

基于多层二部图的高光谱快速谱聚类算法

李思远¹, 郑致远², 杜晓颜³, 刘彤^{1*}, 杨晓君¹

¹广东工业大学信息工程学院, 广东 广州 510006;

²中国人民解放军火箭军工程大学, 陕西 西安 710025;

³中国人民解放军 96962 部队, 北京 102206

摘要 大规模高光谱图像聚类算法广泛应用于遥感领域, 主要包括 K 均值 (K-means) 聚类、谱聚类算法等。然而谱聚类算法仍然有局限性, 由于其计算复杂性高所以不适用于大规模问题。基于锚图的谱聚类算法在一定程度上能够减少计算的成本, 然而在处理大规模高光谱图像数据时, 锚点需要足够密集, 否则无法获得合理的精度, 这使得该聚类算法的计算成本急剧增加。为了克服这些问题, 提出了一种新的基于多层二部图的高光谱快速谱聚类算法。该算法首先使用二叉树选点方式选取锚点, 然后选择多层锚点构建多层锚点图, 接着构造一个多层二部图, 最后对该图进行谱分析。实验证明了提出算法的高效性。

关键词 图像处理; 高光谱图像; 多层图; 二叉树; 二部图; 谱聚类

中图分类号 TP751

文献标志码 A

DOI: 10.3788/LOP202259.1210009

Hyperspectral Fast Spectral Clustering Algorithm Based on Multi-Layer Bipartite Graph

Li Siyuan¹, Zheng Zhiyuan², Du Xiaoyan³, Liu Tong^{1*}, Yang Xiaojun¹

¹College of Information Engineering, Guangdong University of Technology, Guangzhou 510006, Guangdong, China;

²Rocket Force University of Engineering, Chinese People's Liberation Army, Xi'an 710025, Shaanxi, China;

³Chinese People's Liberation Army 96962 Troops, Beijing 102206, China

Abstract Large-scale hyperspectral image clustering algorithms are widely used in the field of remote sensing, including K-means clustering and spectral clustering algorithms. However, the spectral clustering algorithm still has its limitations. Because of its high computational complexity, it is not suitable for large-scale problems. The spectral clustering algorithm based on the anchor graph can reduce the computational cost to a certain extent. However, in the large-scale hyperspectral image data processing, the anchor points need to be dense enough, otherwise reasonable accuracy cannot be obtained. This makes the computing cost of the clustering algorithm increase sharply. In order to overcome these problems, a new fast spectral clustering algorithm based on multi-layer bipartite graph is proposed. Firstly, the anchor points are selected by the binary tree, and the multi-layer anchor points are selected to construct the multi-layer anchor point graph. Then a multi-layer bipartite graph is constructed, and finally the spectrum of the graph is analyzed. The high efficiency of the proposed algorithm is proved by experiments.

收稿日期: 2021-04-22; 修回日期: 2021-05-30; 录用日期: 2021-06-11

基金项目: 国家重点研发计划(2018YFB1802100)、广东省重点领域研发计划(2018B010115001)、广东省自然科学基金(2021A1515011141)、广东省珠江人才计划本土创新科研团队资助项目(2017BT01X168)

通信作者: liutong@gdut.edu.cn

Key words image processing; hyperspectral image; multi-layer graph; binary tree; bipartite graph; spectral clustering

1 引言

随着成像技术和光谱技术的发展和成熟,高光谱遥感可获得空间和光谱连续的高光谱图像(HSI)数据^[1],它提供了丰富的光谱和空间信息,使人们能够全面了解地球的变化。因此,HSI被广泛应用于许多遥感领域,如精确农业、海洋研究和灾害监测^[2]。HSI数据量的增加影响了HSI聚类算法的复杂度和聚类精度,因此如何快速有效地处理HSI数据成为许多研究者关注的焦点^[3-6]。

