遗传增强蚁群优化算法

梁豪默,王智学,刘 艺
(解放军理工大学 指挥信息系统学院,江苏 南京 210007)

摘 要: 针对蚁群优化算法易陷入局部最优的问题,提出遗传增强蚁群优化算法.在算法迭代过程中,维护历史最优解,将历史最优解与当前部分较好解作为遗传算法的初始化染色体,采用遗传算法在更广阔的空间中搜索较好解.将蚂蚁搜索解与遗传算法搜索解合并选择当前最优解,防止蚁群优化陷入局部最优.以二分类问题中的特征选择为例,与粒子群优化算法与差分进化算法进行对比,在3个标准测试数据上进行试验,结果表明算法的有效性及其优越性.

关键词: 蚁群优化算法;遗传算法;局部最优;粒子群优化;差分进化算法
中图分类号: TP301.6 **文献标识码:** A **文章编号:** 1000-7180(2018)08-0101-04

Enhanced Ant Colony Optimization Combined with Genetic Algorithm

LIANG Hao-mo, WANG Zhi-xue, LIU Yi
(College of Command Information Systems, PLA University of Science and Technology, Nanjing 210007, China)

Abstract: In order to solve the problem that ant colony optimization is easily trapped into local optimum, a new algorithm called enhanced ant colony optimization based on genetic algorithm was proposed. It maintained a history best optimal solution which was combined with part of current best solutions as the initial chromosomes during iterations, and adopted genetic algorithm to search better solutions at a larger space. It selected the current best solution from results of ants and chromosomes to prevent ant colony optimization from being trapped into local optimum. It was tested on feature selection problem of binary classification and compared with particle swarm optimization and differential evolution, the experiments were implemented on three benchmark datasets, and the results show the efficiency and superiority of the proposed algorithm.

Key words: ant colony optimization; genetic algorithm; local optimum; particle swarm optimization; differential evolution algorithm

1 引言

蚁群优化是一类优秀的进化算法.

文献[1]为了更为有效地解决约束满足问题,采用了一种新的信息素增强策略,利用进化过程中的动态信息对从未选择的边进行信息素增强,提高了算法的收敛性.文献[2]将问题转化为最小背包问题,并提出采用蚁群优化求解该问题,并在真实数据集上验证了方法的可行性与优越性.为了避免生成不可行解,文献[3]提出采用经过修正的蚁群系统解决该问题,该方法能够确保生成所有的高效用项集,并且避免对相同可行解的重复评估,降低时间开销.本文提出遗传增强蚁群优化算法(Enhanced Ant Colony Optimization,EACO-GA).实验结果表明了EACO-GA算法的有效性与优越性[4].

2 ECO-GA

2.1 蚁群优化算法

蚁群优化的伪代码算法1如下所示.

算法1 蚁群优化伪代码

收稿日期: 2017-10-06; **修回日期:** 2017-11-16
基金项目: 国家自然科学基金(61273210)

BEGIN
初始化信息素值,启发式信息
WHILE(停止条件未满足)
FOR(每只蚂蚁)
蚂蚁搜索解
END FOR
更新当前最好解
蚂蚁更新信息素值
END WHILE
END

算法 1 中,蚂蚁根据路径条件转移概率搜索解,蚂蚁 k 访问节点 i 之后访问节点 j 的路径转移概率计算如式(1) 所示.

$$p_{ij}^{k}=\frac{(\tau_{ij}^{\alpha})(\eta_{j}^{\beta})}{\sum_{b\in allowed_{i}}(\tau_{ib}^{\alpha})(\eta_{b}^{\beta})} \tag{1}$$

式中,allowed_i 表示在选择节点 i 之后可选择的节点候选集;τ_j 表示节点 j 的信息素值;η_j 表示节点 j 的启发式信息,其值的计算方式由具体问题决定;α 和 β 表示信息素值与启发式信息值的重要程度.

