

新的混合小生境鱼群聚类算法

王培崇^{1,2*}, 钱旭¹, 雷凤君¹

(1. 中国矿业大学(北京) 机电与信息工程学院, 北京 100083; 2. 石家庄经济学院 信息工程学院, 石家庄 050031)

(* 通信作者电子邮箱 wpeichong@126.com)

摘要:针对 K-Means 算法对于初始 k 值较敏感和容易过早收敛的问题, 提出基于人工鱼群机制的 K-Means 聚类算法 (NAFS)。首先, 利用先验知识随机产生待求解问题的若干个聚类中心, 组成一个鱼群环境; 其次, 利用鱼群个体的协作、竞争机制寻找满意的结果。鉴于人工鱼群算法后期容易陷入局部最优, 根据鱼群聚集度引入小生境算法, 改善种群的多样性, 提高了算法的求解精度。在 KDDCUP99 数据集上的实验结果表明, 该算法具有较高的聚类精度, 适用于高维数据的聚类问题。

关键词:聚类; 人工鱼群算法; 小生境; 排挤机制; 聚集因子; 算法融合

中图分类号: TP311.13; TP183 **文献标志码:** A

New clustering algorithm based on hybrid niching artificial fish swarm

WANG Pei-chong^{1,2*}, QIAN Xu¹, LEI Feng-jun¹

(1. School of Mechanical Electronic and Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China;

2. School of Information Engineering, Shijiazhuang University of Economics, Shijiazhuang Hebei 050031, China)

Abstract: To overcome the shortcomings such as being sensitive to initial value of k , premature convergence in K-Means algorithm, this paper presented an improved K-Means algorithm based on artificial fish swarm mechanism named NAFS. Firstly, priori knowledge was exploited to randomly generate some cluster centers for the problems to be solved, then composing the fish swarm environment. Secondly, the cooperation and competition mechanism of fish individuals was utilized to search satisfied outcome. In view of the deficiency that artificial fish swarm is prone to fall in local optimum, niching algorithm was introduced according to the fish crowding density to ameliorate the diversity of population and improve its solution accuracy. The results of experiments on KDDCUP99 show NAFS has higher clustering accuracy and is appropriate to solve clustering problems with high dimensionality.

Key words: clustering; Artificial Fish Swarm (AFS); niching; exclusion mechanism; crowding factor; algorithm fusion

0 引言

聚类^[1-6]识别是指在无先验知识的前提下, 将具有相似度的数据聚集在一起。该技术广泛应用于模式识别、数据挖掘、智能信息数据处理中。在众多的聚类算法中, K-Means 是一个受到广泛关注和应用的算法。该算法原理简单, 计算速度快, 但是对于初始 k 值的设置比较敏感, k 值在初始设置时, 很难界定初始范围。而且, 该算法由于采用爬山法进行, 导致计算结果容易陷入局部最优。在处理非凸形状的簇, 以及数据差别较大的簇时, 容易受到孤立点的影响, 导致出现误差。针对 K-Means 算法的缺点, 众多学者对此进行了研究和改进。文献[1]为了克服原始 K-Means 算法对于初始聚类中心的依赖, 在高密度数据区域选择 k 个相距最远的点作为初始聚类中心; 文献[2]提出了基于遗传算法改进 K 均值算法, 但是该算法在数据集和数据的维度较大时, 往往容易过早收敛; 为了求解具有流形特征的高维数据聚类, 文献[3]引入针对流形数据的谱正则化技术, 将低维数据嵌入转换为数据流形上平滑函数的一种线性组合, 然后将低维空间中聚类类间的散度与总体散度的比值最大化^[3]; 文献[4]提出了一个自

组织模糊 K-Means 聚类算法, 通过对目标函数施加惩罚因子的方式, 降低聚类过程对于初始类中心的敏感程度; 文献[5]通过引入小概率随机变异改进粒子群优化算法的种群多样性, 并将其与 K-Means 算法结合, 最后证实了该算法的有效性。

