

# 基于 AP 聚类的高光谱波段选择

任智伟\*, 吴玲达

航天工程大学航天信息学院, 北京 101416

**摘要** 波段选择在降维的同时能够保留高光谱数据的物理意义, 在很多方面具有一定应用。近邻传播(AP)算法根据数据点之间的相关性进行聚类, 将所有数据点视为潜在的聚类中心。提出了一种基于 AP 聚类的波段选择方法, 利用光谱信息散度和光谱相关角(SID-SCA)与光谱信息散度和光谱梯度角(SID-SGA)改进 AP 算法中相似度的计算。将降维结果输入支持向量机(SVM)分类器进行分类, 计算分类准确性, 并通过数据集 Indiana Pines 进行验证。实验结果表明: 所提方法能够更好地提取波段的信息, 得到更高的分类精度。

**关键词** 光谱学; 波段选择; 近邻传播聚类; 高光谱图像

中图分类号 O433.4

文献标识码 A

doi: 10.3788/LOP55.103002

## Hyperspectral Band Selection Based on Affinity Propagation Clustering

Ren Zhiwei\*, Wu Lingda

School of Space Information, Space Engineering University, Beijing 101416, China

**Abstract** Band selection can preserve the physical meaning of hyperspectral data while reducing dimension, and has application in many aspects. The cluster of affinity propagation (AP) algorithm is according to the correlation of data points, and the AP algorithm regards all data points as potential clustering centers. We propose a band selection method based on AP clustering, which uses spectral information divergence and spectral correlation angle (SID-SCA), and spectral information divergence and spectral gradient angle (SID-SGA) to improve the similarity calculation in AP algorithm. The reducing dimension results are input into the support vector machine (SVM) classifier to classify, and the classification accuracy is calculated and verified using the data set Indiana Pines. The experimental results reveal that the proposed method can better extract the information of the band and obtain a high classification accuracy.

**Key words** spectroscopy; band selection; AP clustering; hyperspectral image

**OCIS codes** 300.6320; 100.4145

## 1 引言

高光谱图像能够描述地物的空间形态与分布, 具有光谱分辨率高、光谱连续以及图谱合一等特点。但高光谱图像的光谱分辨率过高, 从而导致数据量大、各波段间相关性高、信息冗余现象严重等问题。因此, 高光谱图像处理的难度较彩色图像及多光谱图像大大提高, “Hughes”现象<sup>[1]</sup>明显。为提高处理效率, 并避免可能出现的维数灾难问题, 在对高光谱图像进行分类分析之前, 通常需要对其进行降维处理。

目前, 用于高光谱图像降维处理的方法主要有

两种: 光谱特征提取以及光谱特征选择<sup>[2]</sup>。光谱特征选择也被称为波段选择, 即针对特定对象选取光谱特征空间中的一个子集。基于聚类的波段选择是实现波段选择的方法之一, 也是实现波段选择的可行思路。但传统的聚类方法需要人为确定类数, 聚类结果可能受随机产生的初始中心影响而不稳定。此外, 通过传统聚类方法得到的结果不是真实的波段, 因此最终的波段选择结果与聚类之间存在很大误差。2007年, Frey等<sup>[3]</sup>提出了一种近邻传播(AP)聚类算法, 该方法解决了以上问题。在处理大规模、多类数据时, AP算法的处理效果较好。目前, AP算法已经成功

收稿日期: 2018-03-03; 修回日期: 2018-04-23; 录用日期: 2018-04-28

基金项目: 国家重点实验室基础研究项目

\* E-mail: juimer@foxmail.com

应用于图像分割、目标识别等领域<sup>[4]</sup>。

本文提出了一种改进的 AP 聚类波段选择方法,该方法属于无监督的波段选择。在 AP 聚类的基础上,利用光谱信息散度和光谱梯度角(SID-SGA)以及光谱信息散度和光谱相关角(SID-SCA)的计算方法对聚类中相似矩阵  $\mathbf{S}$  的计算进行改进,将聚类结果输入支持向量机(SVM)分类器中进行地物分类。将本文所提方法的分类结果与最大信息(MI)法、自适应波段选择(ABS)法、最大方差主成分分析法(MVPCA)等波段选择方法的分类结果进行比较,证明了本文所提方法的有效性。

