

烟花算法优化的软子空间 MR 图像聚类算法*

范虹¹, 侯存存¹, 朱艳春², 姚若侠¹



¹(陕西师范大学 计算机科学学院, 陕西 西安 710062)

²(中国科学院 深圳先进技术研究院 生物医学与健康工程研究所, 广东 深圳 518055)

通讯作者: 范虹, E-mail: fanhong@snnu.edu.cn

摘要: 现有的软子空间聚类算法在分割 MR 图像时易受随机噪声的影响, 而且算法因依赖于初始聚类中心的选择而容易陷入局部最优, 导致分割效果不理想. 针对这一问题, 提出一种基于烟花算法的软子空间 MR 图像聚类算法. 算法首先设计一个结合约束与噪声聚类的目标函数, 弥补现有算法对噪声数据敏感的缺陷, 并提出一种隶属度计算方法, 快速、准确地寻找簇类所在子空间; 然后, 在聚类过程中引入自适应烟花算法, 有效地平衡局部与全局搜索, 弥补现有算法容易陷入局部最优的不足. EWKM, FWKM, FSC, LAC 算法在 UCI 数据集、人工合成图像、Berkeley 图像数据集以及临床乳腺 MR 图像、脑部 MR 图像上的聚类结果表明, 所提出的算法不仅在 UCI 数据集上能够取得较好的结果, 而且对图像聚类也具有较好的抗噪性能, 尤其是对 MR 图像的聚类具有较高的精度和鲁棒性, 能够较为有效地实现 MR 图像的分割.

关键词: 烟花算法; 软子空间聚类; 噪声聚类; MR 图像; 图像分割

中图法分类号: TP18

中文引用格式: 范虹, 侯存存, 朱艳春, 姚若侠. 烟花算法优化的软子空间 MR 图像聚类算法. 软件学报, 2017, 28(11): 3080-3093. <http://www.jos.org.cn/1000-9825/5335.htm>

英文引用格式: Fan H, Hou CC, Zhu YC, Rao RX. Soft subspace algorithm for MR image clustering based on fireworks optimization algorithm. Ruan Jian Xue Bao/Journal of Software, 2017, 28(11): 3080-3093 (in Chinese). <http://www.jos.org.cn/1000-9825/5335.htm>

Soft Subspace Algorithm for MR Image Clustering Based on Fireworks Optimization Algorithm

FAN Hong¹, HOU Cun-Cun¹, ZHU Yan-Chun², RAO Ruo-Xia¹

¹(School of Computer Science, Shaanxi Normal University, Xi'an 710062, China)

²(Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, The Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract: The existing soft subspace clustering algorithm is susceptible to random noise when MR images are segmented, and it is easy to fall into local optimum due to the choice of the initial clustering centers, which leads to unsatisfactory segmentation results. To solve these problems, this paper proposes a soft subspace algorithm for MR image clustering based on fireworks algorithm. Firstly, a new objective function with boundary constraints and noise clustering is designed to overcome the shortcomings of the existing algorithms that are sensitive to noise data. Next, a new method of calculating affiliation degree is proposed to find the subspace where the cluster is located quickly and accurately. Then, adaptive fireworks algorithm is introduced in the clustering process to effectively balance the local and global search, overcoming the disadvantage of falling into local optimum in the existing algorithms. Comparing with EWKM, FWKM,

* 基金项目: 国家自然科学基金(11471004); 陕西省自然科学基金(2014JM2-6115); 陕西省科学技术研究发展计划(2012K06-36)
Foundation item: National Natural Science Foundation of China (11471004); Shaanxi Provincial Natural Science Foundation (2014JM2-6115); Shaanxi Province Science and Technology Research and Development Plan (2012K06-36)

本文由复杂环境下的机器学习研究专刊特约编辑张道强教授推荐.

收稿时间: 2017-01-24; 修改时间: 2017-04-11; 采用时间: 2017-06-16

FSC and LAC algorithms, experiments are conducted on UCI datasets, synthetic images, Berkeley image datasets, as well as clinical breast MR images and brain MR images. The results demonstrate that the proposed algorithm not only can get better results on UCI datasets, but also has better anti-noise performance. Especially for MR images, high clustering precision and robustness can be obtained, and effective MR images segmentation can be achieved.

Key words: fireworks algorithm; soft subspace clustering; noise clustering; MR image; image segmentation

磁共振成像(magnetic resonance imaging,简称 MRI)是断层扫描的一种方法.MR 图像所显示的解剖结构逼真,能使病变组织和正常组织清晰可见,有很高的软组织对比分辨率,近年来在临床上得到了普及.然而在图像采集过程中,由于受到外界不确定因素的影响,图像中始终存在随机噪声,且呈现出灰度不均匀、边界模糊等一系列问题,给图像分割带来一定的难度.为了分割出病变区域,进一步提取病灶特征,近年来,越来越多的学者致力于有效分割 MR 图像的方法研究.

目前,用于医学图像分割的方法主要有区域法、边缘检测法、结合区域与边界技术、神经网络方法、模糊集技术等,其中,基于模糊集技术^[1]的不确定性,可以较好地克服 MR 图像数据获取时的内在不确定性(包括时间、空间和分辨率),被广泛应用于 MR 图像分割中.模糊聚类是一种典型模糊集技术的应用,其中最具代表性的算法是模糊 C 均值算法(FCM 算法)^[2-4].随后,围绕 FCM 算法展开了诸多研究:Ahmed 等人在 FCM 算法的目标函数中加入邻域像素信息,提出了 FCMS 算法^[5];针对 FCMS 效率低下的问题,陈松灿等人运用均值滤波和中值滤波计算像素邻域信息,提出了 FCMS1 和 FCMS2 算法^[6];由于 FCMS,FCMS1 和 FCMS2 算法中参数较多、不易确定,Krinidis 等人提出了无需任何参数的 FLICM 算法^[7]对此加以改进;文献^[8]还提出了混合 FCM 与水平集方法分割 MR 图像,算法中,水平集方法将 FCM 的聚类结果作为初始轮廓演化,取得了更高的分割精确度.然而,上述各方法仅试图在全空间中寻找簇类,事实上,某些簇类可能仅与部分属性相关,存在于一些子空间中.为此,Agrawal 等人于 1998 年提出了子空间聚类的概念^[9].子空间聚类一经提出便受到众多学者的青睐,逐渐演变为硬子空间聚类与软子空间聚类.软子空间聚类反映了属性与簇的相关程度,同时也反映了各属性在相关性上的差异,应用更加灵活,可扩展性更高,但受初始聚类中心的影响极易陷入局部最优,一段时间内,很多学者都在寻觅一种可以解决这一问题的方法.烟花算法自提出后,由于其可以很好地平衡局部与全局搜索,从而得到最优解,所以学者们均尝试将其运用到各自的领域中.

针对现有软子空间聚类算法分割 MR 图像时易陷入局部最优,且对噪声数据敏感的问题,本文在改进软子空间聚类算法的目标函数与隶属度计算方法的同时,将自适应烟花算法引入到聚类搜索过程中以获取最优解,提出了一种基于烟花算法的软子空间 MR 图像聚类算法.

1 相关工作

1.1 软子空间聚类算法

子空间聚类的概念自从 1998 年由 Agrawal 等人提出之后,因其可以有效地减少数据冗余与不相关属性对聚类过程的干扰等优点而得到广泛应用,软子空间聚类属于其中一个分支.软子空间聚类是传统聚类向高维聚类的扩展,是在相关维度子空间按照属性对簇类形成的贡献大小赋予其一个 $[0,1]$ 的权值寻找簇类,而非整个特征空间.按照特征加权方式的异同,可将软子空间聚类分为模糊加权软子空间聚类与熵加权软子空间聚类.