人工鱼群 (Artificial Fish Swarm, AFS) 算法^[7-11]是一种新颖的群智能算法。在人工鱼群算法中, 待处理数据向量使用鱼个体代表, 全部鱼个体组成一个鱼群。算法通过四个算子: 自由游动、觅食、聚群和追尾操作, 快速实现鱼群的聚集, 从而达到求解的目的。

该算法在近几年受到了国内众多专家、学者的重视。文献[7]利用混沌机制初始化鱼群, 同时在聚群和追尾操作之后进行混沌遍历, 使算法容易摆脱极值点; 在文献[8]中, 作者在鱼群算法的后期每隔一定的代数, 引入一次单纯形算子, 替换掉原来大量聚集在非极值点附近的人工鱼个体, 有效提高了局部搜索精度, 提高了解质量; 文献[9]将人工鱼群算法应用于研究求解 ABC 支持的 QoS 单播路由机制; 文献[10]设计了基于人工鱼群技术的 Web 用户聚类分析算法, 并将该算法应用于 Web 用户聚类分析; 文献[11]应用鱼群算法求解

收稿日期: 2012-02-13; 修回日期: 2012-03-29。

基金项目: 河北省科技攻关项目 (11213525D); 石家庄经济学院博士科研基金资助项目。

作者简介: 王培崇 (1973 -), 男, 河北辛集人, 副教授, 博士, 主要研究方向: 智能信息处理、对等计算; 钱旭 (1962 -), 男, 江苏南京人, 教授, 博士生导师, 博士, 主要研究方向: 信息融合、智能信息处理; 雷凤君 (1988 -), 女, 河南南阳人, 硕士研究生, 主要研究方向: 智能信息处理。

0/1 背包问题,根据 0/1 背包问题的特点,直接在编码空间上实施鱼群行为,并设计了优、劣质解随机游走寻优和优质解替换劣质解三个算子,较好地实现了 0/1 背包问题求解。

人工鱼群算法与众多的其他群智能算法(如:粒子群优化算法、蚁群算法、人工蛙跳算法等)一样,通过模拟生物个体的进化、生存等行为,实现寻优。同样,它们的缺点也基本一致:由于缺少严格数学模型和理论上的支持,算法很容易陷入局部最优。同时,研究发现人工鱼群算法在获取问题的精确解上能力相对较差。

为了改善人工鱼群算法的不足,在基本人工鱼群算法中引入基于排挤机制的小生境算法^[12-16]以提高鱼群算法中的种群多样性,使人工鱼群算法的求解能力和求解精度得以提高。最后将改进算法与聚类数学模型进行结合以实现聚类识别。

1 人工鱼群算法

在该算法中引入如下符号:

n 表示整个鱼群中鱼的数量;

$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ 表示人工鱼的状态, $\mathbf{x}_i (i = 1, 2, \dots, n)$ 为欲寻优的变量;

$f(\mathbf{x})$ 表示鱼所处当前位置的食物浓度,即目标函数值;

$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ 表示鱼群内鱼个体之间的距离;

$visual$ 表示人工鱼的感知范围,鱼个体在该范围内实施探测;

$step$ 表示鱼个体每次移动的步长;

δ 表示聚集鱼群的拥挤度因子,用以描述该群体的拥挤程度;

Try_number 表示鱼向聚集鱼群尝试进入次数;

g_{best} 表示当前最优鱼个体的状态值,记录于公告板。

算法运行步骤如下:

步骤 1 初始化各项参数($n, visual, step, \delta, Try_number$)及迭代次数 $n_c = 0$ 。

步骤 2 人工鱼在自己的视野范围内随机游动,其当前状态为 \mathbf{x}_i , 当发现食物的时候向食物靠近。

步骤 3 觅食操作。

1) 设 $m = 0$, 人工鱼的当前状态设为 \mathbf{x}_i 。

2) 执行 $\mathbf{x}_j = random(n(\mathbf{x}_i, visual))$ 生成一个新状态, $m = m + 1$ 。

3) 如果 $f(\mathbf{x}_i) < f(\mathbf{x}_j)$, 则将人工鱼的状态修改为: $\mathbf{x}_i = \mathbf{x}_i + random(step)(\mathbf{x}_j - \mathbf{x}_i) / \|\mathbf{x}_j - \mathbf{x}_i\|$; 否则, 如果 $m < Try_number$, 返回 2) 执行。

