

基于粒子群优化的直觉模糊核聚类算法研究

余晓东, 雷英杰, 岳韶华, 王睿

(空军工程大学 防空反导学院, 陕西 西安 710051)

摘要: 针对现有基于核方法的直觉模糊聚类算法对初始值敏感、收敛速度慢等缺陷, 利用粒子群优化算法全局搜索能力强、收敛速度快的优势, 对直觉模糊核聚类算法的初始聚类中心进行优化, 并提出了一种基于粒子群优化的直觉模糊核聚类算法。该算法在提升聚类性能的同时, 有效增强了算法的收敛速度。在实验阶段, 采用4组标准数据集对该算法进行了分类实验及有效性测试, 并将其与模糊 c 均值聚类算法及直觉模糊 c 均值聚类算法的分类效果及运行时间进行对比, 实验结果充分表明了该算法的有效性及其优越性。

关键词: 直觉模糊集; 核方法; 模糊聚类; 粒子群优化

中图分类号: TP182, TP391

文献标识码: A

Research on PSO-based intuitionistic fuzzy kernel clustering algorithm

YU Xiao-dong, LEI Ying-jie, YUE Shao-hua, WANG Rui

(School of Air and Missile Defense, Air Force Engineering University, Xi'an 710051, China)

Abstract: The intuitionistic fuzzy kernel c -means clustering algorithm has several problems such as sensitivity to the initial value, low convergence speed, etc. To overcome these shortages, the particle swarm optimization (PSO) algorithm with powerful ability of global search and quick convergence rate is applied to intuitionistic fuzzy clustering. Firstly, PSO is used to optimize the initial clustering centers. Then, the approach of intuitionistic fuzzy kernel clustering based on PSO, namely PS-IFKCM, is proposed. This algorithm can enhance both the clustering ability and the convergence speed. Finally, experiments based on four measured datasets are carried out to illustrate the performance of the proposed method. Compared with results from FCM and IFKCM, PS-IFKCM is of great efficiency for classification.

Key words: intuitionistic fuzzy sets; kernel method; fuzzy clustering; particle swarm optimization

1 引言

1969年,著名学者 Ruspini^[1]首先提出了模糊划分的概念,将模糊集理论引入到聚类分析中来。随后,研究者提出了多种模糊聚类分析方法,主要包括基于模糊等价关系的传递闭包方法、基于相似性关系和模糊关系的方法以及基于模糊图论的最大树方法等,但是这些方法由于计算复杂度较高,难以应用于大数据问题及实时性要求较高的领域,在实际应用与研究中已逐步减少^[2,3]。模糊 c 均值(FCM, fuzzy c -means)算法是由 Dunn^[4]、Bezdek^[5]提出的一种基于目标函数的聚类方法,它能够通过最小化目标函数得到各样本点相对各聚类中心的隶属度,从

而达到自动分类的目的,而广泛应用于模式识别、图像处理、网络安全等领域^[6]。

直觉模糊集(IFS, intuitionistic fuzzy sets)理论^[7]作为 Zadeh 模糊理论最重要的拓展形式之一,因其增加了犹豫度属性参数,可更加细腻地刻画客观现象的模糊不确定性本质,引起了相关研究领域的广泛关注^[8,9]。因此,多位学者对如何将 FCM 算法拓展到直觉模糊领域进行了研究,贺正洪^[10]将聚类对象及聚类中心点用直觉模糊集表示,提出基于直觉模糊集合的模糊 c 均值算法。申晓勇^[11]将聚类对象和聚类中心点及两者间的关系均推广到直觉模糊领域,提出了一种基于目标函数的直觉模糊集合数据的聚类方法。这些算法虽然取得了较好的应用效

收稿日期: 2014-01-25; 修回日期: 2014-03-31

基金项目: 国家自然科学基金资助项目(61272011, 61309022)

Foundation Item: The National Natural Science Foundation of China (61272011, 61309022)

果，但同样也遗留了经典 FCM 算法的一些缺点，即对数据原型有较强的依赖性，对复杂的数据结构显得无能为力。

1995 年，Vapnik 和 Cortes^[12]提出了支持向量机(SVM, support vector machine)理论，SVM 在很多领域都体现出比传统分类器更好的性能，使该方法逐渐受到重视并被应用到机器学习领域的各个方面^[13-15]。Girolami^[16]和张莉等^[17]提出了基于核函数的模糊聚类算法(FKCM, fuzzy kernel c-means)，解决 FCM 算法不能发现非凸聚类结构的问题。文献[18]对模糊核聚类算法进行了直觉化扩展，提出一种基于核化距离的直觉模糊聚类(IFKCM, intuitionistic fuzzy kernel c-means)算法，提高了算法的聚类精度。但由于其使用的迭代下降算法只具备局部寻优能力，因此其收敛速度较慢，且对初始值敏感、易陷入局部最优解等缺陷依然没有得到解决^[19-22]。

鉴于此，群体智能优化算法被引入到模糊聚类领域中来，以期达到全局最优，快速收敛的目的。粒子群优化算法(PSO, particle swarm optimization)由 Kennedy 等^[22]模拟鸟群觅食过程提出的一种全局寻优算法，由于该算法收敛速度快、且易于编程实现，因而近年来受到广泛关注^[23]。王纵虎等^[24]提出了一种基于粒子群优化的快速 K 均值算法，有效克服了 K 均值算法对初始聚类中心敏感

