

# 基于改进粒子群优化的移动界面模式聚类算法

贾 伟<sup>1,2</sup> 华庆一<sup>1</sup> 张敏军<sup>1</sup> 陈 锐<sup>1</sup> 姬 翔<sup>1</sup> 王 博<sup>1,3</sup>

(西北大学信息科学与技术学院 西安 710127)<sup>1</sup> (宁夏大学新华学院 银川 750021)<sup>2</sup>

(西安邮电大学计算机学院 西安 710121)<sup>3</sup>

**摘 要** 聚类是一种非常有效的信息分析方法。针对现有基于粒子群优化的模糊 C 均值(Fuzzy C-means,FCM)聚类算法的聚类效果不佳的问题,提出一种基于改进粒子群优化的模糊 C 均值聚类算法,并将该聚类算法应用到移动界面模式的聚类中。首先,利用直觉模糊熵的几何解释和约束构造合理的直觉模糊熵;然后,在粒子群优化中使用直觉模糊熵判断种群的多样性程度,并引入混沌反向学习策略来提高全局搜索能力;最后,为了增强聚类算法的非线性处理能力,在聚类算法中加入高斯核函数,并将该聚类算法应用到移动界面模式的聚类中。移动界面模式聚类的实验表明,与现有聚类算法相比,文中所提聚类算法具有更好的聚类效果。

**关键词** 粒子群优化,移动界面模式,聚类,直觉模糊熵,混沌反向学习

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.04.037

## Mobile Interface Pattern Clustering Algorithm Based on Improved Particle Swarm Optimization

JIA Wei<sup>1,2</sup> HUA Qing-yi<sup>1</sup> ZHANG Min-jun<sup>1</sup> CHEN Rui<sup>1</sup> JI Xiang<sup>1</sup> WANG Bo<sup>1,3</sup>

(School of Information Science and Technology, Northwest University, Xi'an 710127, China)<sup>1</sup>

(Xinhua College, Ningxia University, Yinchuan 750021, China)<sup>2</sup>

(School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)<sup>3</sup>

**Abstract** Clustering is a very efficient method for analyzing information. Focusing on the issue that clustering results of the existing fuzzy C-means clustering algorithms based on particle swarm optimization are not good, a fuzzy C-means clustering algorithm based on improved particle swarm optimization was proposed and applied in mobile interface pattern clustering. Firstly, reasonable intuitionistic fuzzy entropy is constructed by using the geometric interpretation and the constraints of intuitionistic fuzzy entropy. Secondly, in the improved particle swarm optimization, the intuitionistic fuzzy entropy is used to measure the state of particle swarm, and chaotic opposition-based learning is used to improve the global search ability. Finally, the proposed algorithm employs the Gauss kernel function for enhancing nonlinear processing capability, and then it is applied in mobile interface pattern clustering. Experimental results show that the proposed clustering algorithm has better performance in mobile interface pattern clustering than the existing clustering algorithms.

**Keywords** Particle swarm optimization, Mobile interface pattern, Clustering, Intuitionistic fuzzy entropy, Chaotic opposition-based learning

## 1 引言

模糊 C 均值聚类作为一种经典的聚类算法,由于能够使用模糊逻辑有效处理聚类数据的重叠情况<sup>[1]</sup>,引起了研究者的广泛关注。为了克服 FCM 聚类对初始值敏感和易陷入局部极小值的缺点,可以利用粒子群优化算法确定最佳的初始聚类中心<sup>[2]</sup>。为了避免传统粒子群优化算法在寻优过程中出

现局部极值点的问题,一些研究用改进的粒子群优化算法来确定初始聚类中心。文献[3]给出的 FCM 聚类算法使用自适应调节粒子群参数的方法来提高粒子群搜索的方向性和收敛速度。但是该算法没有考虑在粒子群搜索过程中如何增强种群的多样性,导致优化的结果不准确,影响了 FCM 聚类的准确性。文献[4]给出的 FCM 聚类算法利用直觉模糊熵判断粒子群的种群状态,提高种群的多样性。然而,其使用的

到稿日期:2017-02-14 返修日期:2017-05-11 本文受国家自然科学基金资助项目(61272286),高等学校博士学科点专项科研基金资助项目(20126101110006),陕西省工业科技攻关项目(2016GY-123),西北大学科学研究基金资助项目(15NW31)资助。

贾 伟(1980—),男,博士,副教授,CCF 会员,主要研究方向为人机交互、用户界面工程;华庆一(1956—),男,博士,教授,CCF 会员,主要研究方向为人机交互、用户界面工程,E-mail:huaqy@nwu.edu.cn(通信作者);张敏军(1979—),男,博士,讲师,主要研究方向为人机交互、用户界面工程;陈 锐(1979—),男,博士,CCF 会员,主要研究方向为人机交互、用户界面工程;姬 翔(1979—),女,博士,讲师,CCF 会员,主要研究方向为人机交互、用户界面工程;王 博(1976—),男,博士,讲师,主要研究方向为人机交互、用户界面工程。

直觉模糊熵无法合理度量数据的模糊程度,在判断种群状态时还存在一些误判,遗漏了一些应该扩大搜索范围的种群,容易出现局部极小值;此外,该算法的搜索策略所产生的搜索范围较小,影响了粒子群优化算法的全局搜索能力。为了提高 FCM 聚类的准确性,本文给出一种改进粒子群优化的 FCM 聚类算法,该算法利用直觉模糊熵的几何解释和约束构造合理的直觉模糊熵,在粒子群优化中将直觉模糊熵和混沌反向学习策略相结合,进一步扩大了种群搜索范围,避免了粒子群陷入局部极值点。为了验证本文算法的有效性,将其与现有聚类算法应用到移动界面模式的聚类中,从聚类效果和收敛速度两个方面对各聚类方法进行比较。

本文第 2 节介绍使用到的相关知识;第 3 节利用几何解释构造直觉模糊熵,并给出一个基于直觉模糊熵和混沌反向学习的粒子群优化算法;第 4 节将基于改进粒子群优化的聚类算法应用到移动界面模式聚类中;第 5 节将本文提出的聚类算法与现有的聚类算法进行比较;最后总结全文并展望后续研究方向。