为了后续方便,现做如下说明.

- $Z=[z_{ik}]_{C \times D}$ 为聚类中心矩阵,记录每个簇类的中心位置;
- $W=[w_{ik}]_{C \times D}$ 为权值矩阵,记录每个簇的属性权值;
- $U=[u_{ij}]_{C \times N}$ 为划分矩阵,记录每个样本对各个簇的隶属度.

其中, C,N,D 分别是聚类数、样本总数、样本维度.

属性加权算法(attribute weighting algorithm,简称 AWA)^[10]属于最早提出的模糊加权子空间聚类算法,其目标函数如公式(1)所示.

$$\left. \begin{aligned} J_{AWA} &= \sum_{i=1}^C \sum_{j=1}^N u_{ij} \sum_{k=1}^D w_{ik}^r (x_{ik} - v_{ik})^2 = \sum_{i=1}^C \sum_{k=1}^D w_{ik}^r \sum_{j=1}^N u_{ij} (x_{ik} - v_{ik})^2 \\ \text{s.t. } u_{ij} &\in \{0,1\}, \sum_{i=1}^C u_{ij} = 1, 0 < w_{ij} < 1, \sum_{k=1}^D w_{ik} = 1 \end{aligned} \right\} \quad (1)$$

算法通过结合模糊概念,给每个维度赋予一个[0,1]的模糊权值,寻找隐藏在子空间中的簇类.为了规避公式(1)中距离计算出现0的情况,Jing等人^[11]在目标函数中增加一个较小的正数 σ ,提出了模糊加权K-means算法(FWK M算法). σ 的计算如公式(2)所示.

$$\sigma = \frac{\sum_{j=1}^N \sum_{k=1}^D (x_{jk} - o_k)^2}{N \times D}, o_k = \frac{\sum_{j=1}^N x_{jk}}{N} \quad (2)$$

Deng等人^[12]率先意识到类间分散度信息的重要性,结合类内紧凑度与类间分散度信息提出了ESSC算法.该算法性能虽好,但用户所需输入参数较多,在实际运用中,对于不同的数据集需要输入不同的参数,参数训练可能耗费大量时间与精力,无形中增加了算法的复杂度.陈黎飞等人^[13]提出的自适应的软子空间聚类算法根据数据集及划分信息动态地规划算法参数.Zhu等人^[14]在前人研究的基础上,提出了竞争合并模糊软子空间聚类算法以及基于竞争合并的熵加权软子空间聚类算法,无需事先确定聚类个数,在很大程度上简化了算法.Frey等人^[15]提出的近邻传播聚类算法将数据看作是一个由节点与边构成的网络,节点之间通过相互传递近邻信息聚类,直到网络相似度目标函数最大化或者算法已经趋于稳定.算法通过一个参数Preference间接影响最终聚类个数的大小,Preferences越大,聚类个数越多;反之则越少.在AP聚类算法提出后,Yu等人^[16]利用AP算法精简算法去除噪声属性.Gan等人^[17]进一步引入属性加权思想对算法进行改进,提出了SAP算法.该算法迭代速度快,且准确度高.

与模糊加权软子空间聚类算法不同,熵加权软子空间聚类算法通过熵值控制权值分布,其中,典型算法是EWKM^[18].EWKM目标函数如下:

$$\left. \begin{aligned} J_{EWKM} &= \sum_{i=1}^C \sum_{j=1}^N u_{ij} \sum_{k=1}^D w_{ik} (x_{jk} - z_{ik})^2 + \lambda \sum_{i=1}^C \sum_{k=1}^D w_{ik} \ln w_{ik} \\ \text{s.t. } u_{ij} &\in \{0,1\}, \sum_{i=1}^C u_{ij} = 1, 0 < w_{ij} < 1, \sum_{k=1}^D w_{ik} = 1 \end{aligned} \right\} \quad (3)$$

熵值的引入,可以更好地控制维度权值的分布,算法的性能得到有效提升.大多数算法加权都是针对数据的维度,忽略了特征之间的相似性.文献[19]将数据特征按相似性分为特征组,每个特征组与一个组特征权值关联,组内的个体特征与一个个体特征权值关联,聚类过程迭代地更新这两种权值来发现每个簇中相对重要的特征组和个体特征进行聚类,特征组有效地平滑噪声数据与缺失数据的敏感度,消除由于种群差异造成的不平衡现象.其目标函数如公式(4)所示.

$$\left. \begin{aligned} J_{FG-K-means} &= \sum_{l=1}^K \left[\sum_{i=1}^N \sum_{t=1}^T \sum_{j \in G_t} u_{it} w_{lj} v_{lj} (x_{jk} - z_{ik})^2 + \lambda \sum_{t=1}^T w_{lt} \log(w_{lt}) + \eta \sum_{j=1}^D v_{lj} \log(v_{lj}) \right] \\ \text{s.t. } \sum_{l=1}^K u_{il} &= 1, u_{il} \in \{0,1\}, \sum_{t=1}^T w_{lt} = 1, w_{lt} > 0, \sum_{j \in G_t} v_{lj} = 1, v_{lj} > 0 \end{aligned} \right\} \quad (4)$$

其中, λ 和 η 是两个正整数,调节权值的分布.该算法需要在聚类开始前确定特征组,在实际应用中,并不能事先知晓数据集的属性分类信息,为了更好地实现特征组聚类数据,Gan等人^[20]重新设计目标函数,提出了AFG-K-means算法,目标函数包含了通过个体特征进行特征分组的部分,实现特征的自动分组.该算法的精度以及在参数的选取上性能更好,与其他一些算法相比,AFG-K-means算法优于FSC^[21],LAC^[22]和EWKM算法.其目标函数如公式(5)所示.

$$J_{AFG-K-means} = \sum_{l=1}^K \sum_{i=1}^N u_{il} \sum_{j=1}^D w_{lj}^2 (x_{ij} - z_{lj})^2 + \varepsilon_1 \sum_{l=1}^K \sum_{j=1}^D w_{lj}^2 + \beta \left(\sum_{t=1}^T \sum_{j=1}^D g_{it} \sum_{l=1}^K \gamma_{li}^2 (w_{lj} - v_{li})^2 + \varepsilon_2 \sum_{l=1}^K \lambda_{li}^2 \right) \quad (5)$$

针对高维数据流聚类分析存在的问题,朱林等人^[23]提出了EWSSC实现流数据的有效分析方法.为了提高聚类性能,在Vesterström等人^[24]证明经过足够多代的演变,GA的全局收敛性与PSO的快速收敛性可以混合后,Lin等人^[25]运用GA与PSO的可混合性提出了GA-PSO混合聚类算法.该算法采用GA作为全局搜索策略,PSO作为局部搜索策略对子空间进行充分优化,聚类效果更佳.不足之处在于,该算法的实现需要输入诸多参数,无形中为算法增加了一定的复杂度.Yu等人提出的基于网格的聚类算法GCA^[26]考虑实时低维子空间来降

低算法的运行时间.Lu 等人^[27]通过结合粒子群优化算法重新设计目标函数,提出了 PSOvw 算法.算法将对权值的约束放松为界约束.为了提高算法的性能,Yu 等人^[28-30]努力将模糊划分与硬划分有机结合,提高算法精度.毕志升等人^[31]借鉴 PSOvw 引入复合差分演化算法作为搜索策略更新权值矩阵 W ,提出了算法 DESC.该算法可以迅速、准确地找到簇类所在子空间,优势明显.