及易陷入局部最优等缺点。于德亮等^[25]提出一种基于粒子群的模糊核聚类算法，并将其应用于电梯群交通模式识别。因此，本文尝试采用 PSO 算法对 IFKCM 算法的初始聚类中心进行优化来实现数据聚类，并提出了一种基于粒子群的直觉模糊核聚类算法(PS-IFKCM, intuitionistic fuzzy kernel c-means clustering based on particle swarm optimization)。实验选取 3 组 UCI 数据及一组实际数据 Motorcycle 进行仿真实验，并将其与 FCM、IFKCM 的分类效果及运行时间进行比较，充分表明了 PS-IFKCM 算法的优越性和有效性，以展示该算法的性能。

2 基于核的直觉模糊聚类算法

依据核方法的思想，利用非线性映射 $\Phi(\cdot)$ 把输入的样本集合 $X=\{x_1, x_2, \dots, x_n\}$ 、聚类原型 $P=\{p_1, p_2, \dots, p_c\}$ 映射到一个高维核空间 R^q 上，然后在这个高维核空间 R^q 上扩展直觉模糊聚类算法。算法无需知道映射函数 $\Phi(\cdot)$ 的具体形式，通过采用符合 Mercer 条件的核函数代替特征空间内的内积运算，从而实现聚类样本在高维空间线性可分的目的。

2.1 基于核的直觉模糊欧式距离

定义 1 (直觉模糊欧式距离度量^[7]) 若样本 $x=(x_1, x_2, \dots, x_n)$ 和样本 $y=(y_1, y_2, \dots, y_n)$ 均可用直觉模糊集表示，则它们之间的直觉模糊欧式距离可定义如下

$$D_{IF}(x, y) = \sqrt{\frac{1}{2n} \sum_{i=1}^n (\mu_x(x_i) - \mu_y(y_i))^2 + (\gamma_x(x_i) - \gamma_y(y_i))^2 + (\pi_x(x_i) - \pi_y(y_i))^2} \quad (1)$$

尝试将样本间的直觉模糊欧式距离投影到特征核空间，并基于核方法

$$\|\Phi(x) - \Phi(y)\|^2 = K(x, y) - 2K(x, y) + K(x, y) \quad (2)$$

给出基于核的直觉模糊欧式距离。

定义 2 (基于核的直觉模糊欧式距离度量) 若样本 $x=(x_1, x_2, \dots, x_n)$ 和样本 $y=(y_1, y_2, \dots, y_n)$ 均可用直觉模糊集表示，则它们之间基于核的直觉模糊欧式距离可定义如下

$$D_K(x, y) = \sqrt{\frac{1}{2n} \sum_{i=1}^n \omega(i) \left\{ \begin{aligned} &K(\mu_x(x_i), \mu_x(x_i)) + K(\mu_y(y_i), \mu_y(y_i)) + K(\gamma_x(x_i), \gamma_x(x_i)) \\ &+ K(\gamma_y(y_i), \gamma_y(y_i)) + K(\pi_x(x_i), \pi_x(x_i)) + K(\pi_y(y_i), \pi_y(y_i)) \\ &- 2K(\mu_x(x_i), \mu_y(y_i)) - 2K(\gamma_x(x_i), \gamma_y(y_i)) - 2K(\pi_x(x_i), \pi_y(y_i)) \end{aligned} \right\}} \quad (3)$$

考虑到实际应用中，样本各维特征可能对分类贡献不均匀的情况，定义 2 中对各维特征进行了加权处理。其中， $\omega(i) \geq 0, i=1, 2, \dots, n$ 为加于第 i 维特征的权值， $\omega(i)$ 满足归一化条件

$$\frac{1}{n} \sum_{i=1}^n \omega(i) = 1 \quad (4)$$

注意到式(3)并没有选择特定的核函数，因此任何满足 Mercer 条件的核函数 $K(x, y)$ 都适用于该

式。Mercer 条件^[14]可描述为：对任意的平方可积函数 $g(x)$ ，都满足 $\iint K(x,y)g(x)g(y)dxdy \geq 0$ 。下面是常见的 2 个常用的 Mercer 核函数。

高斯核函数： $K(x,y) = \exp(-\|x-y\|^2 / \sigma^2)$

多项式核函数： $K(x,y) = (xy+b)^d$

由于高斯核函数对应的是无穷维的特征空间，而有限容量的样本在无穷维的特征空间内一定是线性可分的。这里采用高斯核函数，代入式(3)得

$$D_{GK} = \sqrt{\frac{1}{2n} \sum_{i=1}^n \omega(i) \begin{Bmatrix} 3 - 2K(\mu_x(x_i), \mu_y(y_i)) \\ -2K(\gamma_x(x_i), \gamma_y(y_i)) - \\ 2K(\pi_x(x_i), \pi_y(y_i)) \end{Bmatrix}} \quad (5)$$

