

基于新型阴影集的模糊 C 均值聚类算法

国栋凯¹, 张钦然¹, 李小南², 易黄建^{1*}

(1. 西北大学信息科学与技术学院, 陕西 西安 710127; 2. 西安电子科技大学数学与统计学院, 陕西 西安 710126)

摘要: 提出一种基于五区域阴影集的模糊 C 均值(fuzzy C -means, FCM)算法, 利用 FCM 算法得到对象簇的隶属度, 引入五区域阴影集, 将对象划分为核心区域、次核心区域、阴影区域、次边缘区域和边缘区域, 分析次核心区域得到阈值 ω , 通过核心区域和次核心区域中隶属度 $\mu \geq \omega$ 的对象簇得到聚类结果, 在 8 个公开数据集中进行实验。本文提出的算法相比于其余 3 种算法在 7 个数据集上取得了最佳的聚类结果。

关键词: 三支决策; 模糊聚类; 三支聚类; 五区域阴影集

中图分类号: TP301.6 **文献标志码:** A

引用格式: 国栋凯, 张钦然, 李小南, 等. 基于新型阴影集的模糊 C 均值聚类算法[J]. 山东大学学报(理学版), 2025, 60(1): 74-82.

Fuzzy C -means clustering algorithm based on new shadowed sets

GUO Dongkai¹, ZHANG Qinran¹, LI Xiaonan², YI Huangjian^{1*}

(1. School of Information Science and Technology, Northwest University, Xi'an 710127, Shaanxi, China; 2. School of Mathematics and Statistics, Xidian University, Xi'an 710126, Shaanxi, China)

Abstract: A fuzzy C -means(fuzzy C -means, FCM) clustering algorithm based on five-region shadowed sets is proposed in this paper. The membership degree of the object to the cluster is obtained by the FCM algorithm. The object is divided into core region, semi-core region, shadow region, semi-negative region and negative region according to the membership degree by introducing the five-region shadowed sets. Then, a threshold value ω is obtained by analyzing the semi-core region. The objects whose membership degree $\mu \geq \omega$ in the core region and semi-core region are classified into this cluster to get the final clustering result. Experiments are carried out on 8 public data sets with other 3 clustering algorithms, compared with the other 3 algorithms, the algorithm proposed in this paper achieves the best clustering results on 7 data sets. The experimental results show that the proposed algorithm in this paper is superior to 3 other algorithms.

Key words: three-way decision; fuzzy clustering; three-way clustering; five-region shadowed sets

0 引言

传统的聚类任务很多都是硬聚类, 使得每一个须要被划分的对象都要划分到一个边界分明的簇里, 然而在实际应用中, 很多对象的特征都不是那么明显, 因此, 对具有不确定性的对象进行划分是值得研究的^[1]。模糊聚类分析的发展为分析大量不确定性的信息带来了较好的解决方案, 并成功地应用于计算机技术、生命科学、工程技术、社会科学, 以及医学图像处理等领域^[2]。

模糊 C 均值(fuzzy C -means, FCM)算法在模糊聚类算法中应用广泛^[3-4], 它通过优化目标函数得到对象与簇中心的隶属度, 将隶属度高的对象划分到对应的簇中并对样本中对象分类, 在数据挖掘、图像分割、模式识别、数据分析等领域。然而 FCM 算法只考虑隶属度最高的对象, 导致错误划分了很多对象, 而引入阴

收稿日期: 2023-06-27; 网络出版时间: 2024-05-16 10:18:24

基金项目: 国家自然科学基金资助项目(61906154); 陕西省教育厅青年创新团队资助项目(21JQ123)

第一作者: 国栋凯(1996—), 男, 硕士研究生, 研究方向为三支决策、阴影集、模糊聚类分析等。E-mail: 1159044933@qq.com

* 通信作者: 易黄建(1985—), 女, 副教授, 博士, 研究方向为光学图像三维重建、三支决策模型与应用、智能信息处理。

E-mail: yhj2014@nwu.edu.cn

影集可以解决这种问题。基于三支决策思想构建的阴影集,通过一对阈值 $[\alpha, \beta]$ 将传统的模糊集划分为核心区域、阴影区域、边缘区域3个区域,这3个区域分别表示对象一定属于该簇、对象可能属于该簇、对象一定不属于该簇,引入阴影集来处理隶属度避免了单一的划分模式,减少了FCM算法的错误划分。

与传统阴影集的三划分相比,五区域阴影集是一种新型的阴影集,通过减小近似误差实现不确定性平衡,找到更合适的阈值。本文将五区域阴影集引入到FCM算法中,有效地提升FCM算法的精确度,提高聚类效果。