近年来,谱聚类由于其优异的性能^[7-8]被认为是现实生活中聚类任务的有力工具。然而,高光谱数据量的增长给谱聚类带来了更多的挑战。幸运的是,人们已经提出了许多基于锚图的模型来处理快速增长的数据,例如可扩展的半监督学习^[7]、多视图大规模谱聚类^[8]等等。基于锚图最重要的步骤是锚生成,有两种最常用的策略来生成锚,即随机选择和K均值生成^[9]。一般来说,K均值生成策略通过生成有代表性的锚来获得良好的性能,但计算成本较高^[9];随机策略是有效的,但性能不能得到保证。因此,为了解决这个问题,文献[10]提出了一种新的锚生成策略,称为基于平衡K均值的分层K均值,又称为二叉树选点方法,与K均值策略相比,该方法能够以较小的计算代价生成具有代表性的锚,适合处理大规模数据。基于单层锚点的谱聚类方法在一定程度上能够减少计算的成本,但在处理大规模的高光谱图像数据时,为了便于建立有效的邻接关系,锚点需要足够密集,否则无法获得合理的精度,从而使分类性能下降。然而增加锚点的数量,又会导致计算成本急增,甚至使问题变得非常棘手。

针对上述问题,本文提出一种基于多层二部图的高光谱快速谱聚类(FCMBG)算法。先使用二叉树选点方式从高光谱数据集中选取锚点,构造一种金字塔结构的多层锚点图,然后将多层锚点图与原始数据点构造成二部图,既可以保证数据点之间的特性,也可以大大降低计算复杂度。最后针对获得的相似图进行谱分解获得聚类结果。通过 Indian Pines 数据集和 Salinas 数据集的

仿真实验,验证了FCMBG算法能够有效地处理大规模高光谱图像。

2 基本原理

2.1 二叉树选点方法

使用二叉树选点方式从HSI数据中选出锚点,先从两类K均值开始。基于二叉树选取锚点方法^[10-11]如下所示:

$$\min_{G \in \text{Ind}, I^T G = [\kappa, \iota]} \|X - GC^T\|_F^2, \quad (1)$$

式中: $C \in \mathbb{R}^{n \times 2}$ 为聚类中心矩阵; $G \in \mathbb{R}^{n \times 2}$ 为指示矩阵由0和1组成; $X \in \mathbb{R}^{n \times d}$ 为原始高光谱图像数据; n 为原始数据的个数; d 为数据集的维度; I 为单位矩阵; κ 和 ι 是每次生成的两个簇的样本数,有 $\kappa + \iota = n$; I 代表所有为1的列向量矩阵; $\|\cdot\|_F$ 是F范数。

子聚类中心就是锚点,最终选取 m 聚类中心作为 m 个锚点,用 $U = [u_1, u_2, \dots, u_m]^T \in \mathbb{R}^{m \times d}$ 代表产生的锚点集,其中 m 为锚点集中锚点的个数, d 为每个锚点的特征维度。

2.2 构建多层锚点图

受半监督学习中多层锚图^[12]的启发,在此构建了一个用于无监督学习的多层二部图。在图的基础上,提出了一种新的基于多层二部图的快速高光谱图像聚类方法。为了构造多层二部图,首先引入基于锚的分层图的定义,该定义受半监督学习的启发,让 $G = \{X, U, \zeta\}$ 表示基于锚的分层图, X 是数据矩阵, U 是锚点集, ζ 是所有相邻层之间连接关系的集合。假设底层(H_0)表示原始数据点 $X \in \mathbb{R}^{n \times d}$,其余层($H_a, a = 1, \dots, h$)由多个锚点集合 U 组成,其中 U_a 的大小($U_a \in \mathbb{R}^{m_a \times d}, a = 1, \dots, h$)逐渐减少,也就是 $m_1 > \dots > m_h$,其中 m_a 是 H_a 中的点数。 $\zeta = \{Z_{0,1}, \dots, Z_{h-1,h}\} \in \mathbb{R}^{(n \times m_1, \dots, m_{h-1} \times m_h)}$,其中 $Z_{a-1,a}$ 是 H_{a-1} 和 H_a 点之间连接关系的度量。为了使得描述构建分层锚点图的过程更加清晰易懂,这里展示一个构建分层锚点图的例子,如图1所示,其中 $H_0 = 5000$ 为数据点, H_1, H_2, H_3, H_4 为锚点层,采用二叉树选点方法选择了 $H_1 = 1000, H_2 = 500, H_3 = 250$ 和 $H_4 = 100$ 的多层锚基层。

