

# 一种遗传量子粒子群的属性约简算法

周丽娟<sup>1</sup>, 王加阳<sup>2</sup>, 谢颖<sup>2</sup>

(1. 湖南工业大学 科技学院, 湖南 株洲 412008; 2. 中南大学 信息科学与工程学院, 湖南 长沙 410083)

**摘要:** 针对粒子群算法收敛速度不佳和易陷入局部最优的问题, 提出了一种遗传量子粒子群优化(GQPSO)的属性约简算法, GQPSO算法利用量子系统较大的搜索范围, 并借鉴遗传算法的选择、变异等操作, 从而避免了算法过早收敛至局部最优, 且能得到可观的收敛速度。实验结果表明, GQPSO算法具有更快的收敛速度和全局搜索能力, 提高了属性约简的效率。

**关键词:** 属性约简; 遗传算法; 量子; 粒子群; 收敛

**中图分类号:** TP391

**文献标志码:** A

**文章编号:** 1673-9833(2010)06-0049-04

## An Attribute Reduction Algorithm Based on Genetic Quantum Particle Swarm

Zhou Lijuan<sup>1</sup>, Wang Jiayang<sup>2</sup>, Xie Ying<sup>2</sup>

(1. College of Science and Technology, Hunan University of Technology, Zhuzhou Hunan 412008, China;

2. School of Information Science and Engineering, Central South University, Changsha 410083, China)

**Abstract:** To solve the problems of the poor convergence speed and being easy to fall into the local optimum in the particle swarm algorithm, an attribute reduction algorithm based on genetic quantum particle swarm (GQPSO) is presented. GQPSO takes advantage of the wide search range of quantum system and utilizes the selection and variation of the genetic algorithm to avoid algorithm premature convergence local optimum and get considerable convergence speed. The experiment shows that GQPSO has a faster convergence rate and global search capabilities, which improves the efficiency of the attribute reduction.

**Keywords:** attribute reduction; genetic algorithm; quantum; particle swarm; convergence

## 0 引言

粗糙集理论是一种处理模糊和不确定性知识的数学工具<sup>[1]</sup>。属性约简是粗糙集理论中的核心内容, 是一种有效的特征选择方法, 已被广泛应用于特征选择和知识库约简<sup>[2]</sup>。属性约简的本质是通过在属性集中寻找一个最小属性集合, 以保持原决策信息系统的分辨能力不变, 但提高系统潜在知识的清晰度和发现效率。人们总期望找到最小的属性约简, 但这已被证明是一个NP-hard问题<sup>[3]</sup>, 因而, 寻求最有效的属性约简算法是粗糙集理论研究的主体方向。

传统的属性约简算法大多无法保证在广阔搜索空间下获得最小属性约简。目前, 很多学者对此进行了深入研究, 并提出了一些改进的属性约简算法。如文献[4]中, 作者提出了一种基于蚁群优化的属性约简算法, 并利用该算法求得了多个约简结果, 但该算法存在搜索空间大、时间复杂度偏高等缺点; 文献[5-6]提出了一种粒子群优化算法与属性约简相结合的算法, 将最小属性约简问题转化成一个适合粒子群优化算法求解的多目标优化问题, 能一次运算找到更多的最优解, 但收敛速度还有待提高; 袁可红等人<sup>[7]</sup>提出了一种基于粗集约简群智能算法的储层识别方法, 将粗糙

收稿日期: 2010-09-13

通信作者: 周丽娟(1974-), 女, 湖南浏阳人, 湖南工业大学讲师, 硕士, 主要从事数据库, 数据挖掘和智能计算方面的研究,

E-mail: zlj706@163.com

集属性约简用于储层属性的特征提取,该方法节省了用于属性探测的成本;文献[8]的作者提出了粒子群属性约简算法,该算法求解效率高,但容易陷入局部最优;文献[9]的作者提出了基于量子粒子群优化的最小属性约简算法,其具有较好的寻优搜索能力,但收敛速度不佳。本文针对量子粒子群收敛速度不够快的不足,借鉴遗传算法中的选择、变异等操作,提出一种遗传量子粒子群优化(genetic quantum particle swarm optimization,简称GQPSO)的属性约简算法。经过理论分析与实验验证,该算法是可行的,且具有更快的收敛速度和全局搜索能力。