1 相关工作

聚类任务分为硬聚类和软聚类(模糊聚类)2种方式,而模糊聚类相比于传统的聚类可以更好地处理不确定性信息^[5]。学者们将三支决策理论应用于模糊聚类算法中^[6-7]。模糊聚类问题可以分为基于模糊关系、基于目标函数的优化、基于模糊规则学习3个分支,Bezdek^[8]研究了模糊C均值模型的一般情况并提出了FCM算法。如何优化FCM算法得到更好的聚类结果成为一个热点课题。Fan等^[9]提出了抑制模糊C均值聚类算法;Lin等^[10]提出了一种基于大小不敏感积分的模糊C均值方法来处理聚类大小敏感性问题;Zhou等^[11]提出了一种基于多颗粒近似区域和阴影集的粗略可能C均值聚类方法。

阴影集作为三支决策理论的一个重要应用在近年来成为了一个研究热点^[12-13]。Ibrahim等^[14]提出了阴影集的五维近似;Pedrycz等^[15]提出了一种基于不确定性平衡原理的阈值选择方法;Tahayori等^[16]介绍了一种利用熵的信息论阴影集近似方法来评价阴影集划分的性质;Ibrahim等^[17]利用熵优化原理对阴影集划分进行了研究;Zhang等^[18]提出了一种基于误差的优化方法,通过平衡阴影集中3个区域产生的误差来找到最优阈值。

2 预备知识

2.1 模糊C均值算法

FCM算法是一种模糊聚类算法,通过对对象与簇之间隶属度的关系完成划分。

在FCM算法中,通过最小化目标函数进行聚类。假设数据集 $X = \{x_1, x_2, \dots, x_n\}$ 由 k 个簇所构成,目标函数 J 为

$$J = \sum_{i=1}^n \sum_{j=1}^k \mu_{ij}^m x_i - c_j^2, \quad (1)$$

式中, m 为模糊化参数, $m=2$;簇的聚类中心为 c_j , μ_{ij} 为对象 x_i 与第 j 个类之间的隶属度,即

$$c_j = \frac{\sum_{i=1}^n \mu_{ij}^m x_i}{\sum_{i=1}^n \mu_{ij}^m}, \quad (2)$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^j (x_i - c_j / x_i - c_k)^{2/(m-1)}}. \quad (3)$$

初始化聚类中心 c_j 和隶属度矩阵 u 。根据式(2)、(3)更新聚类中心和隶属度,判断目标函数 $|J^{(t)} - J^{(t-1)}| < \epsilon$,满足条件终止并得到最终的隶属度矩阵 u ,否则重复更新聚类中心和隶属度,最后把对象划分到对应隶属度最高的簇。

2.2 三支决策理论

三支决策是基于接受、拒绝和待定的概念构建的一种决策理论^[19-21]。不同于以往的二元决策模型,在“非对即错”之间又增加了第3个选项。三支决策理论更加符合人类的思维方式,是一种合理的决策方式,三支决策理论的提出为研究划分边界模糊的数据提供了新的思路,使用三支决策思想构建的数学模型阴影集得到的三支聚类可以更好地处理含有不确定性的数据。

假定一个信息系统 $U = \{x_1, x_2, \dots, x_n\}$, $U = X$, 其中 $\mu(x)$ 为隶属度, 给定一对决策阈值 $[\alpha, \beta]$, 三支决策中的 3 个区域定义为:

$$P_{(\alpha, \beta)}(u) = \{x \in U \mid \mu(x) \geq \alpha\}, \quad (4)$$

$$N_{(\alpha, \beta)}(u) = \{x \in U \mid \mu(x) \leq \beta\}, \quad (5)$$

$$B_{(\alpha, \beta)}(u) = \{x \in U \mid \beta < \mu(x) < \alpha\}. \quad (6)$$

集合中大于等于 α 的部分划分到接受区域 P , 小于等于 β 的部分划分到拒绝区域 N , 小于 α 并且大于 β 的部分划分到待定区域 B 中。

2.3 阴影集及其五区域

阴影集是基于三支决策思想构建的一种数学模型, 通过确定 α, β 将模糊集的隶属度分为 3 个部分, 如图 1 所示。隶属度大于 α 的区域为 1, 表示该成属于该集群, 它是阴影集中区域 C 。隶属度小于 β 的区域减少到 0, 表示该成员不属于该集群, 它是阴影集中区域 N 。将隶属度大于 β 小于 α 的部分近似为 $[0, 1]$, 表示这些要划分的成员属于阴影区域, 即这 3 个区域的定义为:

$$C(S_3) = \{x \in f(x) \mid \mu(x) \in [\alpha, 1]\}, \quad (7)$$

$$S(S_3) = \{x \in f(x) \mid \mu(x) \in (\beta, \alpha)\}, \quad (8)$$

$$N(S_3) = \{x \in f(x) \mid \mu(x) \in [0, \beta]\}. \quad (9)$$

传统的阴影集通过不确定性平衡原理确定阈值 α 和 β ($\alpha > \beta$), 如图 1 所示, 阴影提升区域 Ω_3 为

$$\Omega_3 = \int_{x: f(x) \geq \alpha} (1 - f(x)) dx, \quad (10)$$

阴影区域 Ω_2 为

$$\Omega_2 = \int_{x: \beta < f(x) < \alpha} dx, \quad (11)$$

降低区域 Ω_1 为

$$\Omega_1 = \int_{x: f(x) \leq \beta} f(x) dx. \quad (12)$$

提升区域 Ω_3 对象的隶属度增大为 1, 降低区域 Ω_1 对象的隶属度减小为 0, 将这 2 个区域对象的不确定性消除。根据不确定性平衡原理, 生成一个阴影区域 Ω_2 , Ω_2 区域对象的隶属度近似为 $[0, 1]$ 。 Ω_2 区域不确定性平衡原理为

$$\int_{x: f(x) \geq \alpha} (1 - f(x)) dx + \int_{x: f(x) \leq \beta} f(x) dx = \int_{x: \beta < f(x) < \alpha} dx. \quad (13)$$

为了求解阈值 α 和 β , 构建目标函数

$$V(\alpha, \beta) = \left| \int_{x: f(x) \geq \alpha} (1 - f(x)) dx + \int_{x: f(x) \leq \beta} f(x) dx - \int_{x: \beta < f(x) < \alpha} dx \right|, \quad (14)$$

a, b, c 分别为提升区域、降低区域和阴影区域的离散形式, 即

$$a = \sum_{\mu(x) \leq \alpha} \mu(x),$$

$$b = \sum_{\mu(x) \geq \beta} (1 - \mu(x)),$$

$$c = \text{card}(\{x \in f(x) \mid \beta < \mu(x) < \alpha\}),$$

$$V(\alpha, \beta) = |a + b - c|.$$

当目标函数 $V(\alpha, \beta)$ 取最小值, 即阴影集中 3 个区域的不确定性达到平衡时, 对应的 $[\alpha, \beta]$ 是这 3 个区域的最优阈值, 即

$$\alpha, \beta = \text{argmin} \{V_{(\alpha, \beta)}\}, \quad (15)$$

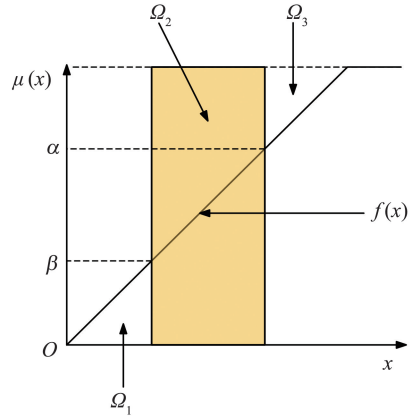


图 1 阴影集的示意图

Fig. 1 Schematic diagram of the shadowed sets

在实际情况中 $V(\alpha, \beta)$ 的值很难取到 0, 即很难达到不确定性平衡。利用五区域阴影集可以减小这种误差。根据图 2 所示, 在五区域阴影集中, 通过阈值 $[\alpha, \beta]$ 和 $[\gamma, \rho]$ (其中 $\alpha > \gamma > \rho > \beta$) 将模糊集划分为 5 个部分, 分别为核心区域 $R_{\text{cor}}(S_5)$ 、次核心区域 $R_{\text{scor}}(S_5)$ 、阴影区域 $R_{\text{sha}}(S_5)$ 、次边缘区域 $R_{\text{sneg}}(S_5)$ 、边缘区域 $R_{\text{neg}}(S_5)$, 即

$$R_{\text{cor}}(S_5) = \{x \in f(x) \mid \mu(x) \in [\alpha, 1]\}, \tag{16}$$

$$R_{\text{scor}}(S_5) = \{x \in f(x) \mid \mu(x) \in (\gamma, \alpha)\}, \tag{17}$$

$$R_{\text{sha}}(S_5) = \{x \in f(x) \mid \mu(x) \in [\rho, \gamma]\}, \tag{18}$$

$$R_{\text{sneg}}(S_5) = \{x \in f(x) \mid \mu(x) \in (\beta, \rho)\}, \tag{19}$$