4) 执行随机游动行为前进一步: $\mathbf{x}_i = \mathbf{x}_i + random(step)$ 。

步骤 4 聚群操作。

1) 在其视野范围 $visual$ 之内生成其同伴的集合 $k = \{\mathbf{x}_j | \mathbf{x}_j - \mathbf{x}_i \leq visual\} (i, j = 1, 2, \dots, n)$ 。

2) 如果 k 不是空集的话, 求取其中心位置: $\mathbf{x}_c = \sum \mathbf{x}_j / n_f$, 其可见区域内的同伴个数为 n_f , 且 $n_f \geq 1$ 。

3) 如果 $n_f / n < \delta (0 < \delta < 1)$, 并且此时 $f(\mathbf{x}_i) < f(\mathbf{x}_c)$, 则人工鱼向中心位置 \mathbf{x}_c 前进一步: $\mathbf{x}_{next} = \mathbf{x}_i + random(step) * (\mathbf{x}_c - \mathbf{x}_i) / \|\mathbf{x}_c - \mathbf{x}_i\|$, 否则执行步骤 2 行为。

步骤 5 追尾操作。

1) 探测邻域内是否存在状态最优的伙伴 \mathbf{x}_{max} , 如果没有则继续执行步骤 2 操作, 否则执行下一步;

2) 如果 $f(\mathbf{x}_i) < f(\mathbf{x}_{max})$, 并且 \mathbf{x}_{max} 的邻域内伙伴的数目 n_f 满足 $n_f / n < \delta (0 < \delta < 1)$, 则该鱼儿向最优伙伴 \mathbf{x}_{max} 的位置前进一步: $\mathbf{x}_i = \mathbf{x}_i + random(step)(\mathbf{x}_{max} - \mathbf{x}_i) / \|\mathbf{x}_{max} - \mathbf{x}_i\|$ 。

步骤 6 迭代次数或算法终止条件是否满足, 没有满足, 则执行步骤 2, 满足则输出结果。

步骤 7 各人工鱼在每次迭代完成后, 检查自身状态值 \mathbf{x}_i , 如果 $f(\mathbf{x}_i) \geq g_{best}$, 则 $g_{best} = \mathbf{x}_i$ 。

2 聚类模型

当前, 学术界对于聚类并没有一个公认的定义。一般情况下, 可以简单地认为聚类就是将已知的数据集在没有先验知识的前提下, 通过数据划分方法将其分成多个组(group)或簇(cluster), 并使数据簇或组达到如下条件: 同一簇内部对象间达到最大相似度, 不同组对象间的相似度最小。

定义 1 设存在一样本集 $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n\}$, \mathbf{x}_i 为其中某向量, 维度为 d , 该样本集的一个划分为 $C = \{c_1, c_2, \dots, c_i, \dots, c_m\}$, 聚类识别即寻找数据集 D 的一个类划分 C^* , 满足如下条件, 并且使类内离散度的和最小。

$$\begin{aligned} D &= C_1 \cup C_2 \cup \dots \cup C_m \\ C_i &\neq \emptyset; i = 1, 2, \dots, m \\ C_i \cap C_j &= \emptyset; i, j = 1, 2, \dots, m \end{aligned} \quad (1)$$

用于评价类内部数据紧密程度及描述类内部特征的类内离散度定义如下。

定义 2 设存在一数据集的类划分 C^* , 则类内离散度如下表示:

$$J_c = \sum_{i=1}^m \sum_{\mathbf{x}_i \in C_k} d(\mathbf{x}_i, \overline{\mathbf{x}}^{C_k})$$

