

改进的种群分类蚁群算法及其应用^①

刘芳 李义杰 (辽宁工程技术大学 计算机软件与理论 辽宁 葫芦岛 125105)

摘要: 提出了一种改进的种群分类蚁群算法,该算法在种群分类的基础上,引入了蚂蚁的知觉感觉特性等。该算法能明显的防止蚁群算法可能出现早熟的问题,从而解决了传统蚁群算法加速收敛与早熟、停滞现象的矛盾。为了说明该算法的性能,将该算法应用到聚类分析算法中,设计了算法的模型以及算法步骤,并通过仿真实验证明了本算法的可行性和有效性。

关键词: 蚁群算法 聚类分析 种群分类蚁群算法 蚂蚁的感觉知觉性 K均值算法

Improved Character-Base Ant Colony Algorithm and Its Application

LIU Fang, LI Yi-Jie

(Liaoning Technical University, Huludao 125105, China)

Abstract: An improved Character-base Ant Colony Algorithm based on Character-base Ant Colony Algorithm and sentience consciousness characters is presented in this paper. It can significantly prevent precocity, then make a balance between accelerating convergence and averting precocity as well as stagnation. The algorithm is applied to clustering analysis and the algorithm steps of the model have been designed to illustrate the performance of the algorithm in this paper. The results of simulation on clustering analysis indicate that the new algorithm is feasible and effective.

Keywords: ant colony clustering; Population Classification ant colony algorithm; Ant's sense and perceptual; K-means algorithm

1 引言

蚁群算法是由 M.Dorigo 于 1992 年提出的一种新型进化算法,其原理是从生物进化的机理中受到启发,模拟自然界中真实蚁群的觅食行为而形成的一种模拟进化算法^[1]。该方法被广泛的应用在网络路由问题,下料问题、图像分割、边缘提取、图像分类、分类规则发现、模式识别、旅行商问题(TSP)、指派问题、Job-shop 调度问题等。该算法运用了正反馈的原理,而且是一种本质上并行的算法,因此具有很强的发现较好解的能力。然而任何一种算法都不可能是最完美的,蚁群算法也存在一些缺点,如需要较长的搜索时间,当问题规模较大时也易陷入局部最优解,即产生过早收敛等问题。本文改进了一种基于种群分

类遗传算法,并将其应用到聚类分析以展现其优良的效果。

聚类分析的基本思想是根据“物以类聚”的原理,对样本进行分类。聚类分析方法是一种无监督的学习过程,它按照事物的某些属性将其聚集成类,使不同类之间的相似性尽量小,相同类之间的相似性尽量大^[2],从而实现对数据的分类。聚类分析即可以作为一个单独的算法使用,也可以作为其他数据挖掘算法的一个预处理步骤,因此,其是数据挖掘领域的一个重要研究课题。目前用得比较多的聚类算法有 K-均值算法、模糊 C 均值聚类算法等。然而这类算法本质上是一种局部搜索算法,采用了一种迭代的爬山技术来寻找最优解。因此,存在两

^① 收稿时间:2009-04-15

个致命的问题：一是对初始化敏感；二是容易陷入局部极小值^[3]。将本文改进的蚁群算法应用到聚类分析中，通过实验证明了本文改进的蚁群算法的优良性能。

2 蚁群算法

蚂蚁算法来源于对自然界蚂蚁寻找从蚁巢到食物的最短路径并找到回巢路径方法的研究，是对自然界蚂蚁的寻径方式进行模拟而得出的一种仿生进化算法。蚂蚁算法中采用有记忆的人工蚂蚁，通过个体之间的信息交流与相互协作来找到从蚁穴到食物源的最短路径。研究发现，蚂蚁个体之间是通过一种称为信息素(pheromone)的物质进行信息传递，从而相互协作，完成复杂任务。蚂蚁在运动过程中能够在所经过的路径上留下该种物质，而且蚂蚁在运动过程中能够感知这种物质的存在及其强度，并以此指导自己的运动方向。因此，有大量蚂蚁组成的蚁群，其集体便表现出一种信息的正反馈现象：某一路径上走过的蚂蚁越多，则后来者选择该路径的概率就越大^[4]。蚂蚁个体之间就是通过这种信息的交流达到搜索食物的目的。