2.2 公式推导

设 $X = \{x_1, x_2, \dots, x_n\} \subset \mathcal{R}^s$ 为模式空间内的一组有限观测样本集，假定每个样本的特征均为 s 维的直觉模糊集，可表示为 $x_i = \{\langle x\mu_{i1}, x\gamma_{i1}, x\pi_{i1} \rangle, \langle x\mu_{i2}, x\gamma_{i2}, x\pi_{i2} \rangle, \dots, \langle x\mu_{is}, x\gamma_{is}, x\pi_{is} \rangle\}$ 。将样本集分成 c 类， c 个聚类中心 $P = \{p_1, p_2, \dots, p_c\}$ 也为直觉模糊集， p_i 可表示为 $p_i = \{\langle p\mu_{i1}, p\gamma_{i1}, p\pi_{i1} \rangle, \langle p\mu_{i2}, p\gamma_{i2}, p\pi_{i2} \rangle, \dots, \langle p\mu_{is}, p\gamma_{is}, p\pi_{is} \rangle\}$ 。引入高斯核函数后，直觉模糊核聚类算法的描述形式如下

$$\begin{cases} J(\mathbf{U}_{k\mu}, \mathbf{U}_{k\gamma}, \mathbf{P}) \\ = \sum_{i=1}^c \sum_{j=1}^n \{[(\mu_{ij})^m / 2 + (1 - \gamma_{ij})^m / 2] D_{GK}(x_j, p_i)^2\} \\ \mu_{ij} + \gamma_{ij} + \pi_{ij} = 1 \\ \sum_{i=1}^c \sum_{j=1}^n \mu_{ij} = n \\ m \in [1, \infty), \mu_{ij} \in [0, 1], \gamma_{ij} \in [0, 1], \forall i, \forall j \end{cases} \quad (6)$$

其中， m 为平滑参数， $D_{GK}(x_j, p_i)$ 为样本 x_j 与聚类中心 p_i 之间的基于核的直觉模糊欧式距离， $\mathbf{U}_{k\mu}$ 为直觉模糊核划分隶属矩阵， $\mathbf{U}_{k\gamma}$ 为直觉模糊核划分非隶属矩阵。这是一个关于自变量 $(\mathbf{U}_{k\mu}, \mathbf{U}_{k\gamma}, \mathbf{P})$ 的约束优化问题，由拉格朗日乘法可得目标函数为

$$F = \sum_{i=1}^c \{[(\mu_{ij})^m / 2 + (1 - \gamma_{ij})^m / 2] D_{GK}(x_j, p_i)^2\} + \lambda(\mu_{ij} + \gamma_{ij} + \pi_{ij} - 1) + \beta(\sum_{i=1}^c \sum_{j=1}^n \mu_{ij} - n) \quad (7)$$

3 基于粒子群的直觉模糊核聚类算法

基于核的直觉模糊聚类算法虽然能有效消除对数据分布的依赖性，但是其对初始值敏感，且容易陷入局部最优解等问题并没有得到解决。而通过

先验知识选取合适的初始值则可以有效解决上述问题。因此，本文尝试采用粒子群算法对直觉模糊核聚类算法的初始值进行寻优从而获得更好的聚类效果。

3.1 算法原理

PSO 算法数学描述为： n 维空间中，种群 $X = \{x_1, x_2, \dots, x_m\}$ 由 m 个粒子组成，其中第 i 个粒子的位置为 $x_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$ ，速度为 $v_i = \{v_{i1}, v_{i2}, \dots, v_{in}\}$ ，该粒子的个体极值为 $p_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$ ，整个种群的全局极值为 $p_g = \{p_{g1}, p_{g2}, \dots, p_{gn}\}$ 。搜索到个体极值 $pbest$ 及全局极值 $gbest$ 后，粒子 x_i 将根据如下公式进行更新

$$v_{id}(t+1) = \omega v_{id}(t) + c_1 r_1 (p_{id}(t) - x_{id}(t)) + c_2 r_2 (p_{gd}(t) - x_{id}(t)) \quad (8)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (9)$$

其中， t 为当前迭代次数， $x_{id}(t)$ 为粒子的当前时刻位置， $x_{id}(t+1)$ 为粒子下一时刻位置， $v_{id}(t)$ 、 $v_{id}(t+1)$ 表示粒子当前时刻及下一时刻的速度， ω 为惯性因子， c_1 、 c_2 为加速常数， r_1 和 r_2 为分布于 $[0, 1]$ 之间的随机数。此外，为了使粒子速度不致过大，设置速度上限值 v_{max} ，若 $v_{id} > v_{max}$ ，取 $v_{id} = v_{max}$ ， $v_{id} < -v_{max}$ ，取 $v_{id} = -v_{max}$ 。

本文方法的基本思想为：采用粒子群优化算法全局搜索能力强、收敛速度快的优势，对 IFKCM 的初始值进行优化。样本投影到高维特征空间后，通过对划分隶属矩阵、划分非隶属度矩阵的迭代更新，以及对聚类中心点不断的修正，缩短了聚类的时间，提高了聚类的精度。设样本空间为 $X = \{x_1, x_2, \dots, x_n\}$ ，用一个粒子代表一个聚类中心集合 $V = \{v_1, v_2, \dots, v_c\}$ ，其中， v_j 和 x_i 是同维度的向量。取 PSO 的适应度函数为

$$f(x_i) = \frac{1}{J_{km}(\mathbf{U}_{k\mu}, \mathbf{U}_{k\gamma}, \mathbf{P}) + 1} = \frac{1}{\sum_{i=1}^c \sum_{j=1}^n \{[(\mu_{ij})^m / 2 + (1 - \gamma_{ij})^m / 2] D_{GK}(x_j, p_i)^2\} + 1} \quad (10)$$

