

# 一种基于改进 PSO 的 K-means 优化聚类算法

谢秀华,李陶深

(广西大学 计算机与电子信息学院,广西 南宁 530004)

**摘要:**针对传统的 K-means 算法对初始聚类中心的选取敏感、容易收敛到局部最优的缺点,提出一种基于改进粒子群优化算法(PSO)的 K-means 优化聚类算法。该算法利用 PSO 算法强大的全局搜索能力对初始聚类中心的选取进行优化;通过动态调整惯性权重等参数增强 PSO 算法的性能;利用群体适应度方差决定算法中前部分 PSO 算法和后部分 K-means 算法的转换时机;设置变量实时监控各个粒子和粒子群的最优值变化情况,及时地对出现早熟收敛的粒子进行变异操作,从而为 K-means 算法搜索到全局最优的初始聚类中心,使聚类结果不受初始聚类中心影响,易于获得全局最优解。实验结果表明文中提出的改进算法与传统聚类算法相比具有更高的聚类正确率、更好的聚类质量及全局搜索能力。

**关键词:**聚类;K-means 算法;粒子群优化算法;全局最优

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2014)02-0034-05

doi:10.3969/j.issn.1673-629X.2014.02.008

## An Optimized K-means Clustering Algorithm Based on Improved Particle Swarm Optimization

XIE Xiu-hua, LI Tao-shen

(School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China)

**Abstract:** Aiming at the shortcomings of traditional K-means algorithm which is sensitive to initial clustering centers and easy to converge to local optima, an optimized clustering algorithm of K-means based on improved PSO algorithm is proposed. It takes advantages of the powerful global searching capability of PSO algorithm to improve the selection of the initial centers; updating parameters dynamically, for the inertia weight, so as to strengthen the global searching capability of PSO; determining the occasion of the improved algorithm transferred from PSO to K-means by the fitness variance of the particle swarm; using variants to monitor the optimal value condition of every particle and the particle swarm in real time and executing mutation operations on those particles that converge prematurely on time, which protects the clustering results from being influenced by the initial clustering centers and thus achieves the global optima solution. Experimental results show that the proposed method has higher accuracy rates, better clustering quality and global searching capability.

**Key words:** clustering; K-means algorithm; PSO algorithm; global optimum

## 0 引言

聚类分析是一种无监督分类技术,按照一定的相似性标准将数据集进行分类,使得类内的对象尽可能相似,而不同类之间的对象尽可能相异<sup>[1-2]</sup>。K-means 算法<sup>[3]</sup>是基于划分的经典聚类算法,具有容易理解、实现简单、收敛速度快等许多优点,但同时它也有着如对初始聚类中心选取敏感、容易收敛到局部最优解等不足之处。

初始聚类中心的选择决定了 K-means 算法的主要处理过程以及数据的划分结果,不同的初始聚类中

心可能得到不同结果。聚类中心选取不当,可能会导致算法陷入局部最优解。已有许多研究人员在聚类时引进粒子群优化算法来解决这个问题。文献[4]提出一种自适应非线性惯性权重的粒子群和 K 均值混合的算法。文献[5]提出一种 PSO 混合 K-means 的聚类算法,并对混合聚类算法实现基于 MPI 的并行化,提高算法的执行效率。文献[6]采用经典的粒子群优化算法对 K-means 算法的初始聚类中心进行改进,提高了聚类结果的精度。文献[7]提出的整合 K 均值和 PSO 的新聚类算法结合了两算法的优点,改善了聚

收稿日期:2013-04-22

修回日期:2013-07-28

网络出版时间:2013-11-29

基金项目:国家自然科学基金资助项目(60973074)

作者简介:谢秀华(1987-),女,硕士,研究方向为数据库系统理论与技术、数据挖掘;李陶深,教授,研究方向为分布式工程数据库、无线 Mesh 网络、网络计算与信息安全、智能信息技术。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20131129.0945.033.html>