试验结果表明，蚁群算法具有很强的发现较好解的能力，而且不容易陷入局部最优。这是因为，算法不仅利用了正反馈原理，在一定程度上加快了进化过程；而且算法是一种本质并行的算法，个体之间不断地进行着信息交流和传递，这就有利于较好解的发现。虽然其单个个体容易收敛于局部最优，但多个个体通过合作，会很快收敛于解空间的某一子集。针对基本蚁群算法存在的缺点有很多改进的方法，其中文献^[5]提出了一种种群分类的蚁群算法。

2.1 种群分类蚁群算法的基本思想

根据信息更新策略和搜索前进策略的不同将蚁群划分为两种类型即智能蚁群和随机蚁群。智能蚁群与基本蚁群算法中的功能基本相同，具有信息感知能力，其搜索前进策略是在尚未走过的路径上选择信息量大的路径前进，以保证算法收敛于最优解，算法最终的结果将是所有智能蚂蚁收敛的路径长度；随机蚁群不具有信息感知能力，其将随机选择一个本次遍历中尚未走过的路径前进，以保证算法搜索更大的解空间。

在信息更新方面，智能蚁群将在其走完全部路径

后对所走路径进行一次全局信息更新，用信息量除以此次迭代中蚂蚁走的总路径长度，以便反映每段路径在此次总路径中的优劣。随机蚁群将对路径信息进行两次更新，走完一段路径即更新一次，更新的信息量与本段路径的优劣有关，以便反映每段路径的优劣，为智能蚁群搜索前进及时提供依据；在走完全部路径后再进行一次信息更新，更新将反映本次所走全局路径的优劣，指导智能蚁群后来的迭代过程。

为了使算法收敛于所能获取的最优解，智能蚁群和随机蚁群的数量将随蚂蚁的迭代次数动态变化。智能蚁群数目将逐渐增大，直至最终所有的蚂蚁都是智能蚂蚁，而随机蚁群数量将逐渐减少直至消失，最终算法将变为基本的蚁群算法，从而能保证算法最终收敛最优解。

2.2 具有感觉知觉特征蚁群算法的思想

心理学研究表明不是所有的刺激都能引起人或者动物的感觉，只有达到一定量的刺激才能引起相应的感觉，这种感受性的差异是用感觉阈限的大小来度量的。感觉阈限是指能引起感觉的、持续了一定时间的刺激量。心理学上把刚刚能引起感觉的最小刺激量，称为绝对感觉阈限(Absolute Sensor Threshold)，记作AST。同时只有当刺激变化到一定量时才能使我们感觉到差别，这种能引起差别感觉的刺激物的最小变化量，称为差别感觉阈限(Contrast Sensor Threshold)，记作CST。根据韦伯定律设I表示原来刺激物的强度， ΔI 表示差别感觉阈限，那么在I小于某个特定的限度IT(intensity-threshold)时，就有 $\frac{\Delta I}{I} = K$ ，其中K为常数，称为韦伯系数。重量感觉的韦伯系数为1/30，听觉为1/10，视觉为1/100。

为了避免蚂蚁从搜索的一开始就失去解的多样性，在路径上信息量的刺激量未达到蚂蚁的绝对感觉阈限(AST)时，让蚂蚁忽视该刺激物的存在，即让蚂蚁在搜索初始阶段的选路不受信息量的影响。只有当信息量积累到超过AST时，蚂蚁才在信息量的刺激下趋于选择信息量较大的路径。这种方法有利于让蚂蚁向最优方向进行搜索并很快收敛，大大节约了计算时间^[6]。

同时蚁群在一定刺激强度范围内也存在一个差别感觉阈限。当路径上信息量的增加或减少的量在差别感觉阈限CST之下时，蚂蚁遵循知觉的恒常性规律，

感受不到该路径上信息量的变化,该条路径被选择的概率主要依据于以前迭代中蚂蚁选路经验形成的潜意识作用;反之,蚂蚁受其显意识控制,按照路径上所有蚂蚁信息量高低决定选路概率的大小。

本文在群分类的蚁群算法的基础上,引入了蚂蚁的感觉知觉系统,提出了一种具有感觉知觉特征的种群分类算法。该算法可以有效的避免蚁群在算法刚开始的时候陷入局部最优,从而扩大搜索空间找到最优解。