## 2 相关知识

### 2.1 模糊 C 均值聚类算法

FCM 聚类算法将  $D$  维欧几里得空间( $x_i \in \mathbb{R}^D$ )中包含的  $n$  个数据  $x_1, x_2, \dots, x_n$  划分到  $c$  个聚类中。FCM 聚类算法优化的目标函数为:

$$J(U, B) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|y_k - b_i\|^2 \quad (1)$$

其中,  $m$  是权重指数,  $B = (b_1, b_2, \dots, b_c)$  是聚类中心集合,  $Y = (y_1, y_2, \dots, y_n)$  是数据向量,  $U = [u_{ik}]_{c \times n}$  是模糊划分矩阵,  $u_{ik}$  是  $y_k$  隶属于以  $b_i$  为中心的聚类簇的隶属度。

### 2.2 粒子群优化算法

粒子群优化算法是一种基于种群智能的随机全局优化算法,能够解决多峰、非凸和不连续问题<sup>[5]</sup>。与遗传算法相比,粒子群优化算法具有记忆功能以及参数设置少和收敛速度快的优点,成为了优化算法领域的研究热点。

在标准的粒子群优化算法中,采用下式更新粒子的速度:

$$v_{ij}(t+1) = \omega v_{ij}(t) + c_1 r_1 (p_{ij} - x_{ij}(t)) + c_2 r_2 (p_{gj} - x_{ij}(t)) \quad (2)$$

其中,  $c_1$  和  $c_2$  是学习因子,  $r_1$  和  $r_2$  是  $[0, 1]$  之间的随机数,  $\omega$  是惯性因子。

更新粒子的速度后,采用下式更新粒子的位置:

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (3)$$

### 2.3 直觉模糊熵

直觉模糊熵<sup>[6-8]</sup>是直觉模糊集的重要研究内容之一,其目的是描述直觉模糊集的模糊程度。直觉模糊熵的具体定义如下:

**定义 1** 设  $X$  为一个给定论域,  $A = \{\langle x, \mu_A(x), \gamma_A(x) \rangle | x \in X\}$  为  $X$  上的一个直觉模糊集,记为  $IFS_S(x)$ ,其中:  $\mu_A: X \rightarrow [0, 1]$  和  $\gamma_A: X \rightarrow [0, 1]$  分别为  $X$  中的元素  $x$  的隶属度和非隶属度。已知信息由隶属度和非隶属度组成,而且对于任意的  $x \in X$ ,有  $0 \leq \mu_A(x) + \gamma_A(x) \leq 1$  成立,则称  $\pi_A(x) = 1 - \mu_A(x) - \gamma_A(x)$  为直觉模糊集  $A$  中元素  $x$  的犹豫度或直觉指数,犹豫度表示未知信息。

如果  $A, B \in IFS_S(x)$ ,则有如下运算:

1)  $A \subseteq B$ , 当且仅当  $\forall x \in X$ , 有  $\mu_A(x) \leq \mu_B(x), \gamma_A(x) \geq \gamma_B(x)$ ;

2)  $A = B$ , 当且仅当  $\forall x \in X, \mu_A(x) = \mu_B(x), \gamma_A(x) = \gamma_B(x)$ ;

3)  $A^c = \{\langle x, \gamma_A(x), \mu_A(x) \rangle | x \in X\}$ 。

为使直觉模糊熵的公理化定义更加合理,一些学者围绕公理化定义展开研究<sup>[9-11]</sup>。本文结合现有的研究,给出如下直觉模糊的熵公理化定义。

**定义 2** 实函数  $E: IFS_S(x) \rightarrow [0, 1]$  满足以下条件,则称为  $IFS_S(x)$  的熵:

1)  $E(A) = 0$ , 当且仅当  $A$  是一个分明集。

2)  $E(A) = 1$ , 当且仅当  $\forall x \in X, \mu_A(x) = \gamma_A(x) = 0$ 。

3) 对于  $\forall x \in X$ , 当  $\pi_A(x) = \pi_B(x)$  时, 有  $\mu_A(x) \leq \mu_B(x) \leq \gamma_B(x) \leq \gamma_A(x)$  或者  $\mu_A(x) \geq \mu_B(x) \geq \gamma_B(x) \geq \gamma_A(x)$ , 则  $E(A) \leq E(B)$ ; 当  $\pi_A(x) \leq \pi_B(x)$  且  $|\mu_A(x) - \gamma_A(x)| = |\mu_B(x) - \gamma_B(x)|$  时, 有  $E(A) \leq E(B)$ 。

4)  $E(A) = E(A^c)$ 。

## 3 改进的粒子群优化算法

### 3.1 直觉模糊熵的改进

近年来,为了有效度量直觉模糊集的模糊程度,很多学者围绕直觉模糊熵的构造展开研究。例如,文献[4]中使用的王毅等<sup>[12]</sup>提出直觉模糊熵  $E_{WY}$  来判断粒子群的多样性程度; Verma 等<sup>[13]</sup>提出了  $E_V$ ; Ye 等<sup>[14]</sup>提出了  $E_Y$ ; 魏翠萍等<sup>[15]</sup>利用余弦函数提出了  $E_{WCP}$ ; 王坚强等<sup>[16]</sup>利用余切函数提出了  $E_{WJQ}$ ; 高明美等<sup>[17]</sup>分析了隶属度和非隶属度的绝对差异的平方项以及犹豫度平方项的作用后,提出了  $E_{GMM}$ ; 刘满凤等<sup>[18]</sup>考虑了隶属度和非隶属度的差异相同且存在犹豫度的情况,提出了  $E_{LMF}$ ; 赵飞等<sup>[19]</sup>在研究了直觉模糊熵的约束条件和一些度量的实际情况后,提出了  $E_{ZF}$ 。然而,现有的构造方法由于没有全面考虑影响模糊程度的因素,对已知信息和未知信息处理得不合理,导致构造出来的直觉模糊熵不能有效且合理地反映直觉模糊集的模糊程度。

直觉模糊熵的几何解释如图 1 所示,任意一个直觉模糊集  $A$  的每一个元素都能映射到等边三角形  $\triangle ABD$  的平面中的某一个点上,可以看出,只有当  $\mu_A(x) = \nu_A(x) = 0, \pi_A(x) = 1$  时,  $E(A) = 1$ ; 只有当  $\mu_A(x) = 1, \gamma_A(x) = 0, \pi_A(x) = 0$  时, 或者当  $\mu_A(x) = 0, \gamma_A(x) = 1, \pi_A(x) = 0$  时,  $E(A) = 0$ 。这与定义 2 的条件 1) 和条件 2) 一致。