3.2 算法步骤

PS-IFKCM 算法的具体步骤如下所示。

输入：样本数据集 X ，聚类类别数 $2 \leq c \leq n$ ，适应度函数 $f(x_i)$ ，权重系数矩阵 W ，平滑参数 m ，核参数 σ 。

输出：划分隶属矩阵 $\mathbf{U}_{k\mu}$ ，划分非隶属矩阵 $\mathbf{U}_{k\gamma}$ ，

聚类中心 \mathbf{P} , 迭代次数 t 、 b 。

过程如下。

Step1 初始化参数。设定种群规模 l , 最大进化代数 T_{\max} , 加速常数 c_1 和 c_2 , 惯性因子 ω , 速度上限值 v_{\max} , 迭代停止阈值 η 、 ε 的取值, 令计数器 $t = b = 0$ 。

Step2 初始化粒子群。按照设定的种群规模产生 l 个粒子 $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_l$ 作为初始种群, 每个粒子 \mathbf{V}_i 代表一个聚类中心集合 $\{\mathbf{v}_{i1}, \mathbf{v}_{i2}, \dots, \mathbf{v}_{ic}\}$ 。

Step3 计算划分隶属矩阵 \mathbf{U}_{ku} 及划分非隶属矩阵 \mathbf{U}_{ky} 。对于 $\forall k, t$, 如果 $D_{kw}(\mathbf{x}_j, \mathbf{p}_i)^{(b)} \neq 0$ 则有

$$\begin{cases} \mu_{ij}^{(b)} = \left\{ \sum_{k=1}^c \sum_{t=1}^n \left(\frac{D_{GK}(\mathbf{x}_j, \mathbf{p}_i)^{(b)}}{D_{GK}(\mathbf{x}_t, \mathbf{p}_k)^{(b)}} \right)^{\frac{2}{m-1}} \right\}^{-1} \\ \gamma_{ij}^{(b)} = 1 - \pi_{ij} - \left\{ \sum_{k=1}^c \sum_{t=1}^n \left(\frac{D_{GK}(\mathbf{x}_j, \mathbf{p}_i)^{(b)}}{D_{GK}(\mathbf{x}_t, \mathbf{p}_k)^{(b)}} \right)^{\frac{2}{m-1}} \right\}^{-1} \end{cases} \quad (11)$$

如果 $\exists k$, 使 $D_{GK}(\mathbf{x}_j, \mathbf{p}_i)^{(b)} = 0$, 则有

$$\begin{cases} \mu_{ij} = 1, \mu_{ij} = 0 & i = k \\ \gamma_{ij} = 0, \mu_{ij} = 1 & i \neq k \end{cases} \quad (12)$$

Step4 用式(10)计算每个粒子的适应度函数值。

Step5 更新最优值: 将粒子的适应值与它的个体最优值 $P_{id}(t)$ 进行比较, 如果粒子的适应值优于 $P_{id}(t)$, 则设 $P_{id}(t)$ 位置为粒子当前位置; 将粒子的适应值与群体最优值 $P_{gd}(t)$ 进行比较, 如果粒子的适应值优于 $P_{gd}(t)$, 则设 $P_{gd}(t)$ 位置为粒子当前位置。

Step6 更新粒子种群: 根据式(8)、式(9)计算粒子新的位置和速度, 并产生下一代粒子群。

Step7 判断是否满足终止条件, 若满足, 则停止迭代, 输出取得 $P_{gd}(t)$ 的粒子为 IFKCM 算法的聚类原型 \mathbf{P} ; 否则, $t = t + 1$, 转至 Step3。结束条件为到达最大进化代数 T_{\max} , 或粒子的适应值达到给定阈值 η 。

Step8 根据式(9)、式(10)计算并更新划分隶属矩阵 \mathbf{U}_{ku} , 划分非隶属矩阵 \mathbf{U}_{ky} 。

Step9 更新聚类原型模式矩阵 $\mathbf{p}_i^{(b+1)}$, 分别按下式求得 $\mathbf{p}\mu_i^{(b+1)}$, $\mathbf{p}\gamma_i^{(b+1)}$ 和 $\mathbf{p}\pi_i^{(b+1)}$ 。

$$\mathbf{p}\mu_i = \frac{\sum_{j=1}^n ((\mu_{ij})^m / 2 + (1 - \gamma_{ij})^m / 2) K(\mathbf{x}\mu_j, \mathbf{p}\mu_i) \mathbf{x}\mu_j}{\sum_{j=1}^n ((\mu_{ij})^m / 2 + (1 - \gamma_{ij})^m / 2) K(\mathbf{x}\mu_j, \mathbf{p}\mu_i)} \quad (13)$$

$$\mathbf{p}\gamma_i = \frac{\sum_{j=1}^n ((\mu_{ij})^m / 2 + (1 - \gamma_{ij})^m / 2) K(\mathbf{x}\gamma_j, \mathbf{p}\gamma_i) \mathbf{x}\gamma_j}{\sum_{j=1}^n ((\mu_{ij})^m / 2 + (1 - \gamma_{ij})^m / 2) K(\mathbf{x}\gamma_j, \mathbf{p}\gamma_i)} \quad (14)$$

$$\mathbf{p}\pi_i = \mathbf{I} - \mathbf{p}\mu_i - \mathbf{p}\gamma_i \quad (15)$$