类算法的质量。以上算法都能在一定程度上提高聚类的效果,缓解 K-means 算法对初值的依赖问题,但是在经典的粒子群算法中粒子群可能出现早熟收敛现象,算法仍可能陷入局部极值。文献[8]研究了基于改进粒子群算法的 K-means 聚类算法,对陷入局部极值的粒子进行处理使其跳出局部最优解,算法虽继承了 PSO 算法的全局搜索能力,却没有充分有效地利用 K-means 算法的局部搜索能力。在分析上述算法的基础上,文中提出一种基于改进 PSO 的 K-means 优化聚类算法。通过改进 PSO 算法中粒子群的演化迭代搜索得到全局最优的  $k$  个聚类中心,然后以此为初始聚类中心执行 K-means 算法来获得理想的聚类划分,充分利用 PSO 算法的全局寻优能力和 K-means 算法的局部搜索能力,动态调整粒子的惯性权重系数以及飞行时间,实时地监控各个粒子和粒子群的状态,以便及时地对陷入局部极值的粒子进行变异,避免 PSO 算法中粒子群出现早熟收敛的现象,消除 K-means 算法对初始聚类中心选取的依赖性,使其脱离局部极值获得近似全局最优的聚类划分,提高聚类算法的正确率并改善聚类结果的质量。

## 1 相关算法的简介

### 1.1 K-means 算法

K-means 算法根据相似性准则将数据集划分到  $k$  个分类中,算法的流程描述如下:

- (1) 从数据集中随机选择  $k$  个初始聚类中心;
- (2) 对数据集中的每个数据对象:计算其与所有聚类中心的距离,并按照最近邻原则将其划分到距离最近的类中;
- (3) 重新计算各个新形成聚类的聚类中心;
- (4) 重复执行(2)、(3)两步操作直到聚类中心不再发生变化则算法结束。

其中,相似性度量采用欧几里得距离计算方法,聚类中心为类内所有数据对象的均值。

### 1.2 PSO 算法

粒子群优化(Particle Swarm Optimization, PSO)算法是由 Kennedy 和 Eberhart<sup>[9]</sup>提出的一种模拟鸟群觅食过程中群体行为的新群体智能算法,是解决聚类 NP 难问题应用最广泛的方法之一。目前已有许多研究将粒子群优化算法应用到聚类分析中,例如文献[10]利用粒子群优化算法对基于滑动窗口模型的数据流聚类算法进行改进,解决了传统数据流聚类算法存在的问题。PSO 算法中粒子根据自身经验(个体最优位置)和群体共享的社会经验(群体最优位置)改变速度和位置朝全局最优解的方向飞行,并通过一个预先定义好的关于当前位置的适应度函数对粒子的性能

进行评价。假设种群规模为  $m$  的粒子群在  $n$  维空间搜索时,若目前为止粒子的个体最优位置为  $Pbest_i = (pbest_{i1}, pbest_{i2}, \dots, pbest_{in})$ ,群体最优位置为  $Gbest = (gbest_1, gbest_2, \dots, gbest_n)$ ,则每个粒子的飞行速度  $V_i = (v_{i1}, v_{i2}, \dots, v_{in})$  和位置  $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$  ( $i=1, 2, \dots, m$ ) 可以根据下面的式子进行调整:

$$V_i(t+1) = w \times V_i(t) + c_1 \times r_1 \times (Pbest_i - X_i) + c_2 \times r_2 \times (Gbest - X_i) \quad (1)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (2)$$

其中,  $t$  是表示迭代次数的变量;  $w$  称为惯性权重系数; 常数  $c_1$ 、 $c_2$  为学习因子;  $r_1$ 、 $r_2$  为均匀分布在  $[0, 1]$  范围内的随机数。

