ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 33,No. 2,Mar. 2018,pp. 351－358
DOI:10. 16337/j. 1004-9037. 2018. 02. 018

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于新相似度的模糊协同聚类改进算法

高翠芳　张　朔　殷　萍　沈菀蔷

(江南大学理学院,无锡,214122)

摘　要:提出一种优化传统协同聚类中模糊点类别归属的改进算法,该算法引入基于清晰半径的新相似性距离公式,用超球体中心区域代替传统算法中的类中心,在各子集初始聚类结果的基础上,对容易导致类别归属错误的模糊点重新计算隶属度,得到较为清晰的聚类结果。实验结果显示,改进算法能很大程度地减少边界上的模糊点个数及纠正分类错误,清晰半径的引入还能弱化各子集之间协同系数的差异,使得参数设置更为简单。

关键词:模糊聚类;FCM 算法;协同聚类;模糊点;清晰半径

中图分类号:TP311　　**文献标志码**:A

Improved Fuzzy Collaborative Clustering Algorithm Based on New Similarity

Gao Cuifang, Zhang Shuo, Yin Ping, Shen Wanqiang

(School of Science, Jiangnan University, Wuxi, 214122, China)

Abstract: An improved algorithm is proposed to correct the assignments of fuzzy points for the previous fuzzy collaborative clustering. The new expression of similarity distance based on clear radius is introduced, and the hypersphere central region is used to represent one cluster instead of the traditional center. In the light of initial results of separated subsets, the membership degrees are recalculated for the fuzzy points in which wrong assignments easily occurred, and finally the more clear-cut partition is obtained. The experimental results show that the improved algorithm can reduce the fuzzy points widely distributed near the boundary and correct quite part of the wrong partitions. Moreover, the method of clear radius can simplify the parameter setting by weakening the difference of collaborative coefficients.

Key words: fuzzy clustering; FCM algorithm; collaborative clustering; fuzzy point; clear radius

引　　言

聚类分析是数据处理的一种重要方法与手段,传统的硬聚类算法对每个样本点的类别归属进行严格的划分,使得每个样本点都具有非此即彼的性质。由于现实中数据样本的特性和类属总是存在一定程度的中间性,于是将模糊理论与聚类算法相结合,出现了模糊聚类算法[1-4]。模糊 C 均值聚类算法(Fuzzy C-means algorithm, FCM)算法是目前研究和应用较为广泛的模糊聚类算法之一[5-7],它在目标函数中建立

基金项目:国家自然科学基金青年基金(61402202)资助项目;高等学校博士学科点专项科研基金(20120093120016)资助项目;江苏省自然科学基金青年基金(BK20130117,BK20150124)资助项目;中国博士后科学基金(2015M581724)资助项目。
收稿日期:2016-07-01;**修订日期**:2016-11-10

对样本类属的模糊描述,并采用模糊隶属度衡量每个数据点属于不同类别的程度。模糊协同聚类[8-11]在算法处理独立数据集的基础上,建立了一种针对不同数据子集之间协同合作关系的综合优化模型,能用一个目标函数同时处理多个既互相联系又各自独立的特征子集。其聚类思想主要是通过计算和调整各数据子集之间的隶属度差异,实现各子集之间的综合聚类,是一种颇有应用潜力的新算法。

但是,无论是使用整体数据集的 FCM 算法还是综合使用各个数据子集的协同聚类算法,都是用一个中心点代表一类(每一维),相比于多中心点的模糊聚类算法[12]和基于中心区域的聚类算法[13],其单一的类中心点和隶属度分类指标,很难充分体现不同类的结构特征。如果各类中数据点的分布疏密差别较大,在边界点的处理上就会存在一定的缺陷,即对于隶属度不清晰的点在分类时会出现偏差。一些基于密度的聚类算法[13-18]利用数据点的疏密信息来减少分类错误,其中基于清晰半径的模糊点二次聚类算法[13]提出了清晰半径的概念,定义了一种基于超球体中心区域的相似性距离公式,该算法以数据点分布的疏密程度为理论依据,对不能清晰确定类别归属的模糊点重新计算隶属度,使模糊点的类别归属更加清晰,从而纠正分类错误,提高传统算法的聚类精度。为了增加模糊协同聚类算法的适用性,特别是改善其在数据点分布差别较大的数据集上的聚类效果,本文引入中心区域的清晰半径,以协同聚类的初始聚类结果为基础,对其中模糊点的隶属度用清晰半径进行重新计算,从而优化传统模糊协同聚类算法中部分模糊点的类属问题。