Step10 比较 $\|\mathbf{p}^{(b+1)} - \mathbf{p}^{(b)}\|$ 和迭代停止阈值 ε 的大小, 如果 $\|\mathbf{p}^{(b+1)} - \mathbf{p}^{(b)}\| \geq \varepsilon$, 则令 $b = b + 1$, 转向 Step8, 如果 $\|\mathbf{p}^{(b+1)} - \mathbf{p}^{(b)}\| \leq \varepsilon$, 则算法终止, 并输出 \mathbf{U}_{ku} 、 \mathbf{U}_{ky} 以及聚类原型 \mathbf{P} 。

3.3 算法复杂度分析

根据算法的时间复杂度及空间复杂度理论分析, 直觉模糊核聚类算法的一次迭代需要运行的计算次数为 nc , 其算法的时间复杂度为 $O(ncl)$, 其中 n 为数据集样本数目, c 为聚类类别数, l 为算法的迭代次数, 因此直觉模糊核聚类算法的时间复杂度也可记为 $O(n)$ 。此外, 直觉模糊核聚类算法需要存储划分隶属矩阵及划分非隶属矩阵, 因此其空间复杂度为 $O(2n^2)$, 也可记为 $O(n^2)$ 。而粒子群算法迭代一次需要计算种群中粒子的位置、速度及适应度值, 计算次数为 $3Nc$, 其中 N 为粒子群的种群规模, 因此粒子群优化算法的时间复杂度为 $O(3Ncl)$, 可记作 $O(N)$ 。同理, 其空间复杂度也为 $O(N)$ 。

根据上文分析, 本文所提出 PS-IFKCM 算法的时间复杂度为 $O(n + N)$, 空间复杂度为 $O(n^2 + N)$, 实际应用中, 粒子的种群规模 N 的取值通常是一个比样本个数 n 远小的值, 因此 PS-IFKCM 算法的时间复杂度可记为 $O(n)$, 空间复杂度为 $O(n^2)$ 。从算法的时间复杂度来看, PS-IFKCM 算法和 IFKCM 算法同阶, 均为 $O(n)$ 。但 PS-IFKCM 算法汲取了 PSO 全局搜索能力强、收敛速度快的优势, 算法迭代次数小于 IFKCM 算法迭代次数, 因此其算法运行时间也小于 IFKCM 算法运行时间。从算法的空间复杂度来看, PS-IFKCM 算法略大于 IFKCM 算法, 但由于二者空间复杂度表达式中的高阶项因子相同, 说明 PS-IFKCM 算法运行所需存储空间的增加在可承受范围之内。

该算法还涉及到平滑参数 m 的数值选取。它是给隶属度所赋的一个权重, 因此又被称为加权因子, 它控制着模糊类间的分享程度。Bezdek 给出了参数 m 的一个经验范围 [1.1, 5], 但没有给出严格的证明^[5]。通常情况下参数 m 取值为 2, 但尚缺乏理论支持。目前来看, 如何对参数 m 进行取值, 还停留在实验阶段, 更多的是依靠经验取值, 缺乏可供

参考的优选及检验算法。

4 实验结果及分析

实验所采用的数据为UCI数据库中的3组标准数据集 Iris, Wine, Breast Cancer Wisconsin(简化为 Wisc)及另外一组常用实际数据 Motorcycle。仿真环境为 Pentium(R) Dual-Core CPU E5500 @2.8 GHz, 内存 2 GB, 操作系统 Windows XP, 仿真软件 Matlab7.6。

4.1 算法聚类实验

先选择 Iris 数据集对 PS-IFKCM 算法进行聚类实验。设置实验参数为: 聚类类别数 $c=3$, 核参数 $\sigma^2=0.03$, 平滑参数 $m=2$, 粒子种群大小为 $l=20$, 学习因子 $c_1=c_2=2$, 惯性权重 $\omega=0.72$, 最大进化代数 $T_{\max}=100$, 迭代停止阈值 $\varepsilon=10^{-5}$ 、 $\eta=0.02$ 。Iris 数据集包含四维特征属性, 聚类效果不容易观测, 因此选择将样本投影到2维空间对算法的聚类效果进行演示, 产生 PCA 图、Fuzzy Sammon 图如图 1 和图 2 所示。3 类样本在图中分别用“·”、“+”和“×”