## 2 基于改进 PSO 的 K-means 聚类算法

K-means 算法用于聚类分析时收敛速度很快,局部搜索能力比较强,但聚类结果受初始聚类中心的影响比较大,不同的初始中心点可能导致不同的聚类结果,得到的聚类结果波动性很大,若聚类初值选择不当算法很容易陷入局部最优解。而 PSO 算法则具有较强的全局搜索能力,但其收敛速度在后期变得缓慢,收敛速度不理想,而且经典的 PSO 算法中粒子群可能存在早熟现象。针对这些问题,文中提出一种基于改进 PSO 算法的 K-means 聚类算法,动态调整粒子的惯性权重系数和飞行时间,增强 PSO 算法的全局搜索性能,通过群体适应度方差来判断两者的结合时机以充分利用 PSO 算法优秀的全局优化能力以及 K-means 算法的快速搜索能力,加快粒子群后期收敛速度,实时监控粒子群的状态变化,对出现早熟收敛的粒子进行及时地变异操作,避免粒子群陷入局部极值,利用粒子群搜索全局最优的初始聚类中心,使 K-means 算法对初始聚类中心选择的敏感性得到抑制,获得理想的聚类划分结果。

### 2.1 动态调整惯性权重和飞行时间

适当地对 PSO 算法中的参数进行调整,有助于增强算法的全局搜索能力。惯性权重  $w$  值越大算法的全局搜索能力越强,  $w$  值越小则算法的局部搜索能力越强,而理想的 PSO 算法是前期全局搜索能力较强而后后期局部搜索能力较强,因此文中 PSO 算法的  $w$  值可采用以下线性调整策略:

$$w(t) = w_{\max} - (w_{\max} - w_{\min})t/t_{\max} \quad (3)$$

其中,  $w_{\max}$  是最大惯性权重,一般取值 1.4;  $w_{\min}$  是最小惯性权重,一般取值 0;  $t$  为当前迭代次数;  $t_{\max}$  是粒子群的最大迭代次数。

经过研究发现<sup>[11]</sup>,在实际情况下鸟群每次位置迁移时的飞行时间都在不断变化,因此可给粒子增加一个飞行时间因子避免粒子在最优解周围来回“振荡”,

加快粒子群的收敛时间。

粒子的位置调整公式可改为:

$$X_i(t+1) = X_i(t) + H_0(1 - t/t_{\max})V_i(t+1) \quad (4)$$

其中,飞行常数  $H_0$  一般取值 1.5;  $t$  为当前迭代次数;  $t_{\max}$  为粒子群的最大迭代次数。

粒子飞行时若脱离搜索空间,可将其位置设为粒子边界上下界之差的一半。其中,粒子的位置边界  $[X_{\max}, X_{\min}]$  为待聚类数据集中各维数据的最大、最小值,速度的范围设为  $[-p * X_{\max}, p * X_{\max}]$ ,若粒子速度超出边界,则将其置为最大速度值的一半。

## 2.2 粒子群的编码和适应度函数

粒子群中的每个粒子均代表解问题空间中的一个潜在可行解,而聚类算法要解决的问题是求出满足条件的聚类中心,因此可将其映射为粒子群中的粒子,即粒子的位置  $x_i$  可表示为由  $k$  个聚类中心  $Z_j (1 \leq j \leq k)$  组成的向量。若聚类数据集为  $q$  维向量,则粒子的位置和速度都是  $q \times k$  维向量。因此粒子可采用如下的实数编码形式:

$$Z_{11} Z_{12} \cdots Z_{1q} Z_{21} Z_{22} \cdots Z_{2q} \cdots Z_{k1} Z_{k2} \cdots Z_{kq} || \\ V_{11} V_{12} \cdots V_{1q} V_{21} V_{22} \cdots V_{2q} \cdots V_{k1} V_{k2} \cdots V_{kq} || f(x)$$

将 PSO 算法引入 K-means 算法时,可将评价聚类质量的准则函数作为粒子群的适应度函数。假设聚类数据集  $S = (S_1, S_2, \cdots, S_n)$  划分为  $k$  类,每个类  $C_j (j=1, 2, \cdots, k)$  的聚类中心为  $Z_j$ ,则粒子的适应度函数可定义为:

$$f(x) = \sum_{j=1}^k \sum_{S_i \in C_j} || S_i - Z_j || \quad (5)$$

粒子的适应度值表示各个类内数据对象之间的相似度,适应度值越小则表明类内数据对象的结合程度越紧密,聚类效果越好。因此改进 PSO 算法的目标是搜索到使粒子的适应度取值最小的粒子位置,则该粒子位置对应的聚类中心即为优化的聚类中心初值。