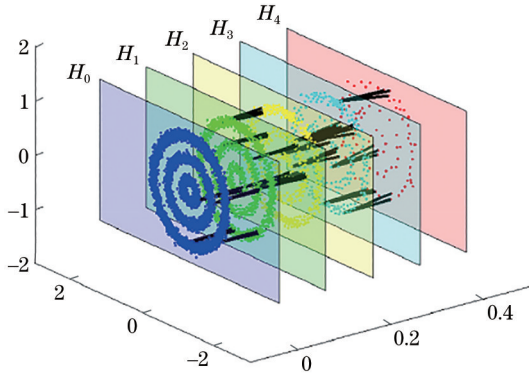


图1 多层锚点图

Fig. 1 Multi-layer anchor graph

2.3 构造基于二部图的相似图

利用原始数据层 H_0 和最后一层锚点 H_h 层构造二部图, 其中二部图的相似度矩阵 \mathbf{W} 可以表示为

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_H \\ \mathbf{Z}_H^T & \mathbf{0} \end{bmatrix}, \quad (2)$$

式中, \mathbf{Z}_H 表示原始数据 H_0 层和最后的锚点 H_h 层之间的累积层间关系的邻接矩阵, 则 \mathbf{Z}_H 可以写为

$$\mathbf{Z}_H = \mathbf{Z}_{0,1}, \dots, \mathbf{Z}_{h-1,h} \in \mathbb{R}^{n \times m_h}, \quad (3)$$

则 $\mathbf{Z}_{a-1,a}$ 表示 H_{a-1} 层与 H_a 层之间的邻接矩阵, $a \in (0, h)$ 。

z_{ij} 为矩阵 $\mathbf{Z}_{a-1,a}$ 的第 i 行、第 j 列元素, 表示 H_{a-1} 层中的第 i 个数据点与 H_a 层中的第 j 个锚点的相邻关系, 即 z_{ij} 可定义为

$$z_{ij} = \frac{K(x_i, u_j)}{\sum_{s \in \varphi_i} K(x_i, u_s)} \quad \forall j \in \varphi_i, \quad (4)$$

式中: $\varphi_i \subset \{1, 2, \dots, m\}$ 代表数据点 x_i 距离最近的 k 个锚点的索引; $K(\cdot)$ 是一种高斯核函数, 高斯核函数通常表示为

$$K(x_i, u_j) = \exp\left(-\frac{\|x_i - u_j\|_2^2}{2\sigma^2}\right), \quad (5)$$

式中: x_i 与 u_j 之间的欧氏距离的平方定义为 $d(x_i, u_j) = \|x_i - u_j\|_2^2$; σ 是热核参数。基于式(2), 对角矩阵可以表示为

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda} \end{bmatrix}, \quad (6)$$

式中: $\mathbf{D}_r \in \mathbb{R}^{n \times n}$ 是一个对角元素为矩阵 \mathbf{Z} 行和的对角矩阵, 即 $D_{r_{ii}} = \sum_{j=1}^n Z_{ij}$; $\mathbf{\Lambda} \in \mathbb{R}^{m_h \times m_h}$ 是一个对角元

素为 \mathbf{Z} 列和的对角矩阵, 即 $\Lambda_{jj} = \sum_{i=1}^n Z_{ij}$ 。

拉普拉斯矩阵可写为

$$\mathbf{L} = \mathbf{D} - \mathbf{W} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda} \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{Z}_H \\ \mathbf{Z}_H^T & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_r & -\mathbf{Z}_H \\ -\mathbf{Z}_H^T & \mathbf{\Lambda} \end{bmatrix}. \quad (7)$$

利用式(3)的累积层间关系的邻接矩阵, 可以得到从 H_0 到 H_h 的类指示矩阵, 其密度-稀疏关系为

$$\mathbf{F}_X = (\mathbf{Z}_{0,1}, \dots, \mathbf{Z}_{h-1,h}) \mathbf{F}_U = \mathbf{Z}_H \mathbf{F}_U, \quad (8)$$

式中: \mathbf{F}_X 为原始数据的类指示矩阵; \mathbf{F}_U 为 H_h 层锚数据集的类指示矩阵, 则这两个类指示矩阵可联合写

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_X \\ \mathbf{F}_U \end{bmatrix} \in \mathbb{R}^{(n+m_h) \times c}, \quad c \text{ 为聚类数。}$$

2.4 多层二部图的谱聚类分析

谱聚类的目标函数为

$$\min_{\mathbf{F}^T \mathbf{F} = \mathbf{I}} \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F}), \quad (9)$$

