

文章编号: 2095-4980(2016)03-0426-06

一种基于信息素的蚁群聚类算法

王 慧^a, 甘 泉^b

(平顶山学院 a.招生就业处; b.计算机与科学技术学院, 河南 平顶山 467000)

摘 要: 提出了一种改进的蚁群聚类分析算法, 通过改进 LF 算法中群体相似度函数, 加入参数的自适应调整策略, 利用短期记忆和网格信息素的局部分布控制蚂蚁的随机移动, 并结合蚂蚁速度动态变化、半径递增、强制放下等特性。采用测试数据和不同的算法进行了对比实验分析, 仿真实验结果表明, 该算法显示出了较高的稳定性和准确率。

关键词: 聚类; 蚁群聚类; 信息素

中图分类号: TN91.3; TP301.6 **文献标识码:** A **doi:** 10.11805/TKYDA201603.0426

An Ant Colony Clustering Algorithm Based on Pheromones

WANG Hui^a, GAN Quan^b

(a.Department of Admission and Employment of Pingdingshan University; b.College of Computer Science and Technology, Pingdingshan Henan 467000, China)

Abstract: An improved ant clustering analysis algorithm is proposed, which modifies the similarity function of the Lumer-Faieta(LF) groups, and adds parameters to adjust the adaptive strategy, controls the random movement of ants by using the short-term memory and grid of the local pheromone distribution, combining with dynamic changing of ants speed, increasing radius, forcing down. Comparison experiments are performed by different algorithms with testing data. The results indicate that the proposed algorithm shows a high stability and a high accuracy.

Key words: clustering; ant colony; pheromone

数据挖掘技术领域中的数据聚类的存在, 是了解事物的重要步骤。20 世纪 90 年代, Dario M 等人提出了蚁群算法^[1], 最早被用于解决旅游商和二度调度问题^[2]。随着蚁群算法的广泛应用^[3-4], 人们发现蚁群模型在某些方面更接近实际的聚类问题。在与其他聚类分析模型进行比较时, 基于蚁群的聚类分析模型有较多的优点: 这种模型需要独自解决相关难题, 自动根据数据对象的特征产生聚类的数量, 更好地处理噪声、异常值^[4]和不同维度的数据集, 能较容易完成数据的可视化。

1 蚁群聚类基本原理

Lumer 和 Faieta 等人^[5]对基本模型进行了扩展, 在数据分析领域广泛应用, 提出了 LF 算法。该算法先将数据任意散放在单独网格中, 在所在地蚂蚁能有效地感受到周围的状况。对象 O_j 在地点 r 与附近事物的相似度由式(1)计算^[6]:

$$f(O_i) = \max\{0, \frac{1}{s^2} \sum_{o_j \in \text{Neigh}_{s \times s}(r)} [1 - \frac{d(o_i, o_j)}{\alpha}]\} \quad (1)$$

式中: $f(O_i)$ 表示对象 O_i 和其邻域内的对象 O_j 之间的相似度的度量^[6-7]; $d(o_i, o_j)$ 代表数据对象 o_i 与 o_j 间的欧氏空间距离。参数 α 代表了相差异的度量, 利用它来确定 2 个数据对象能否放置在一起; $\text{Neigh}_{s \times s}$ 是指以地点 r 为中心的 $s \times s$ 的区域。在 LF 算法中蚂蚁拾起或者放下一个对象 O_i 的概率分别按照式(2)、式(3)计算:

$$P_p = \left(\frac{k^+}{k^+ + f(O_i)} \right)^2 \quad (2)$$

$$P_d = \begin{cases} 2f(O_i), & \text{如果 } f(o_i) < k^- \\ 1, & \text{如果 } f(o_i) \geq k^- \end{cases} \quad (3)$$