粒子  $i$  在第  $t+1$  次迭代时,若  $f(x_i(t+1)) < f(\text{Pbest}_i(t))$ ,则  $\text{Pbest}_i(t+1) = f(x_i(t+1))$ ,否则  $\text{Pbest}_i(t+1) = \text{Pbest}_i(t)$ 。

若  $\min(f(\text{Pbest}_i(t+1))) < f(\text{Gbest}(t))$ ,则  $\text{Gbest}(t+1) = \text{Pbest}_{\min}(t+1)$ ,否则  $\text{Gbest}(t+1) = \text{Gbest}(t)$ 。

## 2.3 粒子变异操作

粒子群在迭代进行的过程中有可能出现早熟收敛的问题而陷入局部最优解,因此文中在 PSO 算法部分,设置两个变量对每个粒子和粒子群的状态进行实时地监控,以便当检测到粒子或粒子群出现早熟收敛现象时,能及时地对其进行变异操作,增加粒子的多样性,使其跳出局部极值的束缚,在多维解空间中开始新的搜索。具体的做法是:

(1) 设置控制变量  $N_{pi} (1 \leq i \leq m, m$  为粒子数量) 来实时监控每个粒子  $i$  的最优值变化情况。若粒子  $i$  的适应度值  $f(x_i)$  没有优于个体最优位置对应的个体极值  $f(\text{Pbest}_i)$ ,表示此次迭代粒子  $i$  无法改善自身的个体最优值,则将表示粒子状态无法得到改善的累计次数的变量  $N_{pi}$  增加一次。若粒子  $i$  的最优值连续无法得到改善的累计次数  $N_{pi}$  超过规定的阈值  $\text{thre}_p$ ,则表明粒子  $i$  出现了早熟现象,应对粒子  $i$  进行变异操作,以使其尽快脱离局部极值。

(2) 设置一个控制变量  $N_g$  来实时监控粒子群的最优值变化情况。若此次迭代所有粒子的个体极值都不小于全局极值,即粒子群的最优值无法得到改善,则将表示粒子群的状态无法得到改善的累计次数的变量  $N_g$  增加一次。若粒子群的最优值连续无法得到改善的累计次数  $N_g$  超过规定阈值  $\text{thre}_g$ ,则表明粒子群很可能陷入了局部最优解。此时可以对所有粒子的适应度值进行排序,选择适应度值最小的前几个粒子对其进行变异操作,增加粒子群的多样性,克服粒子群的早熟收敛问题。

粒子的变异操作具体做法是对变异粒子执行一次 K-means 操作,使粒子的位置、速度以及个体最优位置得到重新初始化。

## 2.4 PSO 算法与 K-means 算法的转换时机

由于算法前面阶段的主要目标是寻找到优化的聚类中心初值,因此仅执行 PSO 算法,以便快速地搜索到 K-means 算法所需的初始聚类中心。粒子群的收敛程度可以通过适应度方差反映出来,群体适应度方差  $\sigma^2$  可定义为:

$$\sigma^2 = 1/m \sum_{i=1}^m [f(x_i) - f_{\text{avg}}]^2 \quad (6)$$

其中,  $m$  为粒子群规模;  $f(x_i)$  为粒子  $i$  的适应度值;  $f_{\text{avg}}$  为所有粒子的适应度均值。当适应度方差  $\sigma^2$  的值很小时,表明粒子群的适应度值的波动性很小,粒子群的状态趋于收敛。因此文中当  $\sigma^2$  小于规定的阈值  $\text{thre}_\sigma$  时,则终止执行改进的 PSO 算法转而开始执行 K-means 算法,以加快 PSO 算法后期的收敛速度。

