

基于半监督信息的截集式可能性C-均值聚类算法

范九伦 高梦飞* 于海燕 陈斌斌

(西安邮电大学通信与信息工程学院 西安 710121)

(西安邮电大学电子信息勘验应用技术公安部重点实验室 西安 710121)

摘要: 截集式可能性C-均值(C-PCM)聚类算法将截集概念引入可能性C-均值(PCM)聚类算法中, 明显改善了PCM的聚类中心重合问题, 并能够对噪声和奇异点的数据进行有效聚类, 但该聚类算法对小目标数据聚类时仍然存在聚类中心偏移的问题。针对此问题, 该文将半监督学习机制引入C-PCM的目标函数中, 通过部分先验信息来指导聚类过程, 提出半监督截集式可能性C-均值(SS-C-PCM)聚类算法。为了提高彩色图像的分割效率和分割准确率, 将差分进化超像素(DES)算法获得的图像空间邻域信息融入SS-C-PCM目标函数中, 并利用彩色直方图重构目标函数, 以降低算法的计算复杂度, 进而提出基于差分进化超像素的半监督截集式可能性C-均值(desSS-C-PCM)聚类算法。通过人造数据和彩色图像分割的仿真并与多种相关算法进行对比, 表明该算法能够有效改善小目标数据的聚类效果, 提高算法的执行效率。

关键词: 截集式可能性C-均值聚类; 半监督; 超像素; 彩色直方图

中图分类号: TN911.73; TP391

文献标识码: A

文章编号: 1009-5896(2021)08-2378-08

DOI: 10.11999/JEIT200757

Cutset-type Possibilistic C-means Clustering Algorithms Based on Semi-supervised Information

FAN Jiulun GAO Mengfei YU Haiyan CHEN Binbin

(School of Communication and Information Engineering, Xi'an University of Posts and
Telecommunications, Xi'an 710121, China)

(Key Laboratory of Electronic Information Application Technology for Scene Investigation of Ministry of
Public Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

Abstract: Cutset-type Possibilistic C-Means clustering (C-PCM) algorithm can significantly reduce the coincident clustering phenomenon of the Possibilistic C-Means clustering (PCM) algorithm by introducing the cut-set concept into the PCM. The C-PCM also has strong robustness to noise and outliers. However, the C-PCM still suffers from the center migration problem for datasets with small targets. In order to solve this problem, a Semi-Supervised Cutset-type Possibility C-Means (SS-C-PCM) clustering algorithm is proposed by introducing the semi-supervised learning mechanism into the objective function of the C-PCM and utilizing some prior information to guide the clustering process. Meanwhile, in order to improve the segmentation efficiency and accuracy of color images, a differential evolutionary superpixel-based Semi-Supervised Cutset-type Possibilistic C-Means (desSS-C-PCM) clustering algorithm is proposed. In the desSS-C-PCM, the Differential Evolutionary Superpixel(DES) algorithm is used to obtain the spatial neighborhood information of an image, which is integrated into the objective function of the semi-supervised C-PCM to improve the segmentation quality. Simultaneously, the color histogram is used to reconstruct the new objective function to reduce the computational complexity of the algorithm. Several experiments of artificial data clustering and color image segmentation show that the proposed algorithm can effectively improve the clustering effect of datasets with small targets and the execution efficiency compared with several related algorithms.

收稿日期: 2020-08-26; 改回日期: 2021-01-22; 网络出版: 2021-02-23

*通信作者: 高梦飞 1143932136@qq.com

基金项目: 国家自然科学基金(62071378, 62071379, 62071380), 西安邮电大学西邮新星团队(xyt2016-01)

Foundation Items: The National Natural Science Foundation of China (62071378, 62071379, 62071380), New Star Team of Xi'an University of Posts & Telecommunications (xyt2016-01)

Key words: Cutset-type Possibilistic C-Means clustering (C-PCM); Semi-supervised; Superpixel; Color histogram

1 引言

图像分割^[1]是计算机视觉领域中的重要技术,是后续进行图像分析与识别的基础,其目的是从图像中将感兴趣的区域与其他部分进行分离并提取出来。常规的图像分割方法包括:基于阈值的分割方法^[2]、基于区域的分割方法^[3]、基于边缘的分割方法^[4]、基于聚类的分割方法^[5]等。作为聚类图像分割的常用方法,模糊C-均值(Fuzzy C-Means, FCM)^[6]对噪声点和奇异点非常敏感。这主要是缘于该算法对隶属度有约束条件:要求各个样本点到所有类的隶属度之和必须为1。