式中: P_p 表示蚂蚁拾起对象 O_i 的概率; P_d 表示蚂蚁放下对象 O_i 的概率; k^+ 和 k^- 是 2 个常量。

2 基于信息素的蚁群聚类算法

在 LF 算法中, 基本模型被扩展到了数据分析范畴, 而且有很大的进展, 由于这种算法的参数设置繁琐, 并且敏感, 同时对于运动的区域进行了限制, 即二维网格的圈定, 对于某些网格的存在失去了意义, 例如没有数据网格。所以, 这种“物以类聚”统计分析情况并不尽如人意。改进算法思想: 首先将有需要的数据对象分布于二维网格, 然后将具有蚂蚁真实的群体行为进行聚类剖析。采用信息素和短期记忆的思想对控制蚂蚁的随意移动进行控制, 把信息熵作为蚂蚁运动状态的转换规则, 用信息熵的计算与比较来替代 LF 算法中的群体相似度计算和概率转换函数, 对拾起和放下数据对象的判断规则进行修改。实验结果证明了改进算法是有效的, 算法取得了较好的聚类结果。

2.1 蚂蚁空间转移策略

在一个二维的平面网格中, 蚂蚁的短期记忆^[8]和网格中信息素的分布决定了蚂蚁从一个位置到下一个位置的空间转移。相应的解决方法为: 让每一个蚂蚁都能够记住近几次所放下的数据对象和相应的位置, 如果新的数据对象被拾起, 分析新对象与某个数据对象之间的关系, 看是否可以归并, 如果不能, 可以按照具体的数据分布来确定下一步的移动方向。

假设蚂蚁在平面网格上的具体位置为 (γ, θ) , 在它走过的痕迹中会留下相应的信息素, 在步长时间段内, 所留下的信息为 η , 伴随时间的推移, 推移过后的信息会相应减少, 设减少的衰减率为 κ 。那么蚂蚁从 k 转移到位置 i 的空间方面的转移概率 p_{ik} 用式(4)~式(5)来确定^[9]:

$$p_{ik} = \frac{W(\sigma_i)\omega(A_i)}{\sum_{j/k} W(\sigma_j)\omega(A_j)} \quad (4)$$

$$W(\sigma) = (1 + \frac{\sigma}{1 + \delta\sigma})^\beta \quad (5)$$

$W(\sigma_i)$ 响应函数(response function)是由关于 i 的信息素(pheromones)浓度表示的, 证明膜翅目昆虫在关于信息素(pheromones)浓度为 σ_i 的位置 i 的相对出现率, 较高的 β 取值使膜翅目昆虫顺着信息素的有关方向位移, 其中 β 是决定膜翅目昆虫所有情况的抗噪比相关的参数数据, $1/\delta$ 证明膜翅目昆虫在具有信息素(pheromones)范围内所能捕捉到信息素改变的能力, j/k 表示在 k 邻域内所有网格单元 j 的和, 集合中的参数 β 与 $1/\delta$ 二者是形成了信息素(pheromones)痕迹的关键因素。 $\omega(A)$ 为到位置 i 的权值因素, A_i 代表在时间为步长时候的位置 i 与位置 k 在方向方面的区别量, 依据 A_i 的值来对 $\omega(A)$ 进行确定^[10], 膜翅目昆虫的移动方向能够事先看成有 8 个紧挨着的网状式单元组成, 向 8 个不同的方向进行移动, 产生 8 个方向, 各个方向之间的夹角为 45° , 那么膜翅目昆虫可事先从这 8 个方向当中, 任意选取 1 个方向, 则膜翅目昆虫迅速地改变方向的概率远远小于缓慢地改变方向的概率, 定义相异改变度数的数值与文献[11]相同, 分别为: $\omega(0^\circ)=1, \omega(\pm 45^\circ)=0.5, \omega(\pm 90^\circ)=0.25, \omega(\pm 135^\circ)=0.08, \omega(180^\circ)=0.05$ 。