式中, Tr 为迹函数, 式(9)的最优解是对拉普拉斯矩阵 \mathbf{L} 的特征值分解。由于 \mathbf{F} 的元素是离散值, 因此很难求解式(9)。为了解决这个问题, 将 \mathbf{F} 从离散值松弛为连续值。通过对矩阵 \mathbf{L} 的特征值分解, 得到松弛连续解, 该松弛连续解由对应于拉普拉斯矩阵 \mathbf{L} 的最小特征值的特征向量组成。故拉普拉斯矩阵 \mathbf{L} 可以转化为

$$\mathbf{L} = \mathbf{D} - \mathbf{W} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}} = \mathbf{I} - \begin{bmatrix} 0 & \mathbf{Z}_H \mathbf{\Lambda}^{-\frac{1}{2}} \\ \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Z}_H^T & 0 \end{bmatrix} = \mathbf{I} - \begin{bmatrix} 0 & \mathbf{B} \\ \mathbf{B}^T & 0 \end{bmatrix}, \quad (10)$$

式中, $\mathbf{B} = \mathbf{Z}_H \mathbf{\Lambda}^{-\frac{1}{2}}$, 因此, 式(9)可以转化为

$$\begin{aligned} & \max_{\mathbf{F}^T \mathbf{F} = \mathbf{I}} \text{Tr} \left(\mathbf{F}^T \begin{bmatrix} 0 & \mathbf{B} \\ \mathbf{B}^T & 0 \end{bmatrix} \mathbf{F} \right) = \\ & \max_{\mathbf{F}_X^T \mathbf{F}_X + \mathbf{F}_U^T \mathbf{F}_U = \mathbf{I}} \text{Tr} \left(\begin{bmatrix} \mathbf{F}_X \\ \mathbf{F}_U \end{bmatrix}^T \begin{bmatrix} 0 & \mathbf{B} \\ \mathbf{B}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{F}_X \\ \mathbf{F}_U \end{bmatrix} \right) = \\ & \max_{\mathbf{F}_X^T \mathbf{F}_X + \mathbf{F}_U^T \mathbf{F}_U = \mathbf{I}} \text{Tr}(\mathbf{F}_X^T \mathbf{B} \mathbf{F}_U), \end{aligned} \quad (11)$$

那么矩阵 \mathbf{B} 的奇异值分解可以写为

$$\mathbf{B} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T, \quad (12)$$

式中: 左奇异矩阵 $\mathbf{U} \in \mathbb{R}^{n \times n}$; 奇异值矩阵 $\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$; 右奇异值矩阵 $\mathbf{V} \in \mathbb{R}^{m \times m}$ 。

本算法不直接对矩阵 \mathbf{L} 进行特征值分解, 而是对矩阵 \mathbf{B} 进行奇异值分解, 得到矩阵 \mathbf{F} 的松弛连续解, 此时还不能够直接得到最终的聚类结果, 需要离散化处理, 所以采用 K 均值 (K-means) 聚类方法

计算离散解,最终获得聚类的结果。

FCMBG 算法的聚类过程进行汇总如表 1 所示。

为更直观地了解提出的 FCMBG 算法,将

表 1 FCMBG 的算法过程

Table 1 Procedure of FCMBG algorithm

Input: Original data \mathbf{X} , the number of classes c , the number of anchors in each layer m_1, \dots, m_h , the number of near neighbor k
1) Generate m anchors by balanced K-means based hierarchical K-means selection;
2) Obtain the matrix $\mathbf{Z}_{0,1}, \dots, \mathbf{Z}_{h-1,h}$, according to Eq. (4);
3) Obtain the matrix \mathbf{Z}_H according to Eq. (3);
4) Calculate the adjacent matrix of bipartite hierarchical graph $\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_H \\ \mathbf{Z}_H^T & \mathbf{0} \end{bmatrix}$;
5) Obtain the relaxed continuous solution of \mathbf{F}_X and \mathbf{F}_U by performing singular value decomposition on matrix $\mathbf{B} = \mathbf{Z}_H \mathbf{\Lambda}^{-\frac{1}{2}}$.
Output: c classes

2.5 时间复杂度分析

表 1 中 \mathbf{X} 是数据矩阵, c 是聚类类别数, m_1, \dots, m_h 分别是 H_1 到 H_n 层的锚点数。为简单起见, H_0 为底层。所以 FCMBG 算法的计算复杂性可以分为以下几个部分:

1) 通过二叉树选取锚点的方式从 H_1 到 H_h 层得到 m 个锚点的计算复杂度为 $O(hnd(\log_m^t))$ 。

2) 获得 $\mathbf{Z}_{0,1}$ 到 $\mathbf{Z}_{h-1,h}$ 矩阵的计算复杂度是 $O(\sum_{i=1}^h dm_{i-1}m_i)$, 因为 $n \gg m_1, \dots, \gg m_h$, 所以计算复杂度近似为 $O(nm_1d)$ 。

3) 计算矩阵 \mathbf{Z}_h 的计算复杂度为 $O(\sum_{i=1}^h m_{h-2} m_{h-1} m_h)$, 因为 $n \gg m_1, \dots, \gg m_h$, 所以计算复杂度近似为 $O(nm_1m_h)$ 。

4) 通过对矩阵 \mathbf{B} 进行奇异值分解得到 F 的松弛连续解的计算复杂度为 $O(nm_hc)$ 。

5) 对最终聚类结果的松弛离散解进行 K-means 聚类的计算复杂度为 $O(nm_hct)$, 其中 t 是迭代数。

该算法的计算复杂度为 $O[hnd(\log_m^t) + ndm_1 + nm_1m_h + nm_hc + nm_hct]$ 。

3 实验结果与分析

3.1 高光谱数据集的介绍

为了验证 FCMBG 算法的高效性, 采用两个高光谱数据集和其他几种流行的聚类算法进行实验, 通过总准确度、平均精度^[13-14]、Kappa 系数^[14]、运行时间和聚类图对实验结果进行评估。高光谱数据集分别是 Indian Pines、Salinas, 详细信息如表 2 所示。

3.2 评价实验结果

为了证明提出的 FCMBG 算法的有效性, 将其

表 2 高光谱数据集的详细信息

Table 2 Details of hyperspectral data sets

Data set	Data volume	Dimension	Class number
Indian Pines	21025	220	16
Salinas	111104	204	16

与 K-means、模糊 C 均值聚类 (FCM)、基于核诱导距离测度的模糊聚类鲁棒图像分割 (FCM_S1)、谱聚类 (SC)、基于锚点图的快速聚类 (FSCBG) 算法在两个高光谱数据集上得到的结果进行了比较。在实验中, 平均精度 (AA)、总准确度 (OA)、Kappa 系数和视觉聚类图被给出以定量评估每种聚类方法的聚类性能。此外, 在实验过程中, 程序设置了循环, 记录下最好的结果及其参数。Indian Pines、Salinas 数据集的所有样本被选为测试数据, 包括所有类别和未标记的像素。

为清楚起见, 在此使用 FCMBG- m_1 - m_2 - \dots - m_h 表示建立在具有 h 锚层的分层锚图上的 FCMBG 方法。例如, FCMBG-2000-200 意味着在其图结构中有两个锚层, 分别包含 2000 和 200 个锚。

3.2.1 Indian Pines 数据集实验结果分析

首先, 在 Indian Pines 数据集上进行聚类实验, 以评估每种聚类方法的性能。FCMBG 中使用的参数分别设置为 $m = [512, 128]$ 、 $k = 15$, 对每种算法进行 15 次实验。每种聚类方法获得的定量评价和聚类图分别示于表 3 和图 2 中。在表 3 中, 最佳结果以粗体显示。从表 3 中可以看到, FCMBG 算法的 OA、AA、Kappa 系数都比 K-means、FCM、SC、FSCBG 高; 关注每种聚类方法的运行时间, 从表 3 可以看出, K-means、FCM、FCM_S1 和 FSCBG 的运行时间是同一数量级的。特别是对于两种基于

表 3 不同算法在 Indian Pines 数据集的实验结果

Table 3 Experimental results of different algorithms in Indian Pines data set

Algorithm	K-means	FCM	FCM_S1	SC	FSCBG-128	FCMBG-512-128
AA / %	36.57	36.83	34.76	31.12	35.25	39.89
OA / %	34.84	35.42	36.58	33.16	35.08	37.80
Kappa	0.2758	0.2859	0.2953	0.2725	0.2866	0.3022
Time / s	3.8601	4.2935	6.2047	39.2537	0.7431	1.3250