其中 $d(\mathbf{x}_i, \overline{\mathbf{x}}^{C_k})$ 为向量 \mathbf{x}_i 与其聚类中心的距离。

3 小生境人工鱼群聚类算法

3.1 小生境技术

小生境技术源于生物学中的小生境概念, 是指在生物界中, 相同种类的生物将会生活在一起, 从而形成一个小的生活环境, 而不同种类的个体则会产生分离。在计算科学中, 将具有相同特征的数据划归一类, 不同特征的数据分离开, 从而避免了大量数据聚集于局部最优解周围, 造成种群多样性降低的问题。

目前, 在小生境技术中存在三种种群选择策略: 基于预选选择机制的选择策略、基于排挤机制的选择策略和基于共享机制的选择策略。

基于排挤机制的选择策略如下:

1) 设置排挤因子 CF (取 2 或 3);

2) 在群体中随机选择 $1/CF$ 个群体参与排挤运算;

3) 产生新个体成员, 计算新个体与排挤成员之间的海明距离;

4) 利用新产生的个体排挤掉一些与排挤成员相类似的个体。

小生境技术能够较好地保持种群多样性,具备较强的跳出局部最优的能力,求解精度高。人工鱼群算法前期收敛速度快,而后期容易陷入局部最优,求解数据精度较差。对比两个算法的优缺点,本文将小生境的排挤选择策略引入人工鱼群算法中,设计了混合小生境人工鱼群聚类算法 NAFS。该算法有效解决了人工鱼群算法后期收敛能力差的问题,提高了算法整体的求解精度。

3.2 NAFS 算法中的一些问题

1) 算法中个体的编码处理。

由于使用二进制编码会造成的反复编码、译码等操作,而且考虑到二进制编码长度受限的问题,本文算法采用实数编码,鱼群中的个体是待求解向量,假设有 C 维,如果该数据集有 m 个聚类,则其编码长度应该是 $C \times m$,这样的编码方式简单,易懂,比较直观。

采用的编码方式如下:

$$p_i(t) = [\bar{X}_{(c_1)}(t), \bar{X}_{(c_2)}(t), \dots, \bar{X}_{(c_m)}(t)] \quad (2)$$

其中 $\bar{X}_{(c_m)}$ 表示第 C 维 m 类的聚类中心。

2) 适应度函数的选择。

如何评价一个聚类的划分是否已经收敛到最佳,是算法设计中的重要考虑点之一。依据聚类的定义可知,一个成功的聚类划分,必然使聚类内部的平均离散度是最小的,故本文采用最小值方式进行设计适应度函数。故适应度函数设计如下:

$$f_c = \sum_{i=1}^m \sum_{x_i \in C_k} |x_i - \bar{X}_k|^2 \quad (3)$$

聚类中心确定之后,数据的划分依据最近邻法进行计算处理,如果 $d(x_i, \bar{X}_{(c_i)}) = \min_{l=1,2,\dots,m} d(x_i, \bar{X}_{(c_l)})$ 成立,则 x_i 划归 C_i 类。在将相关的数据划归入 C_i 类之后,重新计算该类的类中心 \bar{X}_{c_i} ,通过如此反复计算调整。

3) 解空间初始化。

算法采用随机初始化种群的方式产生所需要的参与运算的鱼群。将原始数据随机指定组合形成 m 个类,计算分类之后的各个类的聚类中心,生成一个鱼个体向量,例如第一代鱼个体的编码表示如下:

$$p_i(0) = [\bar{X}_{(c_1)}(0), \bar{X}_{(c_2)}(0), \dots, \bar{X}_{(c_m)}(0)]$$

如此循环,直至生成所需要的全个体。

4) 算法的终止条件。

在发现参与运算的个体适应度函数值连续多次变化范围低于 0.01%,或者小于预先设定的最大迭代次数时,算法终止。

5) 引入小生境机制的时机。

人工鱼群算法早期收敛速度较快,能够迅速找到问题的粗略解。所以,直接引入小生境机制是不合适的。通过分析可以看出,如果鱼群出现聚集现象,则必然是大多数鱼个体的适应度函数值趋于一致,所以采用式(4)来描述鱼个体之间的聚集程度。