## 1 量子粒子群算法基本原理

粒子群算法(particle swarm optimization,简称PSO)的发现是基于对简化的动物社会模型的模拟,其基本思想是通过模拟鱼群、鸟群觅食过程中的迁徙和聚集行为,实现对问题的优化<sup>[10]</sup>。PSO中的每个粒子都代表搜索空间的1个解,先随机初始化每个粒子的位置和速度,接着粒子在自身及整个种群最优位置引导下飞向最优解。PSO算法在迭代时容易陷入局部极值点中,导致得不到全局最优解,且PSO算法的收敛速度比较慢。在量子系统的启发下,一些学者将量子系统融入于PSO算法当中,提出了能保证全局收敛的量子粒子群优化算法(quantum particle swarm optimization,简称QPSO)。QPSO算法是对粒子群优化算法进行搜索策略的改变,是由孙俊等人最先提出来的,他们从量子力学的角度考虑,认为粒子具有量子行为。适用量子粒子群算法的量子系统是一个复杂的非线性系统,它符合状态重叠原理,而且量子系统是一个不确定系统,空间中的粒子移动时没有确定的轨迹,因此,每一个粒子能够以某一确定的概率出现在搜索空间中的任意一个位置,根据这一特性,可寻找全局最优解。因此,该算法具有较强的搜索能力。

在QPSO算法中,不能直接处理粗糙集中的属性,首先得将属性表示成粒子形式。在特征提取过程中,设属性集 $Q=\{q_1, q_2, \dots, q_n\}$ ,将属性集表示成二进制串 $t, t=\{a_1, a_2, \dots, a_n\}$ ,  $k \in [1, n]$ ,其中,  $a_k=1$ 表示 $q_k$ 被选择,  $a_k=0$ 表示 $q_k$ 未被选择,故将 $t$ 看成是量子粒子群优化算法可处理的粒子,粒子 $t$ 所对应的近似精度为 $\gamma_t$ ,条件属性集 $C$ 所对应的近似精度为 $\gamma_C$ 。用适应度函数来求解最小约简,定义如式(1):

$$fitness(t) = \frac{k_1}{k_1 + k_2} \cdot \frac{\gamma_t}{\gamma_C} + \frac{k_1}{k_1 + k_2} \cdot \frac{card(C) - card(t)}{card(C)}, \quad (1)$$

式中: $\gamma_t/\gamma_C$ 用来判断属性集是否为一个约简,当其值为“1”时,表示此刻的属性集为一个约简; $k_1, k_2$ 是随

着粒子群的进化而动态调节的参数; $Card()$ 表示集合的基数, $C$ 为条件属性集。

QPSO算法中的粒子按式(2)进行位置信息更新:

$$X(t+1) = p \pm \beta * |M_{best} - X(t)| * \ln(1/u), \quad (2)$$

式中: $X(t+1)$ 为粒子在 $t$ 时刻的位置;

$u$ 为 $[0, 1]$ 之间产生的随机数;

$\beta$ 为收缩扩张因子,用来控制粒子的收敛速度,已有资料表明, $\beta$ 在0.3~0.8范围内时,能得到很好的最优解;

“ $\pm$ ”是由在迭代过程中 $[0, 1]$ 之间产生的随机数决定的,若随机数小于0.5时取“-”,否则取“+”;

$p$ 表示在全局最优值与局部最优值之间的一个随机值,由式(3)定义:

$$p = (c_1 * p_{besti} + c_2 * g_{best}) / (c_1 + c_2), \quad (3)$$

其中, $p_{besti}$ 是粒子 $i$ 的局部最优值; $g_{best}$ 为所有粒子中的全局最优值; $c_1, c_2$ 为 $[0, 1]$ 之间产生的随机数。