蚁群优化根据当前解在优化目标上的评估值更新当前最好解,并用评估值更新信息素值,在 t 时刻对最优路径上的信息素更新方式如式(2) 所示.

$$\tau_i(t)=\begin{cases}\rho\tau_i(t-1)+\Delta/Q & \text{if } i\in route\\ \rho\tau_i(t-1) & \text{else}\end{cases} \tag{2}$$

式中,route 表示当前最好路径解;ρ 是信息素挥发系数;Δ 是当前最好解在目标上的评估值;Q 是常数,控制信息素增量.

2.2 遗传算法

遗传算法是一种模仿生物遗传进化的演化算法,它将问题的解编码成染色体,通过交叉、变异生成较好解,并从中选择精英解作为下一代父本. 遗传算法具有很强的全局搜索能力,EACO-GA 采用遗传算法在更广阔的空间中搜索较好解,提高蚁群优化的全局探索能力,防止陷入局部最优. 典型的遗传算法流程如算法 2 所示[5].

算法 2 遗传算法伪代码

BEGIN
初始化染色体
WHILE(停止条件未满足)
基于精英解选择父本
父本间执行交叉操作生成新解
新解执行变异操作
将新解与父本合并选择精英解
END WHILE
END

遗传算法的交叉操作有多种形式,如单点交叉、两点交叉、多点交叉等,本文采用单点交叉方式.

2.3 算法描述及时间复杂度分析

本文提出的 EACO-GA 算法是一种结合遗传算法与蚁群优化算法的混合增强优化算法,在蚁群优化算法过程中,保存迄今为止评估值最好的历史解,当前迭代蚂蚁寻优完成后,将历史最好解与本次迭代蚂蚁生成个体解的前 $n\%$ 作为遗传算法的初始化染色体,充分利用历史最优信息和当前最优信息. 遗传算法迭代完成后,将生成的解与本次迭代蚂蚁生成的个体解合并,评估选择当前蚁群优化的最好解. 通过将蚁群优化的历史最好解与当前部分较好解作为遗传算法初始染色体并进行优化,可以利用遗传算法的优点提高发现较好解的概率,防止蚁群优化算法陷入局部最优,提高蚁群优化算法的全局探索能力.

EACO-GA 算法的伪代码算法 3 如下所示:

算法 3 EACO-GA 伪代码

BEGIN
初始化信息素值,启发式信息
WHILE(停止条件未满足)
FOR(每只蚂蚁)
蚂蚁搜索解
END FOR
当前历史最好解和蚂蚁个体解前 $n\%$ 作为初始化染色体
遗传算法搜索解
将遗传算法解与蚂蚁个体解合并
更新当前最好解和历史最好解
蚁群算法信息素值更新
END WHILE
END

此外,如何将蚁群优化的解编码成染色体,以及如何将遗传算法的染色体解表示为蚂蚁的路径解是需要解决的关键问题. 我们提出 2 种操作算子完成染色体到路径解的双向转换. 操作算子 1 完成路径解到染色体的转换,如算法 4 所示.

算法 4 操作算子 1 伪代码

BEGIN
输入:路径解、染色体长度、染色体取值范围
输出:染色体解
FOR(每个路径解)
IF 当前染色体位置为路径解中的路径,将当前位置取值为染色体最大值的 0.6

ELSE 将当前位置取值为染色体最大值的 0.4
END IF
END

操作算子中对路径解在染色体中的位置取值为最大值的 0.6，这是为了防止出现超强个体，导致遗传算法从开始就陷入局部最优. 同理对不在路径解中的位置取值为最大值的 0.4，这是为了防止出现长尾效应，导致遗传算法无法探索到可能存在的较好路径解，影响算法的全局探索能力.

操作算子 2 完成染色体到路径解的转换，如算法 5 所示.

算法 5　操作算子 2 伪代码

BEGIN
输入：染色体、路径解长度
输出：路径解
FOR(每个染色体)
将当前染色体的位置按照取值从大到小排列，选择路径长度的位置信息作为路径解
END IF
END

设 EACO-GA 的迭代次数为 num_c，蚂蚁种群规模为 num_a，遗传算法的迭代次数为 num_g，路径长度为 pl. EACO-GA 中蚂蚁搜索路径解的时间复杂度为 $O(num_c * num_a * pl^2)$，执行遗传算法的时间复杂度为 $O(num_c * num_g * log(n * num_a))$，因此 EACO-GA 的时间复杂度为 $O(max(num_c * num_a * pl^2, num_c * num_g * log(n * num_a)))$.

