#### doi: 10.6046/gtzyyg.2016.03.06

引用格式: 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 im – age[J]. Remote Sensing for Land and Resources, 2016, 28(3): 31 – 36.)

## 空间 - 光谱约束的图半监督高光谱影像分类算法

何浩<sup>1,3</sup>,沈永林<sup>1</sup>,刘修国<sup>1</sup>,马丽<sup>2</sup>

 (1. 中国地质大学(武汉)信息工程学院,武汉 430074; 2. 中国地质大学(武汉)机械与电子信息学院, 武汉 430074; 3. 新疆大学建筑工程学院,乌鲁木齐 830047)

摘要:高光谱影像数据的类标签样本获取困难,而在少量标签点情况下的分类精度通常不理想。为此,提出了一种改进的空间 – 光谱约束的图半监督分类算法(spatial – spectral constrained graph – based semi – supervised classification, SS – GSSC)。首先,以欧氏距离结合 RBF(radial basis function)核函数确定空间相似性权值;采用光谱相关角 (spectral correlation angle, SCA)计算光谱相似性权值;然后,将2种权值以乘积的形式进行组合,对相似性测度进 行约束;最后,利用标签传递算法对测试数据进行标签预测,获得分类结果。通过分别对 Indian Pines 影像和 DC Sub 影像进行分类实验的结果表明,该算法较之以往的分类算法,能更好地消除同类地物图斑中夹杂异类地物散点 的现象,在少量标签点(每类 25 个)情况下,取得了较高的分类精度。

关键词:半监督分类;图;空间-光谱;高光谱遥感

中图法分类号: TP 79 文献标志码: A 文章编号: 1001-070X(2016)03-0031-06

## 0 引言

高光谱遥感影像分类一直是高光谱遥感领域较 为热门的研究方向,从传统的监督分类、非监督分 类,到近年来兴起的半监督分类,国内外研究人员已 发展了多种分类算法。然而,多数算法仅利用光谱 相似性测度进行分类,对地物的空间信息利用不足。 针对此问题,国内外学者进行了大量的相关研究,提 出了多种空间 - 光谱分类算法<sup>[1]</sup>。在引入空间信 息辅助分类时,通常利用影像的纹理、形状等特征作 为辅助信息。赵银娣等<sup>[2]</sup>提出了一种适用于多光 谱纹理影像分类的广义马尔可夫随机场模型,该算 法提高了纹理影像的分类精度;黄昕等[3]提出了一 种像元形状指数及基于形状与光谱特征融合的高空 间分辨率遥感影像分类算法,该算法计算简便且能 有效表达高分辨率影像的地物特征; Xia 等<sup>[4]</sup>利用 4种特征提取方法,结合旋转森林进行特征提取,再 采用空间上下文信息、先验马尔可夫场等方法进行 高光谱影像分类,显著提高了分类精度; Wang 等<sup>[5]</sup> 提出了一种新的基于空间 - 光谱标签传递分类算法 用于高光谱图像的半监督分类; Ji 等<sup>[6]</sup>采用构建超 图结构形式将像元光谱特征与空间约束相结合,进 行高光谱图像分类; Ghamisi 等<sup>[7]</sup>提出了一种基于 隐马尔可夫随机场分割和支持向量机(support vector machine,SVM)相结合的分类算法; Li 等<sup>[8]</sup>提出 了一种非局部联合协同表示分类算法与局部自适应 字典相结合的高光谱图像分类算法。这些算法在标 签数量较多的情况下具有很高的分类精度,但同时 也都具有计算较复杂的问题; 当标签数量少、无标 记样本数量庞大时,分类精度通常不理想。分类结 果也常伴随着同类地物图斑中夹杂其他类别地物散 点的现象。

本文在图半监督分类方法的基础上,顾及空间 信息约束,提出一种改进的空间 – 光谱约束的图半 监督分类算法(spatial – spectral constrained graph – based semi – supervised classification, SS – GSSC)。 该算法基于2点假设<sup>[6]</sup>:①具有相似光谱特征值的 像元点可能具有相同的类别标签;②空间邻近的像 元点可能具有相同的类别标签。算法采用乘积的方 式组合空间信息与光谱信息的权值,从而扩大像元 点之间的相似性差异。能够在少量标签情况下,取 得较高的分类精度。同时,也在一定程度上解决了 同类地物图斑中夹杂其他类别地物散点的问题。