1 模糊协同聚类算法

设有 p 个特征子集 $D[1]$, $D[2]$, …, $D[p]$(各子集中的数据点个数相同),每个子集都具有独立的分类结构及隶属度矩阵 $\boldsymbol{U}[1]$, $\boldsymbol{U}[2]$, …, $\boldsymbol{U}[p]$,各子集之间通过协同系数 $\alpha[i, j]$($j=1, 2, \cdots, p$, $i\neq j$)建立信息交流与合作机制,把聚类结果综合到某个特征子集 $D[i]$上,最终得到全局聚类结果。算法的结构原理如图 1。

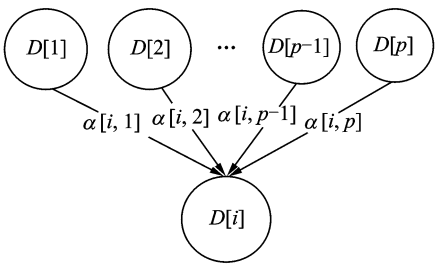


图 1 模糊协同聚类算法示意图

Fig. 1 Scheme of the fuzzy collaborative clustering

模糊协同聚类算法[10]定义了一个包含协同惩罚项的目标函数,采用统一优化的方式来融合 p 个特征子集上的局部结果。惩罚项通过协同系数确定子集之间的关系强度(协同系数应为非负实数),还考虑了局部结果对协同后的隶属度矩阵和聚类中心的影响。算法的目标函数为

$$J_{\text{HC-FCM}}[i]=\sum_{k=1}^{C}\sum_{r=1}^{N}\boldsymbol{u}_{kr}^{2}[i]\left\|\boldsymbol{x}_r[i]-\boldsymbol{v}_k[i]\right\|^2+\sum_{\substack{j=1\\ j\neq i}}^{p}\alpha[i,j]\sum_{k=1}^{C}\sum_{r=1}^{N}(\boldsymbol{u}_{kr}[i]-\boldsymbol{u}_{kr}[j])^2\left\|\boldsymbol{x}_r[i]-\boldsymbol{v}_k[i]\right\|^2 \tag{1}$$

式中:$\boldsymbol{x}_r[i]$为第 i 个数据集上的第 r 个数据点;$\boldsymbol{u}_{kr}[i]$为第 i 个数据集上的第 r 个数据点对于第 k 类的隶属度;$\boldsymbol{v}_k[i]$为第 i 个数据集上的第 k 个类中心;$\alpha[i,j]$为第 i 个数据集与第 j 个数据集的协同系数。

在循环迭代计算时,类中心的更新公式为

$$\boldsymbol{v}_k=\frac{\sum_{r=1}^{N}\boldsymbol{u}_{kr}^{2}[i]\boldsymbol{x}_r[i]+\sum_{\substack{j=1\\ j\neq i}}^{p}\alpha[i,j]\sum_{r=1}^{N}(\boldsymbol{u}_{kr}[i]-\boldsymbol{u}_{kr}[j])^2\boldsymbol{x}_r[i]}{\sum_{r=1}^{N}\boldsymbol{u}_{kr}^{2}[i]+\sum_{\substack{j=1\\ j\neq i}}^{p}\alpha[i,j]\sum_{r=1}^{N}(\boldsymbol{u}_{kr}[i]-\boldsymbol{u}_{kr}[j])^2} \tag{2}$$

隶属度的更新公式为

$$\boldsymbol{u}_{kr}[i]=\frac{1-\sum_{t=1}^{C}\frac{\phi_{tr}[i]}{1+\psi[i]}}{d_{kr}^{2}[i]\sum_{t=1}^{C}\frac{1}{d_{tr}^{2}[i]}}+\frac{\phi_{kr}[i]}{1+\psi[i]} \tag{3}$$