$M_{best}$ 为整个粒子群的中心位置,定义如式(4):

$$M_{best} = \frac{1}{m} \sum_{i=1}^m p_{besti} = \frac{1}{m} \left( \sum_{i=1}^m p_{besti1}, \frac{1}{m} \sum_{i=1}^m p_{besti2}, \dots, \frac{1}{m} \sum_{i=1}^m p_{bestid} \right). \quad (4)$$

其中, $m$ 为种群粒子个数。

## 2 遗传量子粒子群属性约简算法

QPSO算法提高了粒子约简算法的空间搜索能力,但在收敛速度方面仍存在很多不足。遗传算法是一种借鉴自然界选择和遗传机制的随机化自适应函数搜索算法,主要由选择、交叉和变异算子组成,分别模拟达尔文进化过程中的自然选择、群体遗传过程中发生的交配和突变等现象<sup>[11]</sup>。其收敛速度快、染色体间信息共享较充分,因此,借鉴遗传算法的选择、变异操作,将遗传算法的思想融入QPSO算法中,能充分利用量子粒子群搜索到的有用信息,从而提高种群的收敛速度。

遗传量子粒子群属性约简算法主要针对粒子的更新,通过对当前种群使用选择、变异等遗传操作,产生新一代的群体。即在每一次粒子群迭代更新之后,首先计算每个粒子的适应度值 $fitness$ ,得出整个种群的平均适应度值 $A_F$ 。然后以此 $A_F$ 为标准,将种群中每一个粒子的适应度值与 $A_F$ 进行比较,将大于 $A_F$ 的粒子保留下来作为下一代种群中较好的粒子,同时对小于 $A_F$ 的粒子进行变异处理,即以概率 $P_m$ 对粒子的每一位进行取反处理,把1变为0,0变为1。

遗传量子粒子群属性约简算法执行过程如下:

第一步 根据差别矩阵计算核属性 $Core(Q)$ ,判断

$\gamma_{Core(Q)}$  与  $\gamma_Q$  是否相等, 若相等则核属性  $Core(Q)$  为最小属性约简, 否则执行第二步。

第二步 初始化粒子。确定种群规模  $M$  和粒子维数  $D$ , 根据  $weight = \gamma_{Core(c)} \cup q_k - \gamma_{Core(c)}$  来确定每个属性  $q_k$  的权重, 根据粒子的权重来随机初始化  $m$  个粒子; 设置迭代次数  $T$ 。

第三步 计算各粒子的适应度值, 记录群体中每个粒子的局部最优解  $p_{besti}$  和种群全局最优解  $g_{best}$ ; 计算  $m_{best}$ 。

第四步 粒子群更新。

1) 根据式 (2) ~ (4) 以一定概率取加或减, 更新每个粒子的位置;

2) 计算  $t$  时刻第  $i$  个粒子的适应度值  $fitness(i)$ , 根据公式  $p_{besti} = \max(p_{besti}, fitness(i))$  更新局部最优值  $p_{besti}$ , 根据公式  $g_{best} = \max(p_{besti}, g_{best})$  计算种群当前的全局最优粒子;

3) 计算当前种群平均适应度值  $A_F$ ;

4) 进行选择、变异操作。比较种群中每个粒子的适应度值  $fitness$  与当前种群平均适应度值  $A_F$ , 保留适应度值比  $A_F$  大的粒子作为下一代种群中较好的粒子; 将比  $A_F$  小的粒子, 以概率  $P_m$  对其每一位进行取反变异, 即 0 变为 1, 1 变为 0。

