谱和丰富的空间信息中可用于分类的有标记样本远少于无标记样本的数据特性, 提出了一种融合光谱度量标记迁移和 Tri-training 的高光谱遥感图像半监督光谱-空间分类算法。该算法提出了一种基于光谱度量的标记迁移方法, 通过结合迁移标记和 Tri-training 预测标记进行扩充样本标记预测, 提高了扩充样本标记的准确性。同时, 该算法基于空间相关性选择扩充样本, 综合运用光谱和空间特征提升图像分类的精度。在两个公开的高光谱遥感图像数据集上进行了实验, 结果表明该算法优于基于 Tri-training 算法的高光谱遥感图像的分类性能。

## 关键词

高光谱图像分类; 半监督分类; 纹理特征; 光谱度量; Tri-training 算法

中图分类号: TP751

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2022084

## *Semi-supervised classification algorithm for hyperspectral remote sensing images fusing spectral measure-based label transfer and tri-training*

CAO Feng<sup>1</sup>, LI Wentao<sup>1</sup>, LUO Jiancheng<sup>2</sup>, LI Deyu<sup>1,3</sup>, QIAN Yuhua<sup>1,3,4</sup>, BAI Hexiang<sup>1</sup>, ZHANG Chao<sup>1</sup>

1. School of Computer and Information Technology (School of Big Data), Shanxi University, Taiyuan 030006, China
2. State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
3. Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education, Shanxi University, Taiyuan 030006, China
4. Institute of Big Data and Industry, Shanxi University, Taiyuan 030006, China

### Abstract

Aimed at the problem that a large number of hyperspectral remote sensing images were rich in spectral and spatial information, and the labeled samples available for image classification were far less than unlabeled samples, a semi-supervised spectral-spatial classification algorithm was proposed by fusing spectral measure-based label transfer and Tri-training. A spectral measure-based label transfer method was proposed for our algorithm. The transferred labels and predicted labels for Tri-training algorithm were used to predict the labels of expanded unlabeled samples, which can promote the prediction accuracies of labels for expanded unlabeled samples. Meanwhile, our algorithm selected expanded samples based on spatial correlation, and used spectral and spatial features to improve the accuracy of image classification. Experimental study was executed on two public hyperspectral remote sensing image datasets, and the results showed that the proposed algorithm outperforms tri-training algorithm.

### Key words

classification of hyperspectral image, semi-supervised classification, texture feature, spectral measurement, tri-training algorithm

## 0 引言

随着高光谱遥感技术的飞速发展,高光谱遥感图像数据日益丰富,面向大规模高光谱遥感图像进行信息处理和信息提取,已成为大数据时代对地观测的重要手段。高光谱遥感图像分类作为高光谱遥感图像信息处理和信息提取的主要研究方向之一,已被广泛应用于地质制图、植被调查、城市规划、军事侦查和环境监测等领域。高光谱遥感图像包含丰富的光谱和空间信息<sup>[1]</sup>,可以更加细致地区分不同类型地物之间的差异,实现地物的精准分类。因此,结合光谱和空间特征的光谱-空间分类算法研究已成为当前高光谱遥感图像分类研究的热点<sup>[2]</sup>。

高光谱遥感图像光谱-空间分类算法主要包括3类:有监督、无监督以及半监督<sup>[3-4]</sup>。有监督分类算法基于有标记像元的光谱和空间特征构建分类器,并对未标记像元进行分类。无监督分类算法(通常称为聚类)不对像元进行类别标注,仅通过计算不同

像元之间光谱和空间特征的差异来实现像元的分类<sup>[5]</sup>。因为有监督分类算法利用了像元的标记信息,所以往往可以获得比无监督分类算法更优的分类性能。但是,对像元进行正确标记需要借助专家经验,甚至需要进行实地调研,因此会付出大量的人力和时间成本。尤其是当面向大数据量的高光谱遥感图像分类任务时,获取大量标记样本变得尤为困难。半监督分类算法利用少量有标记像元和大量无标记像元进行分类器训练,既可以弥补有监督分类算法的不足之处,又可以获得更优的分类性能。

高光谱遥感图像光谱-空间半监督分类研究受到了众多学者的普遍关注,并取得了一系列的研究成果。当前,光谱-空间半监督分类研究主要包括基于图、基于自训练和基于集成学习的半监督分类算法。基于图的半监督分类以所有样本为节点、以样本相似度为边的权重构造图模型,类标记通过图结构逐步传递至图中的其他样本。程志会等<sup>[6]</sup>提出了基于空间特征与纹理特征的高光谱图像半监督分类算法,将空间特征与纹理特征相结合,并利用图进行半监督分类。何浩等<sup>[7]</sup>提出了空间-光

谱约束的图半监督高光谱图像分类算法,将样本的空间相似性与光谱相似性相结合,并用相似性测度进行约束,进而基于图进行半监督分类。基于自训练的半监督分类算法,通过有标记样本训练出来的分类器对无标记样本进行预测,选择高置信度的“伪标记”样本加入训练集并不断迭代更新分类器,直至收敛。王俊淑等<sup>[8]</sup>在自训练迭代训练过程中,用最近邻域规则的数据剪辑策略对误标记样本进行过滤,提出了高光谱遥感图像DE-self-training半监督分类算法。王春阳<sup>[9]</sup>提出了基于信息熵的自训练半监督高光谱遥感影像分类算法,利用多分类Logistic算法得到类别概率信息,然后利用Renyi熵理论对无标记标签进行标注,扩充训练集后再进行迭代并预测分类。基于集成学习的半监督分类算法训练若干个基分类器,并通过一定的集成策略获得分类能力更强的学习器。Romaszewski M等<sup>[10]</sup>受到视频物体跟踪算法的启发,基于光谱特征和空间特征分别设计了两个分类器,共同组成Co-training模型。Tri-training作为一种经典的基于分歧的集成学习方法,十分适用于高光谱图像分类等多分类问题。王立国等<sup>[11]</sup>提出了基于改进Tri-training的高光谱图像半监督分类算法,在高光谱图像处理的各个环节均引入了空间信息。在基于Tri-training的光谱-空间高光谱遥感图像半监督分类研究中,研究者发现差异化的Tri-training基分类器即组合3种基于不同理论或方法的基分类器,可以获得稳定性和泛化性较好的分类模型<sup>[11-12]</sup>,因此得到了广泛使用。

Tri-training算法试图通过少量的初始有标记样本来学习较为完整的分类规则,故在扩充训练集的过程中难免会产生错分的情况,带有错误标记的样本进入训练集后会进一步影响正确分类规则的

学习,从而严重影响分类结果。基于此,本文提出了一种融合光谱度量(spectral measure, SM)标记迁移和Tri-training的高光谱遥感图像光谱-空间半监督分类算法(简记SMT)。该算法充分利用高光谱遥感图像的光谱和空间特征,通过光谱度量标记迁移修正Tri-training算法来扩充样本的类标记,降低扩充样本的错分率,进而提高半监督分类算法的精度。

## 1 Tri-training 半监督分类算法

Tri-training算法又被称为三重训练算法<sup>[13]</sup>,是在Co-training算法基础上发展起来的半监督协同训练分类算法。该算法重复训练3个基分类器,直到3个基分类器的分类结果不再变化时,算法的训练过程停止。Tri-training算法利用集成学习的思想进行未标记样本扩充,这种策略提高了未标记样本标记预测的置信度。与Goldman S提出的仅靠一个分类器来检测另一个分类器的经典Co-training算法<sup>[14]</sup>相比, Tri-training算法不但更加可靠有效,且无须进行交叉验证。Tri-training算法对基分类器的选择并没有特定要求,但研究表明,当3个基分类器不同时,该算法的预测精度和泛化性能往往更优。Tri-training算法通过计算样本错误率来改善错误分类的问题,然而当错误率满足更新条件时依然存在错分样本进入训练集的风险<sup>[15]</sup>。Tri-training算法的伪代码如**算法1**所示:

**算法1** Tri-training

**Input:** L:初始有标记样本集 U:无标记样本集 Learn:机器学习算法

**Output:**  $h(x)$

1 for  $i \in \{1, \dots, 3\}$  do

2  $S_i \leftarrow \text{BootstrapSample}(L); h_i \leftarrow \text{Learn}(S_i);$

```

 $e'_i \leftarrow 0.5$ ;  $l'_i \leftarrow 0$ ;
3 end for
4 repeat until none of  $h_i (i \in \{1, \dots, 3\})$ 
changes
5 for  $i \in \{1, \dots, 3\}$  do
6  $L_i \leftarrow \emptyset$ ; update $_i \leftarrow$  FALSE;  $e_i \leftarrow$  Measure
Error( $h_j$  &  $h_k$ )( $j, k \neq i$ );
7 if ( $e_i < e'_i$ ) then for every  $x \in U$  do
8 if  $h_j(x) = h_k(x)$  ( $j, k \neq i$ ) then
 $L_i \leftarrow L_i \cup \{(x, h_j(x))\}$ 
9 end if
10 end for
11 if ( $l'_i = 0$ ) then  $l'_i \leftarrow \left\lfloor \frac{e_i}{e'_i - e_i} + 1 \right\rfloor$ 
12 end if
13 if ( $l'_i < |L_i|$ ) then if ( $e_i |L_i| < e'_i l'_i$ ) then
update $_i =$  TURE
14 else if ( $l'_i > \frac{e_i}{e'_i - e_i}$ )
15 then  $L_i \leftarrow$  Subsample  $\left( L_i, \left\lfloor \frac{e'_i l'_i}{e_i} - 1 \right\rfloor \right)$ ;
update $_i =$  TURE
16 end if
17 end if
18 end if
19 end if
20 end for
21 for  $i \in \{1, \dots, 3\}$  do
22 if update $_i =$  TURE then  $h_i \leftarrow$  Learn
( $S_i \cup L_i$ );  $e'_i \leftarrow e_i$ ;  $l'_i \leftarrow |L_i|$ 
23 end for
24 end repeat
25 for rest  $x \in U$  do
26  $h(x) \leftarrow \arg \max_{y \in \text{label}} \sum_{i: h_i(x) = y} 1$ 
27 end for

```

高光谱遥感图像数据具有自身的一些特点,如丰富的光谱信息和空间信息;同类地物光谱特征相似,异类地物光谱特征不同;存在空间相关性和异质性等。当Tri-

training算法应用于高光谱遥感图像分类任务时,如何充分利用图像自身特性并结合算法的强大分类性能来设计更加高效的半监督分类方法,并进一步提升高光谱遥感图像的分类效果成为一项有意义的研究工作。

## 2 融合光谱度量标记迁移和Tri-training的高光谱遥感图像半监督分类算法

高光谱遥感图像同类地物之间具有较强的光谱相似性,而不同地物之间的光谱差异性则相对较大。另外,不同的光谱特征包含的信息量也具有差异性,因此,光谱特征在区分不同地物时发挥的作用不同。除了光谱特征,高光谱遥感影像还具有丰富的空间特征。基于此,本文充分利用光谱相似性、光谱信息量的差异性以及高光谱遥感影像的空间特征,通过集成Tri-training算法,提出了融合光谱度量标记迁移和Tri-training的高光谱遥感图像半监督分类算法,通过定义光谱度量来刻画样本之间的相似性,将基于光谱度量的标记迁移和Tri-training的基分类器相结合,减少错分样本进入训练集的风险。SMT算法将提高基于Tri-training算法的高光谱遥感图像分类的精度。融合光谱度量标记迁移和Tri-training的半监督分类算法流程如图1所示。

### 2.1 基于相关性度量的光谱特征重要性排序

光谱特征是高光谱遥感图像区分地物类别的主要特征,然而不同的光谱特征对分类的作用存在差异性。为了更好地区分这种差异,本文引入了基于相关性度量特征的重要性指标Merit(式(1))<sup>[6]</sup>,用来评

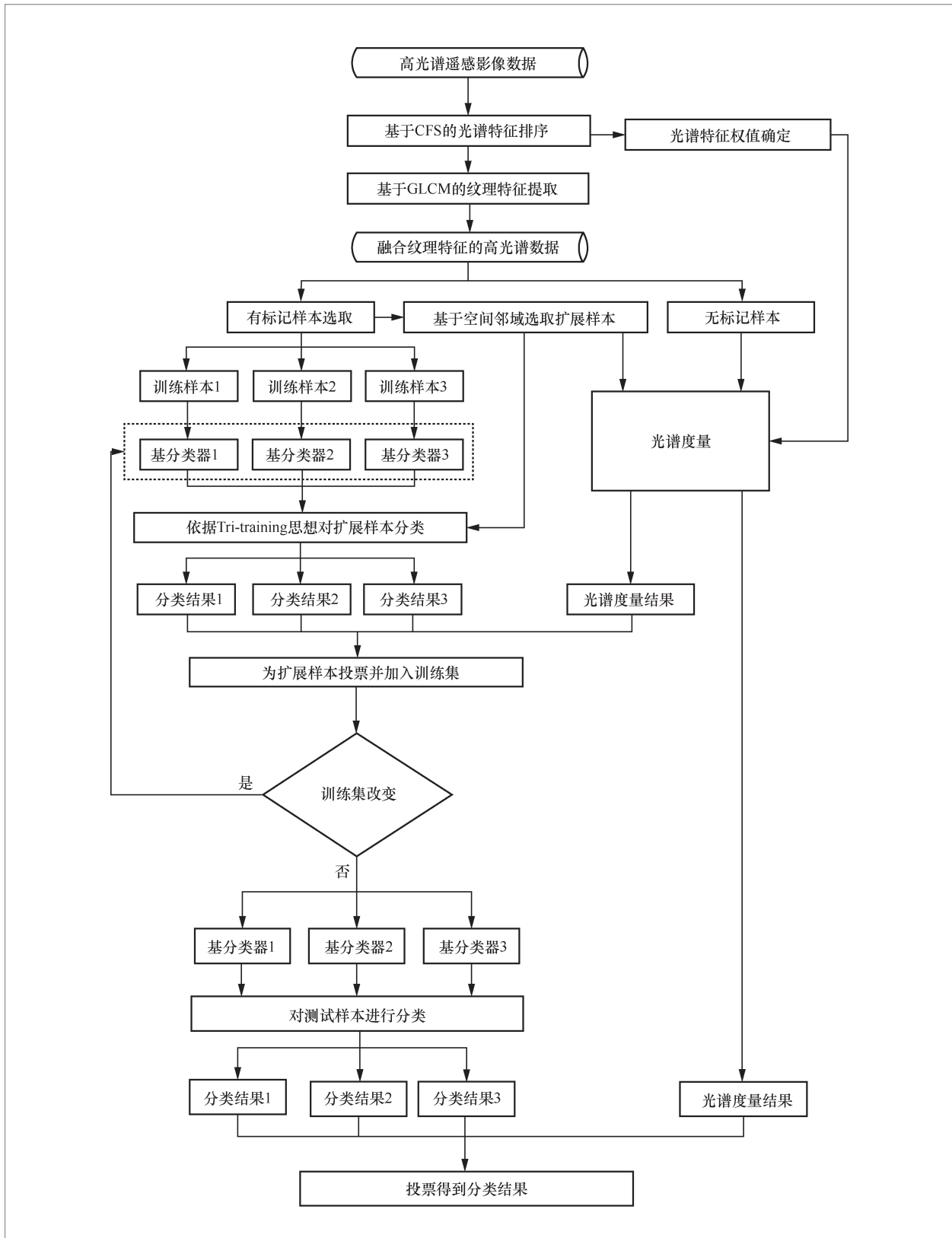


图 1 融合光谱度量标记迁移和 Tri-training 的半监督分类算法流程

价不同光谱特征的重要性并据此进行光谱特征排序。一个重要的特征应该与类标记之间具有较强的相关性,与其他特征具有较弱的相关性。Merit基于该思想,通过计算特征与类标记以及特征与特征之间的相关性来评价不同的特征对分类的重要性。

$$\text{Merit}(f) = \frac{k\overline{r_{cf}}}{\sqrt{k+k(k+1)\overline{r_{ff}}}} \quad (1)$$