式中

$$\phi_{kr}[i]=\sum_{\substack{j=1\\ j\neq i}}^{p}\alpha[i,j]u_{kr}[j] \tag{4}$$

$$\psi[i]=\sum_{\substack{j=1\\ j\neq i}}^{p}\alpha[i,j] \tag{5}$$

2 清晰半径对模糊协同聚类的改进

2.1 清晰半径

当不同类中数据点分布的疏密程度差别较大时,传统聚类算法难免出现对边界点分类模糊甚至错误的情况,因此采用新的相似性度量,对这些模糊点进行二次处理,成为改善聚类效果的一种新途径。图 2 中的部分真核细胞,就是现实中遇到的该类型数据集,其中 A 类表示可分泌蛋白质,B 类表示不分泌蛋白质,从图 2 中可以看出,A 类中的数据点分布较为分散,B 类中的数据点分布相对集中,数据点 X 表示一种可分泌蛋白质,位于两个模糊子集 A,B 的边界上,由于 X 到 A 类中心点的距离略大于 X 到 B 类中心点的距离(即 $L_1>L_2$),是一个模糊点,若根据传统聚类算法采用一个类中心点来代表一个类,则会把 X 点归属于 B 类,显然不正确。

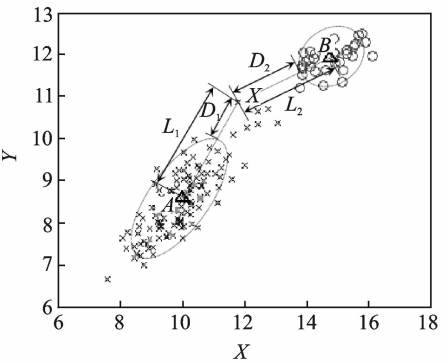


图 2 模糊点 X 的类别归属示意图

Fig. 2 Scheme of partition of fuzzy point X

这种情况下,利用清晰点构成的超球体中心区域代替传统的类中心点,可以得到更为准确的聚类结果[13]。清晰半径的定义为

$$R_k=\frac{1}{N_k}\sum_{i=1}^{N_k}d_{ik} \tag{6}$$

式中:N_k 为第 k 类中清晰点的个数,d_{ik} 为第 k 类中第 i 个清晰点与第 k 类中心的欧氏距离,用式(6)可以计算出第 k 类的清晰半径。

对于类似于 X 的模糊点,用基于清晰半径的中心区域来重新计算相似性距离(即 X 到类中心点的距离与该类清晰半径的差),则有 $d_{AX}=L_1-R_A=D_1$,$d_{BX}=L_2-R_B=D_2$,由于新的相似性距离 $D_1<D_2$,用该相似性距离重新计算隶属度以后,X 点应该归属于 A 类,更符合数据点的实际分布情况。

2.2 清晰半径对模糊协同聚类算法的改进

首先将协同聚类得到的隶属度矩阵进行分类,把数据点分为清晰点、模糊点和次清晰点。根据模糊划分的隶属度评价[19],有意义的隶属度取值应大于 0.5,若将隶属度看作[0.5, 1]区间内的平均分布,则清晰点应分布在[1−δ, 1],模糊点应分布在[0.5, 0.5+δ]。δ 的取值会影响清晰半径的大小以及模糊点的改善效果,取值过小时模糊点得不到改善,算法退化为基础算法,取值过大则会增加计算量。在对数据集掌握部分先验信息的情况下,可根据数据集的实际情况确定 δ 的取值,清晰点和模糊点的取值范围可以不同。

根据上述分析,本文取 $\delta=0.15$,设最大隶属度大于 0.85 的点为清晰点,用于计算每一类的清晰半

径;隶属度小于 0.65 的点为模糊点,这些模糊点需要重新计算相似性距离和隶属度矩阵;介于 0.65～0.85 之间的点,称为次清晰点,不做特殊处理。基于清晰半径的模糊协同聚类算法框架如图 3。

综上所述,用清晰半径改进的协同聚类算法的主要步骤为:

(1) 用基于清晰半径的 FCM 聚类算法,以每个数据子集上产生的独立划分作为初始聚类。

(2) 定义协同系数矩阵 $\boldsymbol{\alpha}[i,j]$(具体取值见实验与结果分析)。

(3) 按照式(2)计算协同聚类中心 $\boldsymbol{v}_k[i]$, $i=1,2,\cdots,p$。

(4) 按照式(3)计算协同到第 i 个数据集上的新隶属度 $\boldsymbol{u}_{kr}[i]$。

(5) 将协同聚类得到的隶属度矩阵分为模糊点、清晰点和次清晰点(改进过程只对模糊点的隶属度进行重新计算,清晰点与次清晰点的隶属度保持其原有值)。

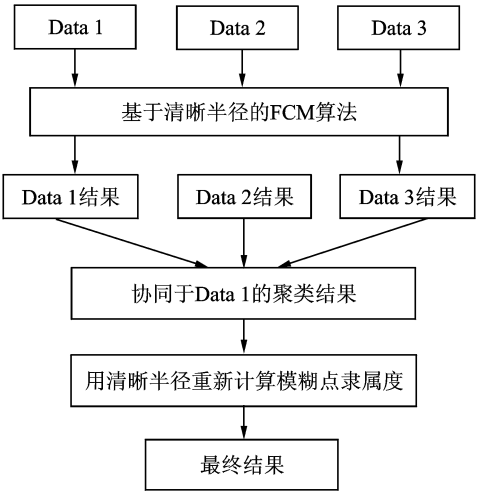


图 3 基于清晰半径的模糊协同聚类算法结构

Fig. 3 Algorithm framework of clear radius-based fuzzy collaborative clustering

(6) 按照式(6)计算每一类的清晰半径 R_k。

(7) 重新计算所有模糊点与每一类的相似性距离,计算公式为

$$d'_{ik}=d_{ik}-R_k \qquad k=1,2,\cdots,C \tag{7}$$

(8) 用式(8)定义的新相似性距离重新计算模糊点的隶属度为

$$\boldsymbol{u}_{ik}=\frac{[d'^{2}_{ik}]^{-1}}{\sum_{r=1}^{c}[d'^{2}_{ir}]^{-1}} \tag{8}$$

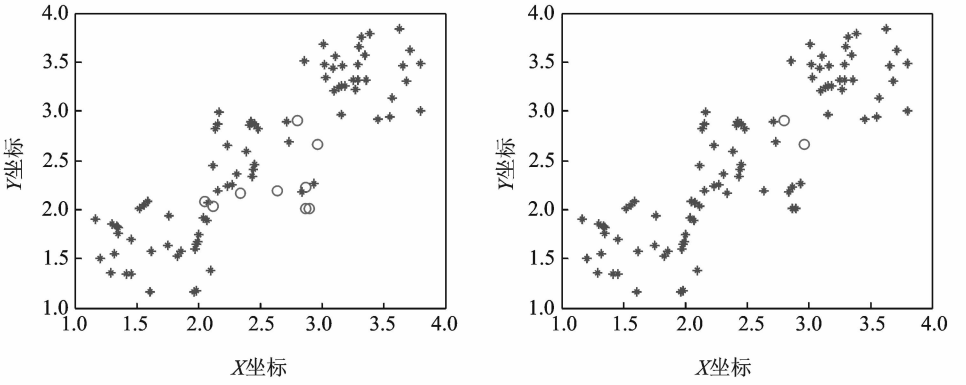
(9) 重复步骤(3)～(8)直至满足最优条件,清晰点与次清晰点的原有隶属度与重新计算的模糊点隶属度组合在一起,作为最终结果。

3 实验与结果分析

3.1 改进算法的聚类性能分析

本文首先使用 3 个人工数据子集 Data 1,Data 2,Data 3 进行协同聚类实验。数据集包括 90 个数据点,分为 3 类(平面分布从左到右依次为第 A,B,C 类),每类 30 个数据点。每个数据点由六维特征向量表示,前两维特征分量构成数据子集 Data 1,中间两维分量构成子集 Data 2,后两维分量构成子集 Data 3。进行协同计算时,将数据子集 Data 2 和 Data 3 协同到 Data 1 上,实验结果如图 4 所示,圆圈代表分类时产生的模糊点。从图 4 可以看到,在 A 类与 B 类交界处的模糊点已经全部优化为清晰点与次清晰点。由于 B 类与 C 类的清晰半径比较接近,重新计算后的相似性距离与之前距离相比变化不大,故位于这两类交界处的模糊点仍然存在,但此处的模糊点类别归属没有错误(归属 B 类),不会影响最终的聚类结果。