$$R_{\text{neg}}(S_5) = \{x \in f(x) \mid \mu(x) \in [0, \beta]\}。 \tag{20}$$

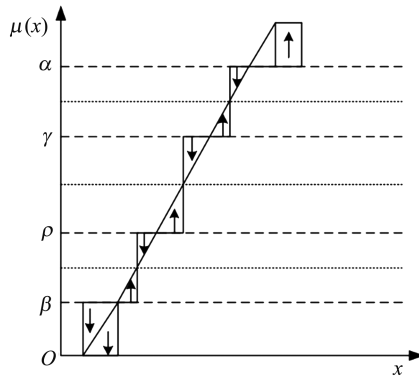


图 2 五区域阴影集
Fig.2 Five-region shadowed sets

核心区域中对象的隶属度增大为 1, 次核心区域中对象的隶属度近似为 0.75, 阴影区域中对象的隶属度近似为 0.5, 次边缘区域中对象的隶属度近似为 0.25, 边缘区域中对象的隶属度减小到 0, 即

$$R_{\text{cor}} = \{\mu(x) = 1 \mid \mu(x) \geq \alpha\},$$

$$R_{\text{scor}} = \{\mu(x) = 0.75 \mid \alpha > \mu(x) > \gamma\},$$

$$R_{\text{sha}} = \{\mu(x) = 0.5 \mid \gamma \geq \mu(x) \geq \rho\},$$

$$R_{\text{sneg}} = \{\mu(x) = 0.25 \mid \rho > \mu(x) > \beta\},$$

$$R_{\text{neg}} = \{\mu(x) = 0 \mid \mu(x) \leq \beta\}。$$

根据不确定性平衡原理, 五区域阴影集中不确定性的关系为

$$\sum_{x \in R_{\text{cor}}(S_5)} \mu(x) + \sum_{x \in R_{\text{scor}}(S_5)} \mu(x) + \sum_{x \in R_{\text{sha}}(S_5)} \mu(x) + \sum_{x \in R_{\text{sneg}}(S_5)} \mu(x) + \sum_{x \in R_{\text{neg}}(S_5)} \mu(x) = \varphi(x),$$

$$\varphi(x) = \varphi(R_{\text{scor}}(S_5)) + \varphi(R_{\text{sha}}(S_5)) + \varphi(R_{\text{sneg}}(S_5)),$$

其中

$$\varphi(R_{\text{sneg}}(S_5)) = \frac{1}{2} \text{Card}(R_{\text{sneg}}(S_5)),$$

$$\varphi(R_{\text{scor}}(S_5)) = \frac{1}{2} \text{Card}(R_{\text{scor}}(S_5)),$$

$$\varphi(R_{\text{sha}}(S_5)) = \frac{1}{2} \text{Card}(R_{\text{sha}}(S_5))。$$

由不确定性平衡原理构造目标函数

$$V(\alpha, \beta, \gamma, \rho) = \left| \sum_{x \in f(x)} \mu(x) - \varphi(x) \right|,$$

当目标函数的值最小时, 对应的阈值就为最优阈值, 即

$$\alpha, \beta, \rho, \gamma = \text{argmin} \{V_{(\alpha, \beta, \rho, \gamma)}\}。$$

3 基于五区域阴影集的 FCM 算法

3.1 基本框架

基于五区域阴影集的模糊 C 均值(fuzzy C -means based on five-regions shadowed sets, FCM-5SS)算法的流程框图如图3所示。通过 FCM 算法得到对象与簇之间的隶属度矩阵,由五区域阴影集通过对象的隶属度得到 2 对阈值 $[\alpha, \beta]$ 和 $[\gamma, \rho]$,通过 2 对阈值与隶属度对应的对象划分到核心区域、次核心区域、阴影区域、次边缘区域、边缘区域 5 个区域中,最后,将核心区域和次核心区域中隶属度 $\mu \geq \omega$ 的对象划分到该簇中,直到所有的对象被划分。