其中,  $k$  是已选特征子集中的特征数;  $\overline{r_{cf}}$  表示特征和类标记之间相关性的平均值;  $\overline{r_{ff}}$  表示特征和特征之间相关性的平均值。

特征和类标记以及特征和特征之间的相关性定义如下:

$$r_{cf} = 2.0 \times \frac{I(f; C)}{H(f) + H(C)} \quad (2)$$

$$r_{ff} = 2.0 \times \frac{I(f; f_s)}{H(f) + H(f_s)} \quad (3)$$

$I(\bullet)$  表示两个随机变量之间的互信息,  $H(\bullet)$  表示单个随机变量的信息熵。

## 2.2 基于光谱度量的标记迁移

高光谱遥感图像具有极高的光谱相似性,一般情况下,光谱特征相近的两个像元极有可能属于同类地物。本文通过定义光谱度量来评价两个像元之间光谱特征的差异,实现有标记像元到无标记像元的标记迁移。无标记像元  $u$  与类标记为  $c$  的像元  $v$  的光谱度量定义如下:

$$\text{SM}(c) = \sum_{i=1}^k (\sigma_i |d_{vi}^c - d_{ui}|) \quad (4)$$

其中,  $k$  为波段数,  $\sigma_i$  为第  $i$  个波段的权重,  $d_{ui}$  为像元  $u$  在第  $i$  个波段的取值,  $d_{vi}^c$  表示类标记为  $c$  的像元  $v$  在第  $i$  个波段的取值。

考虑到光谱特征具有不同的分类能力,本文在定义光谱度量时对不同波段的光谱特征进行加权。对分类能力较强的波段赋予较大的权值,而对分类能力较弱的赋予较小的权值。对  $k$  个波段的光谱特征按照式 (1) 的相关性度量进行重要性排序后,其权值设定为:

$$\left\{ \left[ \frac{k}{2} \right], \left[ \frac{k}{2} \right] - 1, \dots, 0, \dots, \frac{1}{\left[ \frac{k}{2} \right] - 2}, \frac{1}{\left[ \frac{k}{2} \right] - 1} \right\} \quad (5)$$

当类标记取不同值时,未标记像元  $u$  的类  $y$  定义为:

$$y = \arg \min \text{SM}(c) \quad (6)$$

即选择光谱度量值最小的类作为待分类像元的类标记。

## 2.3 光谱-空间特征的结合

高光谱遥感图像不仅包含丰富的光谱信息,还包含丰富的空间信息。融合了光谱和空间特征的分类算法输入的分类信息更加丰富,因此可以获得更优的分类性能。纹理特征是高光谱遥感图像相邻像元或相邻区域内几何位置等相互关系的表征,是一种重要的空间特征。灰度共生矩阵通过计算图像各像元灰度级之间的联合条件概率表示纹理,可以提取多种纹理特征<sup>[17]</sup>。本文利用灰度共生矩阵提取了对比度、熵、二阶矩和反差分矩阵4种重要纹理特征,结合光谱特征和纹理特征,构建基于光谱-空间特征的高光谱遥感图像半监督分类算法。

对比度 (contrast) 反映了图像的清晰度和纹理的沟纹深浅,如式 (7) 所示。纹理越清晰、反差越大,对比度也就越大。

$$\text{Con} = \sum_i \sum_j (i-j)^2 P(i,j) \quad (7)$$

熵 (entropy) 度量了图像包含的随机性, 表现了图像的复杂程度, 如式 (8) 所示。当灰度共生矩阵所有值均相等或像素值表现出最大的随机性时, 熵最大。

$$\text{Ent} = - \sum_i \sum_j P(i,j) \log(P(i,j)) \quad (8)$$

二阶矩 (angular second moment, ASM) 又称能量, 是图像灰度分布均匀程度和纹理粗细的一个度量, 如式 (9) 所示。当图像纹理均一旦规则时, 能量值较大; 反之, 灰度共生矩阵的元素相近时, 能量值较小。

$$\text{ASM} = \sum_i \sum_j P(i,j)^2 \quad (9)$$

反差分矩阵 (inverse differential moment, IDM) 又称逆方差, 反映了纹理的清晰程度和规则程度, 如式 (10) 所示。纹理清晰、规律性较强时, 逆方差值较大。

$$\text{IDM} = \sum_i \sum_j \frac{P(i,j)}{1+(i+j)^2} \quad (10)$$

## 2.4 基于空间相关性的样本扩充

Tri-training 算法在扩充训练样本的过程中, 采用随机策略选择未标记样本, 并不区分未标记样本之间的差异。高光谱遥感图像数据是空间地理实体的描述数

据, 具有很强的空间相关性。因此, 本文基于高光谱遥感图像数据的空间相关性, 以相近相似原理为理论依据, 优先选择已标记样本的邻接像元进行标记扩充, 增强样本扩充过程中标记的准确性。图2展示了基于空间相关性的样本扩充过程, 可以看出, 标记像元  $n$  个邻域的全部未标记像元优先进入扩充样本集。

## 2.5 融合光谱度量标记迁移和 Tri-training 的半监督分类算法

当初始有标记样本数量较少时, Tri-training 算法会存在两个分类器出现相同错分结果的现象, 此时的待测样本会被标记为错误的类别, 并进入训练集参与分类器的训练, 从而影响分类效果。本文将光谱度量标记迁移和 Tri-training 算法相融合, 当两个分类器与光谱度量的预测类别一致时, 待测样本才会被标记为该类别, 并进入训练集, 直到训练集不再发生变化, 迭代过程停止。本文所提算法的伪代码如算法2所示。