改进前后模糊点的具体隶属度见表 1,除了点 2 与点 6(位于 B 类与 C 类的交界处)清晰半径对其改进效果有限,其他点均从原来的模糊点变为清晰点或次清晰点。更重要的是点 3 与点 5,重新计算隶属度以后不仅变得更加清晰,还改变了类别归属,实现了正确分类。



(a) 传统协同聚类算法的结果 (b) 改进协同聚类算法的结果

(a) Result of traditional collaborative clustering (b) Result of improved collaborative clustering

图 4 清晰半径对模糊协同聚类的改进效果

Fig. 4 Improved results of collaborative clustering by clear radius

表 1 清晰半径对模糊点隶属度的改进效果

Tab. 1 Improved results of fuzzy memberships by the clear radius

模糊点	模糊协同聚类的隶属度			清晰半径改进后的隶属度		
	A 类	*B* 类	*C* 类	*A* 类	*B* 类	*C* 类
1	0.172 9	0.622 1	0.204 9	0.110 3	0.778 9	0.110 6
2	0.100 7	0.631 4	0.267 7	0.053 4	0.540 6	0.405 8
3	0.304 6	0.591 8	0.103 5	0.728 5	0.247 0	0.024 4
4	0.130 8	0.615 9	0.253 2	0.092 1	0.861 7	0.046 1
5	0.294 3	0.589 2	0.116 4	0.687 0	0.285 7	0.027 2
6	0.122 6	0.608 2	0.269 0	0.064 6	0.619 6	0.315 7
7	0.189 2	0.619 6	0.191 1	0.208 2	0.664 9	0.126 8
8	0.217 0	0.623 2	0.159 6	0.180 0	0.788 0	0.031 9
9	0.226 2	0.577 5	0.196 2	0.209 8	0.653 7	0.136 3

由于改进算法中引入了新相似度计算,与基础算法相比,在提高聚类准确度的同时也会使改进后的运行时间有所增加。通常在迭代初始阶段存在较多的模糊点,对这些点的二次计算代价会相应增加迭代时间,但是随着最优划分的逼近,模糊点数量越来越少,迭代后期的计算量与传统协同聚类基本相同。算法在不同阶段的运行时间见图 5。

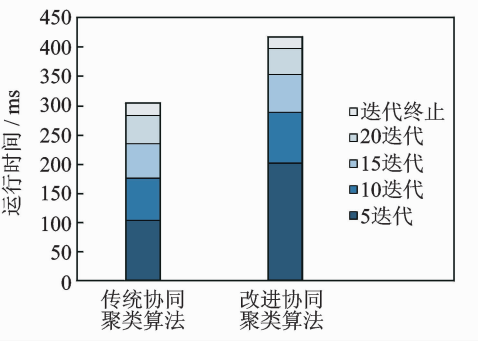


图 5 运行时间对比

Fig. 5 Comparative results of running time

3.2 协同系数对改进算法的影响

聚类结果中,协同系数取值为 $\alpha[1,2]=0.2$,$\alpha[1,3]=0.2$(根据文献[10]的结果,协同系数的最优取值在0.25

左右，因此首先在该值附近进行实验验证）。为了进一步验证其他协同系数对改进算法将产生什么样的影响，本文还进行了下面的实验。首先设各子集之间的协同系数相等，在[0.05, 0.5]的区间范围内，取不同的协同系数值，得到图 6 的结果。

可以看出，协同系数的变化不仅会影响传统协同聚类的结果，而且会影响改进算法的结果。在协同系数都为 0.1 时，传统协同聚类结果中有非常多的模糊点，经过清晰半径改进后的模糊点数量大幅度下降，随着协同系数逐渐增大，模糊点的数量逐渐减少。与文献[10]结果略有不同的是，由于本文重点对各类之间的模糊点进行处理，数据子集之间的差异不是很大，因此在协同系数大于 0.25 后，聚类效果还会有所改善，但是改善程度有限，当协同系数增大到 0.5 时，不仅模糊协同聚类得到的结果较好（只有 4 个模糊点），引入清晰半径改进的模糊协同聚类的结果更好（模糊点全部消失）。因此对于本文所采用的数据集，[0.25, 0.5]的取值区间可视为最优范围。当子集之间的协同系数不同时，令协同系数之和 $\varphi[i]=0.8$，得到图 7 所示的一组实验结果。