收稿日期: 2015-03-23;修订日期: 2015-05-18

基金项目:中国博士后科学基金面上项目"基于分形的玉米作物物候动态检测算法研究"(编号:2013M542086)和中央高校新青年教师科研启动基金项目"基于结构推理的农作物多参量旱情监测方法研究"(编号:CUGL140834)共同资助。

1 空间-光谱约束的图半监督分类算法

传统的分类方法大体可分为监督分类和非监督 分类。监督分类常利用有类别标签的训练样本数据 训练获得一个分类器,对无标签的测试数据进行标 签预测;而非监督分类则是直接对没有类别标签的 数据,利用数据的聚类特性进行分类;近年来在模 式识别领域兴起的半监督分类算法,同时利用了少 量标签数据的类别信息和无标签数据中的隐含信 息,对小样本分类具有一定的优势。基于图的半监 督分类(graph - based semi - supervised classification, CSSC)<sup>[9]</sup>最大的特点就是用图来表示数据之间 的关系。

综合考虑高光谱影像中目标像元与邻近相似像 元在空间和光谱上的关系,将2种相似性测度组合 定权,改进图半监督分类算法。该算法包括:相似 性测度选择、组合和标签传递计算。

#### 1.1 空间相似性测度

若将高光谱影像某波段中像元点的行列号(R, C)看作是二维平面中的横纵坐标(x,y),则高光谱 影像中像元点之间的距离可由欧氏距离公式求得。 按照上文的假设,"相近即相似"的原则,结合 RBF (radial basis function)核函数,可得到像元i与像元j在空间上的相似性度量  $W_{ij}^{s}$ (边权值),即

$$W_{ij}^{\rm s} = e^{-d_{ij}^2/2\sigma^2}, \forall i \neq j \quad , \qquad (1)$$

式中:  $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ ; x 和 y 分别代 表像元点的行号和列号;  $\sigma$  为函数的宽度参数, 控 制函数的径向作用范围;  $W_{ij}^{s}$ 取值范围是(0,1), 值 越大空间 2 点越相似。

#### 1.2 光谱相似性测度

光谱相关角(spectral correlation angle, SCA)是 光谱向量间 Pearson 相关系数(R)的余弦角。它反 映了光谱相对均值的变化,能很好地消除负相关的 影响。取 SCA 的倒数作为光谱相似性权值  $W_{ii}^{\mathbb{F}}$ ,则

$$W_{ij}^{\rm F} = SCA^{-1} = \cos\left(\frac{R_{ij}+1}{2}\right)$$
 (2)

*W*<sup>*i*</sup><sub>*i*</sub>值越大表明2个光谱向量相似度越高。2个光谱向量的*R*介于-1和1之间,其计算公式为

$$R_{ij} = \frac{\left(n\sum_{k=1}^{n}r_{k}t_{k} - \sum_{k=1}^{n}r_{k}\sum_{k=1}^{n}t_{k}\right)}{\left[\sqrt{n\sum_{k=1}^{n}r_{k}^{2} - \left(\sum_{k=1}^{n}r_{k}\right)^{2}}\sqrt{n\sum_{k=1}^{n}t_{k}^{2} - \left(\sum_{k=1}^{n}t_{k}\right)^{2}}\right]},$$
(3)

式中: *n* 表示波段数; *k* 为波段; *r*,*t* 分别代表第*i* 和 第*i* 个像元点的光谱值。

#### 1.3 相似性测度的组合方法

空间相似性权值和光谱相似性权值从不同的角度描述了像元点之间标签相似程度。实验了多种权值组合方法,发现采用权值相乘的方法组合定权,分类效果最好。这从理论上也很容易得到解释,如果2个像元点之间的2种相似性权值都很小或都很大,以乘积的形式组合可使这种相似性差异进一步扩大;当2种相似性权值出现一大一小时,取乘积可使其相似性权值保持适中,从而有效地避免了光谱相似而空间距离较远的异类地物被误分为同一类地物的情况。则组合定权公式为

$$W_{ij} = W_{ij}^{\rm S} W_{ij}^{\rm F} \quad , \tag{4}$$

式中: $W_{ii}$ 为像元i与像元j之间的组合边权值;W为对称矩阵;为了避免像元自相似性,令 $W_{ii}=0$ 。

#### 1.4 标签传递算法

用图表示数据之间的关系,图的结点表示数据 点(包括标签数据和无标签数据),点与点之间的边 权值采用上述相似性测度组合方法计算获得。则预 测标签值可表示为

$$\boldsymbol{F}^* = \lim_{t \to \infty} F(t) = (1 - \alpha) (\boldsymbol{I} - \alpha \boldsymbol{S})^{-1} \boldsymbol{Y} , (5)$$