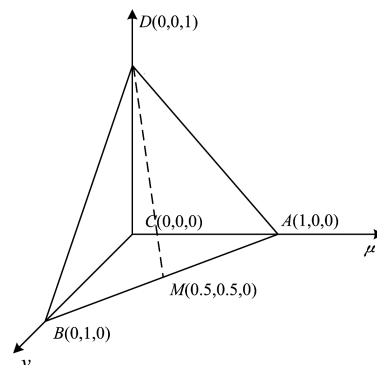


图 1 直觉模糊熵的几何解释

Fig. 1 Geometric interpretation of intuitionistic fuzzy entropy

过  $D$  点向线段  $AB$  做垂线,并相交于  $M$  点,得到  $\triangle ABD$  的中垂线  $DM$ ,此时  $DM$  上的任意一点的  $\mu_A(x)$  和  $\gamma_A(x)$  相同, $DM$  的左右两边互为补集。由于已知信息,犹豫度、隶属度和非隶属度之间的差异度这 3 个方面的值都会影响直觉模糊熵,只有当  $\triangle ABD$  中的两个点的以上 3 个值都分别相等时,两个点的直觉模糊熵才相等,显然只有  $DM$  的左右两边互为补集的两个点的直觉模糊熵相等。随着  $M$  点向  $D$  点方向移动,犹豫度逐渐增大,直觉模糊熵的值也逐渐增大。

通过上述分析,得到以下 3 点直觉模糊熵的约束:

1) 在  $\triangle ABD$  中,只存在一个直觉模糊熵等于 1 的点,即  $\mu_A(x) = \gamma_A(x) = 0, \pi_A(x) = 1$  时,  $E(A) = 1$ ;

2) 在等边三角形  $\triangle ABD$  的平面中,与  $E(A)$  的值相等的只有  $E(A^c)$ ;

3) 在中垂线  $DM$  中,越接近  $D$  点,直觉模糊熵越大。

此外,从图 1 的几何解释中还可以看出隶属度、非隶属度和犹豫度对模糊程度有以下 3 点影响:

1) 犹豫度越大,表明不确定性越大,模糊程度越强;

2) 隶属度和非隶属度越大,表明已知信息越大,模糊程度越弱;

3) 隶属度与非隶属度的差异越大,表明不确定性越小,则模糊程度越弱。

因此,一个有效的直觉模糊熵应该同时满足以下 3 个条件:

- 1) 满足公理化定义;
- 2) 符合几何解释中的约束;
- 3) 符合隶属度、非隶属度和犹豫度对模糊程度的影响。

现有的直觉模糊熵虽然满足公理化定义,但是在形式上不满足条件 2) 或条件 3),造成无法有效区分不同数据的模糊程度。本文充分考虑上述 3 个条件,提出一种改进的直觉模糊熵度量方法。

**定理 1** 设论域  $X = \{x_1, x_2, \dots, x_n\}, A = \{\langle x, \mu_A(x_i), \gamma_A(x_i) \rangle | x \in X\} \in IFS(X)$ , 则

$$E(A) = \frac{1}{n} \sum_{i=1}^n \frac{2 - 2|\mu_A(x_i) - \gamma_A(x_i)| + \pi_A^2(x_i)}{3 + \mu_A(x_i) + \gamma_A(x_i) + |\mu_A(x_i) - \gamma_A(x_i)|} \quad (4)$$

是直觉模糊集  $A$  的直觉模糊熵。

证明:下面证明式(4)满足直觉模糊熵的公理化定义。

1) 当  $A$  是分明集时,  $\mu_A(x_i) = 1, \gamma_A(x_i) = 0, \pi_A(x_i) = 0$ , 或者  $\mu_A(x_i) = 0, \gamma_A(x_i) = 1, \pi_A(x_i) = 0$ , 得到  $E(A) = 0$ 。

反之,当  $E(A) = 0$  时, 则  $(2 - 2|\mu_A(x_i) - \gamma_A(x_i)| + \pi_A^2(x_i)) / (3 + \mu_A(x_i) + \gamma_A(x_i) + |\mu_A(x_i) - \gamma_A(x_i)|) = 0$ , 得到  $2 - 2|\mu_A(x_i) - \gamma_A(x_i)| + \pi_A^2(x_i) = 0$ , 即  $2(1 - |\mu_A(x_i) - \gamma_A(x_i)|) + \pi_A^2(x_i) = 0$ 。因为  $0 \leq 2(1 - |\mu_A(x_i) - \gamma_A(x_i)|) \leq 2, 0 \leq \pi_A^2(x_i) \leq 1$ , 所以  $2(1 - |\mu_A(x_i) - \gamma_A(x_i)|) = 0$  且  $\pi_A^2(x_i) = 0$ , 得到  $\mu_A(x_i) = 1, \gamma_A(x_i) = 0, \pi_A(x_i) = 0$ , 或者  $\mu_A(x_i) = 0, \gamma_A(x_i) = 1, \pi_A(x_i) = 0$ , 则  $A$  是分明集。

2) 当  $E(A) = 1$  时,  $(2 - 2|\mu_A(x_i) - \gamma_A(x_i)| + \pi_A^2(x_i)) / (3 + \mu_A(x_i) + \gamma_A(x_i) + |\mu_A(x_i) - \gamma_A(x_i)|) = 1$ , 化简后得到  $1 + \mu_A(x_i) + \gamma_A(x_i) + 3|\mu_A(x_i) - \gamma_A(x_i)| = \pi_A^2(x_i)$ 。因为  $1 \leq 1 + \mu_A(x_i) + \gamma_A(x_i) + 3|\mu_A(x_i) - \gamma_A(x_i)|, 0 \leq \pi_A(x_i) \leq 1$ , 所以  $\mu_A(x_i) = \gamma_A(x_i) = 0$ 。