为解决FCM聚类算法对噪声敏感的问题,Krishnapuram等人^[7]提出可能性C-均值(Possibilistic C-Means, PCM)聚类算法,PCM聚类算法放弃了FCM聚类算法中隶属度和为1的约束条件,能明显改善FCM聚类算法的噪声敏感问题。然而PCM聚类算法完全释放了类间关系,只有PCM目标函数在全体聚类中心重合时才能获得真正的全局最优解,导致PCM聚类算法非常不稳定,容易出现聚类中心重合的问题^[8,9]。除此之外,PCM聚类算法还存在对初始化敏感、参数设置敏感等一系列问题。为解决这些问题,学者对PCM聚类算法进行了一些改进,Pal等人^[10]提出了可能性-模糊混合聚类算法,以克服聚类中心重合现象。Memon等人^[11]将核PCM聚类算法和FCM聚类算法相结合,提出了结合局部信息的核PCM聚类算法,该算法提高了PCM对噪声的鲁棒性。

针对PCM的聚类中心重合问题,从隶属度的修改入手,文献^[12]提出了截集式可能性C-均值(Cutset-type Possibilistic C-Means, C-PCM)聚类算法,C-PCM聚类算法将截集门限引入PCM聚类算法中,并选择部分样本的隶属度进行修改,从而引入类间关系,该算法在一定程度上能够克服PCM的聚类中心重合问题。C-PCM聚类算法在对目标大小相近的数据聚类的时候,能够找到正确的聚类中心,但是对小目标数据进行聚类时仍然存在聚类中心重合问题。具体来说,小目标的聚类中心会严重偏离到大目标的数据区域,从而影响聚类效果。

半监督聚类^[13-15]使用半监督信息来指导聚类过程,一定程度上改善了聚类效果和运行效率。2009年,Yasunori等人^[13]将监督隶属度引入FCM聚类算法中,提出了半监督模糊C-均值(Semi-Supervised Fuzzy C-Means, SS-FCM)聚类算法。

2012年,Yin等人^[14]提出将监督隶属度引入熵正则化FCM聚类算法中,提出了熵正则化半监督模糊聚类算法,避免了模糊因子人为确定的问题。本文将半监督引入C-PCM聚类算法中,提出半监督截集式可能性C-均值(Semi-Supervised Cutset-type Possibilistic C-Means, SS-C-PCM)聚类算法,通过将标签数据引入目标函数中,防止C-PCM聚类算法的中心偏移问题,从而改善了小目标数据的聚类效果。

超像素最早由Ren等人^[16]提出,它由一系列位置相邻且颜色、亮度、纹理等特征相似的像素点组成子区域,这些子区域保留了进一步进行图像分割的有效信息,并且不会破坏图像中物体的边界信息。同时超像素图像的子区域内的像素具有相近的图像特征,针对小区域进行处理可以大大减少计算量,提高分割效率^[17-20]。鉴于此,本文将超像素子区域信息以彩色直方图^[21]的形式引入SS-C-PCM聚类算法的目标函数中,进一步提出了一种基于差分进化超像素的半监督截集式可能性C-均值(differential evolutionary superpixel-based Semi-Supervised Cutset-type Possibilistic C-Means, desSS-C-PCM)聚类算法,提高了聚类算法的运行效率。

2 可能性C-均值聚类算法

为了改善FCM聚类算法的噪声敏感性,Krishnapuram等人^[7]提出了PCM聚类算法,该算法通过改变FCM聚类算法的约束条件进而减少噪声对聚类的影响。PCM算法的目标函数为

$$J_{\text{PCM}}(\mathbf{T}, \mathbf{V}) = \sum_{k=1}^c \sum_{i=1}^n t_{ki}^m d^2(x_i, v_k) + \eta \sum_{k=1}^c \sum_{i=1}^n (1 - t_{ki})^m \quad (1)$$

其中, $\mathbf{T} = [t_{ki}]_{c \times n}$ 为可能性划分矩阵, $\mathbf{V} = \{v_1, \dots, v_k, \dots, v_c\}$ 表示聚类中心集合, t_{ki} 为可能性隶属度值,表示第 i 个样本属于第 k 个聚类的可能性大小,也称为典型值,满足 $\sum_{i=1}^n t_{ki} > 0, k = 1, 2, \dots, c$, m 是加权指数,一般取 $m = 2$ (下同), d^2 表示欧氏距离, η 为惩罚因子, $\eta = K \sum_{i=1}^n t_{ki}^m d^2(x_i, v_k) / \sum_{i=1}^n t_{ki}^m$, 其中 K 一般取1。

PCM聚类算法通过释放隶属度和为1的约束条件,提高了聚类算法对噪声的鲁棒性。但是PCM

对隶属度的约束过于宽泛,使得PCM缺乏类间关系,目标函数(式(1))在全体聚类中心重合时才能获得真正的全局最优解,最终导致算法容易产生聚类中心重合问题^[8,9],也称为一致性聚类问题。此外PCM聚类算法存在对初始化、参数设置敏感的问题。