## 2.5 文中算法的描述

文中提出的基于改进 PSO 算法的 K-means 聚类算法的描述如下:

输入: 待聚类数据集  $S$ , 聚类数目  $k$ , 粒子群的种群规模  $m$ , 最大迭代次数  $t_{\max}$ 。

输出: 聚类数据集的聚类中心不再变化的  $k$  个聚类划分。

算法步骤如下:

Step1: 对粒子群进行初始化操作。从数据集  $S$  中随机选择  $k$  个中心点, 将其作为粒子位置  $X_i$  的初值。



同时,初始化粒子的速度  $V_i$ 、个体最优位置  $Pbest_i$  及其对应的个体极值  $f(Pbest_i)$ 、群体最优位置  $Gbest$  及其对应的全局极值  $f(Gbest)$ 。这一过程循环进行  $m$  次,即可完成粒子群的初始化构造。

- Step2:执行 PSO 算法进行粒子群迭代搜索。
- 对粒子群中的每个粒子执行以下操作:
- (1)根据式(3)动态调整惯性权重,并根据式(1)和(4)分别更新粒子的速度和位置。
- (2)依照最近邻原则划分数据集,计算粒子的适应度值。

(3)若粒子的适应度值小于个体极值,则更新粒子的个体极值以及个体最优位置  $Pbest_i$ 。否则将表示粒子的状态连续无法得到改善累计次数的变量  $N_{pi}$  值加 1。

(4)若所有粒子的个体极值中的最小值小于粒子群的全局极值,则更新粒子群的全局极值以及群体最优位置  $Gbest$ 。否则将表示粒子群最优值无法得到改善累计次数的变量  $N_g$  值加 1。

(5)若  $N_{pi}$  大于规定的阈值  $thre_p$ ,对粒子  $i$  执行变异操作。

(6)若  $N_g$  大于规定的阈值  $thre_g$ ,对适应度值最小的前几个粒子执行变异操作。

(7)若通过群体适应度方差  $\sigma^2$  判断粒子群已趋向收敛或者循环达到最大迭代次数  $t_{max}$ ,终止粒子群的迭代并将  $Gbest$  对应的聚类中心作为 Step3 的初值,否则转(2)继续迭代执行。

Step3:执行 K-means 算法,按类输出最终的聚类结果。

### 3 实验与分析

实验的环境为:操作系统 Windows 7,CPU 3.20 GHz Pentium(R) Dual-Core,内存 2 GB,编译软件为 Eclipse3.2.2,用 Java 语言编程实现。

实验中算法的各个参数取值为:粒子群的种群规模大小  $m=20$ ,粒子群的最大迭代次数  $t_{max}=100$ , $thre_p=4$ , $thre_g=5$ , $p=0.4$ ,粒子群的适应度方差阈值  $thre_\sigma=0.1$ 。实验采用 UCI<sup>[12]</sup> 三个著名的实际数据集 Iris、Wine 以及 BreastCancer 对算法的正确性和有效性进行验证。其中,Iris 数据集包含 150 个数据样本,每个样本有 4 个属性。Wine 数据集包含 178 个数据样本,每个样本包含 13 个属性。BreastCancer 数据集包含 569 个数据样本,每个样本包含 30 个属性。实验将 K-means 算法、pso-km 算法<sup>[7]</sup>和文中算法进行比较,实验中每种算法分别执行 20 次,测试结果取其平均值。算法的聚类准确率以及适应度值的对比结果分别如表 1、表 2 所示。

表 1 K-means、pso-km 以及文中 算法的聚类准确率比较					%
实际数据集	准确率	K-means	pso-km	文中算法	
Iris	最高	89.33	89.33	91.33	
	最低	51.33	88.67	89.33	
	平均	80.17	89.13	90.10	
Wine	最高	70.79	70.79	73.60	
	最低	51.69	70.79	70.79	
	平均	68.88	70.79	71.66	
BreastCancer	最高	85.41	86.29	87.35	
	最低	85.41	85.41	85.59	
	平均	85.41	85.74	86.59	