**算法2** 融合光谱度量标记迁移和 Tri-training 的 SMT 算法

**Input:**  $L$ : 初始标记样本集,  $U$ : 无标记样本集, Learn: 机器学习算法, More: 空间邻域样本扩充方法, CFS: 基于 CFS 的特征排序算法,  $k$ : 光谱维数,  $M$ : 光谱特

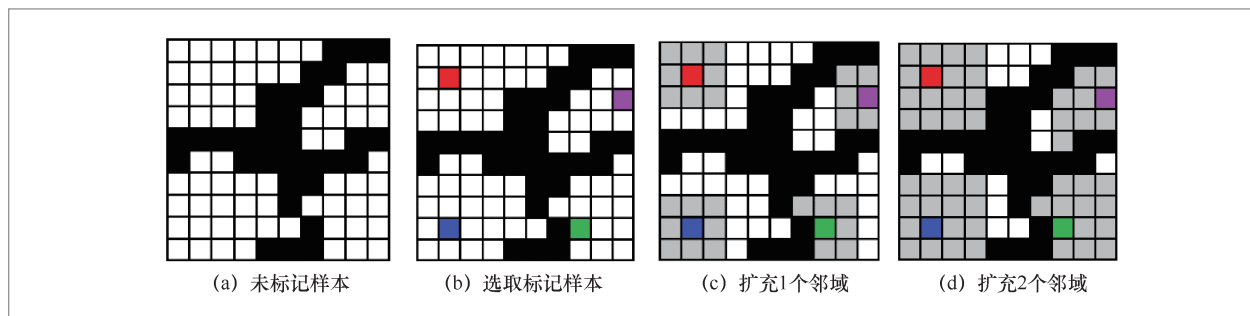


图2 基于空间相关性的样本扩充方法示意

征集,  $d_p$ : 光谱特征  $p \in (0, k)$ ,

SM: 光谱度量算法, GLCM: 灰度共生矩阵算法,  $M'$ : 排序后的光谱特征集,  $\sigma_p$ : 光谱权值,

Con、Ent、ASM、IDM: 纹理特征, TT: Tri-training算法各分类器分类结果, Vote: 绝对多数投票法

```

Output: C
1  $M' \leftarrow \text{CFS}(M)$ ,  $d_p \in M'(p \in \{0, \dots, k\})$ 
2 Con, Ent, ASM, IDM  $\leftarrow$  GLCM
( $d_p$ ) ( $p \in \{0, \dots, 4\}$ )
3  $M' \leftarrow M' \cup \text{Con} \cup \text{Ent} \cup \text{ASM} \cup \text{IDM}$ 
4 for  $p \in [0, k]$ 
5     if ( $p \in [0, \lfloor \frac{k}{2} \rfloor]$ ) then
 $\sigma_p \leftarrow \lfloor \frac{k}{2} \rfloor - p$ 
6     else  $\sigma_p \leftarrow \frac{1}{p - \lfloor \frac{k}{2} \rfloor}$ 
7     end if
8     end for
9 for  $i \in \{1, \dots, 3\}$  do
10      $S_i \leftarrow \text{BootstrapSample}(L)$ ;  $h_i \leftarrow \text{Learn}(S_i)$ 
11 end for
12  $U_L \leftarrow \text{More}(S_1, S_2, S_3)$ 
13 repeat until none of  $h_i (i \in \{1, \dots, 3\})$ 
changes
14 for  $x \in U_L$  do
15  $c_j, c_k \leftarrow \text{TT}(x)$ ;  $y \leftarrow \text{SM}(x)$ 
16  $c_i \leftarrow \text{Vote}(c_j, c_k, y)$ 
17  $L_i \leftarrow L_i \cup \{(x, c)\}$ 
18 end for
19 for  $i \in \{1, \dots, 3\}$  do
20  $h_i \leftarrow \text{Learn}(S_i \cup L_i)$ 
21 end for
22 end repeat
23 for every  $x \in U$  do

```

24  $c_i, c_j, c_k \leftarrow \text{TT}(x)$ ;  $y \leftarrow \text{SM}(x)$

25  $c \leftarrow \text{Vote}(c_i, c_j, c_k, y)$

26 **end for**

## 3 实验设计与结果分析

### 3.1 实验数据

Indian Pines数据集是由AVIRIS传感器拍摄的一幅美国西北部印第安纳州的高光谱遥感图像的一部分,如图3所示。其图像大小为 $145 \times 145$ (空间分辨率为200 m),图像共有16个地物类别、220个波段。其中,第104~108个、第150~163个和第220个波段不能被水反射,在经过修正后,去除了其中20个低信噪比的波段,最终共有200个波段参与实验。

Pavia University数据集是由ROSIS光学传感器拍摄的意大利帕威亚大学的高光谱遥感图像,如图4所示。图像大小为 $610 \times 340$ (空间分辨率为1.3 m),图像共有9个地物类别、115个波段。在经过修正后,去除了12个受噪声影响严重的波段,最终共有103个波段参与实验。

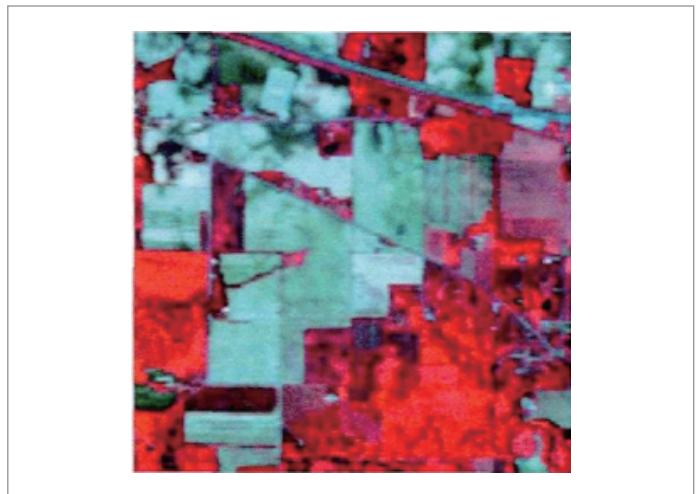


图3 Indian Pines 高光谱遥感图像

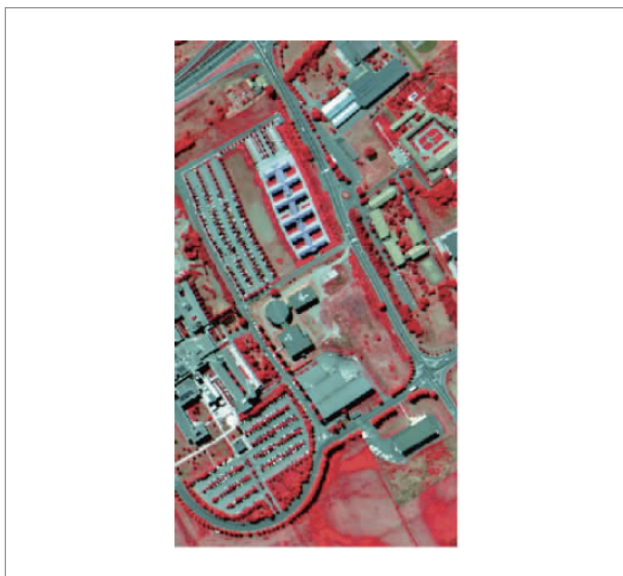


图4 Pavia University 高光谱遥感图像

### 3.2 实验设置

使用本文提出的融合光谱度量与Tri-training的高光谱遥感图像半监督分类算法对以上两个数据集进行分类。为了确保实验的准确性,对每个数据集分别进行10次独立重复实验,获得分类结果的总体分类精度(overall accuracy, OA)、平均分类精度(average accuracy, AA)和Kappa系数的平均值并将其作为算法性能的评价指标。

本文实验中SMT-RSK表示所提算法中Tri-training算法的基分类器为SVM(支持向量机)、RF(随机森林)和KNN(最近邻分类器)。SMT-RF、SMT-SVM和SMT-KNN表示SMT算法中, Tri-training算法分别使用的是3个RF、3个SVM和3个

KNN基分类器。T-RF、T-SVM和T-KNN分别表示使用3个RF、3个SVM和3个KNN基分类器的Tri-training算法。实验对比了SMT与Tri-training算法当基分类器不同时分类性能的差异。