## 2 近邻传播算法

假设数据集中包含  $n$  个样本  $\{x_1, x_2, \dots, x_n\}$ ,

通过每两个样本之间的相似度计算吸引度和归属感,结合吸引度和归属感两方面信息找到最优的类代表点集合,最终使所有数据点到其最近的类代表点的相似度之和最大。

$n \times n$  维的相似度矩阵  $\mathbf{S}$  为工作的基础,它的每个元素均为负值,可根据不同任务选择相应指标作为测度。当  $i=j$  时,  $s(i, j)$  代表偏向参数  $p(i)$ 。  $p(i)$  越大,点  $x_i$  作为聚类中心的可能性就越大。在无监督条件下,  $p$  通常取相似矩阵的中值。  $r(i, j)$  表示数据点  $j$  对点  $i$  的吸引度,描述点  $j$  适合作为点  $i$  的类代表的程度。  $a(i, j)$  表示数据点  $i$  对点  $j$  的归属感,表示数据  $i$  选择点  $j$  作为类代表的程度。  $r(i, j)$  和  $a(i, j)$  越大,表明点  $j$  作为最终聚类中心的可能性越大。吸引度和归属度的计算公式如下

$$r(i, j) = \begin{cases} s(i, j) - \max_{j' \neq j} [a(i, j') + s(i, j')], & i \neq j \\ p(j) - \max_{j' \neq j} [a(j, j') + s(j, j')], & i = j \end{cases}, \quad (1)$$

$$a(i, j) = \begin{cases} \min\{0, r(j, j) + \sum \max[0, r(i', j)]\}, & i \neq j \\ \sum_{i' \neq j} \max[0, r(i', j)], & i = j \end{cases}, \quad (2)$$

在更新  $r(i, j)$  和  $a(i, j)$  时,为避免聚类结果发生振荡,通常引入阻尼系数  $\lambda \in [0, 1)$  对  $r(i, j)$  和  $a(i, j)$  进行缩放。缩放公式如下

$$r(i, j)_{\text{new}} = (1 - \lambda) \cdot r(i, j) + \lambda \cdot r(i, j)_{\text{old}}, \quad (3)$$

$$a(i, j)_{\text{new}} = (1 - \lambda) \cdot a(i, j) + \lambda \cdot a(i, j)_{\text{old}}. \quad (4)$$

## 3 SID-SGA 与 SID-SCA

### 3.1 光谱信息散度

SID 测量方法将光谱相似性的计算转化为两个光谱向量之间信息冗余度的计算,将每个光谱向量看作信息源。该方法由 Chang 等<sup>[5-6]</sup> 提出。SID 的具体计算方法为

$$\text{SID}(\mathbf{S}_i, \mathbf{S}_j) = D(\mathbf{S}_i \parallel \mathbf{S}_j) + D(\mathbf{S}_j \parallel \mathbf{S}_i), \quad (5)$$

式中:  $\mathbf{S}_i$  和  $\mathbf{S}_j$  为两个光谱向量;  $\text{SID}(\mathbf{S}_i, \mathbf{S}_j)$  为  $\mathbf{S}_i$  和  $\mathbf{S}_j$  之间光谱信息散度的值;  $D(\mathbf{S}_i \parallel \mathbf{S}_j)$  与  $D(\mathbf{S}_j \parallel \mathbf{S}_i)$  分别为  $\mathbf{S}_i$  和  $\mathbf{S}_j$  之间以及  $\mathbf{S}_j$  和  $\mathbf{S}_i$  之间的相对熵,前者的表达式为

$$D(\mathbf{S}_i \parallel \mathbf{S}_j) = - \sum_{k=1}^N p_i^k \ln \frac{p_i^k}{p_j^k}, \quad (6)$$

$D(\mathbf{S}_j \parallel \mathbf{S}_i)$  的计算公式与(6)式相似。其中,  $k$  为波段位置,  $N$  为波段数,  $p_i^k$  与  $p_j^k$  分别为光谱向量  $\mathbf{S}_i$  和  $\mathbf{S}_j$  在第  $k$  个波段的概率。