式中:参数 α 取值范围为(0,1); I 为单位矩阵;

$$S = D^{-1/2} W D^{-1/2} , \qquad (6)$$

其中 **D** 为对角矩阵,元素  $D_{ii}$ 为 **W** 的第 *i* 行之和;  $Y^* = (Y_l, Y_u)^T = \{y_1, y_2, \dots, y_l, y_{l+1}, \dots, y_{l+u}\}^T$ , (7) 其中  $Y_l$  为带标签的训练数据向量,  $Y_u$  为无标签的 测试数据向量,由向量  $Y^*$ 构造标签矩阵  $Y; F_{ij}^*$ 表示 数列 $\{F(t)\}$ 的极值,每一个点的预测标签采用数列 的极大值

$$y_i = \arg \max_{i \le C} (F_{ij}^*) \tag{8}$$

获得,式中C为类别数。

空间-光谱约束的图半监督分类步骤如下:

1)输入带标签的训练数据  $Y_i$  和无标签的待分 类数据  $Y_a$ ; 设置 RBF 核函数的带宽超参数  $\sigma$  和标 签传递调节参数  $\alpha$ ; 设置训练数据标签个数。

2) 计算边权值。首先,遍历所有像元利用公式 (1) 计算像元之间的  $W_{ij}^{s}$ ; 然后,利用公式(2) 计算 所有像元之间的  $W_{ij}^{s}$ ; 最后,由公式(3)将  $W_{ij}^{s}$ 和  $W_{ij}^{s}$ 组合构建像元点的边权值矩阵 **W**。

3)设置标签矩阵。步骤为由  $Y_i$  和  $Y_u$  (无标签 数据的标签值均为 0)的标签构成  $N \times C$  的标签矩 阵Y,其中N为像元总数。

4)标签传递算法。首先,构建由训练数据和测 试数据组成的全部数据的全连接图,边权值矩阵采 用上述已计算出的 W; 然后,利用公式(6)计算对 称矩阵 S; 最后,利用公式(5)计算预测标签矩阵 F<sup>\*</sup>,取极大值对应类别作为该像元的类别。

5)输出预测标签矩阵及分类结果图。

2 实验结果及分析

#### 2.1 实验数据

为了方便比较算法的分类效果,选取了2种不同传感器的高光谱遥感影像。

1) Indian Pines 数据集是 AVIRIS 传感器采集的 数据,采集于 1992 年美国印第安纳州的一个农场, 空间分辨率为 20 m,去除水汽吸收波段,剩余 200 个波段。农田区域具有比较规整的几何形状,作物 类别主要包括玉米和大豆,由于耕种情况和土壤湿 度的不同,可以细分为多种类别。根据 Landgrebe 的报告<sup>[10]</sup>,选择所有的16个类别进行实验。

2) Washington DC 数据集为机载传感器 HYDICE 采集,影像由 210 个波段组成,波长范围 400~2 400 nm,去除水汽吸收波段,剩余 191 个波 段。实验选择了 Washington DC 的一个子场景,命 名为 DC Sub。该影像为行号 655—762 和列号 67— 177 组成的 108 像元×111 像元的影像,包含草地、 树、屋顶、道路和阴影等 6 类地物。

实验中分别采用以光谱特征值作为相似性测度 的 GSSC 方法和 SS - GSSC 方法进行对比实验。设 置参数 σ = 10,α = 0.1,标签数量选择时,每一类地 物随机选择 3 ~ 25 个标签,例如标签数量选择为 {3,5,10,15,20,25},分别测试各数据集在不同标 签数量情况下的分类效果。为了避免随机选取标签 带来实验结果的偶然性,对每种参数情况进行 10 次 测试,取平均值进行分类精度评价。

2.2 Indian Pines 数据集的分类结果

当每类地物标签数量取 25 时,16 类地物的分 类精度及总体分类精度如表 1 所示。

表 1 Indian Pines 数据集训练与测试样本数及各类别分类精度

Tab.1 Number of train data and test data and classification accuracy for each class of Indian Pines data set