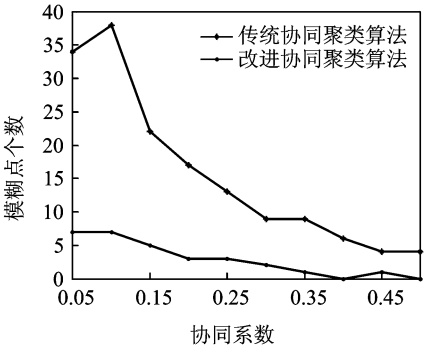


图 6 两个协同系数取值相同时的聚类效果对比

Fig. 6 Comparative results with the same collaborative coefficients

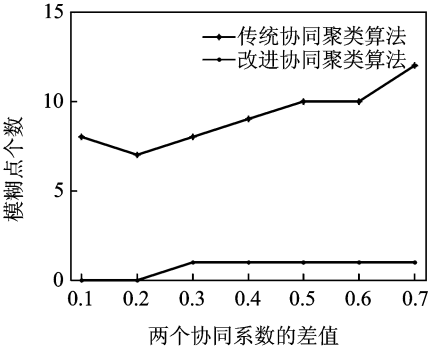


图 7 两个协同系数取值不同时的聚类效果对比

Fig. 7 Comparative results with the different collaborative coefficients

从图 7 可以看出，当两个协同系数比较接近时，聚类结果较好，且仅对传统模糊协同聚类算法产生较大影响，而对清晰半径改进的协同聚类的结果影响不大，该组实验结果显示，改进算法对协同系数差异的敏感性降低，这就意味着改进算法在参数设置上更为简单。

3.3 改进算法对 UCI 标准数据集及革兰氏阴性真细菌的聚类效果

使用 UCI 提供的 IRIS（鸢尾花）数据集，这个数据集中包含 3 个品种的鸢尾花，每个数据点包含 4 个维度（花萼长度、花萼宽度、花瓣长度和花瓣宽度），其中第 2 类与第 3 类之间的数据点存在交叉和重叠，会产生较多模糊点。实验中使用协同聚类算法将花瓣数据协同到花萼数据上进行聚类，聚类结果如图 8 所示，可以看出清晰半径的引入使模糊点的数量大幅减少。

革兰氏阴性真细菌（Gram-bacteria）蛋白序列数据来源于网站 http://www.cbs.dtu.dk/ftp/signalp。数据集中包含 451 条序列片段，分为两种不同的蛋白质，其中可分泌蛋白质 265 条，不分泌蛋白质 186 条。采用压缩感知理论提取蛋白序列的 20 维特征向量[20]，前 10 维特征分量作为数据子集Ⅰ，后 10 维特征分量作为数据子集Ⅱ，采用协同聚类算法将数据子集Ⅱ协同到数据子集Ⅰ上，聚类结果如图 9 所示，可以看出清晰半径的引入同样起到了明显的改进效果。