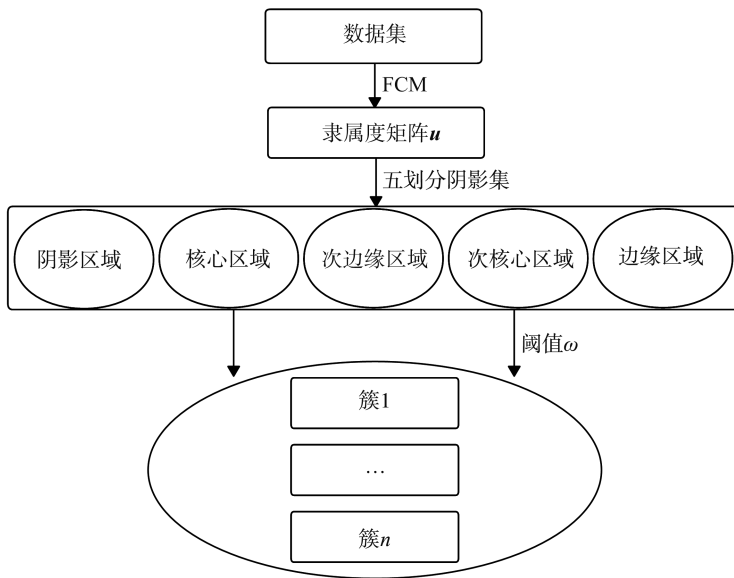


图3 FCM-5SS 算法的基本框架

Fig.3 Basic framework of FCM-5SS algorithm

3.2 将对象划分到 5 个区域

通过 FCM 算法可以得到对象与簇之间的隶属度关系,通过比较对象与每个簇之间隶属度的大小,最终选择相似度最高的簇作为划分依据。这种划分方式是一种绝对的划分,有时候对象对于几个簇之间的隶属度是相近的,导致很多对象被划分到错误的簇中,引入五区域阴影集可以解决这个问题。

五区域阴影集通过阈值 $[\alpha, \beta]$ 和 $[\gamma, \rho]$ 将对象划分到核心区域、次核心区域、阴影区域、次边缘区域、边缘区域 5 个区域中,隶属度 $\mu \geq \alpha$ 的区域划分为核心区域,隶属度 $\alpha > \mu > \gamma$ 的区域划分为次核心区域,隶属度 $\gamma \geq \mu \geq \rho$ 的区域划分为阴影区域,隶属度 $\rho > \mu > \beta$ 的区域划分为次边缘区域,隶属度 $\mu \leq \beta$ 的区域划分为边缘区域。根据以上规则,待划分的对象划分到 5 个区域中。

3.3 阈值 ω 的求解方法

核心区域中的对象是一定属于该簇,而次核心区域中的对象是大概率属于该簇,阴影区域中的对象可能属于该簇,次边缘区域中的对象小概率属于该簇,边缘区域中的对象一定不属于该簇。基于这种性质,将核心区域与次核心区域中的对象划归该簇是可行的,但次核心区域中的对象不一定完全属于该簇,因此对次核心区域进行再划分也是很有必要的。

根据五区域阴影集中次核心区域的定义可知,在次核心区域中对象的隶属度近似为 0.75,因此可以认为次核心区域中对象属于该簇的概率为 0.75,将次核心区域中靠近核心区域的 75%的区域划归核心区域,将剩余 25%的区域划归阴影区域,即

$$\omega = \alpha - 0.75(\alpha - \gamma). \quad (21)$$

3.4 基于阴影集的模糊 C 均值(fuzzy C -means based on shadowed sets, FCM-SS)算法和 FCM-5SS 算法

算法 1 FCM-SS 算法。

输入 聚类数量 k ,数据集 D ;

输出 聚类结果。

- ① 初始化 u ;
- ② 利用式(2)计算簇中心;
- ③ 根据式(3)更新 u ;
- ④ 根据式(1)计算 J ;
- ⑤ 如果目标函数 $|J^{(i)} - J^{(i-1)}| < \varepsilon$,终止并得到最终的隶属度矩阵,否则重复步骤②、③;
- ⑥ 利用式(7)—(9)将对象划分到 3 个区域中;
- ⑦ 将核心区域中的对象划归该簇,直到所有簇完成划分。

FCM-SS 算法使用传统阴影集对 FCM 算法进行改进,下面给出 FCM-5SS 算法。

算法 2 基于五区域阴影集的模糊 C 均值聚类算法。

输入 聚类数量,数据集 ;

输出 聚类结果。

- ① 初始化 u ;
- ② 利用式(2)计算簇中心;
- ③ 根据式(3)更新 u ;
- ④ 根据式(1)计算 J ;
- ⑤ 如果目标函数 $|J^{(i)} - J^{(i-1)}| < \varepsilon$,终止并得到最终的 u ,否则重复步骤②、③;
- ⑥ 利用式(16)—(20)将对象划分到 5 个区域中;
- ⑦ 将核心区域中的对象划归该簇,利用式(21)求解 ω ,将次核心区域中隶属度 $\mu \geq \omega$ 的对象划归该簇。

4 实验设置、结果及分析

4.1 实验设置

为了验证本文所提出的方法的有效性,使用了 8 个加州大学欧文分校 (University of California Irvine, UCI) 数据集如表 1 所示。