$$\alpha^2 = - \sum_{j=1}^m \left(\frac{f_j - f_{avg}}{f} \right)^2 \quad (4)$$

其中: f_i 是鱼个体的适应度函数值, f_{avg} 是鱼个体的适应度平均值。显然, α^2 的值越小,说明鱼的聚集程度越高,当其为 0 时,则算法要么是已经全局收敛完毕,要么早熟收敛于局部最优。

设阈值 λ 为较小的一个值,当 $\alpha^2 \leq \beta$ 时,引入小生境机制,随机生成现有鱼个体数目 20% 的新鱼个体,利用小生境

技术的排挤机制,将不符合条件的鱼个体排挤掉,以保持种群的多样性。

3.3 NAFS 算法流程

通过以上的分析,NAFS 算法流程描述如图 1 所示。

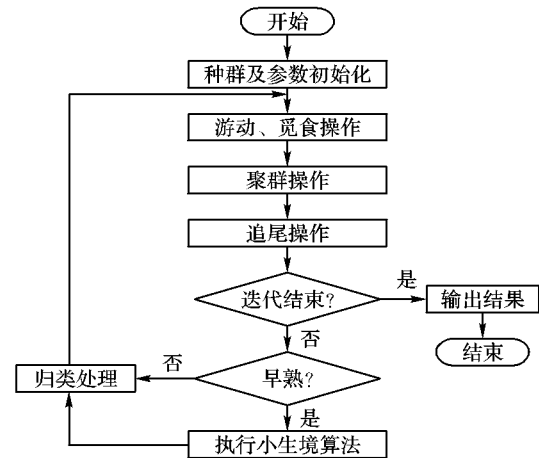


图1 NAFS 算法流程

4 实验结果与分析

为了检验算法的有效性,选择了三个数据集进行实验和对比。首先从 KDDCUP99 中随机选择其中 10% 的数据量,由于该数据集维度较高,实验时仅保留了其中 7 个数据特征。将该数据集与 K-Means 算法、AFS 聚类算法进行实验对比,三个算法在该数据集上均运行 50 次,取最小聚类精度、最大聚类精度以及平均聚类精度三个指标,实验结果列于表 1。同时还对比了三个算法在不同的数据集大小和不同聚类数下,执行时间上的差异,实验结果见图 2、图 3。

从表 1 的数据可以看出,本文所设计的算法 NAFS 在三个数据集上的性能都要比 K-Means、AFS 聚类算法优越。并且通过纵向对比可以看出随着数据的不断增大,三个算法的三个计算指标均有所下降。但是,NAFS 的下降幅度要比 K-Means 和 AFS 聚类算法小。