### 3.3 实验结果和分析

#### 3.3.1 Indian Pines数据集的实验结果与分析

为了验证本文所提SMT算法的有效性,首先对不同集成方式下的多种半监督分类算法在Indian Pines数据集上的分类性能进行了对比。分类结果如图5所示,分类结果的OA、AA和Kappa系数见表1所示。由表1可以看出,融合光谱度量的SMT-RSK算法的OA值比未融合光谱度量的Tri-training算法T-RSK的OA值提高了8.98%。融合了光谱度量和单一基分类器Tri-training的SMT-RF、SMT-SVM和SMT-KNN算法的OA值比未融合光谱度量的T-RF、T-SVM和T-KNN算法分别提高了8.3%、9.8%、9.28%。SMT-RSK算法的AA值较T-RSK算法提高了8.97%,融合了光谱度量的SMT-RF、SMT-SVM和SMT-KNN算法的AA值比未融合光谱度量的T-RF、T-SVM和T-KNN算法分别提高了7.64%、9.33%、9.7%。对比Kappa系数, SMT-RSK比T-RSK提高了0.108, SMT-RF、SMT-SVM和SMT-KNN比T-RF、T-SVM和T-KNN算法分别提高了0.1389、0.1348、0.1231。由此可见,融合光谱度量和Tri-training算法的

表1 Indian Pines 数据集在不同集成方式下的多种半监督分类算法分类结果

评价指标	SMT-RSK	SMT-RF	SMT-SVM	SMT-KNN	T-RSK	T-RF	T-SVM	T-KNN
OA	90.81%	88.13%	85.24%	86.40%	81.83%	79.83%	75.44%	77.12%
AA	92.22%	89.58%	86.88%	88.19%	83.25%	81.94%	77.55%	78.49%
Kappa	0.8691	0.8277	0.7892	0.8036	0.7611	0.6888	0.6544	0.6805

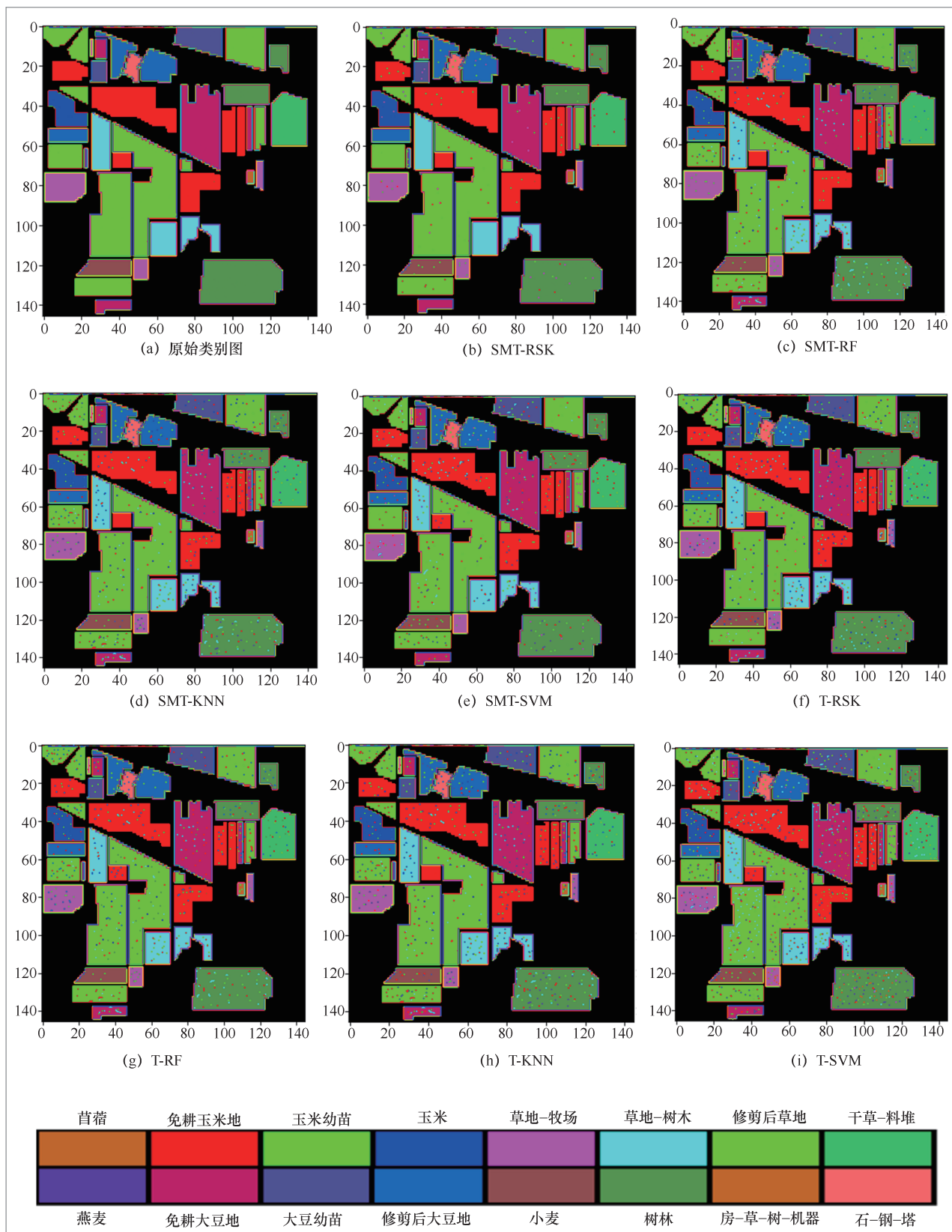


图 5 Indian Pines 数据集在不同集成方式下的多种半监督分类算法分类结果

半监督分类算法在Indian Pines数据集上取得了良好的分类性能。

此外,本文还将所提算法与其他高光谱图像半监督分类算法,如CMSI<sup>[18]</sup>、3CLL<sup>[19]</sup>、LGC<sup>[6]</sup>、SFCM<sup>[20]</sup>、MT<sup>[21]</sup>算法在Indian Pines数据集上的分类性能进行了对比,结果见表2,本文算法得到的分类结果在各方面大多优于CMSI、LGC、SFCM、MT算法。虽然在总体分类精度(OA)上略低于3CLL算法,但本文算法的各类别平均分类精度(AA)高于3CLL算法(约2%)。

### 3.3.2 Pavia University数据集的实验结果与分析

为了进一步验证本文所提算法的有效性,对不同集成方式下的多种半监督分类算法在Pavia University数据集上的

分类性能进行了同样的对比实验。分类结果如图6所示,分类结果的OA、AA和Kappa系数见表3。从表3可以看出,融合光谱度量后,SMT-RSK的OA值、AA值和Kappa系数比未融合光谱度量的Tri-training算法T-RSK分别提高了8.88%、8.65%和0.1059。同样,融合了光谱度量的SMT-RF、SMT-SVM和SMT-KNN算法比T-RF、T-SVM和T-KNN算法的OA值、AA值和Kappa系数均有不同程度的提高。由此可知,融合光谱度量和Tri-training的半监督分类算法在Pavia University数据集上同样取得了良好的分类性能。

本文还将所提算法与其他高光谱图像半监督分类算法CMSI、FWA<sup>[22]</sup>、SDME<sup>[23]</sup>、SFCM、MT在Pavia University数据集上的分类性能进行了比较,对比结果见表4。可以看出,本文算法得到的分类结果在3项评价指标上均优于对比算法。

表2 Indian Pines 数据集在不同半监督分类算法下的分类结果对比

评价指标	SMT-RSK	3CLL	CMSI	SFCM	MT	LGC
OA	90.81%	95.64%	85.71%	88.00%	85.27%	87.32%
AA	92.22%	90.17%	88.88%	90.14%	87.41%	—
Kappa	0.8691	—	0.8318	0.883	0.8255	0.8410