2.2 群体相似度

群体相似度代表的是一个需要聚类对象和其所处位置 r 处的 $s \times s$ 区域内全部数据对象的相似度。本文对经典 LF 算法中的群体相似度函数进行改进, 定义为:

$$f(O_i) = \begin{cases} \frac{1}{s^2} \sum_{o_i \in Neigh_{s \times s}(r)} [1 - \frac{d(o_i, o_j)}{\alpha \mu (1 + (v-1)/v_{\max})}], & \text{如果 } f > 0 \\ 0, & \text{其他} \end{cases} \quad (6)$$

式中: μ 为数据集合中所有数据对象的欧氏空间距离的平均值, $d(o_i, o_j)/\mu$ 则表明了数据对象 o_i 与 o_j 相对于总体的空间距离的偏离程度; v 表示蚂蚁的速度, 即在每个时间单元内一只蚂蚁沿着给定的网格方向移动的网格数, 表示一只蚂蚁通过函数 $f(O_i)$ 或者拾起或者放下一个对象的趋势, 假设蚂蚁的速度任意地分布在 $[1, v_{\max}]$ 上。在判断一个对象与其邻域对象之间的群体相似度时, 慢速的蚂蚁比快速移动的蚂蚁具有很好的选择性, 同时速度较快的蚂蚁形成大规模粗粒度的聚类, 而速度慢的蚂蚁则能更准确地将放置对象放在小规模的小粒度的区域; α 是自适

应相异度参数, 取值范围为[0,1], 决定聚类的尺度。同时 α 的取值对产生的聚类结果影响明显, 根据数据对象的本身特点进行自适应的调整。假定给定 1 只蚂蚁的操作次数的限定值为 N_{action} , 由于 α 的取值导致蚂蚁在相对应的邻域内所放下的数据对象的失败次数记为 N_{fail} , 当失效的次数达到某个限定的值时, 表明 α 的取值不合适, 要进行调整, 自动调整 α 的取值公式为:

$$\alpha = \begin{cases} \alpha - 0.01, & \text{如果 } r_{\text{fail}} \leq 0.99 \\ \alpha + 0.01, & \text{如果 } r_{\text{fail}} > 0.99 \end{cases} \quad (7)$$

式中 r_{fail} 是失败率, $r_{\text{fail}} = \frac{N_{\text{fail}}}{N_{\text{action}}}$ 。

2.3 概率转换函数

概率转换函数是把群体相似度转换成蚂蚁个体运动状态的函数, ILFBP 算法的概率转换函数可定义为, 概率转换函数是 $f(O_i)$ 的一个函数, 它把数据对象的平均相似性转化为拾起概率或者放下概率。它的转换规则为: 如果数据对象与它邻域的平均相似性越小, 表明该数据对象属于此邻域的可能性就越小, 则拾起概率就越高, 放下的概率越低, 反之亦然。鉴于此, 选择对称式 *sigmoid* 函数作为概率转换函数, 该函数具有非线性, 运算中只需要调整一个参数, 与 LF 算法中的二次函数相比具有更快的收敛性。

一个未负载的随机运动的蚂蚁拾起一个对象的拾起概率定义为:

$$p_p = 1 - \text{sigmoid}(f(O_i)) \quad (8)$$

同样, 一个负载的随机运动的蚂蚁放下一个对象的放下概率定义为:

$$p_d = \text{sigmoid}(f(O_i)) \quad (9)$$

其中 *sigmoid* 函数的定义为:

$$\text{sigmoid}(x) = \frac{1 - e^{-cx}}{1 + e^{-cx}} \quad (10)$$

该函数为自然指数形式, 参数 c 的值越大, 曲线的饱和速度就越快, 同时算法的收敛速度也就越快。同时, 在进行聚类的过程中, 存在一些被称之为孤立点的对象与其他对象不相似的情况, 当蚂蚁拾起它们后, 很难快速地放下它们, 从而影响了算法的收敛速度。于是采用在算法后期增大 c 的值, 尽快地放下孤立点。

2.4 算法步骤描述

根据前面对基于信息素的改进 LF 算法关键特性的讨论, 并结合蚁群聚类分析基本模型算法流程图, 下面给出算法的描述为:

步骤 1: 参数设置, 包括网格宽度和高度、迭代次数 t_{max} 、蚂蚁个数 N_{ant} 、蚂蚁最大速度 v_{max} 、初始网格信息素大小 Initcap 、蚂蚁释放信息素 η 、信息素挥发因子 κ 等, 初始化网格、数据对象和蚁群。

步骤 2: 随意选择 1 只蚂蚁, 蚂蚁速度是区间 $[1, v_{\text{max}}]$ 的任意的值, 用来判定 α 是否需要自适应的调整, α 值的调整根据式(7)计算。

步骤 3: 经过计算蚂蚁当前位置周围的数据对象数目 n , 按照蚂蚁运动状态分配蚂蚁任务。

1) 若蚂蚁负载而且当前位置为空的话, 按照式(6)和式(9)计算群体相似度和放下概率, 判断是否能在当前位置放下数据对象。若不能就根据当前位置局部信息素的分布转移到新的位置。

2) 若蚂蚁空载而且当前位置具有数据对象, 根据式(6)和式(8)来计算群体的相似度和拾起概率, 如果拾起对象, 则判断能否归并到其短期记忆中的数据对象中所在的类, 如果能则蚂蚁转移至该位置时就放下数据对象。

步骤 4: 蚂蚁根据其所在位置局部信息素的分布按照式(4)和式(5)计算信息素响应因子和空间转移概率 p_{ik} , 并且转移至新的位置, 经过与在[0,1]之间的随机数比较来判断是否在该位置区域内增加信息素, 并且计算该位置周边数据对象的个数, 在这个局部区域内增加信息素。

步骤 5: 按照迭代进行的次数判断是否根据挥发因子更新网格上的所有信息素。

在聚类形成的过程中, 当一些比较小聚类形成后, 负载的蚂蚁通过多次寻找放下的位置, 在进行了大量计算群体相似度和放下概率之后都不能将负载的数据对象全部放下, 另外, 还有些异常值因为和其他的数据相似度差异特别大, 蚂蚁拾起后也很难放下, 使蚂蚁在很长迭代过程中都不能达到聚类的作用, 影响算法的收敛性。为了避免这种情况的出现, 采用让蚂蚁在放下数据的对象失败次数达到某个设定值时强行放下数据对象, 并且采用在后期增大蚂蚁速度的方式使其快速放下数据对象, 从而加快了聚类的进程。

3 实验分析

本文用 Java 语言实现一个简单的基于蚁群算法的聚类分析平台,开发环境为 Windows+JDK1.5+Eclipse。主要功能类包括表示虚拟的蚂蚁的类 Ant、代表蚂蚁短期记忆的类 Memory、表示蚁群的类 AntColony、代表二维平面网格的类 Grid、表示数据对象的类 DataObject 和数据集类 Data。

3.1 LF 算法和 ACCABP 算法的比较

本节使用蚁群聚类分析平台分析了 2 个数据集,初始值设置为 $t_{\max}=100$, $\alpha=0.2$, $\beta=0.8$, $\delta=0.4$, $\eta=0.2$, $\kappa=0.01$, 数据集的描述和运行结果如下:

1) 正态分布测试数据集。该数据集为 4 组正态分布数据,每组选用 200 个数据,共有 800 个数据,各组数据服从均值为 μ , 方差为 $\sigma=2$ 的正态分布。具体为:类 1 为 $\{x \sim N[0,2], y \sim N[0,2]\}$, 类 2 为 $\{x \sim N[0,2], y \sim N[8,2]\}$, 类 3 为 $\{x \sim N[8,2], y \sim N[0,2]\}$, 类 4 为 $\{x \sim N[8,2], y \sim N[8,2]\}$ 。数据集随机地分布在 60×60 的网格上,邻域大小为 3, 蚁群大小为 20, 蚂蚁速度限制在 1~10 之间。聚类结果如图 1、图 2 所示。