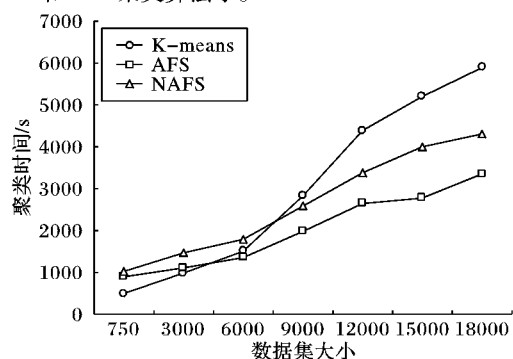


图2 不同数据集下算法执行时间比较

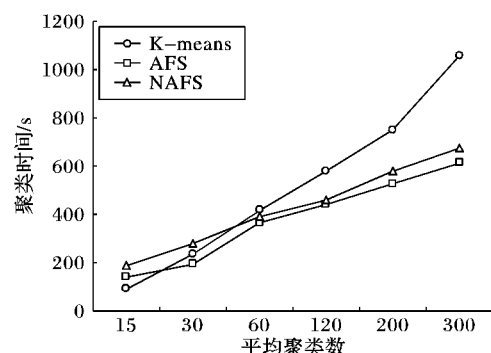


图3 不同聚簇数目下算法执行时间比较

表 1 NAFS、AFS 和 K-Means 算法在不同大小数据集下的聚类结果

算法	20% 数据集			50% 数据集			100% 数据集		
	平均聚类精度/%	最大聚类精度/%	最小聚类精度/%	平均聚类精度/%	最大聚类精度/%	最小聚类精度/%	平均聚类精度/%	最大聚类精度/%	最小聚类精度/%
NAFS 聚类算法	98.73	99.99	97.86	97.33	99.06	96.18	95.29	97.16	92.82
AFS 聚类算法	96.01	99.99	91.32	94.07	95.15	90.12	91.65	93.74	86.18
K-Means 算法	95.21	99.99	92.10	94.93	96.15	91.10	92.57	95.35	89.67

从图 2、3 可以看出,由于 NAFS 算法中增加了小生境算法,其计算时间比 AFS 聚类算法要长一些,但是考虑到其求解精度上要优于 AFS 聚类算法,所以这种时间上的开销还是值得的。而对比 K-Means 算法可以看出,在数据量较小或聚簇数目较小时,该算法的运算时间要比 K-Means 算法稍高,但是随着数据量的增大和聚簇数目的增加,由于群智能算法特有的启发式搜索机制,其运算时间增长速率明显小于 K-Means 算法。

5 结语

本文研究了人工鱼群算法和聚类识别技术,给出了应用人工鱼群算法实现聚类识别的思路。在算法后期根据鱼群个体的聚集程度引入小生境算法,新算法克服了简单人工鱼群算法容易过早收敛的弱点,有效提高了算法的求解精度和求解能力。通过在 KDDCUP99 数据集上的实验表明,NAFS 算法较 K-Means 算法和简单人工鱼群算法的求解精度更高,稳定性也更强,为聚类识别提供了新思路。但是,由于群智能算法缺少严格的数学模型支持,根据聚集因子很难定性说明人工鱼群算法所处状态,如何定量、定性确定算法的收敛性,将是进一步的研究内容。

参考文献:

- [1] 傅德胜,周辰.基于密度的改进 K 均值算法及实现[J].计算机应用,2011,31(2):432-434.
- [2] KRISHNA K, MURTY M N. Genetic K-means algorithm[J]. IEEE Transactions on System, Man, and Cybernetics: Part B, 1999, 29(3):433-439.
- [3] 高丽平,周雪燕,詹宇斌.流形上的非线性判别 K 均值聚类[J].

计算机应用, 2011, 31(12): 3247-3251.

- [4] LI M J, NG M K, CHEUNG Y-M, et al. Agglomerative fuzzy K-means clustering algorithm with selection of number of clusters[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(11): 1519-1534.
- [5] 陶新民,徐晶,杨立标,等.一种改进的粒子群和 K 均值混合聚类算法[J].电子与信息学报,2010,32(1):92-97.
- [6] 张涤,杨艳,唐瑞雪.基于文化算法的混合聚类方法[J].计算机工程与应用,2009,45(4):159-161.
- [7] 曲良东,何登旭.一种混沌人工鱼群优化算法[J].计算机工程与应用,2010,46(22):40-42.
- [8] 张红霞,罗毅,师瑞峰.基于单纯形法的改进型人工鱼群算法[J].计算机应用,2011,31(5):1321-1323.
- [9] 王兴伟,秦培玉,黄敏.基于人工鱼群的 ABC 支持型的 QoS 单播路由机制[J].计算机学报,2010,33(4):718-723.
- [10] 臧文科,刘希.基于人工鱼群算法的 Web 用户聚类分析[J].计算机工程与应用,2011,47(30):12-14.
- [11] 廖向阳,朱命昊,赵亚敏.求解 0/1 背包问题的改进人工鱼群算法研究[J].计算机工程与应用,2011,47(21):43-46.
- [12] 拓守信,汪文勇.求解高维多模优化问题的正交小生境自适应差分演化算法[J].计算机应用,2011,31(4):1094-1098.
- [13] 华杰,崔杜武.基于个体优化的自适应小生境遗传算法[J].计算机工程,2010,36(1):194-196.
- [14] 张宝吉.基于自适应小生境遗传算法的船型优化[J].计算机工程,2011,37(8):207-209.
- [15] 李红婵,朱颖东.基于小生境遗传算法的排课问题研究[J].计算机工程,2011,37(16):194-196.
- [16] 黄平,于金杨,元泳泉.一种改进的小生境多目标粒子群优化算法[J].计算机工程,2011,37(18):1-3.