3 截集式可能性C-均值聚类算法

为了克服PCM最突出的问题——中心重合问题, Yu等人^[12]将截集概念引入PCM聚类算法中,通过截集门限选择部分样本点并修改其典型值,进而引入类间关系,从而提出了截集式可能性C-均值(C-PCM)聚类算法,能够在一定程度上克服PCM的聚类中心重合问题。C-PCM的目标函数为

$$J_{C-PCM}(T, V) = \sum_{k=1}^c \sum_{i=1}^n t_{ki}^m d^2(x_i, v_k) + \eta \sum_{k=1}^c \sum_{i=1}^n (1 - t_{ki})^m \quad (2)$$

典型值 t_{ki} 和聚类中心 v_k 的更新公式与PCM算法的更新公式相同,惩罚因子 η 为提前设置,在算法中取固定值。其中,截集采用基于阴影集的理论^[22,23]自适应选取。C-PCM聚类算法通过将样本点的非获胜典型值进行修改,进而增大获胜典型值与非获胜典型值的差值,从而引入类间关系。该算法在对各类大小相近的数据进行聚类时,能够很好地克服PCM的聚类中心重合问题。但是在各类样本容量大小相差较大,即包含小目标的数据进行聚类时,仍然存在聚类中心偏移、部分中心重合问题。

4 半监督模糊C-均值聚类

当样本不平衡时,C-PCM算法对包含小目标的数据进行聚类时,存在聚类效果不理想的问题。具体来说如果数据集中存在一个类的样本容量很大,而另一类的样本容量很小,那么样本容量较小的类中心很容易就偏移 to 样本容量大的数据区域中。针对这一问题,本文借鉴Yasunori等人^[13]提出的半监督思想,将监督隶属度引入C-PCM聚类算法中。Yasunori等人^[13]提出半监督模糊C-均值(Semi-Supervised Fuzzy C-Means, SS-FCM)聚类算法,其目标函数为

$$J_{SS-FCM}(U, V) = \sum_{k=1}^c \sum_{i=1}^n |u_{ki} - \bar{u}_{ki}|^m \|x_i - v_k\|^2 \quad (3)$$

其中, u_{ki} 表示样本点 x_i 对第 k 类的隶属度,满足约束条件 $\sum_{k=1}^c u_{ki} = 1 (i = 1, 2, \dots, n)$ 。监督隶属度

\bar{u}_{ki} 代表样本点 x_i 对第 k 类的监督隶属度,假设被人为标记的样本点 x_i 隶属于第1类,那么监督隶属度 $\bar{u}_{1i} = 1, \bar{u}_{2i} = 0, \dots, \bar{u}_{ci} = 0$;未被标记的样本点的监督隶属度设置为0,因此 $\sum_{k=1}^c \bar{u}_{ki} \leq 1$ 。

5 本文算法

SS-FCM聚类算法利用已知类别标签的少量样本指导模糊聚类过程,一定程度上提高了运行效率。本文引入该算法的思想,将监督信息融入C-PCM聚类算法的目标函数中,提出半监督截集式可能性C-均值(Semi-Supervised Cutset-type Possibilistic C-Means, SS-C-PCM)聚类算法;同时将SS-C-PCM聚类算法应用到彩色图像中:为了降低计算复杂度,提高分割效果,利用差分进化超像素算法对图像进行预处理的同时引入空间信息,进而提出了一种基于差分进化超像素的半监督截集式可能性C-均值聚类算法desSS-C-PCM。

5.1 半监督截集式可能性C-均值聚类算法

针对C-PCM聚类算法对包含小目标的数据进行聚类时出现中心偏移、部分中心重合现象,本文借鉴半监督模糊C-均值聚类算法^[13]的思想,将半监督信息引入C-PCM的目标函数中,利用已知类别标签的少量样本指导C-PCM聚类算法的聚类过程。半监督截集式可能性C-均值聚类算法的目标函数为

$$J_{SS-C-PCM}(T, V) = \sum_{k=1}^c \sum_{i=1}^n |t_{ki} - \bar{t}_{ki}| \|x_i - v_k\|^2 + \eta \sum_{k=1}^c \sum_{i=1}^n (1 - |t_{ki} - \bar{t}_{ki}|)^m \quad (4)$$

其中, t_{ki} 表示样本点 x_i 对第 k 类的隶属度,满足约束条件 $\sum_{i=1}^n t_{ki} > 0, k = 1, 2, \dots, c$;监督典型值 \bar{t}_{ki} 代表样本点 x_i 对第 k 类的监督典型值,假设被人为标记的样本点 x_i 隶属于第1类,那么样本点 x_i 的监督隶属度为 $\bar{t}_{1i} = 1, \bar{t}_{2i} = 0, \dots, \bar{t}_{ci} = 0$;未被标记的样本点的监督典型值设置为0,因此 $\sum_{i=1}^c \bar{t}_{ki} \leq 1$ 。惩罚因子 η 在本文算法中取固定值。

SS-C-PCM算法聚类中心 v_k 更新公式为

$$v_k = \frac{\sum_{i=1}^n |t_{ki} - \bar{t}_{ki}|^m x_i}{\sum_{i=1}^n |t_{ki} - \bar{t}_{ki}|^m} \quad (5)$$