表 2 K-means、pso-km 以及文中 算法的适应度值比较				
实际数据集	适应度值	K-means	pso-km	文中算法
Iris	最高	123.849 8	97.657 5	97.359 2
	最低	97.325 9	97.222 1	97.222 1
	平均	103.730 0	97.390 1	97.300 3
Wine	最高	18 776.94	16 960.21	16 952.28
	最低	16 960.20	16 940.28	16 940.28
	平均	17 067.07	16 944.78	16 940.89
BreastCancer	最高	152 647.31	151 762.02	151 735.22
	最低	152 647.27	151 735.12	151 735.12
	平均	152 647.29	151 740.43	151 735.14

由表 1 可知,无论是对低维 Iris 数据集、中等维度的 Wine 数据集,还是高维 BreastCancer 数据集,K-means 算法的聚类准确率最低,聚类结果波动性最大,这是 K-means 算法本身对初始聚类中心选取的敏感性造成的,聚类中心初值不同,得到的结果也可能不同。与 K-means 算法相比,pso-km 算法利用了粒子群算法优秀的全局寻优能力,较好地消除了聚类中心选择的随机性对聚类结果的影响,使得算法的聚类准确率得到了提高。而文中算法的聚类准确率在一定程度上要优于 pso-km 算法,这是由于文中算法对粒子的状态进行实时监控,及时地对出现早熟收敛的粒子进行变异操作,使粒子脱离局部极值的束缚,获得重新搜索全局最优解的机会,增加算法获得正确分类的概率。

从表 2 可以看到,K-means 算法的适应度值最大,文中改进算法的适应度值最小,而 pso-km 算法的适应度值则介于两者之间。适应度值是评价聚类质量优劣的有效标准,适应度越小则表明聚类划分得到的类内数据对象之间的结合程度越紧密,由此可知文中改进算法的聚类效果最好。

为了验证算法的收敛性能,对 Iris 数据集进行测试,得到三种算法的收敛曲线图(如图 1 所示)。从图 1 可以看出,K-means 算法的收敛速度最快,而 pso-km 算法由于粒子群的早熟收敛问题,粒子适应度值因局部早熟而无法得到最优解,收敛速度比 K-means 算法慢。文中改进算法由于增加了粒子变异操作,使出现

早熟收敛的粒子及时地跳出局部极值,同时动态地调整粒子的惯性权重和飞行时间加快了粒子群算法的收敛速度,因此文中算法虽然比 K-means 算法的收敛速度慢,但比 pso-km 算法稍快一些。

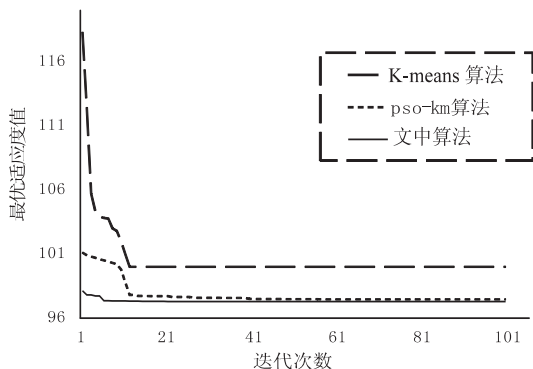


图 1 算法对 Iris 数据集进行测试的收敛曲线图

## 4 结束语

针对 K-means 聚类算法对初始聚类中心的敏感性问题,文中提出了一种基于改进粒子群优化算法的 K-means 优化聚类算法。该算法通过动态调整粒子的惯性权重系数及飞行时间以增强粒子群的全局搜索能力,针对粒子在进化迭代过程中可能会陷入局部最优、出现早熟收敛的问题,实时监控每个粒子和粒子群的状态,并对陷入局部极值的粒子进行变异操作,增加粒子群的多样性,克服粒子群早熟收敛的问题,消除 K-means 算法对聚类初始值的依赖性,改善 K-means 算法由于初值选择不当而使聚类结果陷入局部最优解的问题,并加快了 PSO 算法后期收敛速度。实验结果表明,与传统的聚类算法相比,文中算法的聚类准确率得到了一定程度的提高,聚类的质量得到了改善,能够获得比较理想的聚类划分。然而与传统 K-means 算法相比,文中算法的计算时间有所增加,如何在提高 K-