反之,当  $\mu_A(x_i) = \gamma_A(x_i) = 0$  时,显然  $E(A) = 1$ 。

3) 当  $\pi_A(x_i) = \pi_B(x_i)$  时, 有  $\mu_A(x_i) \leq \mu_B(x_i) \leq \gamma_B(x_i) \leq \gamma_A(x_i)$  或者  $\mu_A(x_i) \geq \mu_B(x_i) \geq \gamma_B(x_i) \geq \gamma_A(x_i)$ , 得到  $1 - |\mu_A(x_i) - \gamma_A(x_i)| \leq 1 - |\mu_B(x_i) - \gamma_B(x_i)|, \mu_A(x_i) + \gamma_A(x_i) = \mu_B(x_i) + \gamma_B(x_i)$ , 则  $2 - 2|\mu_A(x_i) - \gamma_A(x_i)| + \pi_A^2(x_i) \leq 2 - 2|\mu_B(x_i) - \gamma_B(x_i)| + \pi_B^2(x_i)$ , 且  $3 + \mu_A(x_i) + \gamma_A(x_i) + |\mu_A(x_i) - \gamma_A(x_i)| \geq 3 + \mu_B(x_i) + \gamma_B(x_i) + |\mu_B(x_i) - \gamma_B(x_i)|$ , 得到  $E(A) \leq E(B)$ ;

当  $\pi_A(x_i) \leq \pi_B(x_i)$  且  $|\mu_A(x_i) - \gamma_A(x_i)| = |\mu_B(x_i) - \gamma_B(x_i)|$  时, 得到  $\mu_A(x_i) + \gamma_A(x_i) \geq \mu_B(x_i) + \gamma_B(x_i)$ , 则  $2 - 2|\mu_A(x_i) - \gamma_A(x_i)| + \pi_A^2(x_i) \leq 2 - 2|\mu_B(x_i) - \gamma_B(x_i)| + \pi_B^2(x_i)$ , 且  $3 + \mu_A(x_i) + \gamma_A(x_i) + |\mu_A(x_i) - \gamma_A(x_i)| \geq 3 + \mu_B(x_i) + \gamma_B(x_i) + |\mu_B(x_i) - \gamma_B(x_i)|$ , 得到  $E(A) \leq E(B)$ 。

4)  $A^c = \{\langle x_i, \gamma_A(x_i), \mu_A(x_i) \rangle | x_i \in X\}, \pi_{A^c}(x_i) = \pi_A(x_i)$ , 因此  $E(A^c) = \frac{1}{n} \sum_{i=1}^n (2 - 2|\gamma_A(x_i) - \mu_A(x_i)| + \pi_A^2(x_i)) / (3 + \gamma_A(x_i) + \mu_A(x_i) + |\gamma_A(x_i) - \mu_A(x_i)|) = \frac{1}{n} \sum_{i=1}^n \frac{2 - 2|\mu_A(x_i) - \gamma_A(x_i)| + \pi_A^2(x_i)}{3 + \mu_A(x_i) + \gamma_A(x_i) + |\mu_A(x_i) - \gamma_A(x_i)|} = E(A)$ , 即  $E(A) = E(A^c)$ 。证毕。

通过上述证明可以看出,本文提出的直觉模糊熵满足定义 2 中的 4 个条件。

设直觉模糊集  $A_1 = \{\langle x, 0.5, 0.4 \rangle | x \in X\}, A_2 = \{\langle x, 0.2, 0.1 \rangle | x \in X\}, A_3 = \{\langle x, 0.2, 0.6 \rangle | x \in X\}, A_4 = \{\langle x, 0.5, 0.5 \rangle | x \in X\}$ , 从直觉模糊熵的几何解释可以看出,这些直觉模糊集的熵是不相等的。

表 1 是本文提出的直觉模糊熵  $E$  与现有直觉模糊熵度量  $A_1, A_2, A_3$  和  $A_4$  的比较。

表 1 各直觉模糊熵的计算结果

Table 1 Calculation results of different entropy measures

方法	$A_1$	$A_2$	$A_3$	$A_4$
$E_{WY}$	0.8333	0.8889	0.5000	0.5000
$E_V$	0.9905	0.9905	0.8463	0.7588
$E_Y$	0.9881	0.9881	0.8318	0.7391
$E_{WCP}$	0.9898	0.9957	0.8662	0.8662
$E_{WJQ}$	0.8673	0.9124	0.5781	0.5781
$E_{GMM}$	0.5000	0.7400	0.4400	0.5000
$E_{LMF}$	1.0000	0.9999	0.9995	0.9980
$E_{ZF}$	0.8310	0.8358	0.5774	0.5774
$E$	0.4525	0.6735	0.2952	0.3125

通过比较可以看出,现有的直觉模糊熵无法对所有数据进行有效区分,会导致度量结果不合理。而本文提出的直觉模糊熵  $E$  能够有效且合理地度量直觉模糊集的模糊程度,这是因为本文提出的直觉模糊熵在满足公理化定义的同时,在形式上体现出了几何解释中对直觉模糊熵的约束。本文提出的直觉模糊熵将隶属度与非隶属度的差异表示为  $|\mu_A(x_i) - \gamma_A(x_i)|$ , 犹豫度表示为  $\pi_A^2(x_i)$ 。从式(4)可以看出,  $\pi_A^2(x_i)$  的值越大时,得到的直觉模糊熵的值越大;而当隶属度  $\mu_A(x_i)$ 、非隶属度  $\gamma_A(x_i)$  和  $|\mu_A(x_i) - \gamma_A(x_i)|$  的值越大时,得到的直觉模糊熵的值越小,这说明本文提出的直觉模糊熵符合隶属度、非隶属度和犹豫度对模糊程度的影响规律。

### 3.2 粒子群优化算法的改进

粒子群优化算法虽然具有记忆功能以及设置参数少和收敛速度快的优点,但是易陷入局部极值点,从而出现粒子早熟现象。为了解决上述问题,Gao 等<sup>[20]</sup>使用混沌反向学习策略产生粒子群优化算法的初始种群,其目的是增加种群的多样性,提高全局搜索能力。本文将改进的直觉模糊熵和混沌反向学习策略引入到粒子群优化算法中,根据改进的直觉模糊熵对种群多样性程度的判断,使用混沌反向学习策略增加了种群的多样性。鉴于文献<sup>[20]</sup>中使用的 Sinusoidal 映射的混沌范围具有局限性和均匀分布特性较差的问题,本文采用改进的一维混沌映射 Ten-Sine 系统<sup>[21]</sup>:

$$Z_{n+1} = \begin{cases} \text{mod}(\frac{rZ_n}{2} + \frac{(4-r)(\sin(\pi Z_n))}{4}, 1), & Z_n < 0.5 \\ \text{mod}(\frac{r(1-Z_n)}{2} + \frac{(4-r)(\sin(\pi Z_n))}{2}, 1), & Z_n \geq 0.5 \end{cases} \quad (5)$$