表 1 8 个 UCI 数据集的基本信息
Table 1 Basic information of 8 UCI data sets

数据集	特征值数量	对象数	分类数
Ionosphere	34	351	2
Wdbc	30	569	2
Heart	13	270	2
Banknote	4	1 372	2
Sonar	60	208	2
Iris	4	150	3
Wpdc	33	198	2
BCcoimbra	9	116	2

本文使用 3 种经典的聚类算法,包括一种硬聚类算法 (k -means) 和 2 种软聚类算法 (FCM 和 FCM-S) 作为对比算法,3 个聚类评估指标为聚类准确度、归一化互信息、调整兰德系数,验证本文方法的有效性,这 3 个指标越大表示算法的性能越好。当 FCM 算法的模糊参数 $m \in [1.5, 2.5]$ 时,能得到较好的聚类结果,本文中 FCM 算法的模糊参数设置为 2。

4.2 实验结果分析

表 2—4 分别为 4 种算法在 8 个数据集上所得到的聚类精确度、归一化互信息、调整兰德系数, k -means、FCM、FCM-SS、FCM-5SS 算法的聚类准确度分别为 $A_{k\text{-means}}$ 、 A_{FCM} 、 $A_{\text{FCM-SS}}$ 、 $A_{\text{FCM-5SS}}$, 归一化互信息分别为 $N_{k\text{-means}}$ 、 N_{FCM} 、 $N_{\text{FCM-SS}}$ 、 $N_{\text{FCM-5SS}}$, 调整兰德系数分别为 $I_{k\text{-means}}$ 、 I_{FCM} 、 $I_{\text{FCM-SS}}$ 、 $I_{\text{FCM-5SS}}$, 其中数值最大的项使用粗体

表示。FCM-5SS 相比于其余 3 种算法在 7 个数据集上取得了最大的聚类准确度、归一化互信息、调整兰德系数,FCM 算法在数据集 Wdbc 的取得的聚类结果最佳,本文所提出的算法在 8 个数据集上得到的平均聚类准确度、平均归一化互信息、平均调整兰德系数最大。实验结果显示,本文提出的算法优于其余 3 种算法,可以解决 FCM 算法在不能有效划分相似数据的问题。

表 2 8 个 UCI 数据集在 4 种算法中的聚类准确度
Table 2 Cluster accuracy of 8 UCI data sets on 4 algorithms

数据集	$A_{k\text{-means}}$	A_{FCM}	$A_{\text{FCM-SS}}$	$A_{\text{FCM-5SS}}$
Ionosphere	0.675 2	0.672 4	0.675 2	0.732 2
Wdbc	0.855 9	0.864 7	0.773 3	0.845 3
Heart	0.674 1	0.685 2	0.692 6	0.707 4
Banknote	0.650 9	0.658 2	0.653 2	0.680 0
Sonar	0.538 5	0.567 3	0.572 1	0.634 6
Iris	0.946 7	0.946 7	0.960 0	0.973 3
Wpdc	0.686 9	0.697 0	0.697 0	0.707 1
BCcoimbra	0.465 5	0.465 5	0.482 8	0.543 1
平均值	0.686 7	0.694 6	0.688 3	0.727 9

表 3 8 个 UCI 数据集在 4 种算法中的归一化互信息
Table 3 Normalized mutual information of 8 UCI data sets on 4 algorithms

数据集	$N_{k\text{-means}}$	N_{FCM}	$N_{\text{FCM-SS}}$	$N_{\text{FCM-5SS}}$
Ionosphere	0.069 9	0.066 5	0.069 9	0.179 4
Wdbc	0.434 0	0.445 9	0.294 6	0.420 9
Heart	0.112 4	0.095 9	0.102 4	0.122 2
Banknote	0.064 6	0.069 9	0.065 9	0.102 1
Sonar	0.019 0	0.028 8	0.031 6	0.064 3
Iris	0.836 6	0.844 9	0.864 2	0.901 1
Wpdc	0.008 3	0.019 7	0.022 8	0.042 3
BCcoimbra	0.003 5	0.003 5	0.001 3	0.002 8
平均值	0.193 5	0.196 9	0.181 6	0.229 4

表 4 8 个 UCI 数据集在 4 种算法中的调整兰德系数
Table 4 Adjusting the Rand coefficient of 8 UCI data sets on 4 algorithms