概率计算公式为

$$\begin{cases} p_i = [p_i^1 & p_i^2 & \dots & p_i^k & \dots & p_i^N] \\ p_i^k = S_i^k / \sum_{i=1}^N S_i^i \end{cases}, \quad (7)$$

$p_j$  的计算方法与(7)式相似。

### 3.2 光谱梯度角

光谱梯度角测度方法是从光谱角测度(SAM)方法发展而来的。SAM 是一种应用广泛的、基于投影的光谱相似性测度方法,其计算公式为

$$\text{SAM}(\mathbf{S}_i, \mathbf{S}_j) =$$

$$\arccos \left\{ \frac{\sum_{k=1}^N S_i^k S_j^k}{\sqrt{\sum_{k=1}^N (S_i^k)^2 \times \sum_{k=1}^N (S_j^k)^2}} \right\}. \quad (8)$$

可见,两个光谱向量之间的夹角越小,光谱的相似性越高。

SGA 是在 SAM 方法的基础上输入标为  $\mathbf{S}_i$  和  $\mathbf{S}_j$  两个光谱向量的梯度。计算公式为

$$\begin{cases} \text{SGA}(\mathbf{S}_i, \mathbf{S}_j) = \text{SAM}[\text{grad}(\mathbf{S}_i), \text{grad}(\mathbf{S}_j)] \\ \text{grad}(\mathbf{S}_i) = [S_i^2 - S_i^1 \quad S_i^3 - S_i^2 \quad \cdots \quad S_i^k - S_i^{k-1} \quad \cdots \quad S_i^N - S_i^{N-1}] \\ \text{grad}(\mathbf{S}_j) = [S_j^2 - S_j^1 \quad S_j^3 - S_j^2 \quad \cdots \quad S_j^k - S_j^{k-1} \quad \cdots \quad S_j^N - S_j^{N-1}] \end{cases} \quad (9)$$

相对于 SAM 来说,SGA 对光谱畸变的稳健性更强,对光谱曲线局部特征差异更加敏感<sup>[7]</sup>。本研究采用的 SID-SGA 光谱相似性计算方法<sup>[8]</sup>的定义为

$$\text{SIDSGA}_{\tan} = \text{SID}(\mathbf{S}_i, \mathbf{S}_j) \times \tan \frac{\text{SGA}(\mathbf{S}_i, \mathbf{S}_j) + \frac{\pi}{2}}{2} \quad (10)$$

### 3.3 光谱相关角

光谱相关角根据光谱之间的相关性进行相似度测量,其计算公式<sup>[9]</sup>为

$$\begin{cases} \text{SCA}(\mathbf{S}_i, \mathbf{S}_j) = \arccos \frac{r(\mathbf{S}_i, \mathbf{S}_j) + 1}{2} \\ r(\mathbf{S}_i, \mathbf{S}_j) = \frac{N \sum_{k=1}^N S_i^k \cdot S_j^k - \sum_{k=1}^N S_i^k \cdot \sum_{k=1}^N S_j^k}{\sqrt{[N \sum_{k=1}^N (S_i^k)^2 - (\sum_{k=1}^N S_i^k)^2] [N \sum_{k=1}^N (S_j^k)^2 - (\sum_{k=1}^N S_j^k)^2]}} \end{cases} \quad (11)$$

与 SID-SGA 类似, SID-SCA 的计算公式为

$$\text{SIDSCA}_{\tan} = \text{SID}(\mathbf{S}_i, \mathbf{S}_j) \times \tan \frac{\text{SCA}(\mathbf{S}_i, \mathbf{S}_j) + \frac{\pi}{2}}{2} \quad (12)$$

SID-SCA 光谱相似性测度方法能够兼顾光谱整体与局部的状态差异。

## 4 实验验证

将所提方法与文献[10]提出的最大信息量法、文献[11]提出的自适应波段选择法以及基于未改进 AP 聚类的波段选择方法进行对比实验。将各波段

选择方法的输出结果输入 SVM 分类器中进行分类处理,对选出波段子集的相关性以及分类精度进行对比分析。

基于机载成像光谱仪 AVIRIS 在美国印第安纳州西北部某农林混合试验场采集到的 Indiana Pines 数据集进行实验验证。图像空间分辨率为 25 m,图像大小为 145 pixel×145 pixel。原始数据具有 224 个光谱波段,波长范围为 0.4~2.5 μm。最后保留了信噪比较高、质量较好的 200 个波段。最终实验使用的图像大小为 145 pixel×145 pixel×200,包含了 16 个地物类别。Indiana Pines 数据集灰度图像及地物真实标记如图 1 所示。