其中,  $r \in (0, 4]$ 。

本文算法中的粒子适应度函数为:

$$f(x) = \frac{1}{1+J(U, B)} \quad (6)$$

下面给出混沌反向学习算法的描述。

#### 算法 1 混沌反向学习算法

输入:空间维度 D,设置混沌迭代的最大迭代次数 M 和种群大小 S

输出:初始种群 P

begin

for i=1 to S do

for j=1 to D do

随机初始化混沌变量  $Z_{0,j} \in (0, 1)$ ;

for k=1 to M do

$Z_{k,j} = Z_{k-1,j}$ ; //利用式(5)计算混沌映射

end for

$x_{i,j} = x_{\min,j} + Z_{k,j}(x_{\max,j} - x_{\min,j})$ ;

//得到第 j 维第 i 个粒子的值  $x_{i,j}$

end for

for i=1 to S do

for j=1 to D do

$ox_{i,j} = x_{\min,j} + x_{\max,j} - x_{i,j}$ ;

//得到第 j 维第 i 个粒子的反向值  $ox_{i,j}$

end for

end for

利用式(6)从种群和反向种群中选择最优的 S 个粒子作为初始种群 P;

end

本文通过混沌反向学习策略产生粒子群优化算法的初始种群,利用改进的直觉模糊熵判断粒子群的种群多样性程度。若直觉模糊熵低于阈值  $\beta$ ,说明当前的粒子群局限在一个较小的范围内搜索,则利用混沌反向学习策略得到当前种群的反向种群,从当前种群和反向种群中选取最优的 S 个粒子组成新的种群继续搜索。

下面给出采用改进直觉模糊熵和混沌反向学习的粒子群优化算法的描述。

#### 算法 2 改进的粒子群优化算法

输入:设置粒子的最大速度  $v_{\max}$ ,最大惯性权重  $w_{\max}$ ,最小惯性权重

$w_{\min}$ ,最大迭代次数  $l_{\max}$ ,学习因子  $c_1$  和  $c_2$ ,随机数  $r_1$  和  $r_2$ ,  $l=1$

输出:最优聚类中心  $P_g$

begin

利用算法 1 得到 D 维空间中有 S 个粒子的初始种群 P;

计算初始种群每个粒子的适应度函数值,得到粒子的历史最优位置和种群的最优位置;

while( $l \leq l_{\max}$ )

for i=1 to S do

for d=1 to D do

利用式(2)更新粒子的速度;

利用式(3)更新粒子的位置;

end for

end for

利用式(4)计算种群的直觉模糊熵 E;

if  $E \leq \beta$

计算当前种群的反向种群 OP;

利用式(6)从 P 和 OP 中选择最优的 S 个粒子形成新的种群

NP;

end if

利用式(6)计算当前种群每个粒子的适应度函数值;

更新  $P_i$  和  $P_g$ ,  $l=l+1$ ;

end while

end

## 4 聚类算法在移动界面模式中的应用

在移动界面的开发过程中,开发人员以移动界面模式的形式存储解决问题时积累的成功知识和经验<sup>[22-23]</sup>。随着移动界面开发的持续发展,移动界面模式的数量不断增长,这些模式所包含的信息超出了开发人员所能处理的有效范围。聚类算法在数据的分析和使用过程中具有重要作用<sup>[24-25]</sup>,通过聚类可以分析数据属性之间的相互关系,帮助开发人员对不同类别的移动界面模式进行有效的分类和管理。

为了验证本文聚类算法的有效性,将其应用到移动界面模式聚类中,给出基于改进粒子群优化的移动界面模式聚类算法。由于 FCM 聚类不能对非超球体数据结构进行有效聚类,为了获得更好的移动界面模式聚类效果,可以将核函数应用到 FCM 聚类中<sup>[26]</sup>,以提高非线性处理能力。本文在移动界面模式聚类算法中采用高斯核函数将数据非线性映射到高维空间,得到 FCM 聚类的目标函数:

$$J(U, B) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \| \Phi(y_k) - \Phi(b_i) \|^2 \quad (7)$$

高斯函数为:

$$K(x, y) = \exp(-\frac{\|x-y\|^2}{2\sigma^2}) \quad (8)$$

采用核函数取代 FCM 距离中的欧位函数后,得到:

$$\| \Phi(y_k) - \Phi(b_i) \|^2 = K(y_k, y_k) - 2K(y_k, b_i) + K(b_i, b_i) \quad (9)$$

将式(8)代入式(9),得到  $K(y_k, y_k) = K(b_i, b_i) = 1$ 。则式(7)可以转化为:

$$J(U, B) = 2 \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m [1 - K(y_k, b_i)] \quad (10)$$

利用拉格朗日乘子法求解式(10),得到聚类中心  $b_i$  和隶属度  $u_{ik}$  的迭代更新表达式:

$$b_i = \frac{\sum_{k=1}^n u_{ik}^m K(y_k, b_i) y_k}{\sum_{k=1}^n u_{ik}^m K(y_k, b_i)}, 1 \leq i \leq c \quad (11)$$

$$u_{ik} = \frac{(1 - K(y_k, b_i))^{\frac{-1}{m-1}}}{\sum_{j=1}^c (1 - K(y_k, b_j))^{\frac{-1}{m-1}}}, 1 \leq i \leq c, 1 \leq k \leq n \quad (12)$$

下面给出基于改进粒子群的移动界面模式聚类算法。

**算法3 移动界面模式聚类算法**

输入: 根据移动界面模式基本目标类型设置的聚类数量  $c$ , 参数  $\sigma, m$  和  $\eta$  的值, 初始化模糊划分矩阵  $U^{(0)}$