SS-C-PCM算法典型值 t_{ki} 的更新公式分为 $m = 1$ 和 $m > 1$ 两种情况。当 $m > 1$ 时,更新公式为

$$t_{ki} = \bar{t}_{ki} + \frac{1}{1 + (\|x_i - v_k\|^2/\eta)^{1/(m-1)}} \quad (6)$$

当 $m = 1$ 时，典型值 t_{ki} 的更新公式为

$$t_{ki} = \begin{cases} \bar{t}_{ki} + 1 - \sum_{i=1}^c \bar{t}_{ki}, & i = \arg \min_j \|x_k - v_j\|^2 \\ \bar{t}_{ki}, & \text{其他} \end{cases} \quad (7)$$

5.2 基于差分进化超像素的半监督截集式可能性C-均值聚类算法

传统的聚类图像分割算法是在像素层面进行聚类，其计算复杂度高，算法运行时间长，占用空间大。超像素图像是一系列位置相邻且颜色、亮度、纹理等特征相似的像素点组成的子区域，且超像素子区域的个数远小于彩色图像的像素数。差分进化超像素(Differential Evolutionary Superpixel, DES)算法是由Gong等人^[19]在2018年提出的一种新的基于全局优化的进化超像素方法，相对于简单线性迭代聚类超像素算法^[20]，该算法可以更好地捕获图像的边界信息，提高图像的边界依附性。为减少计算量，改善图像的分割效果，本文将差分进化超像素(DES)算法获取的超像素子区域信息引入本文的SS-C-PCM图像分割算法中，并借鉴Lei等人^[21]提出的彩色直方图的方法重新设计目标函数，进而提出了一种基于差分进化超像素的半监督截集式可能性C-均值(differential evolutionary superpixel-based Semi-Supervised Cutset-type Possibilistic C-Means, desSS-C-PCM)聚类算法，该算法的目标函数为

$$J_{\text{desSS-C-PCM}}(T, V) = \sum_{k=1}^c \sum_{l=1}^q S_l |t_{kl} - \bar{t}_{kl}|^m \cdot \left\| \left(\frac{1}{S_l} \sum_{p \in R_l} x_p \right) - v_k \right\|^2 + \eta \sum_{k=1}^c \sum_{l=1}^q (1 - |t_{kl} - \bar{t}_{kl}|)^m \quad (8)$$

其中， l 是颜色等级， $1 \leq l \leq q$ ， q 是超像素区域的个数， $l, q \in N^+$ 。 S_l 为第 l 个超像素区域 R_l 中像素数， x_p 是第 l 个超像素区域的彩色像素值， $1/S_l \sum_{p \in R_l} x_p$ 表示超像素区域的像素平均值，典

型值 t_{kl} 为第 l 个超像素区域的像素平均值到聚类中心 v_k 的典型性， \bar{t}_{kl} 是监督典型值。本文在式(8)的目标函数中引入超像素的子区域信息，将原始图像中的彩色像素 x_p 替换为对应的每一个超像素区域中像素的平均值 $1/S_l \sum_{p \in R_l} x_p$ ，从而有效降低了计算复杂度。

本文算法采用拉格朗日乘子法求出式(8)的最优解，得到聚类中心 v_k 和典型性 t_{kl} 更新公式为

$$v_k = \frac{\sum_{l=1}^q |t_{kl} - \bar{t}_{kl}|^m \sum_{p \in R_l} x_p}{\sum_{l=1}^q S_l |t_{kl} - \bar{t}_{kl}|^m} \quad (9)$$

典型值 t_{kl} 的更新公式分为 $m = 1$ 和 $m > 1$ 两种情况。当 $m > 1$ 时，典型值更新公式为

$$t_{kl} = \bar{t}_{kl} + \frac{1}{1 + \left(S_l \left\| \left(\frac{1}{S_l} \sum_{p \in R_l} x_p \right) - v_k \right\| / \eta \right)^2} \quad (10)$$

当 $m = 1$ 时，典型值 t_{kl} 的更新公式为

$$t_{kl} = \begin{cases} \bar{t}_{kl} + 1 - \sum_{k=1}^c \bar{t}_{kl}, & l = \arg \min_j \|x_l - v_j\|^2 \\ \bar{t}_{kl}, & \text{其他} \end{cases} \quad (11)$$