数据集	$I_{k\text{-means}}$	I_{FCM}	$I_{\text{FCM-SS}}$	$I_{\text{FCM-5SS}}$
Ionosphere	0.116 7	0.112 5	0.116 7	0.213 6
Wdbc	0.497 8	0.524 5	0.275 0	0.466 5
Heart	0.115 1	0.133 3	0.144 7	0.168 8
Banknote	0.090 2	0.099 3	0.092 1	0.129 0
Sonar	0.001 2	0.013 4	0.016 1	0.068 0
Iris	0.851 0	0.851 2	0.886 0	0.922 2
Wpdc	0.047 2	0.077 0	0.082 9	0.116 3
BCcoimbra	-0.003 9	-0.003 9	-0.007 5	-0.001 3
平均值	0.214 4	0.225 9	0.200 8	0.260 4

在五区域阴影集中, α 是划分核心区域和次核心区域的阈值, γ 是划分次核心区域和阴影区域的阈值, ω 是本文所提出的用于划分次核心区域的阈值。为了探究阈值(α 、 ω 、 γ)对聚类准确度的影响并验证本文提出的 ω 是否有效,设计了一组对照实验。对 8 个 UCI 数据集使用 FCM-5SS 算法,分别使用了 3 个阈值作为最终的划分标准。实验结果如图 4 所示。Ionosphere、Wdbc、Wpdc、Heart、Banknotes、BC coimbra 数据集使用阈值 α 得到的平均精度最低,当阈值为 γ 时,Sonor 数据集得到的平均精度最低,当阈值为 ω 时,8 个数据集上均得到最大的平均精度。实验结果证明了五区域阴影集中次核心区域是不能被忽略的,须要一个阈值来划分该区域以提升聚类准确率。

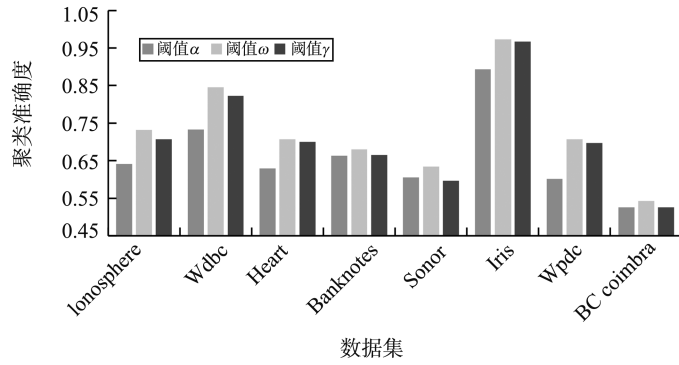


图 4 不同阈值对聚类准确度的影响

Fig.4 Influence of different thresholds on cluster accuracy

5 结论

本文提出了 FCM-5SS 算法,通过 FCM 算法得到对象与簇之间的隶属度关系,对隶属度使用五区域阴影集的方法,通过 2 对阈值 $[\alpha, \beta]$ 和 $[\gamma, \rho]$,将隶属度对应的对象划分到 5 个区域内,分别是核心区域、次核心区域、阴影区域、次边缘区域、边缘区域。其中核心区域中的对象是一定属于该簇的,次核心区域中的对象大概率属于该簇,而对象位于阴影区域、次边缘区域、边缘区域属于该簇的可能性较低,因此对次核心区域的划分进行了研究,找到了一个合适的阈值 ω 划分次核心区域。将划分好的次核心区域和核心区域中的对象划归该簇,得到最终的聚类结果。在 8 个数据集上进行实验,本文提出的 FCM-5SS 算法与 3 种算法进行对比,FCM-5SS 算法的聚类结果的聚类准确度、归一化互信息、调整兰德系数大多优于这 3 种算法。今后对于五区域阴影集中次核心区域的划分、FCM 算法的相似度求解方法等继续研究。

参考文献:

- [1] VERMA H, VERMA D, TIWARI P K. A population based hybrid FCM-PSO algorithm for clustering analysis and segmentation of brain image[J]. Expert Systems with Applications, 2021, 167:114121.
- [2] RUSOPINI E H, BEZDEK J C, KELLER J M. Fuzzy clustering: a historical perspective[J]. IEEE Computational Intelligence Magazine, 2019, 14(1):45-55.
- [3] ALMOMANY A, JARRAH A, AI ASSAF A. FCM clustering approach optimization using parallel high-speed intel FPGA technology[J]. Journal of Electrical and Computer Engineering, 2022, 2022.
- [4] SUGANYA R, SHANTHI R. Fuzzy C-means algorithm; a review [J]. International Journal of Scientific and Research Publications, 2012, 2(11):1-5.
- [5] JIANG Chunmao, LI Zhicong, YAO Jingtiao. A shadowed set-based three-way clustering ensemble approach[J]. International Journal of Machine Learning and Cybernetics, 2022, 13(9):2545-2558.
- [6] 凡嘉琛,王平心,杨习贝. 基于三支决策的密度敏感谱聚类[J]. 山东大学学报(理学版), 2023, 58(1):59-66.
FAN Jiachen, WANG Pingxin, YANG Xibei. Density-sensitive spectral clustering based on three-way decision[J]. Journal of Shandong University(Natural Science), 2023, 58(1):59-66.
- [7] 王君宇,杨亚锋,薛静轩,等. 可拓序贯三支决策模型及应用[J]. 山东大学学报(理学版), 2023, 58(7):67-79.
WANG Junyu, YANG Yafeng, XUE Jingxuan, et al. Extension sequential three-way decision model and its application[J]. Journal of Shandong University(Natural Science), 2023, 58(7):67-79.
- [8] BEZDEK J C. Cluster validity with fuzzy sets[J]. Journal of Cybernetics, 1973, 3(3):58-73
- [9] FAN Jiulun, ZHEN Wenzhi, XIE Weixin. Suppressed fuzzy C-means clustering algorithm[J]. Pattern Recognition Letters, 2003, 24(9/10):1607-1612.
- [10] LIN P L, HUANG P W, KUO C H, et al. A size-insensitive integrity-based fuzzy C-means method for data clustering[J]. Pattern Recognition, 2014, 47(5):2042-2056.
- [11] ZHOU Jie, ZHI Huilai, GAO Can, et al. Rough possibilistic C-means clustering based on multigranulation approximation

- regions and shadowed sets[J]. Knowledge-based Systems, 2018, 160:144-166.
- [12] PEDRYEZ W. Shadowed sets: representing and processing fuzzy sets[J]. IEEE Transactions on Systems, Man, and Cybernetics(Part B: Cybernetics), 1998, 28(1):103-109.
- [13] 姜春茂,赵书宝. 基于阴影集的多粒度三支聚类集成[J]. 电子学报, 2021, 49(8):1524-1532.
JIANG Chunmao, ZHAO Shubao. Multi-granulation three-way clustering ensemble based on shadowed sets[J]. Acta Electronica Sinica, 2021, 49(8):1524-1532.
- [14] IBRAHIM M A, WILLIAN-WEST T O, KANA A F D, et al. Shadowed sets with higher approximation regions[J]. Soft Computing, 2020, 24(22):17009-17033.
- [15] PEDRYEZ W, VUKOVICH G. Granular computing with shadowed sets[J]. International Journal of Intelligent Systems, 2002, 17(2):173-197.
- [16] TAHAYORI H, SADEGHIAN A, PEDRYCZ W. Induction of shadowed sets based on the gradual grade of fuzziness[J]. IEEE Transactions on Fuzzy Systems, 2013, 21(5):937-949.
- [17] IBRAHIM A M, WILLIAM-WEST T O. Induction of shadowed sets from fuzzy sets[J]. Granular Computing, 2019, 4(1):27-38.
- [18] ZHANG Yan, YAO Jingtao. Game theoretic approach to shadowed sets: a three-way tradeoff perspective[J]. Information Sciences, 2020, 507:540-552.
- [19] YAO Yiyu. Three-way decision and granular computing[J]. International Journal of Approximate Reasoning, 2018, 103:107-123.
- [20] LI Xiaonan, WANG Xuan, LANG Guangming, et al. Conflict analysis based on three-way decision for triangular fuzzy information systems[J]. International Journal of Approximate Reasoning, 2021, 132:88-106.
- [21] LI Xiaonan, WANG Xuan, SUN Bingzhen, et al. Three-way decision on information tables[J]. Information Sciences, 2021, 545:25-43.

(编辑:陈丽萍)

(上接第73页)

- [30] XU Bin, CHEN Zhenzhou. Multi-level fusion based 3D object detection from monocular images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018:2345-2353.
- [31] SHI X P, CHEN Z X, KIM T K. Distance-normalized unified representation for monocular 3D object detection[C]//2020 16th European Conference on Computer Vision (ECCV). Berlin: Springer, 2020:91-107.
- [32] CAI Yingjie, LI Buyu, JIAO Zeyu, et al. Monocular 3D object detection with decoupled structured polygon estimation and height-guided depth estimation[C]//2020 34th AAAI Conference on Artificial Intelligence (AAAI-20). New York: AAAI Press, 2020:10478-10485.
- [33] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami: IEEE, 2009:248-255.
- [34] DAI Jifeng, QI Haozhi, XIONG Yuwen, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE Computer Society, 2017:764-773.

(编辑:陈丽萍)