3 具有感觉知觉特征的种群分类蚁群算法

本文改进了种群分类的蚁群算法,提出了一种具有感觉知觉特征的种群分类蚁群算法。其具体思想是:在智能蚂蚁的前进搜索方法中引入感觉知觉特征,来防止算法在刚开始迭代时受信息素影响过大而产生的早熟现象。在种群分类的蚁群算法中,算法的最终结果依赖于智能蚂蚁,如果智能蚂蚁过于依赖路径上的信息量就会陷入局部最优。引入感觉知觉特征可以保证智能蚂蚁搜索有自己的“选择意识”,让智能蚂蚁在搜索初始阶段的选路不受信息量的影响,只有当信息量积累到超过 AST 时,蚂蚁才在信息量的刺激下趋于选择信息量较大的路径。信息量大于 AST 而信息的变化量小于 CST 时,让智能蚂蚁忽略信息的变化量,按自己的潜意识选择。信息量大于 AST 而信息的变化量大于 CST 时,智能蚂蚁将会选择此路径。对于信息更新策略,智能蚂蚁采用一次更新方法,随机蚂蚁采用两次更新法,算法具体过程如下。

3.1 算法的搜索前进策略

① 随机蚁群随机选择一个本次遍历中尚未走过的路径前进,其选择路径的方法如下:

$$p_{ij}^k(t) = \begin{cases} rand() & j \in allowed_k \\ 0 & \text{否则} \end{cases} \quad (1)$$

其中 $p_{ij}^k(t)$ 为 t 时刻蚂蚁 k 从 i 点选择 j 点的概率, $allowed_k$ 表示蚂蚁 k 下一步允许走过的点的集合(尚未遍历的点), $rand()$ 为一个随机函数。

② 智能蚂蚁的搜索前进策略,引入了感觉和知觉特征,其具体实现方法如下:

根据韦伯定律可知:

$$CST = \tau_{ij}(t) \times K$$

$\tau_{ij}(t)$ 表示在 t 时刻 ij 间路径上的信息量, K 为韦

伯系数。可以取 $1/30, 1/40, 1/50$ 。

由于韦伯定律只能在刺激物的强度 I 小于某个特定的限度 IT 时才能适用,在超过 IT 时并不适用,故当路径上信息量达到相当大的程度时,我们就不能用此方法进行搜索。首先估计各个路径上可能的最大信息量 τ_{max} :

$$\tau_{max} = C \times m \times \text{适应度的数量级}$$

其中 C 为预订的蚂蚁搜索的最大遍历次数, m 为蚂蚁的总个数。由于“适应度的数量级”和 AST 相当,因此公式也可以表达为:

$$\tau_{max} = C \times m \times AST$$

$$IT = h \times \tau_{max}$$

h 为一常数,可取 $1/2, 1/3$ 等。

AST 的取值方法:随机产生 m 个初始解,计算这 m 个初始解的适应度,即 m 只蚂蚁遍历完成的路径长度的倒数,记其最大适应度为 f_{max} , $AST = C \times f_{max}$, C 为常数,可以去 $3, 4, 5, 6$ 等。

改进后的智能蚂蚁前进算法为:

$$p_{ij}^k(t) = \begin{cases} rand() & \tau_{ij}(t) < AST \\ 1 & AST \leq \tau_{ij}(t) < IT \text{ 且 } \Delta\tau_{ij}(t) > CST \\ \tau_{ij}^k(t) / \tau_{ij}(t) & AST \leq \tau_{ij}(t) < IT \text{ 且 } \Delta\tau_{ij}(t) \leq CST \end{cases} \quad (2)$$

3.2 算法的信息更新策略

对于智能蚂蚁的信息更新方式,采用一次更新的方法,即当每个智能蚂蚁完成一次全局遍历后,对路径上的信息进行一次更新,全局更新策略反映本次遍历所得全程路径的优劣。具体更新方法如下:

$$\begin{aligned} \tau_{ij}(t) &= (1 - \rho)\tau_{ij}(t-1) + \Delta\tau_{ij} \\ \Delta\tau_{ij} &= \sum_{k=1}^{N_1} \Delta\tau_{ij}^k \\ \Delta\tau_{ij}^k &= \begin{cases} \frac{Q_1}{L_k} \times d_{ij} & \text{智能蚂蚁 } k \text{ 路过 } ij \\ 0 & \text{其他} \end{cases} \end{aligned} \quad (3)$$