基于差分进化超像素的半监督截集式可能性C-均值(desSS-C-PCM)聚类算法步骤如下，流程图如图1所示。

步骤 1 给定聚类数目 c ，超像素个数 k ，模糊因子 m ，惩罚因子 η 。设置最大循环次数 T_{\max} 和算法停止的阈值 ε ；

步骤 2 对彩色图像进行DES超像素处理得到超像素区域 R ，然后计算对应的彩色直方图 $1/S_l \sum_{p \in R_l} x_p$ ；

步骤 3 获得用户对图像的标记信息，构造半监督典型信息 \bar{t}_{kl} ；

步骤 4 随机初始化聚类中心 $v^{(0)}$ ，设置迭代次数 $l = 0$ ；

步骤 5 当 $m > 1$ 时用式(10)更新典型值 t_{kl} ，当 $m = 1$ 时用式(11)更新典型值 t_{kl} ；

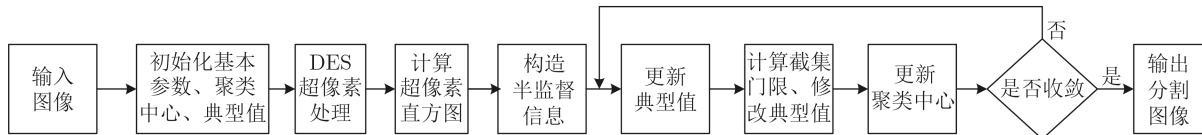


图 1 desSS-C-PCM算法流程图

- 步骤6 采用阴影集^[22,23]计算截集门限 β ;
 步骤7 利用截集门限修正典型值;
 步骤8 用式(9)更新聚类中心;
 步骤9 如果 $\|V^{(l+1)} - V^{(l)}\| < \varepsilon$ 或者迭代次数 $l > T_{\max}$, 则算法结束。否则 $l = l + 1$ 返回步骤5。

6 实验结果分析

为了验证本文所提SS-C-PCM和desSS-C-PCM算法有效性, 本节采用人工合成数据集和Berkeley图像^[24]进行仿真实验, 采用的对比算法有FCM聚类算法、SS-FCM聚类算法、C-PCM聚类算法。实验环境为MATLAB R2014a, 参数设置分别为 $m = 2$, 设置 η 为20, 最大迭代次数 $r_{\max} = 100$, 算法停止的阈值为 $\varepsilon = 0.00001$ 。

6.1 人工合成数据实验

人工合成数据: 3类人工合成数据集 X_{1600} 包含1600个数据点。数据集为高斯分布, 共3类, 分别包含1000, 500, 100个数据点, 它们的均值分别为 $\mathbf{u}_1 = [0, 10]^T$, $\mathbf{u}_2 = [5, 5]^T$, $\mathbf{u}_3 = [10, 10]^T$, 协方差矩阵分别为 $\mathbf{S}_1 = [1, 0; 0, 1]$, $\mathbf{S}_2 = [0.5, 0; 0, 0.5]$, $\mathbf{S}_3 = [0.5, 0; 0, 0.5]$ 。数据集周围分布有100个随机噪声点。实验分别对FCM, C-PCM, SS-FCM和SS-C-PCM算法进行比较, 各算法的实验结果对比如图2所示, 其中“*”代表聚类中心。各算法的中心偏移量和迭代次数如表1所示。

图2显示了各个算法对3类数据集的聚类结果, 可以看出FCM, SS-FCM, C-PCM算法其中一类的

中心发生严重偏移, 而本文所提SS-C-PCM算法利用少量的监督信息来指导聚类过程, 得到的中心只发生了轻微偏移。表1显示了各个算法的中心偏移量。

6.2 Berkeley图像对比实验

对于图3#124084, 其标准分割应为3类: 红色花瓣为第1类, 黄色花蕊为第2类, 除花以外的背景归为第3类。FCM算法在背景处理上出现了较多的错分; C-PCM算法将花瓣和花蕊分割为一类, 没有成功地分割出花蕊, 分割效果较差; SS-FCM算法虽然在FCM的基础上加入了半监督, 提高了该算法的运行效率, 但是分割结果相对于FCM, 并没有得到提升; SS-C-PCM算法相对于C-PCM算法, 分割效果有一定的改善; desSS-C-PCM算法虽然在一些细节上出现了分割错误, 但是能较好地花朵从背景中分割出来。

表2分别给出了FCM, C-PCM, SS-FCM, SS-C-PCM算法以及desSS-C-PCM算法的分割准确率, 从表2可以看出, FCM算法和C-PCM算法的分割准确率相对较低; SS-FCM算法在分割图像#3063, #135069以及#238011时, 分割准确率相对于FCM有比较明显的提升, 对于大部分图像, SS-FCM的分割准确率相对于FCM没有明显的改善; SS-C-PCM算法在C-PCM算法的基础上加入了半监督, 相对于C-PCM算法, SS-C-PCM算法在分割图像#42044时, 分割准确率下降了0.0114, 在