(上接第 2188 页)

5 结语

针对 K-means 算法随机选取初始聚类中心导致聚类准确率低和聚类结果不稳定的缺陷,提出一种改进的初始聚类中心选取算法。改进算法借助 k -dist 图来选择主要密度水平的平缓曲线上 k -dist 值最小的点作为初始聚类中心,能够反映数据的实际分布。由于聚类结果的中心往往在密度较大的区域,本文算法在一个密度水平上选择的 k -dist 值最小,也就是密度相对最大的点作为初始聚类中心。这样的初始聚类中心唯一的,保证了聚类结果的稳定,提高了聚类准确率,减少了迭代次数。

参考文献:

- [1] HAN J, KAMBER M. 数据挖掘概念与技术[M]. 范明, 孟晓峰, 译. 北京: 机械工业出版社, 2001: 223-225.
- [2] 吴凤慧,成颖,郑彦宁,等. K-means 算法研究综述[J]. 知识组织与知识管理, 2011, 27(5): 28-35.
- [3] 刘一鸣,张化祥. 可变阈值的 K-Means 初始中心选择方法[J]. 计算机工程与应用, 2011, 47(32): 56-58.
- [4] 熊忠阳,陈若田,张玉芳. 一种有效的 K-means 聚类中心初始化方法[J]. 计算机应用研究, 2011, 28(11): 4188-4190.
- [5] 全雪姣,孟凡荣,王志晓. 对 K-means 初始聚类中心的优化[J].

计算机工程与设计, 2011, 32(8): 2721-2788.

- [6] 姚跃华,史秀岭.一种优化初始中心的 K-means 粗糙聚类算法[J].计算机工程与应用,2010,46(34):126-128.
- [7] 周海岩,白晓林.基于图的 K-均值聚类法中初始聚类中心选择[J].计算机测量与控制,2010,18(9):2167-2169.
- [8] 黄敏,何中市,邢欣来,等.一种新的 K-means 聚类中心选取算法[J].计算机工程与应用,2011,47(35):132-134.
- [9] 王赛芳,戴芳,王万斌,等.基于初始聚类中心优化的 K-均值算法[J].计算机工程与科学,2010,32(10):105-116.
- [10] 汪中,刘贵全,陈恩红.一种优化初始中心点的 K-means 算法[J].模式识别与人工智能, 2009, 22(2): 299-304.
- [11] 钱线,黄萱菁,吴立德.初始化 K-means 的谱方法[J].自动化学报,2007,33(4):342-346.
- [12] 王玲,薄列峰,焦李成.密度敏感的谱聚类[J].电子学报, 2007, 35(8): 1577-1581.
- [13] 张琳,陈燕,汲业,等.一种基于密度的 K-means 算法研究[J].计算机应用研究, 2011, 28(11): 4071-4085.
- [14] 周董,刘鹏. VDBSCAN: 变密度聚类算法[J].计算机工程与应用, 2009, 45(11): 137-153.
- [15] ESTER M, KRIEDEL H-P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise [C]// KDD-96: Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining. Menlo Park: AAAI Press, 1996: 226-231.