第五步 如果满足终止条件, 则输出全局最优值, 否则转到第四步。

### 3 实验结果与分析

为验证本文所提出的 GQPSO 算法的有效性, 采用几个 UCI 数据库标准数据集作为试验数据, 选取了几个记录数和属性数都较苛刻的数据集 (具体数据参见表 1), 分别采用经典的启发式最小属性约简算法 Hu X. H. 算法 (简记为 Hu 算法)、PSO 算法、QPSO 算法和 GQPSO 算法对所取标准数据集进行约简, 约简后所得到的属性数如表 2 所示, 适应度值的变化曲线如图 1 所示。实验环境的配置如下: 操作系统为 Windows XP, 物理内存为 2G, CPU 为 AMD 2800+, 前台开发工具为 VC++6.0, 后台数据库系统为 SQL Server2000。

表 1 数据集  
Table 1 Data set

数据集名称	记录数	条件属性数	提供者
Wine	178	13	Forina
Sponge	76	45	I. Uriz
Spect_Heart	267	22	J. Cios
Soybean_large	307	36	R. Michalski
Diagnostic	569	32	Dr. William

表 2 属性约简结果

Table 2 The results of attribute reduction

数据集名称	记录数	条件属性数	算法名称	约简后属性数
Wine	178	13	HU 算法	5
			PSO 算法	6
			QPSO 算法	5
			GQPSO 算法	5
Sponge	76	45	HU 算法	8
			PSO 算法	12
			QPSO 算法	8
			GQPSO 算法	8
Spect_Heart	267	22	HU 算法	7
			PSO 算法	7
			QPSO 算法	7
			GQPSO 算法	7
Soybean_large	307	36	HU 算法	10
			PSO 算法	12
			QPSO 算法	9
			GQPSO 算法	9
Diagnostic	569	32	HU 算法	12
			PSO 算法	13
			QPSO 算法	11
			GQPSO 算法	11

实验的性能通过以下两个方面进行评估:

1) 算法获得的约简后属性数, 此数据用来评估该算法的约简效果;

2) 在迭代更新过程中最优粒子适应度值的变化, 以评价该算法的收敛速度。

从表 2 中的约简结果可以看出, 本文所提出的 GQPSO 算法在所选取的 4 种算法中, 取得了属性数目最少的约简结果, 且和 QPSO 算法保持一致, 这主要是由于 GQPSO 算法中遗传算法的选择、变异操作对低适应度值的粒子进行变异, 只保留了高适应度值的粒子, 因而提高了粒子群的整体适应度值, 且没影响其搜索能力。

在最优解的搜索能力上, GQPSO 算法与 QPSO 算法的效果基本相似, 但在收敛速度方面, 二者则有相当大的差异性。

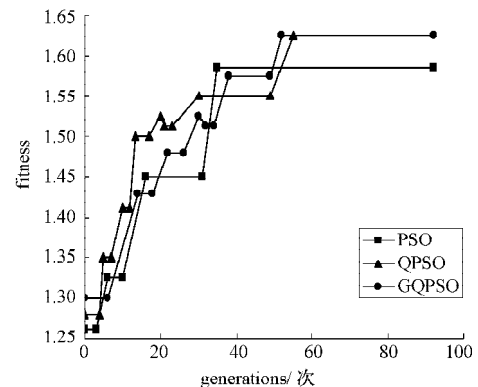


图 1 Wine 迭代更新中适应度值曲线

Fig. 1 The curve of fitness value in Wine iterative updating

从图 1 可以看出, 3 种算法都能找到最小约简, GQPSO 算法找到的最小约简与 QPSO 算法相同, 且这两者都优于 PSO 算法, 但 GQPSO 算法在收敛速度上要快于 QPSO 算法, 这是由于 GQPSO 算法增加了遗传算法中的选择、变异操作后, 能充分利用种群的有效信息, 使得种群整体的适应度值得到了提高, 即收敛速度得到了提高, 这与理论分析预期相吻合。