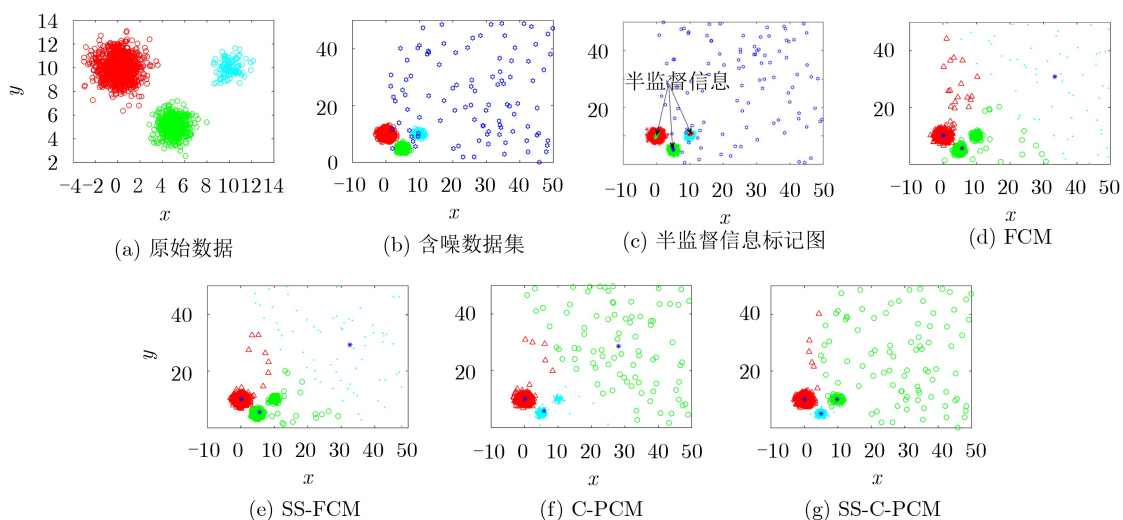


图2 针对数据集 X_{1600} 的聚类结果

表1 针对数据集 X_{1600} 各个算法的中心偏移量以及迭代次数

算法	FCM	SS-FCM	C-PCM	SS-C-PCM
中心偏移量	31.9391	29.8891	23.4551	0.1165
迭代次数	85	60	70	38

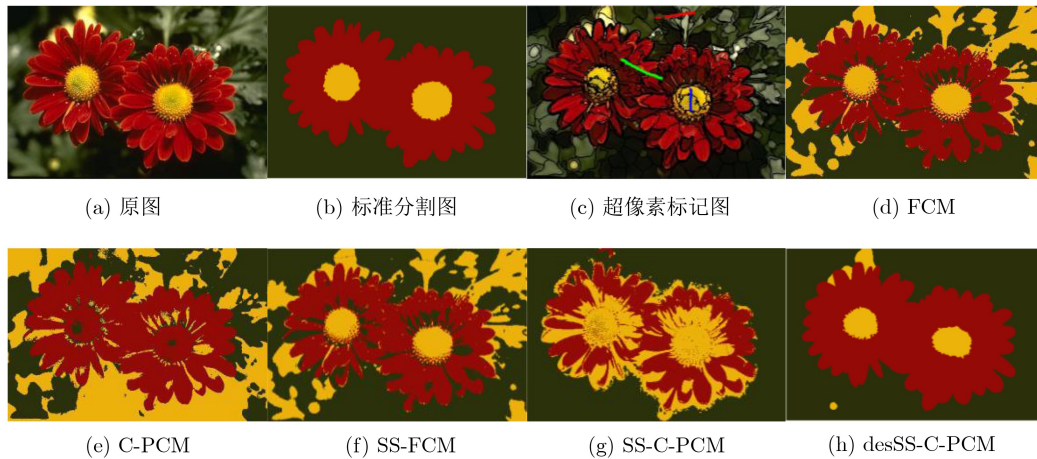


图3 #124084分割结果图

表2 各个算法的分割准确率对比

图像	FCM	C-PCM	SS-FCM	SS-C-PCM	desSS-C-PCM
#3063	0.7292	0.6540	0.9924	0.9931	0.9939
#3096	0.9859	0.6324	0.9860	0.9921	0.9932
#135069	0.7358	0.5624	0.9919	0.9895	0.9905
#118035	0.9342	0.6970	0.9342	0.9457	0.9754
#124084	0.7486	0.6127	0.7487	0.8049	0.9658
#86016	0.8393	0.6126	0.8396	0.9676	0.9931
#161062	0.8846	0.9285	0.8847	0.9452	0.9885
#113044	0.8352	0.6243	0.8354	0.9521	0.9782
#12003	0.7738	0.5811	0.7740	0.9483	0.9740
#238011	0.8123	0.8090	0.9560	0.9282	0.9637
#101027	0.8840	0.6330	0.8846	0.9212	0.9418
#28075	0.4473	0.3987	0.4456	0.5922	0.9842
#24063	0.8919	0.5965	0.8919	0.8976	0.8920
#253036	0.6192	0.5360	0.6195	0.8966	0.9849
#42044	0.7525	0.8181	0.7526	0.8067	0.8884
#299091	0.6960	0.7950	0.6964	0.9941	0.9945
#113016	0.8168	0.5897	0.8185	0.7376	0.9806
#147091	0.9316	0.6690	0.9317	0.9253	0.9413
#67079	0.8274	0.8312	0.8276	0.8936	0.9918