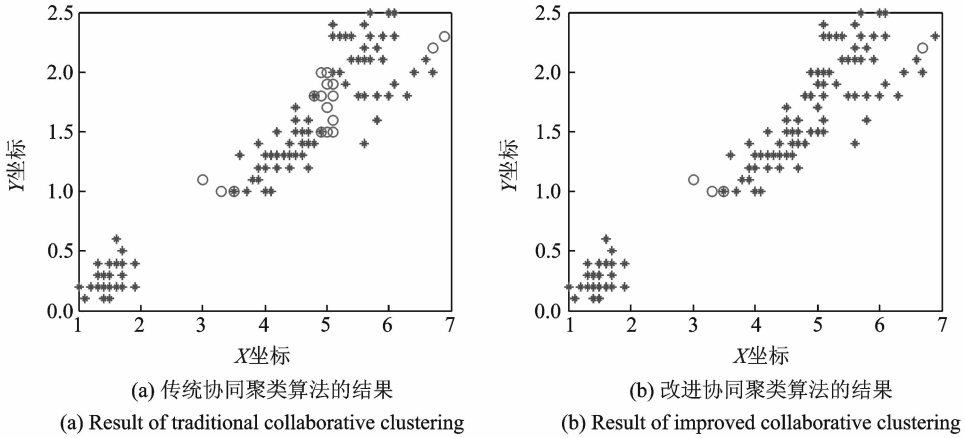


图 8 IRIS 标准数据集的聚类效果

Fig. 8 Clustering results on IRIS benchmark dataset

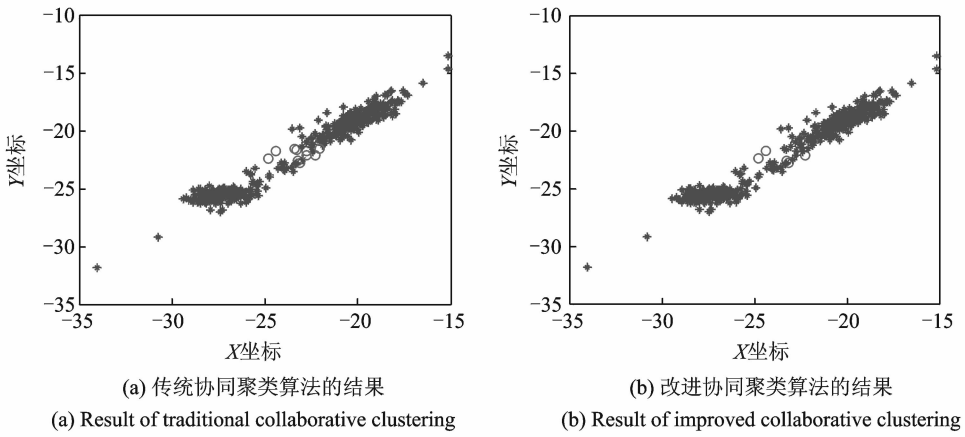


图 9 革兰氏阴性菌的聚类效果

Fig. 9 Clustering results on Gram-bacteria dataset

4 结束语

本文引入清晰半径对传统协同聚类算法进行改进,能在较大程度上减少边界上的模糊点个数,使其分类更加清晰。实验结果显示,对于数据点分布密度差别较大的数据集,由于类内清晰半径的差异,基本上能够消除边界上存在的模糊点,改进效果明显;对于清晰半径较接近的边界上的模糊点,优化效果虽然有限,但也能保持其原有聚类结果。而且,由于改进算法中清晰半径的引入,弱化了各子集之间协同系数的差异,使参数设置变得更为简单,这就使得模糊协同聚类算法的应用范围更加广泛。目前,清晰半径对模糊点隶属度的改进还是作为独立的优化步骤,在初始聚类结果的基础上开展,后续研究能否进一步完善相似性距离的计算公式,以中心区域的清晰半径为主要思想,直接得到聚类结果,有待进一步深入研究。

参考文献:

[1] Rui X, Wunsch D. Survey of clustering algorithms [J]. IEEE Transactions on Neural Networks, 2005, 16(3): 645-678.

[2] 高新波. 模糊聚类分析及其应用[M]. 西安:西安电子科技大学出版社,2004:49-61.
Gao Xinbo. Fuzzy clustering analysis and its applications[M]. Xi'an: Xidian University Press, 2004:49-61.

[3] Kupka J. Some remarks on approximations of the Zadeh's extension[C]// IEEE International Conference on Fuzzy Systems. New York: IEEE, 2013:1-7.

[4] Omran M G H, Engelbrecht A P, Salman A. An overview of clustering methods[J]. Intelligent Data Analysis, 2007, 11

(6):583-605.

[5] Gong Maoguo, Liang Yan, Shi Jiao, et al. Fuzzy C-means clustering with local information and kernel metric for image segmentation[J]. IEEE Transactions on Image Processing, 2013,22(2):573-584.