1.2 噪声聚类

大多数聚类算法都未提出处理界外值的机制,在一个簇中,隶属度仅仅取决于点到簇中心的距离.对于界外值来说,点到所有簇中心的距离均很大,其隶属于任何簇的隶属度均很小,这给聚类带来了一定的难度.为了解决聚类算法对噪声敏感的问题,1991 年,Dave^[32]在聚类过程中引入噪声监测机制,提出了噪声聚类算法.该算法额外增加一个簇,这个簇包含所有远离任何簇类的目标,剩余的簇类则认为是数据集中无界外值的好簇.

噪声簇中, x_j 的隶属度 u_{*j} 定义为

$$u_{*j} = 1 - \sum_{i=1}^C u_{ij} \tag{6}$$

将对好簇的隶属度等式约束放松为 $\sum_{i=1}^C u_{ij} < 1$,界外值在好簇中的隶属度任意小,噪声聚类的目标函数为

$$J_{NC} = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m d^2(x_j - v_i) + \sum_{j=1}^N \delta^2 \left(1 - \sum_{i=1}^C u_{ij}\right)^m \tag{7}$$

其中, d 代表数据点 x_j 到簇中心 v_i 的距离, δ 是数据点到在噪声簇的距离.若 δ 较小,则大多数的点将被认为是噪声数据而错分;反之,若 δ 较大,则几乎没有数据被错分到噪声簇中. δ 的计算公式如下:

$$\delta^2 = \lambda \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^N d^2(x_i - v_i) \tag{8}$$

λ 是确定 δ 时所用的乘子($\lambda=0.1$). δ 在每次迭代中更新.于是,隶属度 U 的更新如下:

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{d^2(x_j - v_i)}{d^2(x_j - v_k)} \right)^{1/(m-1)} + \left(\frac{d^2(x_j - v_i)}{\delta^2} \right)^{1/(m-1)}} \tag{9}$$

当一个点属于噪声数据时, U 分母中的第 2 项变得很大,在好簇中获得一个很小的隶属度.

1.3 烟花算法

受到烟花爆炸在近邻产生火花的启发,2010 年,Tan 等人提出了烟花算法(FWA 算法)^[33].该算法有效地平衡了局部搜索与全局搜索.

算法初始根据求解问题的实际情况,在解空间中放置 N 个烟花,评估每个烟花的适应度值,计算烟花的爆炸半径与爆炸火花数.对于烟花 x_i ,其爆炸半径 A_i 与爆炸火花数 S_i 的计算公式如下:

$$A_i = \hat{A} \times \frac{f(x_i) - y_{\min} + \varepsilon}{\sum_{i=1}^N (f(x_i) - y_{\min}) + \varepsilon} \tag{10}$$

$$S_i = M \times \frac{y_{\max} - f(x_i) + \varepsilon}{\sum_{i=1}^N (y_{\max} - f(x_i)) + \varepsilon} \tag{11}$$

\hat{A}, M 均为常数,用来限制最大的爆炸幅度与产生的火花总数.一般适应度值小的烟花在较小范围内产生较多的火花,局部搜索能力强;反之,适应度值大的烟花在较大范围内产生较少的烟花,具有较强的全局搜索能力.为了增强种群的多样性,算法随机选定一个烟花,对于选择得到的烟花,再随机选择一定数量的维度进行变异操作.对于爆炸操作与变异操作过程中超出可行域的火花,按照一定的策略映射到可行域中.为使种群中的优秀信息传递到下一代,爆炸操作与变异操作结束后,算法在初始烟花以及所有产生的火花中选择 N 个个体形成下一代烟花,直至达到终止条件.烟花算法求解问题的能力强大,受到广泛关注,因此,人们就算法爆炸半径的计算、映射策略等不足之处改进研究.Pei 等人^[34]使用进化算法加速策略加速烟花算法,并估计空间搜索形状,提出了一种基于适应度函数值的搜索空间估计的加速型烟花算法(AcFWA).Zheng 等人^[35]对基本烟花算法的爆炸算子、变异算子、选择策略和映射规则等进行了细致分析,并提出了增强型烟花算法(EFWA).随后,Zheng 等人^[36]和 Li

等人^[37]分别提出了一种自适应半径策略——动态搜索烟花算法(dynFWA)和自适应烟花算法(AFWA).在dynFWA中,烟花爆炸半径的变化过程依据烟花种群产生的火花适应度值是否发生改进(即优化得到适应度值更优的火花)来决定.在AFWA中,烟花的爆炸半径的确定根据当前种群适应度值最优的个体和一个特定个体之间的距离计算.这两种方法由于采用了自适应调整爆炸半径的机制,都极大地提升了增强型烟花算法的性能.算法1给出了AFWA算法的实现过程,其中,UB和LB分别代表搜索空间的上界和下界, A^* 代表最优烟花的半径.

算法1. 自适应烟花算法(AFWA).

1. 从搜索空间随机生成 m 个烟花
2. 评估其适应度值
3. $A^* \leftarrow UB-LB$
4. Repeat
5. 根据公式(11)计算 S_i
6. 根据公式(10)计算 A_i (除了 A^*)
7. 生成 S_i 个爆炸火花
8. 生成高斯火花
9. 评估所有火花的适应度值
10. 计算 A^*
11. 保留最优个体作为一个烟花
12. 在剩下的个体中随机选择其余 $N-1$ 个烟花
13. 直到终止条件达到
14. 返回最优烟花及其适应度值

2 基于烟花算法优化的软子空间聚类算法(FWASSC)

借鉴PSOVW,DESC和噪声聚类(noise clustering),本文设计了新的目标函数,弥补现有算法对噪声数据敏感的缺陷,然后,结合硬聚类、模糊聚类与噪声聚类重新设计隶属度计算方法,旨在快速、准确地寻找簇类所在子空间;最后,在聚类过程中引入AFWA,平衡局部搜索与全局搜索寻找最优子空间,实现聚类.

2.1 目标函数与隶属度计算方法设计

PSOVW与DESC目标函数将对权值的等式约束放松为界约束,在算法寻优过程中,目标函数与搜索过程只需处理界约束而无需处理复杂的等式约束,简化了算法的复杂度且不影响最终结果.本文借鉴这种思想,将对权值的等式约束放松为界约束;同时,为了实现算法对噪声数据的鲁棒性,引入噪声聚类.结合上述算法思想,本文设计了新目标函数,如公式(12)所示.

$$J_{FWASSC} = \left. \begin{aligned} & \sum_{i=1}^C \sum_{k=1}^D \sum_{j=1}^N u_{ij}^m \left(\frac{w_{ik}}{\sum_{l=1}^D w_{il}} \right)^\beta (x_{jk} - z_{ik})^2 + \sum_{j=1}^N \delta^2 (1 - \sum_{i=1}^C u_{ij})^m \\ & \text{s.t } \sum_{i=1}^C u_{ij} < 1 \text{ and } 0 \leq w_{ik} \leq 1 \end{aligned} \right\} \quad (12)$$

目标函数融合了模糊聚类与噪声聚类,保留了PSOVW与DESC将 W 的等式约束放松为界约束的优点.硬聚类可以在算法初始快速找到簇类所在子空间,模糊聚类进一步在较小的子空间内细化聚类.为了实现快速、准确地识别簇类,本文的隶属度计算方法结合了硬聚类、模糊聚类与噪声聚类,计算方法如公式(13)所示.

$$u_{ij} = \alpha(t) u_{fuzzy_ij} + (1 - \alpha(t)) u_{crisp_ij} \quad (13)$$

$$u_{crisp_ij} = \begin{cases} 1, & i = \arg \min_{q=1, \dots, c} d_{qj} \\ 0, & \text{其他} \end{cases} \quad (14)$$

$$u_{fuzzy_ij} = \frac{(d_{ij})^{-1/(m-1)}}{\sum_{i=1}^C (d_{ij})^{-1/(m-1)} + (\delta^2)^{-1/(m-1)}} \quad (15)$$

$$d_{ij} = \sum_{k=1}^D \left(\frac{W_{ik}}{\sum_{k'}^D W_{ik'}} \right)^\beta (x_{jk} - z_{ik})^2 \quad (16)$$

$$\alpha(t) = \left(\frac{t}{MaxIter} \right)^\eta \quad (17)$$

$$z_{ik} = \frac{\sum_{j=1}^N u_{ij}^m x_{jk}}{\sum_{j=1}^N u_{ij}^m} \quad (18)$$