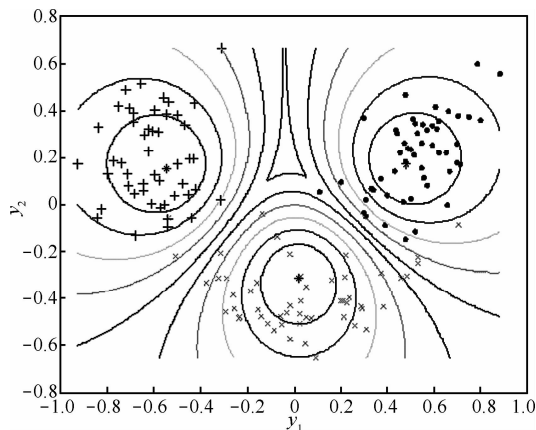


图1 Iris 数据 PCA 映射

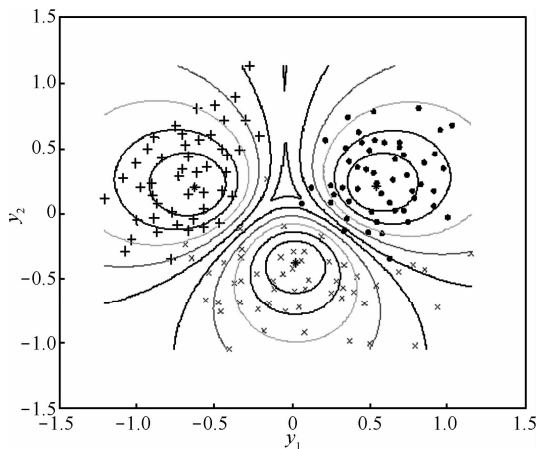


图2 Iris 数据 Fuzzy Sammon 映射

3 种符号进行表示。从图中可以看出, 3 类样本点被清楚地分离开来, 并且几乎不存在重叠分布。可见本文提出的 PS-IFKCM 算法取得了较好的分类效果。

此外, 基于 4 组标准数据集分别对 3 种算法进行 100 次 Monte Carlo 仿真实验, 所得的平均分类正确率如表 1 所示。

数据集	FCM	IFKCM	PS-IFKCM
Iris	92.76	94.74	96.91
Wine	92.36	95.37	96.73
Wisc	93.58	95.46	96.89
Motorcycle	93.26	94.88	97.07

从表 1 中可以看出, FCM 算法的分类正确率最低, 聚类效果最差。IFKCM 算法通过将核方法和模糊聚类方法相结合, 并将研究拓展到直觉模糊领域, 获取了更多的样本的信息, 聚类识别效果较好。而 PS-IFKCM 算法则同时兼具了 PSO 算法全局搜索强以及 IFKCM 算法局部搜索能力强的优势, 分类正确率最高, 聚类识别效果也最好。

4.2 算法有效性实验

实验选取 Motorcycle 数据集对本文算法有效性进行测试。算法首先采用 PSO 算法对 IFKCM 聚类的初始值进行寻优, 在将样本点映射到高维核特征空间后, 通过对划分隶属度矩阵、划分非隶属度矩阵的迭代更新, 以及对聚类中心点持续不断的修正, 如图 3 所示, “·”和“*”分别表示样本点和动态聚类中心点。每次迭代结束后, 都会得到不同的动态聚类中心点, 图 3 所示为实验最后一次迭代结果。

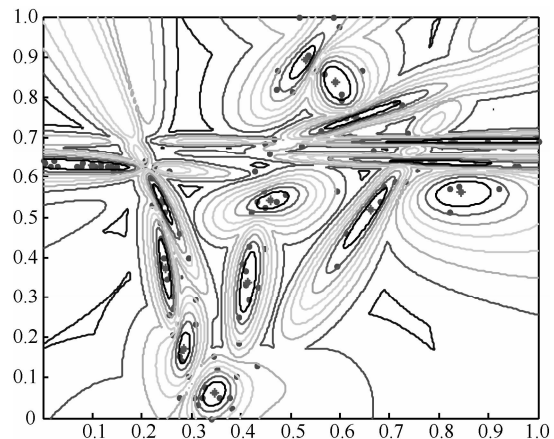


图3 最优动态聚类中心点分布

为了对本文算法的有效性进行检测, 在算法每

次迭代后生成 7 项常用的有效性指标值 (classification entropy(CE), partition coefficient(PC), Xie and Beni's index(XB), separation index(S), partition index(SC), dunn's index(DI), alternative dunn index(ADI)), 其各自所形成的曲线如图 4 所示, 可以看出, 本文算法的 7 项指标值曲线均较为平滑, 因此该算法是有效的。