其中 ρ 为信息的挥发度, $\rho \in (0,1)$ 表示单步时间内路径上信息量的损耗程度。 $\Delta\tau_{ij}$ 表示路径 ij 上的信息总量,它等于所有蚂蚁在路径 ij 上的信息量的总和。 N_1 为所有智能蚂蚁的数目; Q_1 为常数,表示每次全局刷新时智能蚂蚁所用的信息量; L_k 表示第 k 只智能蚂蚁本次遍历所走的路径总长度; d_{ij} 表示 ij 之间的路径长度。

随机蚂蚁的信息更新策略包括局部更新和全部更

新两部分,全局更新策略同上所述的智能蚂蚁信息更新策略。可直接引用上算法,只需将式中智能蚂蚁的量 N_1 改为随机蚂蚁的数量 N_2 。用全局更新策略来反映本次搜索中个路段的优劣程度。局部更新信息策略主要为了及时反映本段路程的优劣,以便为后来的智能蚂蚁能够更大的机会选择信息量大的路径。其局部更新策略如下:

$$\begin{aligned} \tau_{ij}(t) &= (1-\rho)\tau_{ij}(t-1) + \Delta\tau_{ij}^k \\ \Delta\tau_{ij}^k &= \begin{cases} Q_2/d_{ij} & \text{随机蚂蚁}k\text{经过}ij \\ 0 & \text{其他} \end{cases} \end{aligned} \quad (4)$$

其中,常数 Q_2 表示每次刷新时随机蚂蚁用的信息量。 ρ 同智能蚂蚁算法,表示信息的挥发度。

4 聚类分析算法思想

已知模式样本集 $\{X\}$ 有 n 个样本和 K 个模式分类 $\{S_j, j=1,2,\dots,K\}$, 每个样本有 d 个特征指标,因此可以得到一个 $n \times d$ 的矩阵 X 。为了对它们进行分类,矩阵 X 中,每一行为一个样本 $X_{i1}, X_{i2}, \dots, X_{id}$ 为第 i 个样本的 d 个特征指标,以每个模式样本到各自聚类中心的距离之和达到最小为标准,其目标函数为:

$$\begin{aligned} T &= \min \sum_{j=1}^k \sum_{X \in S_j} \|X - m_j\| \\ m_j &= \frac{1}{n} \sum_{i=1}^n y_{ij} X_i \end{aligned}$$

其中 K 为聚类数目, m_j 表示第 j 类样本 (S_j) 的均值向量。若模式样本 i 分配到第 j 聚类中心,则 y_{ij}

$=1$, 否则, $y_{ij}=0$ 。 $\sum_{i=1}^n y_{ij}=1$ 表示模式样本 i 只能分配给一个聚类中心。

聚类分析算法中最常用的是 K 均值算法,其一种基本的划分方法^[7],因其理论可靠、算法简单、收敛速度快、能有效处理大数据集而被广泛使用,其基本算法过程如下:

- ① 任意选择 K 个记录作为初始的聚类中心。
- ② 计算每个记录与 K 个聚类中心的距离,并将距离最近的聚类作为该点所属的类。
- ③ 计算每个聚类的质心(聚集点的均值)以及每个对象与这些中心对象的距离,并根据最小距离重新对

相应的对象进行划分。重复该步骤,直到式(1)不再明显地发生变化。

5 改进的蚁群算法在聚类分析中的应用模型

设待聚类的样本数为 N , 聚类的个数为 K , 聚类中心为 $Z_j(j=1,2,\dots,k)$, 蚂蚁总数为 M , 每个样本点是一个 D 维向量。目标函数为:

$$T = \min \sum_{j=1}^k \sum_{X \in S_j} \|X - m_j\| \quad (5)$$

将聚类的过程看成 M 个蚂蚁搜索食物的过程,待聚类样本看作蚂蚁要寻找的食物源。将 M 只蚂蚁均匀的分配到 K 个聚类中心上,若蚂蚁 k 将样本 i 分配到聚类中心 j 时,会在路径 ij 上留下一定量的信息素。其中智能蚂蚁的信息素浓度按公式(3)计算,随机蚂蚁的信息素浓度按公式(4)计算;随机蚂蚁选择路径的方法按公式(1)计算,智能蚂蚁的选择路径方法按公式(2)计算。

具体的算法过程如下:

① 给定一个精度 $\varepsilon > 0$ 以及总的搜索次数 L 。初始状态,让 M 只蚂蚁中的每一只都将 N 个样本随机分配到各自的 K 个模式分类。设定蚂蚁将样本分配到聚类中心的初始信息素浓度 $p_{ij}^k(0)=0, p_{ij}^k(t)=0$ 表示表示 t 时刻蚂蚁 k 没有将样本 i 分配到类 j 中, $p_{ij}^k(t)=1$ 表示 t 时刻蚂蚁 k 将样本 i 分配到类 j 中。

② 用上述改进的蚁群算法并行的进行聚类,即所有的蚂蚁同时的进行搜索。按照公式(2)(3)选择路径,按公式(4)(5)更新个路径上的信息素浓度。

③ 当所有蚂蚁都完成一遍聚类时,一次循环结束,将 h 加 1。设定蚂蚁的初始聚类中心为 $z^k(0 < j \leq K, 0 < k \leq M)$, 引用 K 均值算法计算每个蚂蚁的聚类中心。

④ 按公式(6)计算目标函数,当 $T < \varepsilon$ 时算法结束,输出算法分配的每个样本 i 及其所属模式分类。若 $T \geq \varepsilon$, 则判断搜索次数 h , 如果 $h = L+1$, 则算法结束;若已搜索次数 $h < L+1$, 转向②继续寻找最优解。

6 实验

为了验证本算法的有效性,本文分别两组数据集在 Matlab 下进行验证。

第一组为森林火灾数据,该数据共有 517 个,每个数据有六个属性,设聚类个数为三。用 matlab 中的 `kmeans.m` 函数来实现 K 均值算法,同时编写本文的算法。实验的参数设置:种群大小 $M=150$,最大循环次数 $L=500$,分类数目 $K=3$,将 150 只蚂蚁均匀的分到三个聚类中心上,且每个聚类中心上的智能蚂蚁和随机蚂蚁的数目相同。

第二组数据我们采用的是一组 Abalone 数据集,该数据来自 <http://archive.ics.uci.edu/ml/datasets.html>,共有 4177 个,每个数据包含八个属性,聚类个数为五。实验参数设置:种群大小 $M=400$,最大循环次数 $L=1000$,分类数目 $K=5$ 。

让两个实验都重复的做 500 次,其聚类结果对比如表 1 所示:

表 1 两种算法的性能比较

聚类算法	数据集	正确聚类次数	错误聚类次数	正确聚类率
K 均值	森林火灾	290	210	58%
	Abalone	480	20	96%
本文算法	森林火灾	255	245	51%
	Abalone	460	40	92%

从上面的两个实验可以看出,本文改进的蚁群算法应用在聚类分析算法中,无论是在少量数据还是海量数据上性能都优于原始的 K 均值算法,从而证明了本算法能有效的改善聚类算法局部最优的缺陷。

7 结语

本文针对蚁群算法可能出现的停滞以及早熟现象,提出了一种“具有感觉知觉特征的种群分类蚁群算法”。

觉与知觉”的思想,有效的避免的原始蚁群算法的缺点,该算法在原始的蚁群算法中引入了“种群分类”和“感让蚂蚁以最大的可能性搜索所有的解空间,从而阻止蚁群算法的停滞和早熟现象。为了说明本算法的优良性能,本文将改进的蚁群算法应用到聚类分析算法中。结合聚类分析的 K 均值算法,编写用本文思想改进的算法。通过实验验证,本算法在性能方面明显的高于原始的 K 均值算法。从而证明了本文提出的“具有感觉知觉特征的种群分类蚁群算法”的有效性。

参考文献

- 1 Dorigo M, Maniezzo V, Colomi A. Ant system: Optimization by a colony of cooperative Agents. IEEE Transactions on Systems, Man, and Cybernetics, 1996,26.
- 2 毛国君,段立娟,王实,等.数据挖掘原理与算法.北京:清华大学出版社,2006.
- 3 王敞,陈增强,袁著祉.基于遗传算法的 K 均值聚类分析.计算机科学,2003,30(2):163-164.
- 4 吴启迪,汪镭.智能蚁群算法及应用.上海:上海科技教育出版社,2004.
- 5 英恒松.蚁群算法及其应用研究[硕士学位论文].无锡:江南大学,2008.
- 6 陈峻,秦玲,陈宏建.具有感觉和知觉特征的蚁群算法.系统仿真学报,2003,11:1418-1425.
- 7 张琦,吴斌,王柏.非平衡数据训练方法概述.计算机科学,2005,32(10):181-186.