表3 Pavia University 数据集在不同集成方式下的多种半监督分类算法分类结果对比

评价指标	SMT-RSK	SMT-RF	SMT-SVM	SMT-KNN	T-RSK	T-RF	T-SVM	T-KNN
OA	92.47%	89.88%	87.55%	88.31%	83.59%	81.93%	76.81%	79.56%
AA	93.92%	91.14%	88.96%	90.23%	85.27%	83.11%	77.99%	81.94%
Kappa	0.8874	0.8362	0.8042	0.8186	0.7815	0.6954	0.6713	0.7001

表4 Pavia University 数据集在不同半监督分类算法下的分类结果对比

评价指标	SMT-RSK	CMSI	FWA	SDME	SFCM	MT
OA	92.47%	87.68%	87.44%	89.4%	91.35%	66.89%
AA	93.92%	92.14%	—	86.6%	—	73.76%
Kappa	0.8874	0.8283	0.83	—	0.8974	0.5802

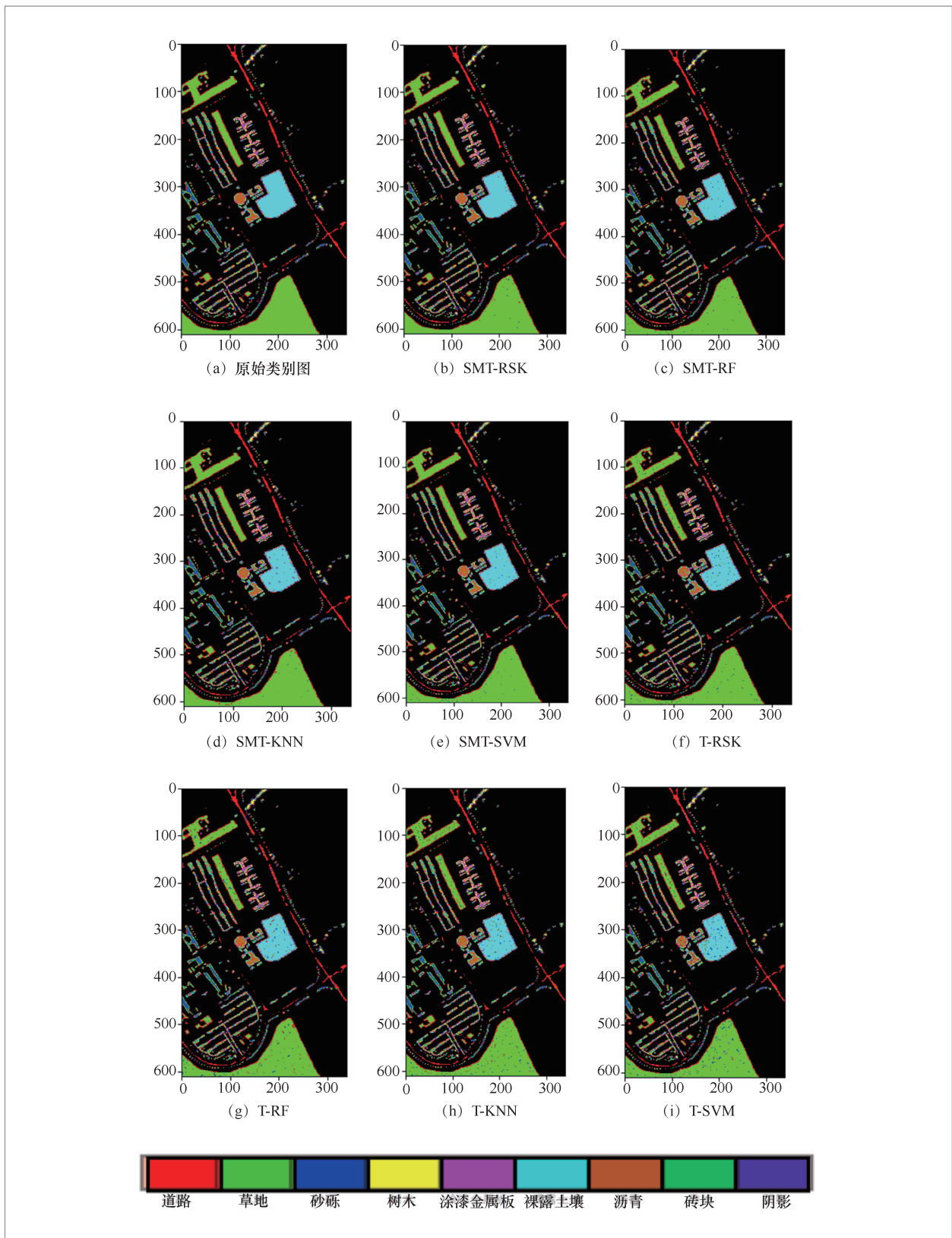


图 6 Pavia University 数据集在不同集成方式下的多种半监督分类算法分类结果

图7对表1和表3中不同集成方式下的多种半监督分类算法在Indian Pines和Pavia University两个数据集上的分类

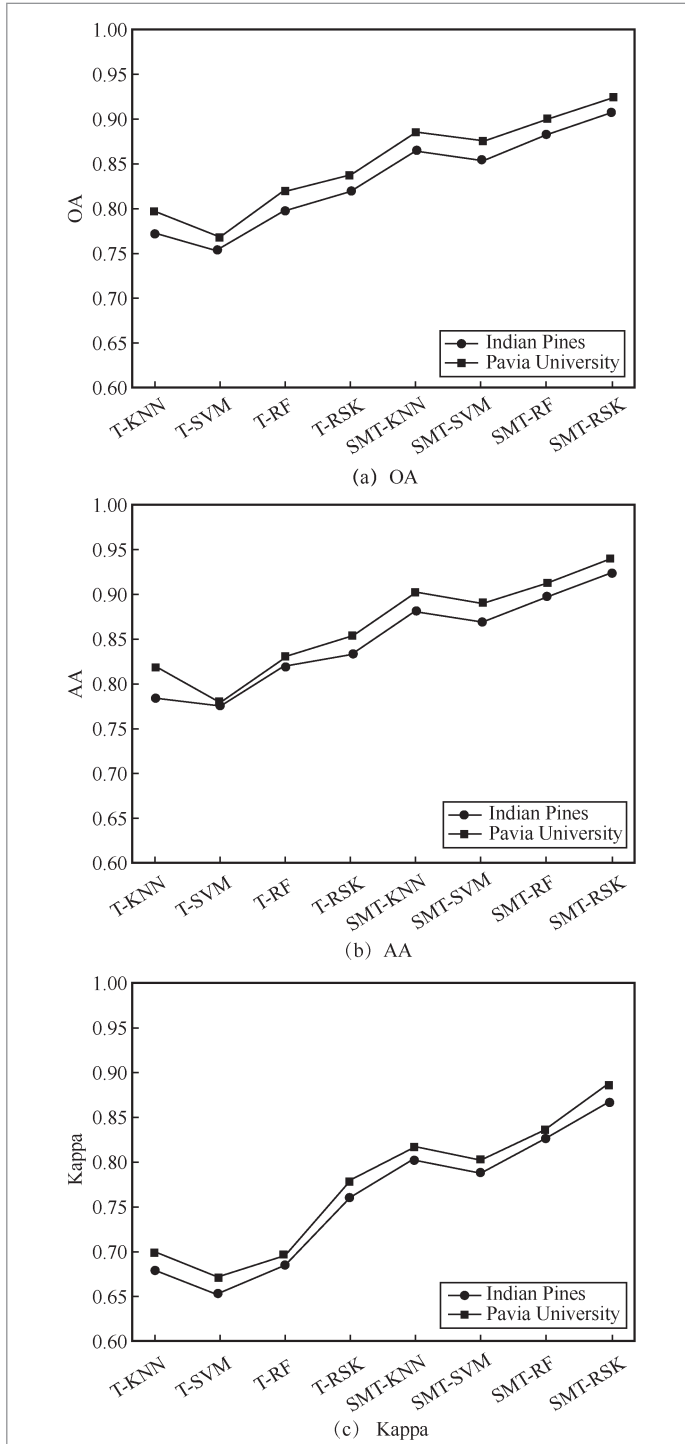


图7 Indian Pines 和 Pavia University 数据集在不同集成方式下的半监督分类结果的总体精度 (OA)、平均精度 (AA) 和 Kappa 系数对比