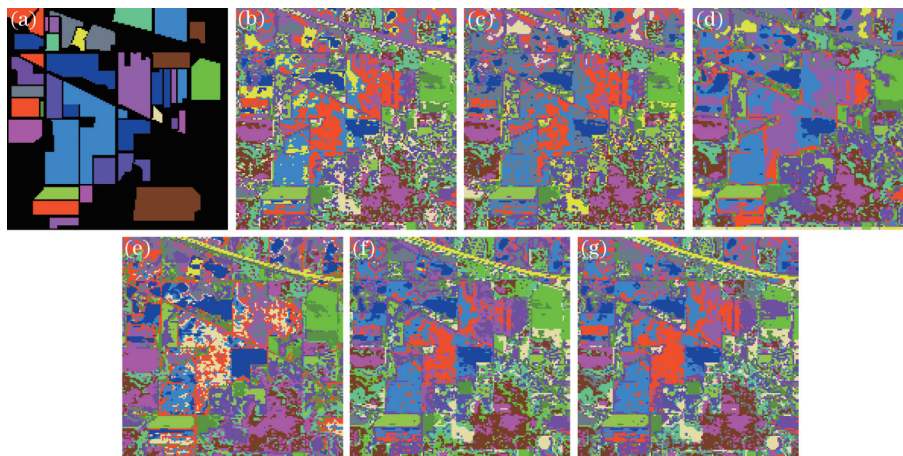


图 2 不同算法在 Indian Pines 数据集所得聚类图。(a)真实图;(b) K-means算法;(c) FCM算法;(d) FCM_S1算法;
(e) SC算法;(f) FSCBG-128算法;(g) FCMBG-512-128算法

Fig. 2 Clustering graphs of different algorithms in Indian Pines data set. (a) Real image; (b) K-means algorithm; (c) FCM algorithm; (d) FCM_S1 algorithm; (e) SC algorithm; (f) FSCBG-128 algorithm; (g) FCMBG-512-128 algorithm

图的聚类方法,FCMBG只需要1.3250 s,比SC方法快了近30倍。注意,SC可以在不属于大规模HSI数据集的Indian Pines数据集上工作,因为样本总数只有21025。FCMBG算法的OA、Kappa系数都比其他的算法好,且得到的分类图中错分点更少,地物分布更加平滑,证明了FCMBG算法的实用性和可行性。实验结果表明选择多层锚点构建多层锚点图,可以获得更好的数据间的邻接关系,层数增加,精度会相应变高,但是需要花费更大的时间代价。

3.2.2 Salinas数据集实验结果分析

在Salinas数据集上进行了聚类实验。参数分别设置为 $m = [1024, 256]$ 、 $k = 5$ 。每种聚类方法获得的定量评价和聚类图分别示于表4和图3中。参考表4,FCMBG获得最佳OA为70.13%,Kappa

系数也最佳为0.6684。从图3中可以看到,FCMBG比其他算法产生更同质的区域和更好的聚类图。表4显示,K-means、FCM、FCM_S1和FSCBG的运行时间在同一数量级。注意,Salinas数据集属于大规模HSI数据集,样本总数为111104。提出的FCMBG只需要11.4740 s,但由于“内存不足(OM)错误”,SC无法在Salinas数据集上工作。

与其他方法相比,FCMBG算法在高维数据中获得了更好的聚类精度,说明FCMBG算法能够有效处理高维数据;FCMBG的聚类精度在相同实验下都优于其他的算法,更加说明了FCMBG算法的有效性和高效性。构造多层锚点图,可获得每一层锚图之间累积层关系的邻接矩阵,多层选点是每层都构建了一个邻接矩阵, $Z_{a-1,a}$ 表示 H_{a-1} 层与 H_a 层

表 4 Salinas数据集的实验结果

Table 4 Results of Salinas data set

Algorithm	K-means	FCM	FCM_S1	FSCBG-256	FCMBG-1024-256
AA / %	64.93	65.79	66.05	64.11	68.41
OA / %	66.02	64.42	66.41	64.90	70.13
Kappa	0.6216	0.6062	0.6264	0.6128	0.6684
Time / s	20.1429	21.5936	35.2432	3.1920	11.4740