割其余18张图像时，SS-C-PCM算法的分割准确率都有所提高；desSS-C-PCM算法通过DES超像素获取空间信息，并通过半监督信息来指导聚类过程，所以分割准确率有明显的提高。

7 结束语

针对C-PCM算法在处理小目标数据时聚类中心严重偏移的问题，本文将半监督信息引入C-PCM算法，提出半监督截集式可能性C-均值(SS-C-PCM)聚类算法。为了改善C-PCM算法对彩色图像

的分割效果以及运行效率，本文通过DES超像素算法获取空间信息，将此空间信息引入SS-C-PCM算法中，提出基于DES超像素的半监督截集式可能性C均值(desSS-C-PCM)聚类算法，并将其用到彩色图像分割。实验结果表明，SS-C-PCM算法提高了C-PCM算法的运行效率以及分割准确率，desSS-C-PCM算法在SS-C-PCM算法的基础上进一步提高了分割准确率以及运行效率。下一步将研究本文算法在噪声环境下对彩色图像的分割效果。

参考文献

- [1] 章毓晋. 图像分割[M]. 北京: 科学出版社, 2001: 1-5.
ZHANG Yujin. Image Segmentation[M]. Beijing: Science Press, 2001: 1-5.
- [2] 范九伦. 灰度图像阈值分割法[M]. 北京: 科学出版社, 2019: 1-7.
FAN Jiulun. Gray Image Thresholding Segmentation Methods[M]. Beijing: Science Press, 2019: 1-7.
- [3] ZHANG Yingchun, GUO He, CHEN Feng, *et al.* Weighted kernel mapping model with spring simulation based watershed transformation for level set image segmentation[J]. *Neurocomputing*, 2017, 249: 1-18. doi: [10.1016/j.neucom.2017.01.044](https://doi.org/10.1016/j.neucom.2017.01.044).
- [4] 胡炎, 单子力, 高峰. 基于增强指数加权均值比的SAR图像边缘检测算法[J]. 电子与信息学报, 2018, 40(5): 1166-1172. doi: [10.11999/JEIT170806](https://doi.org/10.11999/JEIT170806).
HU Yan, SHAN Zili, and GAO Feng. Edge detection algorithm for SAR image based on enhanced ROEWA[J]. *Journal of Electronics & Information Technology*, 2018, 40(5): 1166-1172. doi: [10.11999/JEIT170806](https://doi.org/10.11999/JEIT170806).
- [5] 赵凤, 孙文静, 刘汉强, 等. 基于近邻搜索花粉优化的直觉模糊聚类图像分割[J]. 电子与信息学报, 2020, 42(4): 1005-1012. doi: [10.11999/JEIT190428](https://doi.org/10.11999/JEIT190428).
ZHAO Feng, SUN Wenjing, LIU Hanqiang, *et al.* Intuitionistic fuzzy clustering image segmentation based on flower pollination optimization with nearest neighbor searching[J]. *Journal of Electronics & Information Technology*, 2020, 42(4): 1005-1012. doi: [10.11999/JEIT190428](https://doi.org/10.11999/JEIT190428).
- [6] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms[J]. *Advanced Applications in Pattern Recognition*, 1981, 22(1171): 203-239. doi: [10.1007/978-1-4757-0450-1](https://doi.org/10.1007/978-1-4757-0450-1).
- [7] KRISHNAPURAM R and KELLER J M. A possibilistic approach to clustering[J]. *IEEE Transactions on Fuzzy Systems*, 1993, 1(2): 98-110. doi: [10.1109/91.227387](https://doi.org/10.1109/91.227387).
- [8] BARNI M, CAPPELLINI V, and MECOCCI A. Comments on "A possibilistic approach to clustering"[J]. *IEEE Transactions on Fuzzy Systems*, 1996, 4(3): 393-396. doi: [10.1109/91.531780](https://doi.org/10.1109/91.531780).
- [9] KRISHNAPURAM R and KELLER J M. The possibilistic C-means algorithm: Insights and recommendations[J]. *IEEE Transactions on Fuzzy Systems*, 1996, 4(3): 385-392. doi: [10.1109/91.531779](https://doi.org/10.1109/91.531779).
- [10] PAL N R, PAL K, KELLER J M, *et al.* A possibilistic fuzzy C-means clustering algorithm[J]. *IEEE Transactions on Fuzzy Systems*, 2005, 13(4): 517-530. doi: [10.1109/TFUZZ.2004.840099](https://doi.org/10.1109/TFUZZ.2004.840099).
- [11] MEMON K H, MEMON S, QURESHI M A, *et al.* Kernel possibilistic fuzzy c-means clustering with local information for image segmentation[J]. *International Journal of Fuzzy Systems*, 2019, 21(1): 321-332. doi: [10.1007/s40815-018-0537-9](https://doi.org/10.1007/s40815-018-0537-9).
- [12] YU Haiyan and FAN Jiulun. Cutset-type possibilistic C-means clustering algorithm[J]. *Applied Soft Computing*, 2018, 64: 401-422. doi: [10.1016/j.asoc.2017.12.024](https://doi.org/10.1016/j.asoc.2017.12.024).
- [13] YASUNORI E, YUKIHIRO H, MAKITO Y, *et al.* On semi-supervised fuzzy c-means clustering[C]. 2009 IEEE International Conference on Fuzzy Systems, Jeju Island, South Korea, 2009: 1119-1124. doi: [10.1109/FUZZY.2009.5277177](https://doi.org/10.1109/FUZZY.2009.5277177).
- [14] YIN Xuesong, SHU Ting, and HUANG Qi. Semi-supervised fuzzy clustering with metric learning and entropy regularization[J]. *Knowledge-Based Systems*, 2012, 35: 304-311. doi: [10.1016/j.knosys.2012.05.016](https://doi.org/10.1016/j.knosys.2012.05.016).
- [15] 赵凤, 张咪咪, 刘汉强. 区域信息驱动的多目标进化半监督模糊聚类图像分割算法[J]. 电子与信息学报, 2019, 41(5): 1106-1113. doi: [10.12000/JEIT180605](https://doi.org/10.12000/JEIT180605).
ZHAO Feng, ZHANG Mimi, and LIU Hanqiang. Multi-objective evolutionary semi-supervised fuzzy clustering image segmentation motivated by region information[J]. *Journal of Electronics & Information Technology*, 2019, 41(5): 1106-1113. doi: [10.12000/JEIT180605](https://doi.org/10.12000/JEIT180605).
- [16] REN Xiaofeng and MALIK J. Learning a classification model for segmentation[C]. The Ninth IEEE International Conference on Computer Vision, Nice, France, 2003: 10-17. doi: [10.1109/ICCV.2003.1238308](https://doi.org/10.1109/ICCV.2003.1238308).
- [17] SHI Jianbo and MALIK J. Normalized cuts and image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905. doi: [10.1109/34.868688](https://doi.org/10.1109/34.868688).
- [18] COMANICIU D and MEER P. Mean shift: A robust approach toward feature space analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603-619. doi: [10.1109/34.1000236](https://doi.org/10.1109/34.1000236).
- [19] GONG Yuejiao and ZHOU Yicong. Differential evolutionary superpixel segmentation[J]. *IEEE Transactions on Image Processing*, 2018, 27(3): 1390-1404. doi: [10.1109/TIP.2017.2778569](https://doi.org/10.1109/TIP.2017.2778569).
- [20] ACHANTA R, SHAJI A, SMITH K, *et al.* SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282. doi: [10.1109/TPAMI.2012.120](https://doi.org/10.1109/TPAMI.2012.120).
- [21] LEI Tao, JIA Xiaohong, ZHANG Yanning, *et al.* Superpixel-based fast fuzzy c-means clustering for color image segmentation[J]. *IEEE Transactions on Fuzzy Systems*, 2019, 27(9): 1753-1766. doi: [10.1109/TFUZZ.2018.2889018](https://doi.org/10.1109/TFUZZ.2018.2889018).