[6] 闫晓玲,王黎明,卜乐平.一种基于多维彩色向量空间的(火焰)图像模糊聚类分割算法[J]. 数据采集与处理, 2012,27(3):1-6.
Yan Xiaoling, Wang Liming, Bu Leping. Fuzzy clustering segmentation algorithm of flame image based on multi-dimensional color vector space[J]. Journal and Data Acquisition and Processing, 2012,27(3):1-6.

[7] 孙权森, 纪则轩.基于模糊聚类的脑磁共振图像分割算法综述[J].数据采集与处理,2016,31(1):28-42.
Sun Quansen, Ji Zexuan. Fuzzy clustering for brain MR image segmentation[J]. Journal and Data Acquisition and Processing, 2016,31(1):28-42.

[8] 谭欣, 徐蔚鸿.一种协同的可能性模糊聚类算法[J].计算机工程与应用,2014,50(21):147-151.
Tan Xin, Xu Weihong. Collaborative PCM fuzzy clustering algorithm[J]. Computer Engineering and Applications, 2014,50(21):147-151.

[9] Yu Fusheng, Tang Juan, Cai Ruiqiong. A necessary preprocessing in horizontal collaborative fuzzy clustering[C]// IEEE International Conference on Granular Computing. Washington, DC: IEEE Computer Society, 2007:399-403.

[10] Pedrycz W, Rai P. Collaborative fuzzy clustering with the use of fuzzy C-means and its quantification [J]. Fuzzy Sets and Systems, 2008,159(18):2399-2427.

[11] Coletta S, Vendramin L, Hruschka E R, et al. Collaborative fuzzy clustering algorithms: Some refinements and design guidelines[J]. IEEE Transactions on Fuzzy Systems, 2012,20(3):444-462.

[12] Golsefid S M M, Zarandi M H F, Turksen I B. Multi-central general type-2 fuzzy clustering approach for pattern recognitions [J]. Information Sciences, 2016,328(C):172-188.

[13] 高翠芳,胡权.基于清晰半径的模糊点二次聚类算法[J].计算机应用,2013,33(2):547-549.
Gao Cuifang, Hu Quan. Second clustering algorithm for fuzzy points based on clear radius[J]. Journal of Computer Applications, 2013,33(2):547-549.

[14] 吕宗磊, 王建东.一种基于多维空间超球体的快速聚类算法[J].南京航空航天大学学报,2006,38(6):706-711.
Lü Zonglei, Wang Jiandong. Fast clustering algorithm based on hypersphere of multidimentional space[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2006,38(6):706-711.

[15] Menardi G, Azzalini A. An advancement in clustering via nonparametric density estimation[J]. Statistics & Computing, 2013,24(5):753-767.

[16] Rinaldo A, Wasserman L. Generalized density clustering [J]. Annals of Statistics, 2009,38(5):2678-2722.

[17] 张晓, 张媛媛, 高阳,等.一种基于密度的快速聚类方法[J].数据采集与处理,2015,30(4):888-895.
Zhang Xiao, Zhang Yuanyuan, Gao Yang, et al. Fast density-based clustering approach[J]. Journal and Data Acquisition and Processing, 2015,30(4):888-895.

[18] Kavitha P, Vidhya R S. A modified fuzzy C means algorithm for clustering based on density sensitive measure[J]. Fuzzy Systems, 2015,7(7):185-190.

[19] 于剑,程乾生.模糊划分的一个新定义及其应用[J].北京大学学报(自然科学版),2000,36(5):619-623.
Yu Jian, Cheng Qiansheng. A new definition of fuzzy partition and its application[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2000,36(5):619-623.

[20] Gao Cuifang, Guan Qiang, Zhang Hao, et al. A novel feature extraction method by compressive sensing for signal peptide [J]. Journal of Chemical and Pharmaceutical Research, 2013,5(11):212-218.

作者简介:

高翠芳(1974-),女,副教授,研究方向:模式识别、生物信息学,E-mail: cuifang_gao@163.com。

张朔(1992-),男,硕士研究生,研究方向:模式识别。

殷萍(1981-),女,副教授,研究方向:数值计算、模式识别。

沈莞蔷(1981-),女,副教授,研究方向:模式识别、计算机图形学。

(编辑:陈珺)