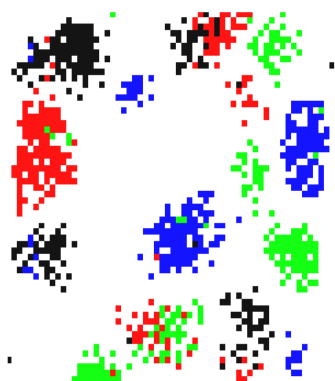


Fig.1 LF algorithm clustering results
图 1 LF 算法聚类结果

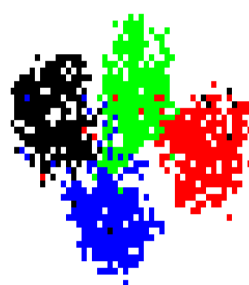


Fig.2 ACCABP algorithm clustering results
图 2 ACCABP 算法聚类结果

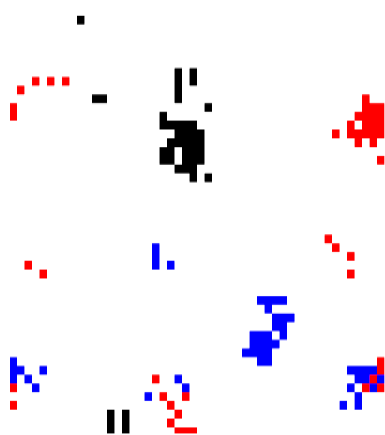


Fig.3 LF algorithm clustering results
图 3 LF 算法聚类结果

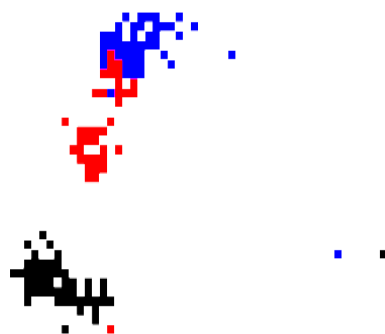


Fig.4 ACCABP algorithm clustering results
图 4 ACCABP 算法聚类结果

2) IRIS 植物样本数据集。该数据集分别属于 3 种植物的 150 个样本,每种植物样本数为 50 个,每个样本均为代表植物 4 种属性特征四维向量。网格大小为 50×50 , 邻域大小为 3, 蚁群大小为 10, 蚂蚁速度在 1~10 之间。在此是按前 2 个属性得到的聚类结果如图 3、图 4 所示。

通过对比 2 组数据集的聚类结果,可以看出 ACCABP 算法性能明显优于 LF 算法,同时在聚类分析中发现,如果开始的一段时间内蚂蚁随机转移,并释放了信息素,然后再采用信息素控制蚂蚁的随机移动,聚类的结果会显著提高,相异度的参数 α 的取值对 2 种算法的聚类结果影响明显,若 α 取值偏大或者偏小都能使产生的聚类质量比较差,通过自适应 α 值的调整可以在一定程度上改善聚类的质量。

3.2 K-means 算法和 ILFBP 算法的比较

1) IRIS 样本集实验(见表 1)

K-means 算法参数设定:聚类簇数 $k=3$, ε 分别取 $\varepsilon=0.1$ 及 $\varepsilon=0.01$ 。ILFBF 的实验参数设定:网格大小为 50×50 , 邻域大小为 3, 蚁群大小为 10, 蚂蚁速度在 1~10 之间, $t_{\max}=100, \alpha=0.2, \beta=0.8$ 。 $\delta=0.4, \eta=0.2, \kappa=0.01$ 。

对于 K-Means 算法的初始数据来说, 初始数据的选择是至关重要的, 如果初始值选择的是不同组的数据, F-measure 的值就比较大, 聚类效果也比较好;但是如果初始值选择的是相同组的数据的话, 会使得到的 F-measure 值很小, 聚类效果比较差。K-Means 算法本身具有易陷入局部最优解的缺点, 所以在结束条件 ε 下, 很难跳出局部最优解, 算法在没有得到比较理想的结果之前就结束了。ILFBP 算法的聚类情况优于 K-Means 算法。在 F-measure 的值和聚类效果方面, ILFBP 算法都强于 K-Means 算法和 LF 算法, 因为所选择的聚类试验数据和 IRIS 数据本身就很接近, 区分度不是很明显, 根据 IRIS 数据本身的聚类情况, 采用这 3 种算法进行聚类时所得的聚类结果都不理想。