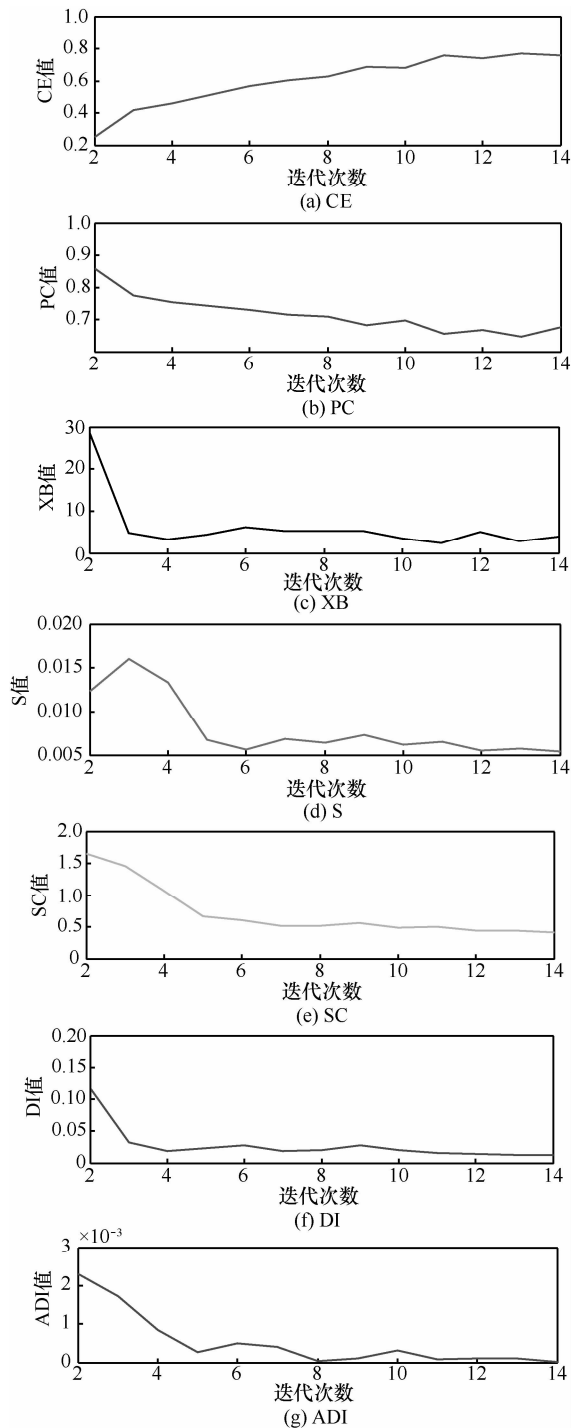


图 4 算法有效性指标迭代曲线

4.3 算法时间复杂度实验

基于 Iris, Wine, Wisc, Motorcycle4 种不同数据集分别进行 100 次 Monte Carlo 仿真实验, FCM、IFKCM、PS-IFKCM 算法的平均运行时间如表 2 所示。

数据集	FCM	IFKCM	PS-IFKCM
Iris	0.106	6.256	2.025
Wine	0.289	9.134	3.131
Wisc	0.365	9.789	3.541
Motorcycle	0.101	4.378	1.416

由表 2 可知, PS-IFKCM 算法运行时间较 FCM 算法高出一些, 但在可接受范围之内, 然而该算法优秀的聚类性能是传统算法不可比拟的。此外, PS-IFKCM 算法的运行时间远远小于 IFKCM 算法运行时间, 可见该算法通过采用 PSO 算法对初始值进行寻优确实获得了更好的聚类效果, 有效地缩短了聚类所需的时间, 提高了聚类精度。因此, 与 IFKCM 算法相比, PS-IFKCM 算法在分类性能和运行时间上均具备明显优势。

5 结束语

本文将粒子群算法与直觉模糊核聚类算法进行有效结合, 并提出了基于粒子群优化的直觉模糊核聚类算法, 克服了传统聚类算法对初始值敏感、易陷入局部最优及收敛速度慢等问题, 在降低算法的时间复杂度的同时获得了更优的聚类效果。因此, 该算法通过改进优化直觉模糊核聚类算法的学习过程, 为模糊聚类方法提供了一次有效的尝试。但是, 该算法仍有一些需要改进及完善的地方, 如平滑参数 m 的确定方法, 以及如何进一步提升算法的泛化性能等, 这些都将是下一步重点研究的问题。

参考文献:

- [1] RUSPINI E H. A new approach to clustering[J]. Information and Control, 1969, 15(1): 22-32.
- [2] CECCARELLI M, MARATEA A. Improving fuzzy clustering of biological data by metric learning with side information[J]. Int'l Journal of Approximate Reasoning, 2008, 47(1): 45-57.
- [3] 张洪美, 徐泽水, 陈琦. 直觉模糊集的聚类方法研究[J]. 控制与决策, 2007, 22(8): 882-888.
ZHANG H M, XU Z S, CHEN Q. On clustering approach to intuitionistic fuzzy sets[J]. Control and Decision, 2007, 22(8):882-888.
- [4] DUMN J C. A graph theoretic analysis of pattern classification via tamura's