输出: 移动界面模式聚类簇

begin

    利用算法2得到初始聚类中心;

$t=1$ ;

    while(  $\|U^{(t)} - U^{(t-1)}\| \geq \eta$  )

        利用式(11)更新聚类中心;

        利用式(12)更新隶属度函数;

$t=t+1$ ;

    end while

end

**5 实验与分析**

为了验证本文算法的有效性, 从聚类效果和收敛速度两个方面对本文算法与现有算法进行实验比较。

**5.1 实验数据**

实验数据来自文献[27]和5个移动界面模式网站<sup>1)</sup>, 包含11种基本目标类型, 共670个移动界面模式。每个移动界面模式的目标属性包括主要功能目标、主要可用性目标、次要功能目标和次要可用性目标。

**5.2 评价指标**

移动界面模式数据聚类的来源多样, 数据分布不均衡, 会出现聚类结果偏斜现象。为了全面评价聚类算法的有效性, 需要从多个角度对移动界面模式的聚类结果进行评价, 因此, 本文采用多个聚类指标来评价聚类结果。使用调整兰德指数(Adjusted Rand Index, ARI)<sup>[28]</sup>是因为, 该指标对聚类数不敏感, 能够评价在不同聚类数情况下聚类结果与实际数据的一致程度; 使用Davies-Bouldin指数(Davies-Bouldin Index, DBI)<sup>[29]</sup>评价类内紧致度和类间离散度; 使用准确性(Accuracy, ACC)<sup>[30]</sup>评价聚类正确的比例; 使用标准化互信息(Normalized Mutual Information, NMI)<sup>[31]</sup>是因为该指标利用互信息来评价聚类结果与实际数据的一致程度。

1) ARI

$$ARI = \frac{\sum_{i=1}^c \sum_{j=1}^D \binom{n_{ij}}{2} - \binom{n}{2}^{-1} \sum_{i=1}^c \binom{n_i}{2} \sum_{j=1}^D \binom{n_j}{2}}{\frac{1}{2} \left[ \sum_{i=1}^c \binom{n_i}{2} + \sum_{j=1}^D \binom{n_j}{2} \right] - \binom{n}{2}^{-1} \sum_{i=1}^c \binom{n_i}{2} \sum_{j=1}^D \binom{n_j}{2}} \quad (13)$$

其中,  $n$  是整个数据集的数据个数,  $n_{ij}$  是同时在类  $i$  和簇  $j$  中的数据个数,  $n_i$  是类  $i$  中数据的个数,  $n_j$  是簇  $j$  中数据的个数。

2) DBI

$$\bar{R} = \frac{1}{N} \sum_{i=1}^N R_i \quad (14)$$

其中,  $R_i$  是从相似度  $R_{ij}$  中选出的最大值。

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \quad (15)$$

其中,  $S_i$  是第  $i$  类中的数据点的分散程度,  $M_{ij}$  是第  $i$  类与第  $j$  类的距离。

3) ACC

$$ACC = \frac{\sum_{i=1}^N \delta(y_i, map(c_i))}{N} \quad (16)$$

其中,  $N$  是数据个数,  $y_i$  是实际的类标签,  $c_i$  是获得的聚类标签。当  $y=c$  时,  $\delta(y, c)=1$ , 否则  $\delta(y, c)=0$ 。  $map(\cdot)$  是一个排列函数, 用于将每一个聚类标签和类标签进行匹配。

4) NMI

$$NMI = \frac{\sum_{i=1}^c \sum_{j=1}^c N_{ij} \log \frac{N N_{ij}}{N_i N_j}}{\sqrt{\sum_{i=1}^c N_i \log \frac{N_i}{N} \sum_{j=1}^c N_j \log \frac{N_j}{N}}} \quad (17)$$

其中,  $N$  是整个数据集的数据个数,  $N_{ij}$  是同时在类  $i$  和簇  $j$  中的数据个数,  $N_i$  是类  $i$  中数据的个数,  $N_j$  是簇  $j$  中数据的个数。

在4项评价指标中, ARI, NMI 和 ACC 的值越大, 表明聚类效果越好; DBI 的值越小, 表明聚类效果越好。

**5.3 结果与分析**

在本文算法中, 对种群多样性阈值  $\beta$  的判断对聚类效果有重要影响。为了确定  $\beta$  的值, 采用聚类评价指标对  $\beta$  值在取不同值时的聚类效果进行评价。在算法的执行过程中, 算法1中参数  $M=100, S=50$ ; 算法2中参数  $v_{max}=4, \omega_{max}=0.9, \omega_{min}=0.1, l_{max}=100, c_1=c_2=2$ ; 算法3中参数  $c=11, m=2, \eta=0.0001$ 。高斯核函数的参数  $\sigma$  由式(18)确定:

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{i=1}^N \|y_{ij} - b_i\|} \quad (18)$$

其中,  $b_i$  为第  $i$  个聚类中心,  $y_{ij}$  为第  $i$  类的第  $j$  个样本数据,  $N$  为第  $i$  个聚类中包含的样本个数。

$\beta$  取不同值时对4个评价指标的影响如图2所示。当  $\beta$  的取值在0.5和0.6之间时, 本文的聚类算法在4个评价指标中都能达到较好的聚类效果, 因此将  $\beta$  的值设置为均值0.55。

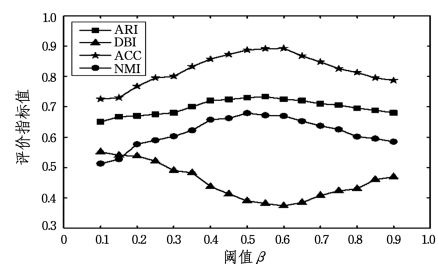


图2 不同  $\beta$  值对聚类的影响

Fig. 2 Effect of  $\beta$  on clustering

在聚类效果方面, 利用4个聚类评价指标将本文提出的

1) <http://www.mobile-patterns.com/>, <http://inspired-ui.com/>, <http://ui-patterns.com/>, <http://pttrns.com/>, <https://www.cocoacontrols.com>