图 1 Indiana Pines 数据集(a)灰度图及(a)地物真实标记

Fig. 1 (a) Grayscale map and (b) real map for Indiana Pines dataset

分别采用 5 种波段选择方法对 Indiana Pines 数据集进行波段选择。将选择的波段子集输入到 SVM 分类器中进行分类,比较各方法的分类精度。其中,使用全部波段进行分类得到的总体分类精度为 89.70%。实验共进行 6 次,将最优与最差结果去

掉后的结果均值作为最终的实验结果。在每次实验中,选择每类地物样本的 10% 像素点作为分类的训练集,其余 90% 的样本点作为实验的测试集。图 2 是 5 种方法选出波段子集后,进行 SVM 分类得到的总体分类精度曲线(SVM 分类器训练样本集规模

为 10%)。其中,MI 为最大信息量法,ABS 为自适应波段选择法,SID-SGA-AP 为基于 SID-SGA 的 AP 聚类方法,SID-SCA-AP 为基于 SID-SCA 的 AP 聚类方法。

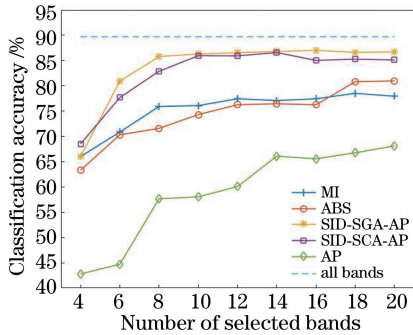


图 2 各方法的分类精度变化曲线

Fig. 2 Classification accuracy curves using different method

从图 2 中可以看出:对采用 SID-SGA-AP 和 SID-SCA-AP 方法选出的波段子集进行分类,得到的地物总体分类精度相近,且比其他方法高出很多,并十分接近用全部波段进行分类的分类精度。这说明 SID-SGA-AP 与 SID-SCA-AP 方法能够更好地挖掘波段的有效信息,选出更具代表性的波段。除 AP 聚类外,当波段数达到一定数量时,分类精度变化缓慢,甚至出现下降。这说明波段之间存在严重的信息冗余,这更加证明了波段选择的重要意义。

表 1 是波段子集数目为 10 时,各方法选出的波段位置。可以看出:MI 及 ABS 方法选择的波段集中在高光谱数据的前半部分,波段选择不全面,难以有效表征原始数据信息;SID-SGA-AP 与 SID-SCA-AP 方法选出的波段更加全面,能够更好地代表原始数据信息,证明了所提方法的有效性、合理性。

表 1 波段选择结果

Table 1 Results of band selection

Method	Band
MI	{1,3,18,20,34,35,42,61,75,90}
ABS	{90,28,27,25,26,74,29,33,24,31}
SID-SGA-AP	{7,15,26,53,99,114,135,165,176,195}
SID-SCA-AP	{7,14,31,64,99,114,134,167,180,195}

图 3(a)~(d)是波段选择结果为 10 时,4 种方法的地物分类示意图。可以看出:对地物面积大且周围环境简单的地物进行分类时,例如图中右下角“木头”及下方“苜蓿”地物等,所提方法的表现较好,错分情况明显少于 MI 与 ABS 方法;对地物面积小且环境复杂的地物进行分类时,例如图中左上角的地物,所提方法的分类精度明显高于 MI 与 ABS 两种方法。

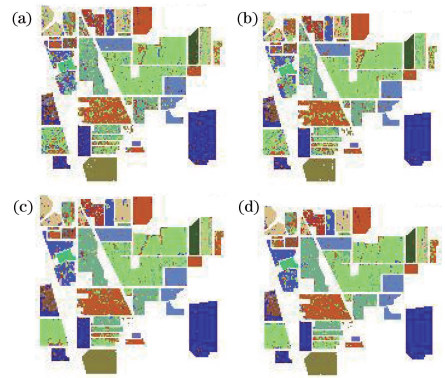


图 3 不同方法的分类结果。(a) MI;(b) ABS;(c) SID-SCA-AP;(d) SID-SGA-AP

Fig. 3 Classification results using different methods. (a) MI; (b) ABS; (c) SID-SCA-AP; (d) SID-SGA-AP