其中, β 是用户输入参数, 控制算法对权值矩阵 W 变化的敏感度; t 是当前迭代次数; $MaxIter$ 是最大迭代次数; $0 < \eta < 1$ 同样是一个控制参数, 用于控制算法从硬聚类过度到模糊聚类的速度. 模糊隶属度计算公式(15)中, 分母第 2 项为噪声聚类部分, 公式中加入噪声聚类, 平滑算法对噪声数据的敏感度. 当一个样本属于噪声时, u_{fuzzy_ij} 的分母变得很大, 使整个模糊隶属度值变得极小, 最终致使隶属度 u_{ij} 变得很小, 隶属于噪声簇. 同时, 算法在迭代寻找一组最优权值矩阵 W 时, 由于烟花算法产生数据的随机性, 使每次产生结果均是本次循环里的最优值而不是所需的最优值; 加入噪声聚类后, 模糊隶属度分母的值变大, 可以有效地减弱 W 对隶属度 U 的影响. 算法初始, 当 $t=0$ 时, 有 $\alpha(t)=0$, $u_{ij}=u_{crisp_ij}$, 是一个硬隶属度; 随着算法迭代 η 逐渐增大, 隶属度 u_{ij} 逐渐向模糊隶属度过渡, 最终成为一个完全模糊隶属度. 算法初始阶段偏向于硬聚类, 可以快速得到较优簇及其所在的子空间; 随着算法的逐步迭代, 逐渐过渡到模糊聚类, 在较小的子空间中进一步得到更好的聚类结果.

2.2 FWASSC 算法及其实现流程

AFWA 在烟花种群中选择最优烟花与一个特定个体之间的距离作为爆炸半径, 并根据搜索结果自适应地进行调节. 但 AFWA 开始并没有评估函数的任何信息, 必须进行全局搜索寻找较优局部区域, 仅当烟花进入某个较优局部区域后才快速搜索, 最后进入微调阶段, 算法进行更精确的搜索. 由于 AFWA 能够同时兼顾全局与局部搜索, 且搜索能力较优, 所以本文将 AFWA 引入软子空间聚类中求解.

运用 AFWA 优化软子空间聚类时, 和 DESC, PSO, V 一样, 让 AFWA 中的一个个体对应于软子空间聚类问题的一个可行解, 并将权值矩阵 W 作为求解问题的解, 算法迭代搜寻一组最优解使目标函数达到最小值. 记 $P_g = \{x_{1,g}, x_{2,g}, \dots, x_{M,g}\}$ 为第 g 代种群, 个体 $x_{i,g} = (x_{i,g,1}, \dots, x_{i,g,d}, \dots, x_{i,g,C \times D})$ 是一个长度为 $C \times D$ 的向量, 对应于权值矩阵 W , 其中, 每 D 维对应于一个簇中的 D 个属性权值. 算法同时为每个个体保留其对应的聚类中心矩阵及划分矩阵 U .

算法伊始, 在数据集中随机选取 C 个样本作为初始聚类中心, 初始化权值矩阵 $W = [w_{ik}]_{C \times D} = 1/D$, 并用权值矩阵初始化第 1 代种群 P_1 , 运用公式(8)和公式(13)计算 δ^2 、划分矩阵 U , 将数据样本划分到最近的簇类中; 然后更新聚类中心 Z , 获得新的种群 P_g , 新的种群通过烟花算法演化而来, 根据公式(1)计算目标函数值, 判断是否满足终止条件, 若不满足, 则返回计算划分矩阵 U . 重复上述步骤, 直至算法达到终止条件. 算法 2 给出 FWASSC 实验步骤, 其中, 种群 P_g 的演化过程如下.

种群 P_g 的演化过程.

1. 根据公式(12)计算每个个体 $x_{i,g}$ 的目标函数值, 作为个体适应度值
2. 按照公式(10)与公式(11)计算烟花的爆炸半径与爆炸火花数
3. 按照 AFWA 计算最优烟花的爆炸半径
4. 高斯变异
5. 选择最优烟花添加到下一代种群中
6. 随机选取 $N-1$ 个火花与最优火花形成下一代种群

算法 2. FWASSC 算法.

1. 初始化 FWASSC 参数: 确定聚类个数 C 以及各个参数 $m, \beta, \gamma, \eta, \lambda$, 随机选定 C 个样本作为初始聚类中心

V ;初始化去权值矩阵 $W=[w_{ik}]_{C \times D}=[1/D]$,确定种群大小 M 、最大评价次数 $maxIter$.

2. For
3. $g=0$
4. 根据公式(8)计算 σ^2
5. 根据公式(13)计算隶属度 U
6. 根据公式(18)更新聚类中心 Z
7. 生成种群 P_0, DO
8. Repeat
9. 对于每一个个体计算适应度值(目标函数值)
10. 根据适应度值,运用公式(10)和公式(11)计算烟花的爆炸半径以及爆炸火花数
11. 根据算法 1 进行种群演化
12. $g=g+1$
13. 直至满足终止条件
14. End for

3 实验结果与分析

本节采用 UCI 数据集与图像数据进行实验,并与 EWKM,FWKM,FSC,LAC 算法结果进行对比,验证本文算法性能.由于 FWASSC 是基于群智能的算法,为了定性比较 FWASSC 算法与对比算法的性能,本文采用最大迭代次数($MaxIter=500$)作为所有算法的迭代终止条件,种群规模 $M=5$.EWKM,FWKM,FSC,LAC 参数设置见表 1.实验平台为 Windows10,硬件为 Intel(R) Core(TM) i5-6600 3.30GHz CPU,8G RAM,所有算法在 MATLAB 2014b 环境下实现.

Table 1 Parameter settings

表 1 参数设置

算法	参数设置
FWASSC	$m=2, \beta=2, \eta=0.2$
EWKM	$\gamma=2$
FWKM	$\beta=2$
FSC	$\beta=2$
LAC	$h=2$

3.1 UCI数据集结果与分析

本节采用来自 UCI 的 5 个数据集对算法性能进行验证,UCI 数据集详细信息参见表 2.选用公式(19)的 Rand Index(RI)^[38]与公式(20)的 Normalized Mutual Information(NMI)^[39]性能指标对聚类结果进行评价.

$$RI = \frac{f_{00} + f_{11}}{N(N-1)/2} \quad (19)$$

$$NMI = \frac{\sum_{i=1}^K \sum_{j=1}^C n_{ij} \log \frac{N \times n_{ij}}{n_i \times n_j}}{\sqrt{\left(\sum_{i=1}^K n_i \log \frac{n_i}{N} \right) \left(\sum_{j=1}^C n_j \log \frac{n_j}{N} \right)}} \quad (20)$$

其中, K 是类别数, C 是聚类个数, N 是样本总数, f_{00} 是不属于同一簇类并被分配到不同簇的样本点对的数量, f_{11} 是属于同一类并被分配到同一簇的样本点对的数量. n_i 是属于类 i 的样本数, n_j 是属于类 j 的样本数, n_{ij} 是属于类 i 并被分配到类 j 的样本点对的数量.两个指标越大,表明聚类结果越好.