- [22] PEDRYCZ W. Shadowed sets: Representing and processing fuzzy sets[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 1998, 28(1): 103–109. doi: [10.1109/3477.658584](https://doi.org/10.1109/3477.658584).
- [23] 雒僖, 范九伦, 于海燕, 等. 基于阴影集的截集式可能性C-均值聚类截集门限的选取[J]. *计算机科学*, 2019, 46(8): 249–254. doi: [10.11896/j.issn.1002-137X.2019.08.041](https://doi.org/10.11896/j.issn.1002-137X.2019.08.041).
LUO Xi, FAN Jiulun, YU Haiyan, *et al.* Selection of cutset threshold for cutset-type possibilistic C-means clustering based on shadowed set[J]. *Computer Science*, 2019, 46(8): 249–254. doi: [10.11896/j.issn.1002-137X.2019.08.041](https://doi.org/10.11896/j.issn.1002-137X.2019.08.041).
- [24] ARBELÁEZ P, MAIRE M, FOWLKES C, *et al.* Contour detection and hierarchical image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(5): 898–916. doi: [10.1109/TPAMI.2010.161](https://doi.org/10.1109/TPAMI.2010.161).
- 范九伦: 男, 1964年生, 教授, 研究方向为模式识别与图像处理、模糊信息处理、图像安全技术。
高梦飞: 男, 1994年生, 硕士生, 研究方向为图像处理。
于海燕: 女, 1982年生, 副教授, 研究方向为模式识别与图像处理。
陈斌斌: 男, 1994年生, 硕士生, 研究方向为图像处理。

责任编辑: 余蓉