结果的OA、AA和Kappa系数进行了直观的对比,可以看出融合了光谱度量标记迁移和Tri-training的分类算法比Tri-training算法的分类性能有比较显著的提升。当所提算法中Tri-training的基分类器不同时,分类效果最优。

### 3.3.3 消融实验

为了研究本文算法中光谱度量标记迁移和Tri-training算法在融合过程中发挥的作用,本文在Indian Pines和Pavia University两个数据集上进行了消融实验,实验结果见表5。由表5可以看出,在两个数据集上,使用基分类器不同的Tri-training进行分类时,分类结果的OA、AA和Kappa系数均高于使用光谱度量标记迁移进行分类时相对应的评价指标的值,表明Tri-training算法的分类性能优于光谱度量标记迁移。而当光谱度量标记迁移和Tri-training进行融合时,分类结果的OA、AA和Kappa系数得到了较大的提升,由此可见二者的融合可以明显改善分类的性能。

### 3.3.4 重要参数对算法性能的影响分析

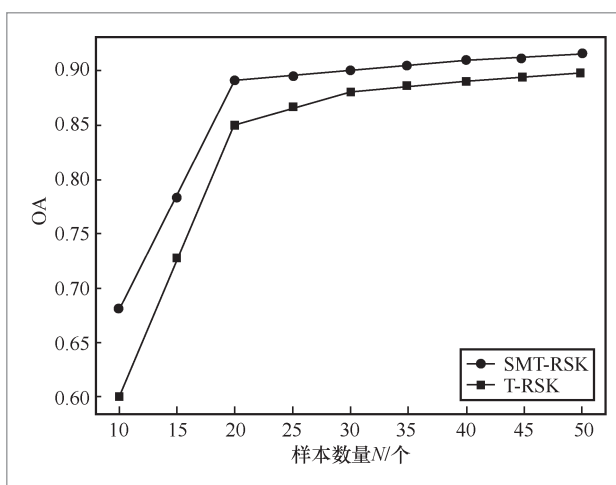
初始样本数量 $N$ 和扩充邻域数 $n$ 是Tri-training算法的两个重要参数,它们的取值将直接影响算法的性能。下面以Indian Pines数据集为例,通过观察样本数量和扩充邻域数的取值对总体分类精度OA的影响,探讨如何获取最优参数值。图8和图9分别描述了初始样本数量 $N$ 和扩充邻域数 $n$ 与OA的关系。通过图8可以看出,随着初始样本数量的增长,SMT-RSK和T-RSK算法的OA值都在不断提升。当初始样本数量从10增长至20时,两种算法的OA值分别提升至0.89和0.85,提升效果显著。之后OA值随着初始样本数量的增加呈缓慢上升趋势。因此,

表5 消融实验对比结果

数据集	SM			T-RSK			SMT-RSK		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
India Pines	71.42%	75.33%	0.6919	81.83%	83.25%	0.7611	90.81%	92.22%	0.8691
Pavia University	75.67%	78.41%	0.7178	83.59%	85.27%	0.7815	92.47%	93.92%	0.8874

初始样本数量的取值设定为20,即按比例提取各类初始样本最多不超过20个,此时总体分类精度较高且所需的有标记样本数量较少。

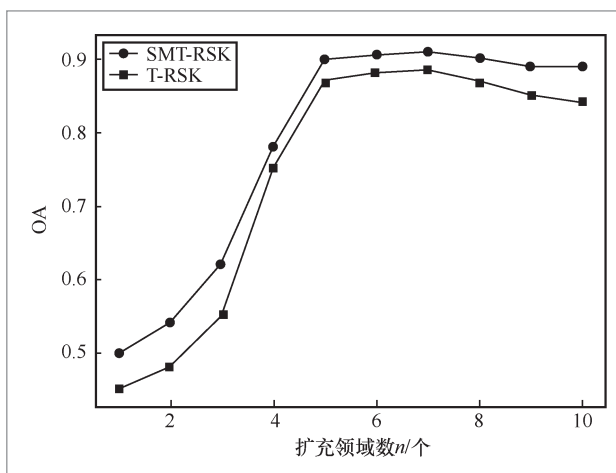
由图9可以看出,从有标记样本向外扩充5~7个邻域时,总体分类精度OA提升至0.9~0.91,提升效果显著。而当扩充8个以上邻域时,OA值有所下降。因此,在同时满足OA值较高且所需标记样本数量较少的条件下,本文设定 $n$ 的取值为5。

图8 初始样本数量  $N$  与 OA 关系

## 4 结束语

本文主要研究基于光谱-空间特征的高光谱遥感图像的半监督分类,提出了一种融合光谱度量标记迁移和Tri-training算法的半监督分类算法。该算法综合运用了高光谱遥感图像的光谱特征和空间特征,以增强分类过程中样本的区分度。在结合Tri-training算法进行有标记样本扩充的过程中,定义了一种新的光谱度量标记迁移方法,并基于该方法增强扩充样本标记的准确性,减少错误标记样本的引入。实验结果证明,本文算法具有良好的分类性能。

形态学特征也是高光谱遥感图像包含的重要空间特征。形态学特征有助于选择更具有光谱代表性及空间合理性的初始有标记像元。当高光谱图像分类中包含的地物类别较多时,可以借助形态学特征将全局

图9 扩充邻域数  $n$  与 OA 关系

的多分类问题转化为局部的二分类问题,更易于提高图像分类的精度。因此,未来笔者将对融合光谱度量标记迁移和Tri-training的高光谱遥感图像分类算法进行改进,通过融合形态学特征进一步提升图像分类的精度。