## 5 结 论

本课题组提出了基于改进 AP 聚类的高光谱图像波段选择方法,利用 SID-SGA 以及 SID-SCA 对波段相似度矩阵进行计算。设计实验对 Indiana Pines 数据集进行降维处理,并将降维结果进行地物分类。实验结果证明:基于 SID-SCA 与 SID-SGA 改进 AP 聚类的高光谱图像波段选择结果优于 MI、ABS 等方法;所提高光谱图像波段选择方法能够更加有效地利用数据集的信息,提高分类的准确性。在今后的研究中,将重点利用其他先进的相似度计算方法,以提高算法效率和精度。

## 参 考 文 献

- [1] Hughes G. On the mean accuracy of statistical pattern recognizers [J]. IEEE Transactions on Information Theory, 1968, 14(1): 55-63.
- [2] Zhang L P, Zhang L F. Hyperspectral remote sensing [M]. Beijing: Surveying and Mapping Press, 2011: 71-74.  
张良培, 张立福. 高光谱遥感[M]. 北京: 测绘出版社, 2011: 71-74.
- [3] Frey B J, Dueck D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976.
- [4] Xia C M, Ni Z W, Ni L P, et al. Affinity propagation clustering algorithm based on density adjustment and manifold distance [J]. Computer Science, 2017, 44(10): 187-192, 215.  
夏春梦, 倪志伟, 倪丽萍, 等. 基于密度调整和流形距离的近邻传播算法[J]. 计算机科学, 2017, 44(10): 187-192, 215.

- [5] Chang C I. Spectral information divergence for hyperspectral image analysis [C]. IEEE 1999 International Geoscience and Remote Sensing Symposium, 1999: 6409333.
- [6] Zhao C H, Tian M H, Li J W. Research progress on spectral similarity metrics [J]. Journal of Harbin Engineering University, 2017, 38(8): 1179-1189.  
赵春晖, 田明华, 李佳伟. 光谱相似性度量方法研究进展[J]. 哈尔滨工程大学学报, 2017, 38(8): 1179-1189.
- [7] Angelopoulou E, Lee S W, Bajcsy R. Spectral gradient: a material descriptor invariant to geometry and incident illumination [C]. Proceedings of the Seventh IEEE international Conference on Computer Vision, 1999: 861.
- [8] Zhang X B, Yuan Y, Jing J J, *et al.* Spectral discrimination method information divergence combined with gradient angle [J]. Spectroscopy and Spectral Analysis, 2011, 31(3): 853-857.  
张修宝, 袁艳, 景娟娟, 等. 信息散度与梯度角正切相结合的光谱区分方法 [J]. 光谱学与光谱分析, 2011, 31(3): 853-857.
- [9] Wu H, Xu Y J, Gao R. Extraction of alteration information from hyperspectral imagery based on SCA and SID [J]. Geography and Geo-Information Science, 2016, 32(1): 44-48.  
吴浩, 徐元进, 高冉. 基于光谱相关角和光谱信息散度的高光谱蚀变信息提取 [J]. 地理与地理信息科学, 2016, 32(1): 44-48.
- [10] Liu X S, Ge L, Wang B, *et al.* An unsupervised band selection algorithm for hyperspectral imagery based on maximal information [J]. Journal of Infrared and Millimeter Waves, 2012, 31(2): 166-170, 176.  
刘雪松, 葛亮, 王斌, 等. 基于最大信息量的高光谱遥感图像无监督波段选择方法 [J]. 红外与毫米波学报, 2012, 31(2): 166-170, 176.
- [11] Liu C H, Zhao C H, Zhang L Y. A new method of hyperspectral remote sensing image dimensional reduction [J]. Journal of Image and Graphics, 2005, 10(2): 218-222.  
刘春红, 赵春晖, 张凌雁. 一种新的高光谱遥感图像降维方法 [J]. 中国图象图形学报, 2005, 10(2): 218-222.