序号	类别名称	训练样本数	测试样本数	GSSC/%	SS - GSSC/%
1	紫花苜蓿	25	21	100.00	100.00
2	免耕玉米地	25	1 403	31.60	87.53
3	玉米幼苗	25	805	52.75	91.46
4	玉米	25	212	76.96	100.00
5	草地、牧场	25	458	79.21	91.90
6	草地、树林	25	705	80.27	95.89
7	修剪过的草地	25	3	100.00	100.00
8	干草、料堆	25	453	82.48	100.00
9	燕麦	15	5	100.00	100.00
10	免耕大豆地	25	947	65.84	90.67
11	大豆幼苗	25	2 430	50.67	90.38
12	整理过的大豆地	25	568	51.98	92.55
13	小麦	25	180	96.01	99.80
14	木材	25	1 240	78.05	99.70
15	大厦 – 草 – 树 – 机器	25	361	42.22	100.00
16	石钢塔	25	68	99.24	100.00
	总体精度			56.20	92.09

从表1可以看出,利用基于光谱特征的GSSC 方法,分类精度较低,总体精度仅为56.20%;加入 空间信息后,采用SS-GSSC算法总体分类精度提 高到了92.09%。由此可见,空间信息的参与大大 提高了影像的分类效果。此外,本文算法SS-GSSC 相对于传统算法GSSC能更好地区分玉米、干草、大 豆、木材和大厦等地物。

#### 2.3 DC sub 数据集的分类结果

在 DC sub 数据集中6 类地物的分类精度如表2 所示。

表 2 DC sub 数据集训练与测试样本数及各类别分类精度 Tab. 2 Number of training data and test data and

classification accuracy for each class of DC sub data set

序号	类别名称	训练 样本数	测试 样本数	GSSC/%	SS – GSSC/ %
1	草地	25	395	99.62	100.00
2	公路	25	256	97.54	100.00
3	屋顶	25	762	92.76	99.20
4	阴影	25	188	95.73	99.56
5	小路	25	177	97.08	100.00
6	树	25	102	97.79	100.00
总	体精度			95.57	99.62



图 1 DC Sub 数据集分类结果 Fig. 1 Classification results of DC Sub

为了便于定量评价本文方法的分类精度,分类 对象全部选择有地面真实数据的区域,如图1(b)所 示。标签数量取25时,仅利用光谱特征的GSSC算 法分类结果,在大块图斑中有较明显的异类散点出 现,如图1(c)所示;采用SS-GSSC算法分类,分类 精度由95.57%提升到了99.62%,且消除了同类图 斑中夹杂异类散点的现象,只在类别边界区域出现 少量误分类点,如图1(d)所示。

进一步验证分类效果,将分类对象扩大到全图 所有像元,标签数量仍取每类 25 个,采用 SS - GSSC 算法分类,总体分类效果良好,分类结果如图1(e) 所示。影像右侧的小路、草地、树和建筑物等都得到 了很好的区分,同类别内异类散点较少。但受标签 点位置的影响,在标签点较少的区域,出现了少量误 分类点。例如影像左侧的部分草地、树被误分成了 道路。

#### 2.4 标签数量对分类精度影响分析

分别选取标签数量为3,5,10,15,20和25个, 采用SS-GSSC算法对数据集 Indian Pines 进行分 类,其分类结果如图2所示。



Fig. 2 - 1 Classification results of Indian Pines in different label numbers





从图 2 中可以发现,每类标签数量取 3 时,在地 类复杂区域有明显的误分类现象,随着标签数量的 增加,分类精度逐渐提高。当标签数量增加到 15 时,分类精度的提升较为缓慢。但整体上,采用本文 算法都能不同程度地消除同类图斑中部夹杂异类散 点的现象,误分类点只在区域边缘出现。

为了进一步比较本文算法的分类效果,在实验 条件相同的情况下与文献[5]中提出的(spatial – spectral label propagation based on the SVM, SS – LPS-VM)算法比较, SS – GSSC 算法在标签数量很少时, 例如标签数量取3和5时,已具有较高的分类精度; 随着标签数量的增加,该算法分类精度仍略高于 SS – LPSVM 算法。分类精度比较如图3所示。



Indian Pines data set

## 3 结论

本文提出的空间 – 光谱约束的图半监督高光谱 影像分类算法克服了以往基于像元光谱特征分类算 法的局限性,增加了相邻像元之间的空间信息,提高 了影像的分类精度,减少了同一类地物图斑中夹杂 异类地物散点的情况,对高光谱影像分类具有一定 的普遍适用性。