2) Zoo 样本集实验

聚类簇数 $k=7$, 维数为 15, 其他实验参数设置如上, 实验结果见表 2。

本组实验中, 使用了维数更高的数据集 Zoo。测试 3 个算法聚类质量。LF 算法的平均 F-measure 值仅为 0.798, 聚类质量出现了较大幅度下滑。ILFBP 算法平均 F-measure 值为 0.856, 仍能维持一个比较满意的聚类结果。ILFBP 算法的聚类效果均优于 LF 算法和 K-Means 算法。由此可知 ILFBP 算法更适合处理高维数据。

表 1 三种算法对 IRIS 样本数据的聚类结果

Table1 IRIS sample data clustering results for three kinds of algorithm

algorithm	number of clusters	average F-measure
K-Means	3	0.659
ACCABP	3	0.932
LF	3	0.913

表 2 三种算法对 Zoo 样本数据的聚类结果

Table2 Zoo sample data clustering results for three kinds of algorithm

algorithm	number of clusters	average F-measure
K-Means	7	0.605
ACCABP	7	0.856
LF	7	0.798

3) 人工数据集

ILFBP 算法应用到任意生成的人工数据集: 在具有 2 个属性、5 个聚类的 500 个样本集 Ant(500,2,5)及具有 2 个属性、9 个聚类的 1 000 个数据 Ant(1 000,2,9)中进行实验。

针对 2 个人工数据样本集, K-Means 算法参数设定聚类的簇数分别为 $k=5$ 和 $k=9, \varepsilon=0.01$ 。ILFBF 的实验参数: 网格大小为 50×50 , 邻域大小为 3, 蚁群大小为 10, 蚂蚁速度在 1~10 之间, $t_{\max}=100, \alpha=0.2, \beta=0.8, \delta=0.4, \eta=0.2, \kappa=0.01$ 。

表 3 三种算法对人工数据集的聚类结果

Table3 Three algorithms for clustering results of artificial data set

parameter	Ant(500,2,5)			Ant(1 000,2,9)		
	K-Means	ACCABP	LF	LF	K-Means	ACCABP
number of clusters	5	5	5	9	9	9
minimum number of passes	16	26	28	31	20	29
the average number of errors	13.16	4.28	6.42	8.56	16.35	5.63
average accuracy	96.35%	98.56%	97.56%	97.36%	96.24%	98.16%

表 3 为 K-Means 算法、ILFBP 算法和 LF 算法对人工数据样本集 Ant(500,2,5)及 Ant(1 000,2,9)的聚类结果。由表 3 可以看出 ACCABP 算法对 LF 算法和 K-Means 来说, 错误分类个数要少, 正确率比较高。表 3 的结果显示这 3 种算法用于区分度比较清晰的人工数据集 Ant(500,2,5)和 Ant(1 000,2,9)比用于 IRIS 数据集聚类的数据集准确率高, 这是因为 IRIS 数据相对于本文采用的人工数据样本集的区分度不是很高, 因此影响了算法的准确率, 由此可得出 ACCABP 算法的可行性。

从表 3 可以看出, ACCABP 算法和 LF 算法与 K-Means 算法相比, 尽管 K-Means 算法在迭代次数很小的情况下, 可以得到较好的解, 但是在错误率上 ACCABP 算法比它要低得多, 而且准确率要高。LF 算法相对于 ILFBP 算法来说, 准确率比较低, 主要是由于定义了蚂蚁在二维网格中是任意移动的, 而且任意移动过程中对某些区域并没有什么意义, 影响了聚类效果。