3　实验分析

3.1　实验数据集与实验方法

为了验证 EACO-GA 算法的有效性与优越性，以分类中的特征选择问题为例，对算法的性能做比较分析. 实验数据来源于 UCI 公开数据集，数据的相关属性如表 1 所示.

表 1　实验数据集属性

数据集	实例	特征	类别
Spect	267	22	2
Sonar	208	60	2
Ionosphere	351	34	2

对比方法采用粒子群优化(Partilce Swarm Optimization，PSO)与差分进化算法(Differential Evolution，DE)，使用分类正确率作为优化目标，采用 k 近邻分类器，粒子群优化与差分进化算法的迭代次数与种群规模设置与 EACO-GA 相同，其他相关参数设置与文献[6]和[7]一致，实验采用 5 重交叉检验.

EACO-GA 的参数设置如下，迭代次数 $num_c = 200$，种群规模 $num_a = 40$，初始信息素值 $\tau = 1$，常数 $Q = 1$，信息素挥发系数 $\rho = 0.05$，$\alpha = 0.5$，$\beta = 0.5$. EACO-GA 中蚁群优化算法的启发式信息采用费舍尔分值计算，如式(3).

$$F(h) = \frac{|\bar{\mu}_{1h} - \bar{\mu}_{2h}|}{\sqrt{\sigma_{1h}^2 + \sigma_{2h}^2}} \tag{3}$$

式中，$\bar{\mu}_{1h}$，$\bar{\mu}_{2h}$ 分别为正类和负类内第 h 个特征值的均值；σ_{1h}^2，σ_{2h}^2 分别为正类和负类内第 h 个相似特征值的方差. 通过费舍尔分值的定义和计算方法可以看出费舍尔分值的大小表明了特征判别能力的强弱，费舍尔分值越大，特征的判别能力越强.

3.2　参数敏感性分析

首先对 EACO-GA 算法中的参数 n 做敏感性分析，以 Spect 数据集为例，n 的取值范围从 0.1 到 0.9，算法的分类正确率与分类正确率标准差变化表明.

算法的分类正确率随着 n 的增加而上升，当 $n = 0.4$ 时，算法的分类正确率达到最好值 0.99，继续增加 n 的取值，算法的分类正确率随之下降. 分类正确率的标准差先上升再下降，在 $n = 0.4$ 时，分类正确率标准差达到最小值 0.022，继续增加 n 的取值，分类正确率标准差随之上升.

通过实验，可以得出结论：n 决定遗传算法初始染色体的数量与质量，n 较大时会引入较多质量较差的路径，导致遗传算法难以搜索到较好解，n 较小时，初始染色体的差异度较小，可能导致遗传算法陷入局部最优. 实验结果表明，当 $n = 0.4$ 时，EACO-GA 算法拥有较好的性能，即能够在开发与探索之间达到较好的平衡. 在其他两个数据集上的实验亦能得出相同的结果.

3.3　ECO-GA 性能分析

本节对 EACO-GA 的性能做比较分析，对 EACO-GA 中的参数 n 取值 0.4，其他相关参数设置见 3.1 节. 图 3、图 4 和图 5 显示算法在实验数据集上特征比例从 0.1 到 0.9 变化时，分类正确率的变化趋势图.