每种算法独立运行 20 次,用 RI 与 NMI 的平均值表示聚类结果.实验结果见表 3、表 4.对比本文算法与其他几种算法可以看到,本文算法具有更加良好的性能.例如在数据集 Iris 上的实验结果,本文算法的 RI 与 NMI

均值远大于其他几种算法.在大多数情况下,相似的结论也可以从其他几个数据集中得到.因此,本文算法性能优良,尤其在维度较多时,本文算法优点更加明显.

Table 2 Dataset information

表 2 数据集信息

数据集	样本数	维数	类别数
Iris	150	4	3
Glass	214	9	6
heart	270	13	2
Sonar	208	60	2
Wine	175	13	3

Table 3 Clustering result of FWASSC, EWKM, FSC, FWKM, LAC runs 20 times (RI)

表 3 FWASSC,EWKM,FSC,FWKM,LAC 运行 20 次的聚类结果(RI)

数据集		FWASSC	EWKM	FSC	FWKM	LAC
Iris	Mean	0.924 6	0.854 9	0.910 8	0.898 3	0.843 5
	Std	0.015 1	0.056 9	0.078 3	0.083 0	0.057 1
Glass	Mean	0.684 5	0.661 0	0.590 3	0.579 8	0.674 3
	Std	0.025 6	0.028 4	0.087 7	0.094 1	0.019 2
heart	Mean	0.644 1	0.515 9	0.551 9	0.547 8	0.513 6
	Std	0.017 6	0.002 5	0.033 1	0.031 5	0.009 9
Sonar	Mean	0.548 2	0.505 1	0.501 4	0.501 4	0.506 4
	Std	0.011 8	0.002 4	0.002 0	0.000 8	0.003 5
Wine	Mean	0.860 9	0.708 3	0.600 1	0.600 2	0.702 9
	Std	0.026 6	0.013 3	0.005 5	0.005 5	0.023 8

Table 4 Clustering result of FWASSC, EWKM, FSC, FWKM, LAC runs 20 times (NMI)

表 4 FWASSC,EWKM,FSC,FWKM,LAC 运行 20 次的聚类结果(NMI)

数据集		FWASSC	EWKM	FSC	FWKM	LAC
Iris	Mean	0.825 0	0.726 2	0.818 2	0.787 5	0.719 3
	Std	0.028 7	0.057 1	0.077 3	0.115 2	0.057 9
Glass	Mean	0.321 6	0.319 0	0.198 0	0.229 3	0.320 4
	Std	0.029 0	0.051 5	0.060 7	0.052 2	0.042 4
heart	Mean	0.268 3	0.021 3	0.093 5	0.092 6	0.022 4
	Std	0.024 5	0.004 0	0.038 2	0.032 2	0.012 0
Sonar	Mean	0.050 0	0.009 9	0.015 0	0.013 6	0.011 8
	Std	0.029 7	0.002 7	0.006 9	0.003 0	0.005 5
Wine	Mean	0.682 5	0.429 3	0.146 9	0.152 2	0.422 6
	Std	0.042 1	0.007 9	0.018 4	0.020 7	0.027 5

3.2 图像数据结果与分析

为了进一步验证本文算法的性能,将算法应用于图像分割.实验时,采用空间灰度相关法(SGLDM)提取图像的纹理特征,进而采用本文算法聚类图像.文献[40]提出了空间灰度相关法(SGLDM)提取图像的纹理特征.该方法是基于一个二阶联合条件概率密度函数估计,其概率函数为 $P(i,j|d,\theta)$,其中, i,j 是分布于空间中两个像元的像素值, d 是两个像元点之间的距离, θ 是两像元点之间的连线与水平轴所形成的夹角.当 $\theta=0^\circ,45^\circ,90^\circ,135^\circ$ 时,联合概率密度函数的定义如下.

$$\begin{cases} p(i,j|d,0^\circ) = \text{No.}\{(k,l),(m,n) \in (N \times M) |_{m-k=0,n-l=d,G(k,l)=i,G(m,n)=j}\} \\ p(i,j|d,45^\circ) = \text{No.}\{(k,l),(m,n) \in (N \times M) |_{m-k=d,n-l=d,G(k,l)=i,G(m,n)=j}\} \\ p(i,j|d,90^\circ) = \text{No.}\{(k,l),(m,n) \in (N \times M) |_{m-k=0,n-l=d,G(k,l)=i,G(m,n)=j}\} \\ p(i,j|d,135^\circ) = \text{No.}\{(k,l),(m,n) \in (N \times M) |_{m-k=d,n-l=d,G(k,l)=i,G(m,n)=j}\} \end{cases} \quad (21)$$

其中, $N \times M$ 是图像大小,且有 $L+1$ 个灰度级;No.表示集合中元素的个数; $G(k,l),G(m,n)$ 是图像中的两个像素点.

选用人工合成图片,并加入高斯噪声、Berkeley 图像分割数据集验证算法的抗噪性能、分割精度与时间复杂度.

3.2.1 算法抗噪性能分析

在如图 1(a)所示的人工合成图像中加入均值为 0.1、方差为 0.003 的高斯噪声后进行实验,并与上述的 EWKM,FWKM,FSC,LAC 这 4 种算法进行对比,聚类结果如图 1(c)~图 1(g)所示.从图中可以看出,对比算法 EWKM,FWKM,FSC 对图像噪声极其敏感,导致结果中背景与目标混淆严重;虽然对比算法 LAC 的结果稍好一些,已基本区分开背景与目标,但仍存在被噪声污染的数据点错分现象;相比之下,本文算法的聚类结果就比较理想,不但区分开了噪声污染的目标和背景部分,只存在个别样本的错分.因此,本文算法具有较好的抗噪性能.

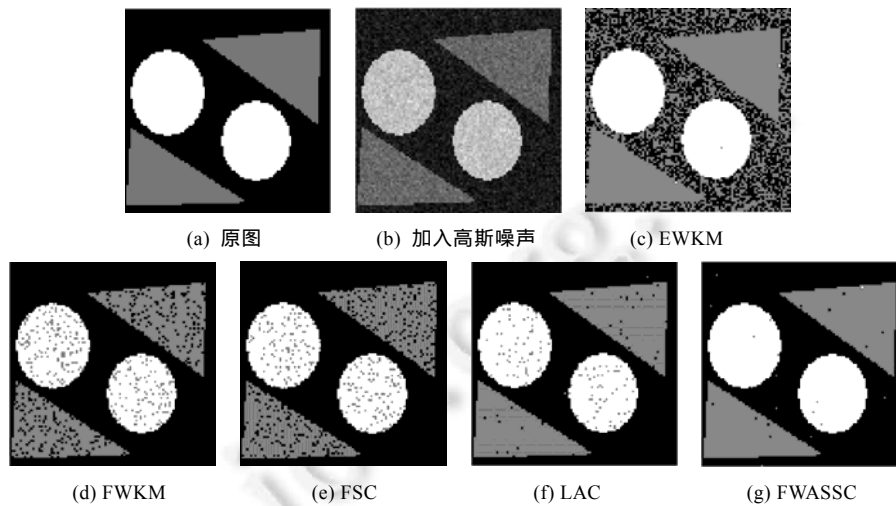


Fig.1 Clustering result of noisy images

图 1 含噪声图像的聚类结果

3.2.2 算法分割精度分析

Berkeley 计算机视觉组公开的图像分割数据集中包括了大量自然图像,为研究提供了便利.选取图 2(a)所示的图像进行实验,聚类个数为 2,背景天空为第 1 类,目标物飞机为第 2 类.实验结果如图 2(b)~图 2(f)所示.