算法 IEPSO-FCM 与标准模糊 C 均值聚类算法 FCM<sup>[1]</sup>、粒子群优化模糊 C 均值聚类算法 PSO-FCM<sup>[2]</sup>、改进的粒子群优化模糊 C 均值聚类算法 IPSO-FCM<sup>[3]</sup>、采用直觉模糊熵判断粒子群多样性<sup>[4]</sup>的模糊 C 均值聚类算法 EPSO-FCM 进行比较,聚类效果的比较如表 2 所列。

表 2 各算法聚类效果的比较

Table 2 Clustering results of different clustering algorithms

方法	ARI	DBI	ACC	NMI
FCM	0.3835	1.6273	0.7522	0.3074
PSO-FCM	0.5067	1.0261	0.8082	0.4582
IPSO-FCM	0.6812	0.7544	0.8163	0.5729
EPSO-FCM	0.7069	0.5026	0.8547	0.6153
IEPSO-FCM	0.7331	0.3781	0.8912	0.6718

从表 2 中的比较可以看出,在对移动界面模式的数据进行聚类后,本文算法的 ARI, NMI 和 ACC 值达到最大,并且 DBI 值达到最小,表明本文算法的聚类效果优于其他聚类算法。

在收敛速度方面,本文算法与现有聚类算法在对移动界面模式数据进行聚类时,式(7)的值在迭代过程中的变化情况如图 3 所示。

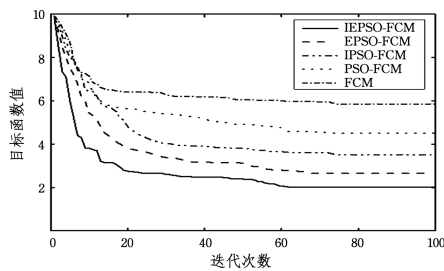


图 3 各算法对目标函数的迭代过程

Fig. 3 Iterative processes of different clustering algorithms on objective function

从图 3 中可以看出,本文算法在迭代到 60 次左右时,目标函数达到最优值,收敛速度比其他聚类算法快,而且利用本文算法得到的最优目标函数值低于其他聚类算法的最优目标函数值,说明本文算法搜索全局最优值的能力较强。

通过上述实验比较可以看出,本文算法在聚类效果和收敛速度两个方面的表现都优于现有的聚类算法。这是因为本文算法对现有粒子群优化算法进行了以下两点改进:

1) 在综合考虑直觉模糊熵的约束和影响因素后,对直觉模糊熵进行改进,提高了直觉模糊熵度量数据模糊程度的准确性,能够准确判断种群是否需要扩大搜索范围。

2) 混沌反向学习策略在利用混沌的遍历性生成初始种群之后生成了初始种群的反向解,然后从初始种群和反向初始种群中选取最优值,该策略可以提高搜索到最优值的概率。将该策略应用在粒子群优化产生初始种群和种群搜索的过程中,不但能够在种群的初始阶段确定较为合理的初始值,防止种群在初始阶段陷入局部寻优,而且可以在种群后续的寻优过程中保证种群的多样性,进一步提高全局搜索能力。

通过上述两点改进,本文算法能够更为合理地确定初始

聚类中心,避免出现聚类算法所更新的聚类中心和隶属度处于局部极值的情况,因此其聚类效果优于已有聚类算法。此外,由于在 FCM 聚类算法运行的初期就已经确定了较合理的初始聚类中心,减少了后续寻找极小值过程中的迭代次数,通过较少的迭代就可以找到全局极小值,因此本文算法的收敛速度较快。

**结束语** 由于现有的粒子群优化算法还存在不足之处,会导致基于粒子群优化的模糊 C 均值聚类算法得到的聚类结果不准确,本文利用直觉模糊熵和混沌反向学习策略对粒子群优化算法进行改进,提出了基于改进粒子群优化的模糊 C 均值聚类算法,并将该算法应用到移动界面模式的聚类中。经过实验的比较和分析,验证了本文算法能够对移动界面模式进行有效聚类。后续将会研究聚类算法在移动界面模式分类和检索中的应用。

## 参 考 文 献

- [1] BEZDEK J C. Pattern Recognition with Fuzzy Objective Function Algorithms[M]. New York: Plenum Press, 1981.
- [2] LI L, LIU X, XU M. A novel fuzzy clustering based on particle swarm optimization[C]// Proceedings of the 1st IEEE International Symposium on Information Technologies and Applications in Education. Kunming: IEEE Press, 2007: 88-90.
- [3] GENG Z K, WANG C B, ZHANG Z G. Fuzzy c-means and adaptive POS based fuzzy clustering algorithm[J]. Computer Science, 2016, 43(8): 267-272. (in Chinese)  
耿宗科, 王长宾, 张振国. 基于模糊 c-means 与自适应粒子群优化的模糊聚类算法[J]. 计算机科学, 2016, 43(8): 267-272.
- [4] WANG Y Z, LEI Y J, ZHOU L, et al. Intuitionistic fuzzy discrete particle swarm algorithm[J]. Control and Decision, 2012, 27(11): 1735-1739. (in Chinese)  
汪禹喆, 雷英杰, 周林, 等. 直觉模糊离散粒子群算法[J]. 控制与决策, 2012, 27(11): 1735-1739.
- [5] EBERHART R, KENNEDY J. Particle swarm optimization [C] // Proceedings of IEEE International Conference on Neural Networks. Perth: IEEE Press, 1995: 1942-1948.
- [6] ZADEH L A. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-356.
- [7] ATANASSOV K T. Intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 1986, 20(1): 87-96.
- [8] BURILLO P, BUSTINCE H, BUSTINCE H. Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets[J]. Fuzzy Sets and Systems, 1996, 78(3): 305-316.
- [9] SZMIDT E, KACPRZYK J. Entropy for intuitionistic fuzzy sets [J]. Fuzzy Sets and Systems, 2001, 118(3): 467-477.
- [10] LV Y C, GUO S Z. Entropy of intuitionistic fuzzy sets and its general forms [J]. Computer Engineering and Applications, 2011, 47(28): 52-55. (in Chinese)  
吕印超, 郭嗣琮. 直觉模糊集的熵及其一般形式[J]. 计算机工程与应用, 2011, 47(28): 52-55.
- [11] GUO K, SONG Q. On the entropy for Atanassov's intuitionistic fuzzy sets[J]. Applied Soft Computing, 2014, 24: 328-340.