## 4 结语

本文将遗传算法的选择、变异操作融入量子粒子群算法中, 提出了一种基于遗传量子粒子群属性约简算法, 该算法发挥了 QPSO 算法较强的寻优能力, 同时弥补了 QPSO 算法收敛速度不足的缺点。经过实验验证, GQPSO 算法具有更快的收敛速度和全局搜索能力, 提高了属性约简的效率。但是, GQPSO 算法在其收敛性能上还需要更进一步优化, 这是作者下一步的研究目标。

## 参考文献:

- [1] 王国胤. Rough 集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001: 10-22.  
Wang Guoyin. Rough Sets Theory and Knowledge Acquisition [M]. Xi'an: Xi'an Jiaotong University Press, 2001: 10-22.
- [2] 张文修, 梁 怡, 吴伟志. 信息系统与知识发现[M]. 北京: 科学出版社, 2003: 56-68.  
Zhang Wenxiu, Liang Yi, Wu Weizhi. Information System and Knowledge Discovery[M]. Beijing: Science Press, 2003: 56-68.
- [3] Wong S K M, Ziarko W. On Optimal Decision Rules in Decision Tables[M]. Poland: Bulletin of Polish Academy of Science, 1985: 693-696.
- [4] Jiang Yuanchun, Liu Yezheng. An Attribute Reduction Method Based on Ant Colony Optimization[C]//Proceedings of the 6th World Congress on Intelligent Control and Automation. Washington: IEEE Computer Society Press, 2005: 3542-3546.
- [5] Wang Xiangyang, Yang Jie, Teng Xiaolong. Feature Selection Based on Rough Sets and Particle Swarm Optimization[J]. Pattern Recognition Letters, 2007, 28(4): 459-471.
- [6] 杨晓燕, 陈国龙, 郭文忠. 基于粒子群优化的最小属性约简算法[J]. 福州大学学报: 自然科学版, 2010, 38(2): 193-197.  
Yang Xiaoyan, Chen Guolong, Guo Wenzhong. Minimum Attribute Reduction Algorithm Based on Particle Swarm Optimization[J]. Journal of Fuzhou University: Natural Science, 2010, 38(2): 193-197.
- [7] 袁可红, 李艳晓, 诸克军. 基于粗集约简的群智能算法的储层识别[J]. 湖南工业大学学报, 2008, 22 (5): 46-48.  
Yuan Kehong, Li Yanxiao, Zhu Kejun. Reservoir Identification of the Swam Intelligence Algorithms Based on Rough Set Reduction[J]. Journal of Hunan University of Technology, 2008, 22 (5): 46-48.
- [8] Dai Jianhuan, Chen Weidong, Guo Hongying, et al. Particle Swarm Algorithm Form Inimal Attribute Reduction of Decision Data Tables[C]//Processing of the 1st International Multi-Symposiums on Computer and Computational Science. Washington: IEEE Computer Society Press, 2006: 3021-3025.
- [9] 王加阳, 谢 颖. 基于量子粒子群优化的最小属性约简算法[J]. 计算机工程, 2009, 35(12): 148-150, 153.  
Wang Jiayang, Xie Ying. Minimal Attribute Reduction Algorithm Based on Quantum Particle Swarm Optimization [J]. Computer Engineering, 2009, 35(12): 148-150, 153.
- [10] 苗 荣, 闫 伟, 李树荣. 基于粒子群位移转移的混合遗传算法及其应用[J]. 计算机工程与应用, 2007, 43(15): 41-44.  
Miao Rong, Yan Wei, Li Shurong. Novel Hybrid GA Based on Position Displacement Idea of PSO and Its Appliction[J]. Computer Engineering and Applications, 2007, 43(15): 41-44.
- [11] 孔宪仁, 秦玉灵, 罗文波. 遗传-粒子群算法模型修正[J]. 力学与实践, 2009, 31(5): 56-60.  
Kong Xianren, Qin Yuling, Luo Wenbo. GA-PSO Algorithm Model Updating[J]. Mechanics in Engineering, 2009, 31(5): 56-60.

(责任编辑: 廖友媛)