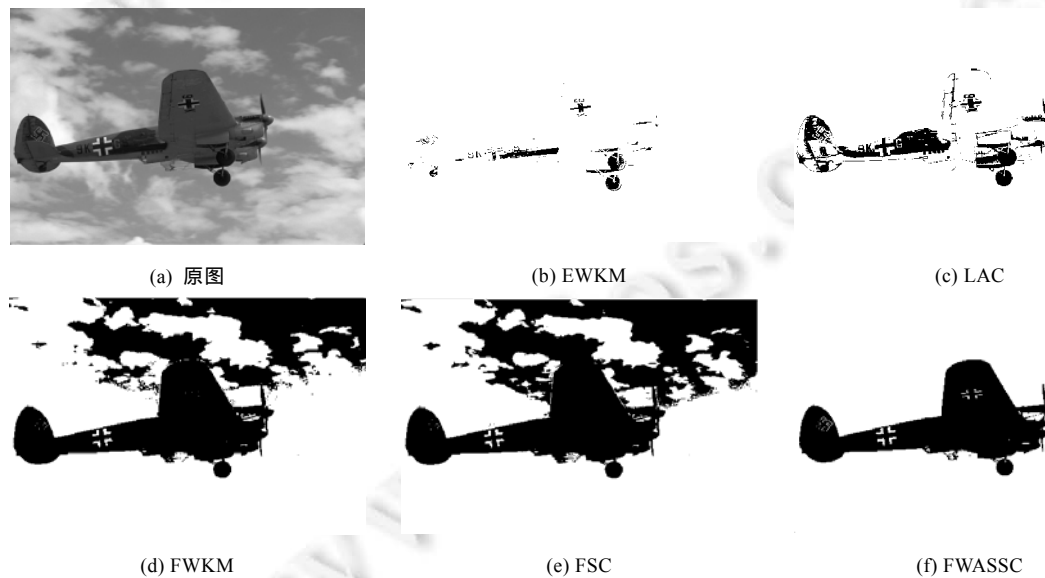


Fig.2 Berkeley natural image clustering results

图 2 Berkeley 自然图像聚类结果

对于结果的评价,主要从视觉评价与 Dice 系数进行.其中,视觉评价主要从图像是否错分、区域是否一致、边缘保持是否完整等几个方面来判断.Dice 系数采用公式(22)的计算方法.

$$DC = \frac{2|M \cap N|}{|M| + |N|} \quad (22)$$

其中, M 是非零的金标准, N 为非零的分割结果. DC 值越大,表示分割结果越好.

从实验结果可以看出,对比算法 EWKM 与 LAC 分割效果相当不理想,出现了严重的过分割现象,尤其是 EWKM,整个飞机只有极少部分划分正确;LAC 分割结果虽然可以大致看清飞机的轮廓,但机身部分错分严重,与天空相近的颜色都被划分到另一类中;虽然 FWKM 和 FSC 分割结果相对于前两种有很大的改善,飞机轮廓清晰可见,但机身与天空的划分并不准确.本文提出的算法 FWASSC 分割结果相对于对比算法,取得了令人满意的分割结果,不但将飞机轮廓清晰的分割出来,而且边缘保持完整,细节划分准确,未出现过分割或欠分割问题.根据 Berkeley 计算机视觉组提供的分割金标准,采用 Dice 系数评价分割结果,即分割结果与分割金标准越相似,Dice 系数就越大.图 2 的分割结果对应的 Dice 系数见表 5.

Table 5 Dice coefficients of Fig.2

表 5 图 2 的 Dice 系数

算法	FWASSC	EWKM	FSC	FWKM	LAC
DC	0.872 6	0.637 4	0.718 4	0.709 7	0.648 3

由表 5 中的数据可以得到:本文算法 FWASSC 的 Dice 系数值为 0.8726,而对比算法 EWKM,FSC,FWKM,LAC 中 Dice 系数最高的也只有 0.7184(FSC 算法).依据 Dice 系数评价不难得得到本文算法分割精度高于其他几种算法的结论.

3.2.3 算法时间复杂度分析

在本文提出的算法 FWASS 中,其时间复杂度主要来源于隶属度矩阵 U 的计算、聚类中心 Z 的计算与维度权值矩阵 W 的更新计算.其中,计算所有样本的隶属度时,时间复杂度为 $O(NCD)$;为每一个个体重新计算聚类中心时,其时间复杂度为 $O(CD)$;对于维度权值的计算,初始化烟花时时间复杂度为 $O(CD)$.假设算法中烟花种群规模为 M ,需要经过 T 次迭代算法收敛,则 FWASSC 算法的时间复杂度为 $O(MNCDT)$.可见,算法的时间复杂度与数据个数、维度、聚类个数以及烟花种群规模有关.

表 6 给出了 EWKM,FSC,FWKM,LAC 算法及本文算法 FWASS 分割图 1 和图 2 的运行时间对比情况.

Table 6 Running time of each algorithm on Fig.1 and Fig.2

表 6 各种算法分割图 1 与图 2 的运行时间

图像	大小	聚类个数	运行时间(s)				
			FSC	FWKM	EWKM	LAC	FWSSC
图 1(加入噪声)	100×100	3	1.36	1.37	1.46	1.35	2.02
图 2	321×481	2	0.95	0.98	0.96	0.95	1.50

从表 6 中的分割时间也可以看出,FWASS 算法中由于烟花算法的引入,导致整个算法的耗时比其他几种算法稍长.如果对于实时性要求不是很高的图像分割领域,用相对较长的时间换取抗噪性能与分割精度较大提高的算法也是有意义的.

4 临床 MR 图像数据实验

4.1 乳腺 MR 图像的分割

虽然以上分析均已说明本文算法具有良好的性能,本节仍将其应用于临床数据分析,进一步验证其聚类 MR 图像的性能.在与上述数据实验同等环境下,对如图 3(a)所示的 DICOM 格式临床乳腺 MR 图像进行聚类,分割出肿瘤所在区域,即图中白色圆圈标注的区域,并与 FSC,FWKM,EWKM 和 LAC 算法的分割结果进行对比.图像数据来源于德国西门子 1.5 T 标准双乳线圈磁共振扫描仪,为注入造影剂增强后的三维 T1 加权梯度回波序

列图.扫描成像时,病人采取俯卧位,双侧乳腺自然悬垂于乳房线圈的空洞内,各参数如下: $TR=5.6\text{ms}$, $TE=2.76\text{ms}$,层间距为 0.3mm ,层厚为 1.2mm , $FOV=34\text{cm}\times 34\text{cm}$,图像大小为 512×512 ,每次扫描时间为 60s .

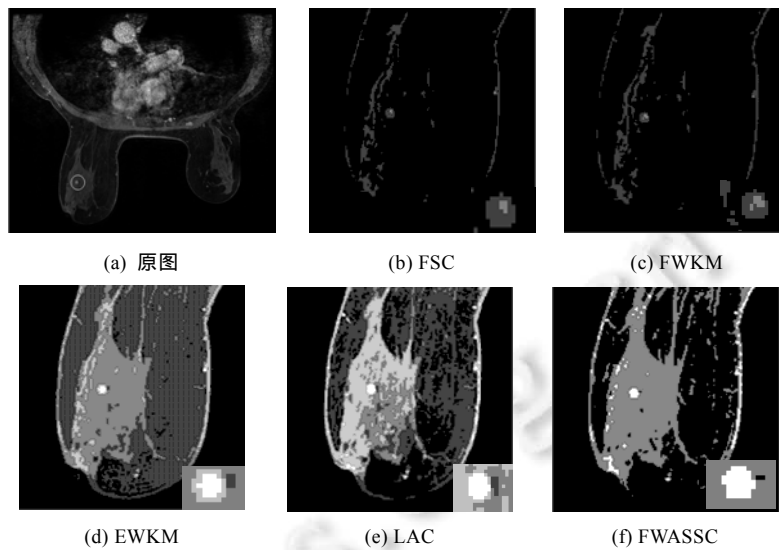


Fig.3 Clustering results of clinicalmammography MR images

图3 临床乳腺 MR 图像聚类结果

图 3(b)~图 3(f)分别为对比算法 FSC,FWKM,EWKM,LAC 与本文算法 FWASSC 对图 3(a)的聚类结果.为了更加清晰地显示结果,图 3(b)~图 3(f)中仅保留了与肿瘤相关的一边乳房的分割情况,且在图的右下角给出了所分割肿瘤形状的放大效果.