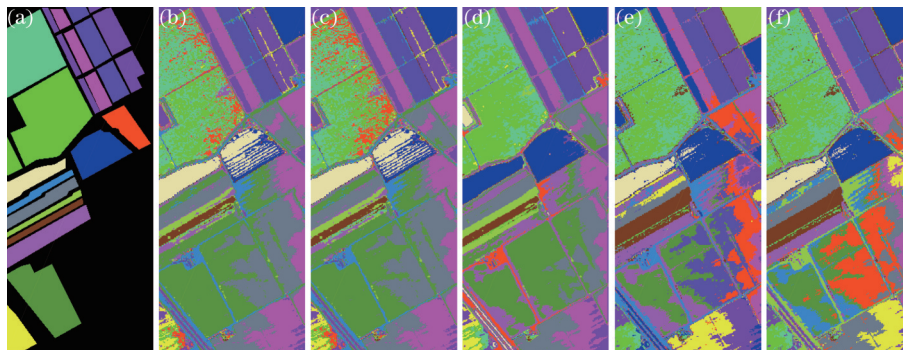


图 3 不同算法在 Salinas 数据集所得聚类图。(a) 真实图; (b) K-means 算法; (c) FCM 算法; (d) FCM_S1 算法; (e) FSCBG-256 算法; (f) FCMBG-1024-256 算法

Fig. 3 Clustering graphs of different algorithms in Salinas data set. (a) Real image; (b) K-means algorithm; (c) FCM algorithm; (d) FCM_S1 algorithm; (e) FSCBG-256 algorithm; (f) FCMBG-1024-256 algorithm

之间的邻接矩阵。最后相乘得到 Z_H , 实验结果可以证明提出算法获得的 Z_H 比单层得到的 Z_H 更好。

4 结 论

提出了一种新的基于多层二部图快速谱聚类算法来有效地处理大规模的 HSI 聚类问题。首先, 通过二叉树选点方法选择锚点, 同时通过探索多层锚结构, 获得更好的数据间的邻接关系; 选择多层锚点; 然后构建多层二部图; 最后对该图进行谱分析。实验结果表明, 该算法的聚类精度结果最优, 比大多数传统方法运行时间短, 比最接近的方法运行时间稍长但聚类效果变好, 从而验证了该算法对于大规模的 HSI 数据具有良好的聚类性能。

参 考 文 献

- [1] Zhang H Y, Zhai H, Zhang L P, et al. Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(6): 3672-3684.
- [2] Tasdemir K, Merenyi E. A validity index for prototype-based clustering of data sets with complex cluster structures[J]. IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society, 2011, 41(4): 1039-1053.
- [3] Xie D Y, Zhang X D, Gao Q X, et al. Multiview clustering by joint latent representation and similarity learning[J]. IEEE Transactions on Cybernetics, 2020, 50(11): 4848-4854.
- [4] Xue Z H, Du P J, Li J, et al. Sparse graph regularization for hyperspectral remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(4): 2351-2366.
- [5] 付青, 郭晨, 罗文浪. 一种利用空谱联合特征的高光谱图像分类方法[J]. 激光与光电子学进展, 2020, 57(20): 202803.
- [6] 张明华, 邹亚晴, 宋巍, 等. GGCN: 基于 GPU 的高光谱图像分类算法[J]. 激光与光电子学进展, 2020, 57(20): 201101.
- [7] Chang X J, Nie F P, Ma Z G, et al. A convex formulation for spectral shrunk clustering[C]// Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA. Menlo Park: AAAI Press, 2015: 2532-2538.
- [8] Chang X J, Nie F P, Yang Y, et al. Convex sparse PCA for unsupervised feature learning[J]. ACM Transactions on Knowledge Discovery from Data, 2016, 11(1): 1-16.
- [9] Cai D, Chen X L. Large scale spectral clustering via landmark-based sparse representation[J]. IEEE Transactions on Cybernetics, 2015, 45(8): 1669-1680.
- [10] Wang R, Nie F P, Yu W Z. Fast spectral clustering with anchor graph for large hyperspectral images[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(11): 2003-2007.

- [11] 许裕雄, 杨晓君, 蔡湧达, 等. 基于二叉树锚点的高光谱快速聚类算法[J]. 激光与光电子学进展, 2021, 58(2): 0210021.
Xu Y X, Yang X J, Cai Y D, et al. Hyperspectral fast clustering algorithm based on binary tree anchor points[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210021.
- [12] Wang M, Fu W J, Hao S J, et al. Learning on big graph: label inference and regularization with anchor hierarchy[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(5): 1101-1114.
- [13] Yuan Y, Lin J Z, Wang Q. Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization[J]. IEEE Transactions on Cybernetics, 2016, 46(12): 2966-2977.
- [14] Richards J A, Jia X. Remote sensing digital image analysis: an introduction[M]. New York: Springer, 1999.