## 参考文献:

- [1] 骆剑承, 胡晓东, 吴田军, 等. 高分遥感驱动的精准土地利用与土地覆盖变化信息智能计算模型与方法研究[J]. 遥感学报, 2021, 25(7): 1351-1373.  
LUO J C, HU X D, WU T J, et al. Research on intelligent calculation model and method of precision land use/cover change information driven by high-resolution remote sensing[J]. Journal of Remote Sensing, 2021, 25(7): 1351-1373.
- [2] 王增茂, 杜博, 张良培, 等. 基于纹理特征和形态学特征融合的高光谱影像分类法[J]. 光子学报, 2014, 43(8).  
WANG Z M, DU B, ZHANG L P, et al. Based on texture feature and extend morphological profile fusion for hyperspectral image classification[J]. Acta Photonica Sinica, 2014, 43(8).
- [3] 王杰, 张松岩, 梁吉业. 融合一致性正则与流形正则的半监督深度学习算法[J]. 大数据, 2022, 8(3): 103-114.  
WANG J, ZHANG S Y, LIANG J Y. A semi-supervised deep learning algorithm combining consistency regularization and manifold regularization[J]. Big Data Research, 2022, 8(3): 103-114.
- [4] SHAHSHAHANI B M, LANDGREBE D A. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon[J]. IEEE Transactions on Geoscience and Remote Sensing, 1994, 32(5): 1087-1095.
- [5] 焦洪赞, 王少宇, 彭正洪. 基于条件随机场的光谱相似性匹配高光谱遥感影像聚类方法[J]. 武汉大学学报(工学版), 2016, 49(6): 937-943, 948.  
JIAO H Z, WANG S Y, PENG Z H. A spectral similarity matching classifier based on conditional random field for hyperspectral remote sensing imagery[J]. Engineering Journal of Wuhan University, 2016, 49(6): 937-943, 948.
- [6] 程志会, 谢福鼎. 基于空间特征与纹理信息的高光谱图像半监督分类[J]. 测绘通报, 2016(12): 56-59, 73.  
CHENG Z H, XIE F D. Semi-supervised classification for hyperspectral image based on spatial features and texture information[J]. Bulletin of Surveying and Mapping, 2016(12): 56-59, 73.
- [7] 何浩, 沈永林, 刘修国, 等. 空间-光谱约束的图半监督高光谱影像分类算法[J]. 国土资源遥感, 2016, 28(3): 31-36.  
HE H, SHEN Y L, LIU X G, et al. Spatial-spectral constrained graph-based semi-supervised classification for hyperspectral image[J]. Remote Sensing for Natural Resources, 2016, 28(3): 31-36.
- [8] 王俊淑, 江南, 张国明, 等. 高光谱遥感图像DE-self-training半监督分类算法[J]. 农业机械学报, 2015, 46(5): 239-244.  
WANG J S, JIANG N, ZHANG G M, et al. Semi-supervised classification algorithm for hyperspectral remote sensing image based on DE-self-training[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(5): 239-244.
- [9] 王春阳. 基于信息熵的自训练半监督高光谱遥感影像分类研究[D]. 焦作: 河南理工大学, 2015.  
WANG C Y. Research on self-training semi-supervised hyperspectral remote sensing image classification based on information entropy[D]. Jiaozuo: Henan Polytechnic University, 2015.
- [10] ROMASZEWSKI M, GLOMB P, CHLOEW M. Semi-supervised hyperspectral classification from a small number of training samples using a co-training approach[J]. ISPRS: Journal of Photogrammetry & Remote Sensing, 2016, 121: 60-76.

- [11] 王立国, 杨月霜, 刘丹凤. 基于改进三重训练算法的高光谱图像半监督分类[J]. 哈尔滨工程大学学报, 2016, 37(6): 849–854.  
WANG L G, YANG Y S, LIU D F. Semi-supervised classification for hyperspectral image based on improved tri-training method[J]. Journal of Harbin Engineering University, 2016, 37(6): 849–854.
- [12] 宋国娇. 高光谱图像降维及半监督分类算法研究[D]. 哈尔滨: 哈尔滨工程大学, 2017.  
SONG G J. Research on dimensionality reduction and semi-supervised classification algorithm of hyperspectral images[D]. Harbin: Harbin Engineering University, 2017.
- [13] ZHOU Z H, LI M. Tri-training: exploiting unlabeled data using three classifiers[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(11): 1529–1541.
- [14] GOLDMAN S, YAN Z. Enhancing supervised learning with unlabeled data[D]. Washington: Washington University, 2001.
- [15] 胡梦林. 基于混沌杜鹃搜索算法的高光谱影像波段选择和半监督分类[D]. 武汉: 武汉大学, 2018.  
HU M L. Band selection and semi-supervised classification of hyperspectral images based on chaotic cuckoo search algorithm[D]. Wuhan: Wuhan University, 2018.
- [16] HALL M A. Feature selection for discrete and numeric class machine learning[C]// Proceedings of the 17th International Conference on Machine Learning. [S.l.]: Morgan Kaufmann Publishers Inc, 2000: 584–596.
- [17] HARALICK R M. Statistical and structural approaches to texture[J]. Proceedings of the IEEE, 1979, 67(5): 786–804.
- [18] 王立国, 马骏宇, 李阳. 联合多种空间信息的高光谱半监督分类方法[J]. 哈尔滨工程大学学报, 2021, 42(2): 280–285.  
WANG L G, MA J Y, LI Y. Hyperspectral semi-supervised classification algorithm considering multiple spatial information[J]. Journal of Harbin Engineering University, 2021, 42(2): 280–285.
- [19] 关世豪, 杨桃, 卢珊, 等. 基于改进阶梯网络的高光谱半监督分类算法[J]. 激光与光电子学进展, 2021, 58(24): 480–490.  
GUAN S H, YANG G, LU S, et al. Hyperspectral semi-supervised classification algorithm based on improved ladder network[J]. Laser & Optoelectronics Progress, 2021, 58(24): 480–490.
- [20] 李彩虹, 赵祎霏. 一种高光谱图像的半监督分类方法[J]. 测绘通报, 2018(2): 46–49.  
LI C H, ZHAO Y F. A semi-supervised classification method for hyperspectral images[J]. Bulletin of Surveying and Mapping, 2018(2): 46–49.
- [21] 崔颖, 王雪婷, 陆忠军, 等. 改进M-training算法的高光谱图像分类[J]. 哈尔滨工程大学学报, 2018, 39(10): 1688–1694.  
CUI Y, WANG X T, LU Z J, et al. Hyperspectral image classification based on improved M-training algorithm[J]. Journal of Harbin Engineering University, 2018, 39(10): 1688–1694.
- [22] 宋国娇. 高光谱图像降维及半监督分类算法研究[D]. 哈尔滨: 哈尔滨工程大学, 2017.  
SONG G J. Research on dimensionality reduction and semi-supervised classification algorithm of hyperspectral images[D]. Harbin: Harbin Engineering University, 2017.
- [23] 罗甫林. 高光谱图像稀疏流形学习方法研究[J]. 测绘学报, 2017, 46(3): 400.  
LUO F L. Sparse manifold learning for hyperspectral imagery[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(3): 400.

## 作者简介



曹峰(1980- ),男,博士,山西大学计算机与信息技术学院(大数据学院)副教授、大数据系主任,主要研究方向为人工智能与数据挖掘。



李文涛(1994- ),男,山西大学计算机与信息技术学院(大数据学院)硕士生,主要研究方向为数据挖掘。



骆剑承(1970- ),男,博士,中国科学院空天信息创新研究院遥感科学国家重点实验室研究员,中国科学院大学教授,主要研究方向为遥感大数据智能计算。



李德玉(1965- ),男,山西大学计算机与信息技术学院(大数据学院)教授,山西大学计算机与信息技术学院党委书记,山西大学计算智能与中文信息处理教育部重点实验室副主任,主要研究方向为人工智能与数据挖掘。



钱宇华(1976- ),男,山西大学计算机与信息技术学院(大数据学院)教授,山西大学科学技术处处长,山西大学大数据科学与产业研究院负责人,山西大学计算智能与中文信息处理教育部重点实验室副主任,主要研究方向为人工智能与数据挖掘。



白鹤翔(1980- ),男,博士,山西大学计算机与信息技术学院(大数据学院)副教授,主要研究方向为人工智能与数据挖掘。



张超(1989- ),男,博士,山西大学计算机与信息技术学院(大数据学院)副教授,主要研究方向为数据挖掘与粒计算。

**收稿日期:** 2022-08-17

**通信作者:** 骆剑承, luojc@aircas.ac.cn

**基金项目:** 国家自然科学基金资助项目(No.62072291, No.42071316, No.62072294, No.61672332, No.41871286, No.61806116); 山西省重点研发计划项目(No.201903D421003, No.201903D421041); 山西省教育厅科技成果转化培育项目(No.2020CG001)

**Foundation Items:** The National Natural Science Foundation of China (No. 62072291, No.42071316, No.62072294, No.61672332, No.41871286, No.61806116), Key Research and Development Plan of Shanxi Province (No.201903D421003, No.201903D421041), Scientific and Technological Transformative Project of Shanxi Provincial Department of Education (No.2020CG001)