在图 3(b)~图 3(f)中,对比算法 FSC,FWKM 分割结果中,肿瘤并未被完全分割出来,肿瘤周围大片区域划分错误.尽管 EWKM 与 LAC 分割效果与前两种算法相比有一定的改进,肿瘤也较为完整地分割出来,但是分割效果同样不理想,肿瘤周围区域本应属于同一个簇类被划分到不同的簇类中.反观本文算法的结果,不仅准确分割出肿瘤部分,而且肿瘤周围区域基本划分正确,轮廓也较为清晰.

4.2 脑部MR图像的分割

本节选取如图 4(a)所示的 DICOM 格式临床脑部 MR 图像来验证本文算法的有效性.图像数据来源于 GE 公司 signa HDxT 3.0T 核磁共振设备,T1 序列图,各参数如下: $TR=3042.13\text{ms}$, $TE=11.77\text{ms}$,层间距为 6mm ,厚度为 5mm , $DFOV=220\text{mm}$,图像大小为 512×512 .聚类个数为 4,如图 4(b)~图 4(f)所示为对比算法与本文算法的聚类结果.从图中可以很清晰地看出各对比算法对于脑部灰质与白质的分割效果均不理想,而本文算法基本能够准确地划分灰质和白质.脑部 MR 图像的分割结果再一次验证了本文算法的性能.

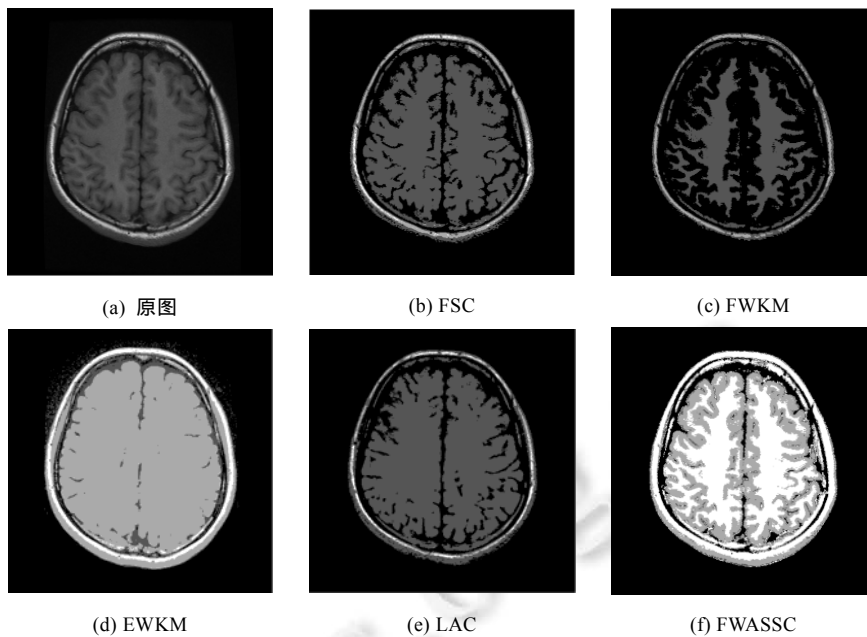


Fig.4 Clustering results of brain MR images

图 4 临床脑部 MR 图像聚类结果

5 结 论

本文提出了一种基于烟花算法的软子空间聚类算法,进行 MR 图像的聚类.算法借鉴 PSO 和 DESC,将对权值的等式约束放松为界约束,并引入噪声聚类重新设计了目标函数,弥补了现有软子空间聚类算法对噪声数据敏感的缺陷;结合硬聚类、模糊聚类与噪声聚类算法思想,提出了新的隶属度计算方法,提高了算法的鲁棒性;在聚类过程中引入自适应烟花算法,有效平衡局部搜索与全局搜索,弥补了现有算法依赖初始聚类中心的选择而易陷入局部最优的缺陷.实验结果表明,本文算法不仅在 UCI 数据集上能够取得较好的聚类结果,而且对图像聚类也具有较好的抗噪性能,对 MR 图像的聚类也具有较高的精度和鲁棒性,能够较为有效地实现 MR 图像的分割.

未来的研究将继续对该算法进行改进,使其参数自适应,适应各种情况下的 MR 图像聚类.

References:

- [1] Montseny E, Sobrevilla P, Romani S. A fuzzy approach to white blood cells segmentation in color bone marrow images. In: Proc. of the 2004 IEEE Int'l Conf. on Fuzzy Systems, Vol.1. IEEE, 2004. 173-178. <http://ieeexplore.ieee.org/document/1375712/> [doi: 10.1109/FUZZY.2004.1375712]
- [2] Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 1973,3(3):32-57. [doi: 10.1080/01969727308546046]
- [3] Bezdek JC. A convergence theorem for the fuzzy ISODATA clustering algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1980,2(1):1-8. [doi: 10.1109/TPAMI.1980.4766964]
- [4] Hathaway RJ, Bezdek JC, Tucker WT. Recent convergence results for the fuzzy C-means clustering algorithms. *Journal of Classification*, 1988,5(2):237-247. [doi: 10.1007/BF01897166]
- [5] Ahmed MN, Yamany SM, Farag AA. Bias field estimation and adaptive segmentation of MRI data using a modified fuzzy C-means algorithm. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 1999. 250-255. [doi: 10.1109/CVPR.1999.786947]