图 1 显示 3 种算法在 Spect 数据集上的分类正确率随特征比例的变化趋势. 通过观察可以看出，PSO 与 DE 在特征比例为 0.2 时达到最好值，但要

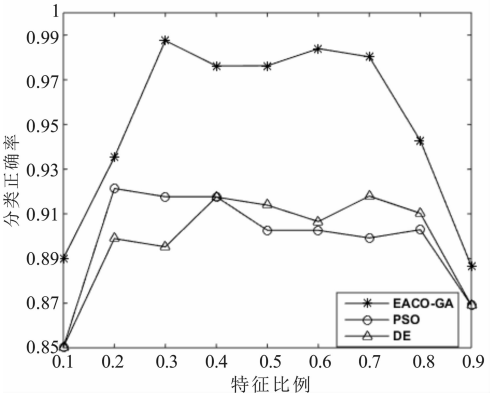

图 1　Spect 数据集分类准确率趋势图

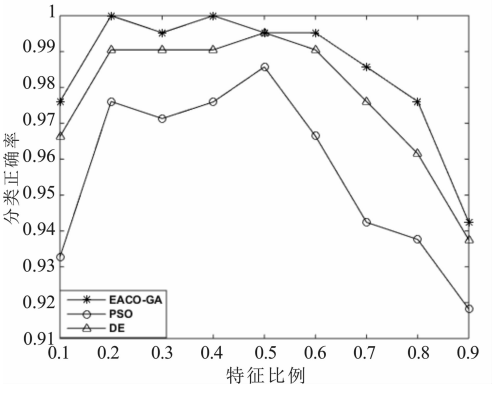

图 2　Sonar 数据集分类准确率趋势图

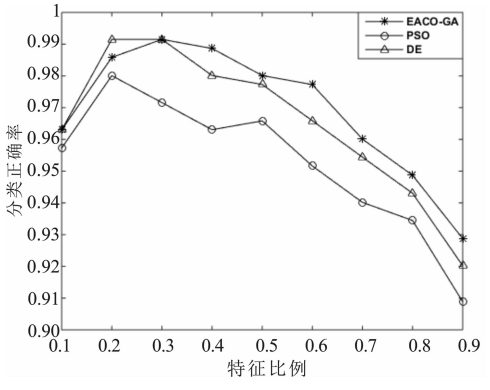

图 3　Ionosphere 数据集分类准确率趋势图

低于 EACO-GA 的分类正确率，而 EACO-GA 在特征比例为 0.3 时达到最好值 0.99. 进一步增加特征比例，3 种算法的性能都有所下降，但 EACO-GA 的性能要好于其他 2 种算法.

如图 2 所示，在 Sonar 数据集上，3 种算法的分类正确率随着特征比例的增加呈现出先增加上升后下降的趋势. 在特征比例为 0.5 时，EACO-GA 的分类正确率与 PSO 相同，在其他情况下 EACO-GA 的算法性能好于 PSO 与 DE 算法.

图 3 是 3 种算法在 Ionosphere 数据集上的分类正确率变化趋势，3 种算法均在特征比例为 0.2 时达到最好值，且 EACO-GA 的分类正确率最好，在特征比例为 0.3 时，EACO-GA 的分类正确率与 PSO 相同. 在特征比例大于 0.2 时，随着特征比例的增加，3 种算法的性能都呈现明显的下降趋势，但是 EACO-GA 算法的分类性能仍然好于 PSO 与 DE.

通过在数据集上的实验表明，EACO-GA 算法的性能在多数情况下要好于 PSO 与 DE 算法，在部分情况下与 PSO 算法相似. 这表明 EACO-GA 算法通过将历史最优信息与当前最优信息结合并采用遗传算法在更广阔的空间中搜索较好解，能够显著提高算法的性能，使得算法能够在开发与平衡之间达到较好的平衡，防止算法陷入局部最优.

同时从实验结果可以看出，随着特征比例的增加，算法的性能出现显著的下降，这是由于特征比例的增加使得算法选择了冗余和不相关特征，导致分类正确率的降低. 这表明，采用 EACO-GA 算法进行特征选择能够确定当前数据集的最佳特征选择比例，并且能够获得较好的分类结果.

4　结束语

为了进一步提高蚁群优化的性能，平衡算法的开发与探索能力，提出了一种采用遗传算法增强的蚁群优化算法 EACO-GA. 该算法充分利用历史最优解与当前部分最优解信息，将该信息编码成染色体，使用遗传算法在更广阔的空间中搜索较好解，避免算法陷入局部最优. 以二分类问题中的特征选择为例，与粒子群优化算法和差分进化算法进行对比，在 3 个典型数据集上进行实验，结果验证了算法的有效性与优越性.