- [12] WANG Y, LEI Y J. A technique for constructing intuitionistic fuzzy entropy [J]. *Control and Decision*, 2007, 22(12): 1390-1394. (in Chinese)  
王毅, 雷英杰. 一种直觉模糊熵的构造方法 [J]. *控制与决策*, 2007, 22(12): 1390-1394.
- [13] VERMA R, SHARMA B D. Exponential entropy on intuitionistic fuzzy sets [J]. *Kybernetika*, 2013, 49(1): 114-127.
- [14] YE J. Two effective measures of intuitionistic fuzzy entropy [J]. *Computing*, 2010, 87(1): 55-62.
- [15] WEI C P, GAO Z H, GUO T T. An intuitionistic fuzzy entropy measure based on trigonometric function [J]. *Control and Decision*, 2012, 27(4): 571-574. (in Chinese)  
魏翠萍, 高志海, 郭婷婷. 一个基于三角函数的直觉模糊熵公式 [J]. *控制与决策*, 2012, 27(4): 571-574.
- [16] WANG J Q, WANG P. Intuitionistic linguistic fuzzy multi-criteria decision-making method based on intuitionistic fuzzy entropy [J]. *Control and Decision*, 2012, 27(11): 1694-1698. (in Chinese)  
王坚强, 王佩. 基于直觉模糊熵的直觉语言多准则决策方法 [J]. *控制与决策*, 2012, 27(11): 1694-1698.
- [17] GAO M M, SUN T, ZHU J J. Revised axiomatic definition and structural formula of intuitionistic fuzzy [J]. *Control and Decision*, 2014, 29(3): 470-474. (in Chinese)  
高明美, 孙涛, 朱建军. 一种改进的直觉模糊熵公理化定义和构造公式 [J]. *控制与决策*, 2014, 29(3): 470-474.
- [18] LIU M F, REN H P. A study of multi-attribute decision making based on a new intuitionistic fuzzy entropy measure [J]. *System Engineering Theory and Practice*, 2015, 35(11): 2909-2916. (in Chinese)  
刘满凤, 任海平. 基于一类新的直觉模糊熵的多属性决策方法研究 [J]. *系统工程理论与实践*, 2015, 35(11): 2909-2916.
- [19] ZHAO F, WANG Q S, HAO W L. Improvement of constraint conditions and new constructional method for intuitionistic fuzzy entropy [J]. *Journal of Computer Applications*, 2015, 35(12): 3461-3464. (in Chinese)  
赵飞, 王青山, 郝万亮. 直觉模糊熵约束条件的改进及新模糊熵构造 [J]. *计算机应用*, 2015, 35(12): 3461-3464.
- [20] GAO W, LIU S, HUANG L. Particle swarm optimization with chaotic opposition-based population initialization and stochastic search technique [J]. *Communications in Nonlinear Science and Numerical Simulation*, 2012, 17(11): 4316-4327.
- [21] ZHOU Y, BAO L, CHEN C L P. A new 1D chaotic system for image encryption [J]. *Signal Processing*, 2014, 97(7): 172-182.
- [22] THONGMOOL G, PHANKOKKRUAD M. Analysis of interaction user interface patterns and usability study in computer assisted instruction for Tablet PC [C] // 4th IEEE International Conference on Control System, Computing and Engineering. Batu Ferringhi; IEEE Press, 2014: 472-477.
- [23] WETCHAKORN T, PROMPOON N. Method for mobile user interface design patterns creation for iOS platform [C] // 12th International Joint Conference on Computer Science and Software Engineering. Songkhla; IEEE Press, 2015: 150-155.
- [24] BIRTOLO C, RONCA D. Advances in clustering collaborative filtering by means of fuzzy c-means and trust [J]. *Expert Systems with Applications*, 2013, 40(17): 6997-7009.
- [25] PAUL A, MAITY S. Kernel fuzzy c-means clustering on energy detection based cooperative spectrum sensing [J]. *Digital Communications and Networks*, 2016, 2(4): 196-205.
- [26] WU Z, XIE W, YU J. Fuzzy c-means clustering algorithm based on kernel method [C] // 5th International Conference on Computational Intelligence and Multimedia Applications. Xi'an; IEEE Press, 1995: 1942-1948.
- [27] NEIL T. *Mobile Design pattern gallery: UI patterns for mobile applications* [M]. Sebastopol; O'Reilly Media, 2012.
- [28] HUBERT L, ARABIE P. Comparing partitions [J]. *Journal of Classification*, 1985, 2(1): 193-218.
- [29] DAVIES L D, BOULDIN W D. A cluster separation Measure [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1979, 1(2): 224-227.
- [30] CHEN W, SONG Y, BAI H, et al. Parallel spectral clustering in distributed systems [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(3): 568-586.
- [31] STREHL A, GHOSH J. Cluster ensembles-Knowledge reuse framework for combining multiple partitions [J]. *Journal of Machine Learning Research*, 2003, 3(3): 583-617.

(上接第 195 页)

- [8] KAI R. *Compositional Scheduling Analysis Using Standard Event Models* [D]. Braunschweig: Technical University Carolus-Wilhelmina of Braunschweig, 2005.
- [9] MONOT A, NAVET N, BAVOUX B, et al. Multisource Software on Multicore Automotive ECUs-Combining Runnable Sequencing With Task [J]. *Scheduling IEEE Transactions on Industrial Electronics*, 2012, 59(10): 3934-3942.
- [10] FERRARI A, NATALE M D, GENTILE G, et al. Time and memory tradeoffs in the implementation of AUTOSAR components [C] // Conference on Design, Automation and Test in Europe. 2009: 864-869.
- [11] FARAGARDI H R, LISPER B, SANDSTRÖM K, et al. A Communication-Aware Solution Framework for Mapping AUTOSAR Runnables on Multi-core Systems [C] // Proceedings of the 2014 IEEE Emerging Technology and Factory Automation. 2014: 1-9.
- [12] HAN J W, KAMBER M, PEI J. *Data Mining: Concepts and Techniques (Third Edition)* [M]. Burlington: Morgan Kaufmann, 2011: 456-461.
- [13] PARTNERSHIP A. Explanation of Application Interfaces of the Body and Comfort Domain (V1. 2. 0 R4. 0 Rev 2) [OL]. <http://www.autosar.org>, 2010.