means 算法的聚类准确率的同时保持较快的计算速度,是笔者下一步需要进一步研究的工作。

## 参考文献:

- [1] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报, 2008,19(1):48-61.
- [2] Han Jiawei, Kamber M. Data mining: Concepts and techniques [M]. 2nd ed. Beijing: China Machine Press, 2006:251-253.
- [3] Zhang Wei, Yang Hen-I, Jiang Hsin-yi, et al. Automatic data clustering analysis of arbitrary shape with K-means and enhanced ant-based template mechanism [C]//Proc of 2012 IEEE 36th international conference on computer software and applications. [s. l.]: [s. n.], 2012:452-455.
- [4] 刘悦婷,李岚. 基于自适应权重的粒子群和 K 均值混合聚类算法研究[J]. 甘肃科学学报, 2010,22(4):106-109.
- [5] 吕奕清,林锦贤. 基于 MPI 的并行 PSO 混合 K 均值聚类算法[J]. 计算机应用, 2011,31(2):428-431.
- [6] 傅涛,孙亚民. 基于 PSO 的 K-means 算法及其在网络入侵检测中的应用[J]. 计算机科学, 2011,38(5):54-55.
- [7] 徐辉,李石君. 一种整合粒子群优化和 K-均值的数据聚类算法[J]. 山西大学学报(自然科学版), 2011,34(4):518-523.
- [8] 陈小全,张继红. 基于改进粒子群算法的聚类算法[J]. 计算机研究与发展, 2012,49(Suppl.):287-291.
- [9] Kennedy J, Eberhart R C. Particle swarm optimization [C]//Proceedings of IEEE international conference on neural networks. Perth: [s. n.], 1995:1942-1948.
- [10] 肖裕权,周肆清. 基于粒子群优化算法的数据流聚类算法[J]. 计算机技术与发展, 2011,21(10):43-46.
- [11] 张学林,丁树良,胡景春,等. 带飞行时间因子的改进粒子群优化算法[J]. 实验技术与管理, 2011,28(5):54-57.
- [12] Center for machine learning and intelligent systems. Machine learning repository [EB/OL]. [2013-03-20]. <http://archive.ics.uci.edu/ml/datasets.html>.

(上接第 33 页)

- [7] Kim T S, May G S. Time series modeling of photosensitive polymer development rate for via formation applications [J]. IEEE transactions on electronics packaging manufacturing, 2002,25(3):203-209.
- [8] 张鑫,温显斌,孟庆霞. 基于颜色特征的图像检索方法研究[J]. 计算机科学, 2012(11):243-245.
- [9] 黄元元,刘宁钟. 一种新的基于颜色特征的图像检索方法[J]. 小型微型计算机系统, 2012,33(3):609-613.
- [10] Berens J, Finlayson G D. Image indexing using compressed color histogram [J]. IEEE trans on vision, image and signal processing, 2000,147(4):349-355.
- [11] 李二森,张保明,周晓明,等. 自适应 Canny 边缘检测算法研究[J]. 测绘科学, 2008,33(6):119-121.
- [12] 肖潇,王宪保,庄德文,等. 基于均匀区域分割的图像检索研究[J]. 计算机科学, 2012,39(6):255-257.

# 一种基于改进PSO的K-means优化聚类算法

作者：[谢秀华](#)，[李陶深](#)，[XIE Xiu-hua](#)，[LI Tao-shen](#)  
作者单位：[广西大学 计算机与电子信息学院, 广西 南宁, 530004](#)  
刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2014(2)

本文链接：[http://d.wanfangdata.com.cn/Periodical\\_wjfz201402009.aspx](http://d.wanfangdata.com.cn/Periodical_wjfz201402009.aspx)