参考文献：

[1] Zhang Q, Zhang C. An improved Ant Colony Optimization Algorithm with Strengthened Pheromone Updating Mechanism for Constraint Satisfaction Problem [J]. Neural Computing & Applications, 2017:1-12.

[2] Shabeera T P, Kumar S D M, Salam S M, et al. Optimizing VM Allocation and Data Placement for Data-intensive Applications in Cloud Using ACO Metaheuristic Algorithm[J]. Engineering Science & Technology An International Journal, 2016. 20(2).

（下转第 108 页）

4 结束语

本文提出一种基于关联规则挖掘的社区网络数据爬虫算法.挖掘社区网络用户行为信息特征量,根据数据的属性特征进行关联规则合并,采用模糊指向性聚类方法进行社区网络的用户行为属性特征聚类处理,采用自相关特模板匹配方法实现信息融合,实现社交网络数据的关联规则挖掘,结合用户行为特征分布属性进行网络信息爬虫,实现社区属性混合推荐.研究表明,本文方法进行社区网络的用户行为属性特征数据爬虫挖掘的准确性较高,特征匹配能力较好,提高社区发现能力.

参考文献:

[1] 刘东,吴泉源,韩伟红,等.基于用户名特征的用户身份同一性判定方法[J].计算机学报,2015,38(10):2028-2040.

[2] 罗磊,陈恳,杜峰坡,等.基于改进型粒子群算法的曲面匹配与位姿获取[J].清华大学学报(自然科学版),2015,55(10):1061-1066.

[3] DURAO F, DOLOG P. Improving tag-based recommendation with the collaborative value of wiki pages for knowledge sharing[J]. Journal of Ambient Intelligence Humanized Computing, 2014, 5(1):21-38

[4] JIA D Y, ZHANG F Z. A collaborative filtering recommendation algorithm based on double neighbor choosing strategy[J]. Journal of Computer Research and Development, 2013, 50(5):1076-1084.

[5] POLATIDIS N, GEORGIADIS C K. A multi-level collaborative filtering method that improves recommendations[J]. Expert Systems with Applications, 2016, 48:100-110.

[6] GUO Huaping, ZHOU Jun, WU Chang´an, FAN Ming. k-nearest neighbor classification method for class-imbalanced problem. Journal of Computer Applications, 2018, 38(4): 955-959.

[7] 孟祥武,刘树栋,张玉洁,等.社会化推荐系统研究[J].软件学报,2015,26(6):1356-1372.

[8] 吴铮,于洪涛,刘树新,等.基于信息熵的跨社区网络用户身份识别方法[J].计算机应用,2017,37(8):2374-2380.

作者简介:

穆 俊 男,(1979-),硕士,副教授.研究方向为计算机软件与理论、计算机应用、数据挖掘.
E-mail:mujun511@126.com.

(上接第 104 页)

[3] Wu M T, Zhan J, Lin C W. Mining of High-Utility Itemsets by ACO Algorithm[C]// The Multidisciplinary International Social Networks Conference. 2016:44.

[4] Liu Y, Diao X, Cao J, et al. Local Enhanced Catfish Bat Algorithm[C]// International Conference on Robots & Intelligent System. 2016:238-245.

[5] Król A. Application of the Genetic Algorithm for Optimization of the Public Transportation Lines[M]// Intelligent Transport Systems and Travel Behavior. Springer International Publishing, 2017.

[6] Wang B, Zheng X, Zhou S, et al. Constructing DNA Barcode Sets Based on Particle Swarm Optimization. [J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2017, PP(99): 1-1.

[7] Meng Y S, Enríquez A C, Hsiao T Y, et al. Enhanced Differential Evolution Algorithm for Coordination of Directional Overcurrent Relays[J]. Electric Power Systems Research, 2017, 143:365-375.

作者简介:

梁豪默 男,(1988-),博士研究生.研究方向为系统工程.

王智学 男,(1961-),硕士,教授.研究方向为需求工程、系统工程.

刘 艺(通讯作者) 男,(1990-),博士研究生.研究方向为进化算法.E-mail:albetliu20th@163.com.