分别对 Indian Pines 影像和 DC Sub 影像进行的

分类实验表明:与传统的基于光谱特征分类算法相 比,本文算法有较明显的精度提升;在 Indian Pines 影像上进一步做标签数量对分类精度的影响分析, 发现即使在标签数据较少的情况下,该算法仍然具 有较高的分类精度;同等实验条件下,与文献[5]中 提出的 SS – LPSVM 算法比较,本文算法分类精度也 略有提高。

但在标签数据分布非常不均匀的情况下,本文 算法容易产生明显的误分类现象。这是该算法的不 足之处,也将是下一步研究的重点。

#### 参考文献(References):

- [1] Fauvel M, Tarabalka Y, Benediktsson J A, et al. Advances in spectral – spatial classification of hyperspectral images [J]. Proceedings of the IEEE, 2013, 101 (3):652 - 675.
- [2] 赵银娣,张良培,李平湘.广义马尔可夫随机场及其在多光谱 纹理影像分类中的应用[J].遥感学报,2006,10(1):123-129.

Zhao Y D,Zhang L P,Li P X. Universal Markov random fields and its application in multispectral textured image classification [J]. Journal of Remote Sensing,2006,10(1):123 - 129.

- [3] 黄 昕,张良培,李平湘.融合形状和光谱的高空间分辨率遥感 影像分类[J].遥感学报,2007,11(2):193-200.
  Huang X,Zhang L P,Li P X. Classification of high spatial resolution remotely sensed imagery based on the fusion of spectral and shape features[J]. Journal of Remote Sensing,2007,11(2):193-200
- [4] Xia J S, Chanussot J, Du P J, et al. Spectral spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(5):2532 – 2546.
- [5] Wang L G, Hao S Y, Wang Q M, et al. Semi supervised classification for hyperspectral imagery based on spatial – spectral label propagation [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 97:123 – 137.
- [6] Ji R R, Gao Y, Hong R C, et al. Spectral spatial constraint hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(3):1811 – 1824.

- [7] Ghamisi P, Benediktsson J A, Ulfarsson M O. Spectral spatial classification of hyperspectral images based on hidden Markov random fields [J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(5): 2565 – 2574.
- [8] Li J Y, Zhang H Y, Huang Y C, et al. Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary[J]. IEEE Transactions on Geoscience and Re-

mote Sensing, 2014, 52(6): 3707 - 3719.

- [9] Zhu X J. Semi Supervised Learning with Graphs [D]. Pittsburgh, Pennsylvania State: Carnegie Mellon University, 2005.
- [10] Landgrebe D. Multispectral Data Analysis: A Signal Theory Perspective [R]. West Lafayette: School of Electrical and Computer Engineering, Purdue University, 1998.

# Spatial – spectral constrained graph – based semi – supervised classification for hyperspectral image

HE Hao<sup>1,3</sup>, SHEN Yonglin<sup>1</sup>, LIU Xiuguo<sup>1</sup>, MA Li<sup>2</sup>

(1. Faculty of Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China; 2. Faculty of

Mechanical and Electronic Information, China University of Geosciences(Wuhan), Wuhan 430074, China;

3. Faculty of Architecture Engineering, Xinjiang University, Urumqi 830047, China)

**Abstract**: It is difficult to obtain labels of samples for hyperspectral data. Few labeled samples usually lead to low classification accuracy. In view of this situation, an improved spatial and spectral constraint graph – based semi – supervised classification algorithm (SS – GSSC) is proposed. First of all, Euclidean distance combined with radial basis function (RBF) is used to construct the spatial similarity edge weight; Spectral correlation angle (SCA) is used to calculate spectral similarity weights; Then, the two kinds of weights are combined to the form of product to restrict the similarity measurement; Finally, the label propagation algorithm is used to predict the test data labels so as to obtain the classification results. Classification experiments on Indian Pines image and DC Sub image show that, compared with the previous classification algorithm, the algorithm designed by the authors can better eliminate the phenomenon of the existence of the same category map spot included in other categories of scattered points , and can achieve higher classification accuracy under the condition of less label points (25 per class).

Keywords: semi – supervised classification; graph; spatial – spectral; hyperspectral remote sensing

**第一作者简介:**何浩(1979-),男,博士研究生,讲师,主要从事高光谱遥感数据处理方面的研究。Email: hehao\_1216@126. com。

(责任编辑:陈理)