- [6] Zhang DQ, Chen SC. A novel kernelized fuzzy *C*-means algorithm with application in medical image segmentation. *Artificial Intelligence in Medicine*, 2004,32(1):37–50. [doi: 10.1016/j.artmed.2004.01.012]
- [7] Krinidis S, Chatzis V. A robust fuzzy local information *C*-means clustering algorithm. *IEEE Trans. on Image Processing*, 2010, 19(5):1328–1337. [doi: 10.1109/TIP.2010.2040763]
- [8] Kumar S, Ray SK, Tewari P. A hybrid approach for image segmentation using fuzzy clustering and level set method. *Int'l Journal of Image Graphics and Signal Processing*, 2012,6:1–7. [doi: 10.5815/ijigsp.2012.06.01]
- [9] Agrawal R, Gehrke J, Gunopulos D. Automatic subspace clustering of high dimensional data for data mining application. In: *Proc. of the ACM SIGMOD Record*. 1998. 94–105. [doi: 10.1145/276304.276314]
- [10] Chan E, Ching W, Ng M. An optimization algorithm for clustering using weighted dissimilarity measures. *Pattern Recognition*, 2004,37(5):943–952. [doi: 10.1016/j.patcog.2003.11.003]
- [11] Jing L, Ng MK, Xu J. Subspace clustering of text documents with feature weighting *K*-means algorithm. In: *Proc. of the Advances in Knowledge Discovery and Data Mining, Pacific-Asia Conf. (PAKDD 2005)*. Hanoi, 2005. 802–812. [doi: 10.1007/11430919_94]
- [12] Deng Z, Choi KS, Chung FL. Enhanced soft subspace clustering integrating within-cluster and between-cluster information. *Pattern Recognition*, 2010,43(3):767–781. [doi: 10.1016/j.patcog.2009.09.010]
- [13] Chen LF, Guo GD, Jiang QS. Adaptive algorithm for soft subspace clustering. *Ruan Jian Xue Bao/Journal of Software*, 2010,21(10): 2513–2523 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3763.htm> [doi: 10.3724/SP.J.1001.2010.03763]
- [14] Zhu L, Cao L, Yang J. Soft subspace clustering with competitive agglomeration. In: *Proc. of the IEEE Int'l Conf. on Fuzzy Systems*. 2011. 691–698. [doi: 10.1109/FUZZY.2011.6007424]
- [15] Frey BJ, Dueck D. Clustering by passing messages between data points. *Science*, 2007,315(5814):972–976. [doi: 10.1126/science.1136800]
- [16] Yu ZW, Li L, Liu JM, Zhang J, Han GQ. Adaptive noise immune cluster ensemble using affinity propagation. *IEEE Trans. on Knowledge and Data Engineering*, 2015,27(12):3176–3189. [doi: 10.1109/TKDE.2015.2453162]
- [17] Gan G, Ng MKP. Subspace clustering using affinity propagation. *Pattern Recognition*, 2015,48(4):1455–1464. [doi: 10.1016/j.patcog.2014.11.003]
- [18] Jing L, Ng MK, Huang JZ. An entropy weighting *k*-means algorithm for subspace clustering of high-dimensional sparse data. *IEEE Trans. on Knowledge and Data Engineering*, 2007,19(8):1026–1041. [doi: 10.1109/TKDE.2007.1048]
- [19] Chen X, Ye Y, Xu X. A feature group weighting method for subspace clustering of high-dimensional data. *Pattern Recognition*, 2012,45(1):434–446. [doi: 10.1016/j.patcog.2011.06.004]
- [20] Gan G, Ng MKP. Subspace clustering with automatic feature grouping. *Pattern Recognition*, 2015,48(11):3703–3713. [doi: 10.1016/j.patcog.2015.05.016]
- [21] Gan G, Wu J. A convergence theorem for the fuzzy subspace clustering (FSC) algorithm. *Pattern Recognition*, 2008,41(6):1939–1947. [doi: 10.1016/j.patcog.2007.11.011]
- [22] Domeniconi C, Gunopulos D, Ma S. Locally adaptive metrics for clustering high dimensional data. *Data Mining and Knowledge Discovery*, 2007,14(1):63–97. [doi: 10.1007/s10618-006-0060-8]
- [23] Zhu L, Lei JS, Bi ZQ, Yang J. Soft subspace clustering algorithm for streaming data. *Ruan Jian Xue Bao/Journal of Software*, 2013, 24(11):2610–2627 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4469.htm> [doi: 10.3724/SP.J.1001.2013.04469]
- [24] Vesterström J, Thomsen R. A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. In: *Proc. of the Congress on Evolutionary Computation (CEC 2004)*, Vol.2. IEEE Xplore, 2004. 1980–1987. [doi: 10.1109/CEC.2004.1331139]
- [25] Lin L, Gen M, Liang Y. A hybrid EA for high-dimensional subspace clustering problem. In: *Proc. of the 2014 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2014. 2855–2860. [doi: 10.1109/CEC.2014.6900313]
- [26] Yu Z, Wong HS. GCA: A real-time grid-based clustering algorithm for large data set. In: *Proc. of the Int'l Conf. on Pattern Recognition*. IEEE, 2006. 740–743. [doi: 10.1109/ICPR.2006.597]
- [27] Lu YP, Wang SR, Li SZ, Zhou CL. Particle swarm optimizer for variable weighting in clustering high-dimensional data. *Machine Learning*, 2011,82(1):43–70. [doi: 10.1007/s10994-009-5154-2]

- [28] Yu Z, Wong HS. Quantization-Based clustering algorithm. Pattern Recognition, 2010,43(8):2698–2711. [doi: 10.1016/j.patcog.2010.02.020]
- [29] Yu ZW, Chen HT, You J, GQ Han, Li L. Hybrid fuzzy cluster ensemble framework for tumor clustering from bio-molecular data. IEEE/ACM Trans. on Computational Biology and Bioinformatics, 2013,10(3):657–670. [doi: 10.1109/TCBB.2013.59]
- [30] Yu ZW, Chen HT, You J, Liu JM, Wong HS, Han GQ, Li L. Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data. IEEE/ACM Trans. on Computational Biology & Bioinformatics, 2015,12(3):568–582. [doi: 10.1109/TCBB.2014.2359433]
- [31] Bi ZS, Wang JH, Yin J. Subspace clustering based on differential evolution. Chinese Journal of Computers, 2012,35(10):2116–2128 (in Chinese with English abstract).
- [32] Dave RN. Characterization and detection of noise in clustering. Pattern Recognition Letters, 1991,12(11):657–664. [doi: 10.1016/0167-8655(91)90002-4]
- [33] Tan Y, Zhu Y. Fireworks algorithm for optimization. In: Proc. of the 1st Int'l Conf. on Advances in Swarm Intelligence (ICSI 2010). Beijing, 2010. 355–364. [doi: 10.1007/978-3-642-13495-1_44]
- [34] Pei Y, Zheng SQ, Tan Y, Takagi H. An empirical study on influence of approximation approaches on enhancing fireworks algorithm. In: Proc. of the 2012 IEEE Congress on System, Man and Cybernetics. 2012. [doi: 10.1109/ICSMC.2012.6377916]
- [35] Zheng S, Janecek A, Tan Y. Enhanced fireworks algorithm. In: Proc. of the Evolutionary Computation. IEEE, 2013. 2069–2077. [doi: 10.1109/CEC.2013.6557813]
- [36] Zheng SQ, Janecek A, Li JZ, Tan Y. Dynamic search in fireworks algorithm. In: Proc. of the Evolutionary Computation. IEEE, 2014. 3222–3229. [doi: 10.1109/CEC.2014.6900485]
- [37] Li J, Zheng S, Tan Y. Adaptive fireworks algorithm. In: Proc. of the 2014 IEEE Congress on Evolutionary Computation. Beijing, 2014. 3214–3221. [doi: 10.1109/CEC.2014.6900418]
- [38] Rand WM. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 1971, 66(336):846–850. [doi: 10.1080/01621459.1971.10482356]
- [39] Strehl A, Ghosh J. Cluster ensembles—A knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research, 2003,3(3):583–617. [doi: 10.1162/153244303321897735]
- [40] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans. on Systems Man and Cybernetics, 1973,SMC-3(6):610–621. [doi: 10.1109/TSMC.1973.4309314]

附中文参考文献:

- [13] 陈黎飞,郭躬德,姜青山. 自适应的软子空间聚类算法. 软件学报, 2010,21(10):2513–2523. <http://www.jos.org.cn/1000-9825/3763.htm> [doi: 10.3724/SP.J.1001.2010.03763]
- [23] 朱林,雷景生,毕忠勤,杨杰. 一种基于数据流的软子空间聚类算法. 软件学报, 2013(11):2610–2627. <http://www.jos.org.cn/1000-9825/4469.htm> [doi: 10.3724/SP.J.1001.2013.04469]
- [31] 毕志升,王甲海,印鉴. 基于差分演化算法的软子空间聚类. 计算机学报, 2012,35(10):2116–2128.



范虹(1969 -),女,宁夏平罗人,博士,副教授,CCF 专业会员,主要研究领域为图像处理,模式识别,智能信息处理.



朱艳春(1986 -),男,博士生,助理研究员,主要研究领域为动态磁共振成像技术,定量磁共振成像算法,快速磁共振成像技术,重建算法.



侯存存(1992 -),女,硕士生,主要研究领域为图像处理.



姚若侠(1968 -),女,博士,教授,博士生导师,主要研究领域为计算复杂性与符号计算,孤立子理论,模式识别,算法分析与设计.