- fuzzy relation[J]. IEEE Trans on Fuzzy System, 1974, 4(3):310-313.
- [5] BEZDEK J C. Pattern Recognition with Fuzzy Objective Function Algorithms[M]. New York Plenum Press, 1981.
- [6] XUE Z K, LI S Y. A multi-model modeling approach to MIMO nonlinear systems[J]. Acta Electronic Sinica, 2005,33(1):52-56.
- [7] 雷英杰, 王宝树, 苗启广. 直觉模糊关系及其合成运算[J]. 系统工程理论与实践, 2005, 25 (2): 113-118, 133.
LEI Y J, WANG B S, MIAO Q G. On the intuitionistic fuzzy relations with compositional operations[J]. Systems Engineering Theory and practice, 2005, 25 (2): 113-118, 133.
- [8] 郑寇全, 雷英杰, 王睿等. 基于 IFSTR 的抽象化空间推理方法[J]. 系统工程与电子技术, 2013, 35(3): 651-654.
ZHENG K Q, LEI Y J, WANG R, *et al.* The method for abstract spatial reasoning based on IFSTR[J]. Systems Engineering and Electronics, 2013, 35(3): 651-654.
- [9] 樊雷, 雷英杰. 直觉模糊统计决策与判决[J]. 控制与决策, 2011, 26(3):357-362.
FAN L, LEI Y J. Intuitionistic fuzzy statistic adjudging and decision-making[J]. Control and Decision, 2011, 26(3):357-362.
- [10] 贺正洪, 雷英杰. 直觉模糊 c 均值聚类算法研究[J]. 控制与决策, 2011, 26(6): 847-850,856.
HE Z H, LEI Y J. Research on intuitionistic fuzzy c -means clustering algorithm[J]. Control and Decision, 2011, 26(6): 847-850, 856.
- [11] 申晓勇, 雷英杰, 李进等. 基于目标函数的直觉模糊集合数据的聚类方法[J]. 系统工程与电子技术, 2009, 31(11):2732-2735.
SHEN X Y, LEI Y J, LI J, *et al.* Clustering technique to intuitionistic fuzzy sets data based on objective function[J]. Systems Engineering and Electronics, 2009, 31(11):2732-2735.
- [12] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [13] CHEN S C, ZHANG D Q. A novel kernelized fuzzy c -means algorithm with application in medical image segmentation[J]. Artificial Intelligence in Medicine, 2004, 32(1): 37-50.
- [14] CAMASTRA F, VERRI A. A novel kernel method for clustering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, 27(5): 801-805.
- [15] FILIPPONE M, CAMASTRA F, MASULLI F, *et al.* A survey of kernel and spectral methods for clustering[J]. Pattern Recognition, 2008, 41(1): 176-190.
- [16] GIROLAMI M. Mercer kernel based clustering in feature space[J]. IEEE Trans on Neural Networks, 2002, 13(3):780-784.
- [17] 张莉, 周伟达, 焦李成. 核聚类算法[J]. 计算机学报, 2002, 25(6): 587-590.
ZHANG L, ZHOU W D, JIAO L C. Kernel clustering algorithm[J]. CHINESE J.COMPUTERS, 2002, 25(6):587-590.
- [18] 范成礼, 邢清华, 付强等. 基于直觉模糊核聚类的弹道中段目标识别方法[J]. 系统工程与电子技术, 2013,35(7):1362-1367.
FAN C L, XING Q H, FU Q, *et al.* Technique for target recognition in ballistic midcourse based on intuitionistic fuzzy kernel clustering[J]. Systems Engineering and Electronics, 2013,35(7):1362-1367.
- [19] 曲福恒, 马驹良, 胡雅婷. 一种基于核的模糊聚类算法[J]. 吉林大学学报, 2008,46(6): 1137-1141.
QU F H, MA S L, HU Y T. A kernel based fuzzy clustering algorithm[J]. Journal of Jilin University, 2008, 46(6): 1137-1141.
- [20] 雷阳, 雷英杰, 周创明. 基于直觉模糊核匹配追踪的目标识别方法[J]. 电子学报, 2011, 39(6): 1441-1446.
LEI Y, LEI Y J, ZHOU C M. Techniques for target recognition based on intuitionistic fuzzy kernel matching pursuit[J]. Acta Electronica Sinica, 2011, 39(6): 1441-1446.
- [21] 雷阳, 孔韦韦, 雷英杰. 基于直觉模糊 c 均值聚类核匹配追踪的弹道中段目标识别方法[J]. 通信学报, 2012,33(11):136-143.
LEI Y, KONG W W, LEI Y J. Technique for target recognition based on intuitionistic fuzzy c -means clustering and kernel matching pursuit[J]. Journal on Communications, 2012, 33(11):136-143.
- [22] KENNEDY J, EBERHART R C. Particle swarm optimization[A]. Proc of IEEE Internal Conference on Neural Networks Perth[C]. 1995. 1942-1948.
- [23] RANA S, JASOLA S, KUMAR R. A review on particle swarm optimization algorithms and their applications to data clustering[J]. Artificial Intelligence Review, 2011, 35(3):211-222.
- [24] 王纵虎, 刘志镜, 陈东辉. 一种改进的粒子群优化快速聚类算法[J]. 西安电子科技大学学报, 2012, 39(5):61-65, 78.
WANG Z H, LIU Z J, CHEN D H. Improved PSO-based fast clustering algorithm[J]. Journal of Xidian University, 2012, 39(5):61-65, 78.
- [25] 于德亮, 唐海燕, 丁宝等. 基于粒子群优化模糊核聚类的电梯群交通模式识别[J]. 哈尔滨工业大学学报, 2012, 44(10):84-88.
YU D L, TANG H Y, DING B, *et al.* Elevator traffic mode identification with kernel fuzzy clustering based on particle swarm optimization[J]. Journal of Harbin Institute of Technology, 2012, 44(10):84-88.

作者简介:



余晓东 (1989-), 男, 江西九江人, 空军工程大学博士生, 主要研究方向为模式识别、智能信息处理等。



雷英杰 (1956-), 男, 陕西渭南人, 空军工程大学教授、博士生导师, 主要研究方向为智能信息处理与智能决策。

岳韶华 (1968-), 女, 湖北黄梅人, 空军工程大学高级实验师、硕士生导师, 主要研究方向为智能信息处理与智能决策。

王睿 (1964-), 女, 陕西三原人, 空军工程大学副教授、硕士生导师, 主要研究方向为智能信息处理与多传感器信息融合。