4 结论

本文在蚁群聚类模型中,引入了信息素,提出了基于信息素的蚁群聚类分析算法,并设计和实现了一个简单的蚁群聚类分析平台,通过测试数据集对改进的算法进行分析,算法显示出了较高的稳定性和准确率,证明了该改进算法的有效性。

参考文献:

- [1] DORIGO M, BONABEAU E, THERAULAZ G. Ant algorithms and stigmergy[J]. Future Generation Computer Systems, 2000, 16(8): 851-871.
- [2] 宋佩莉, 祁飞, 张鹏. 混合蚁群蜂群算法在旅行Agent问题中的应用[J]. 计算机工程与应用, 2012, 48(36): 33-38. (SONG Peili, QI Fei, ZHANG Peng. Application of hybrid ant colony and bee colony algorithm in traveling Agent problem[J]. Computer Engineering and Applications, 2012, 48(36): 34-38.)
- [3] 侯超, 侯永. 一种改进蚁群聚类算法在数据挖掘中的研究[J]. 微计算机信息, 2011, 27(12): 136-138. (HOU Chao, HOU Yong. An improved ant colony clustering algorithm in data mining[J]. Microcomputer Information, 2011, 27(12): 136-138.)
- [4] 宋中山, 周腾, 周晶平. 一种改进的蚁群聚类算法在客户细分中的应用[J]. 中南民族大学学报, 2013, 32(3): 77-81. (SONG Zhongshan, ZHOU Teng, ZHOU Jingpin. An improved ant colony clustering algorithm in customer segmentation[J]. Journal of South-Central University for Nationalities, 2013, 32(3): 77-81.)
- [5] LAMER E, FAJITA B. Diversity and adaptation in populations of clustering ants[C]// From Animals to Animates 3. Proc of the Third International Conference on Simulation of Adaptive Behavior. Cambridge, MA, USA: MIT Press, 1994: 501-508.
- [6] 李丹. 蚁群优化算法在差异工件批调度问题的应用研究[D]. 安徽: 安徽大学, 2014. (LI Dan. Research on batch scheduling problem with non-identical job sizes using ant colony optimization algorithm[D]. Anhui China: Anhui University, 2014.)
- [7] 许智超. 基于蚂蚁系统的基因表达数据聚类分析[D]. 福州: 福建农林大学, 2012. (XU Zhichao. Clustering analysis based on the ant system for gene expression data[D]. Fuzhou China: Fujian Agriculture and Forestry University, 2012.)
- [8] DORIGO M, GAMBARDELLA L M. Ant colony system: a cooperative learning approach to the traveling salesman problem[J]. IEEE Transactions on Evolutionary Computation, 1997, 1(1): 53-66
- [9] 朱峰, 陈莉. 一种改进的蚁群聚类算法[J]. 计算机工程与应用, 2010, 46(6): 133-135. (ZHU Feng, CHEN Li. Improved ant based clustering algorithm[J]. Computer Engineering and Applications, 2010, 46(6): 133-135.)
- [10] CHIALVO D R, MILLONAS M M. How swarms build cognitive maps[C]// In: Luc Steels ed. The Biology and Technology of Intelligent Autonomous Agents. Nato ASI Series: [s.n.], 1995: 439-450.
- [11] RAMOS V, MUGE F, PINA P. Self-organized data and image retrieval as a consequence of inter-dynamic synergistic relationships in artificial ant colonies[C]// In: J Ruiz-del-Solar, A Abraham, M Koppen Eds. Hybrid Intelligent Systems, Frontiers of Artificial Intelligence and Applications. Santiago, Chile: [s.n.], 2002: 253-262.

作者简介:



王 慧(1982-), 女, 河南省平顶山市人, 硕士, 讲师, 主要研究方向为数据挖掘. email: 727383938@qq.com.

甘 泉(1980-), 男, 安徽省灵璧县人, 硕士, 讲师, 主要研究方向为操作系统、算法分析等.