

# 高斯扰动粒子群算法的数据库查询优化<sup>①</sup>

李国芳, 李 静

(广州医科大学 卫生职业技术学院, 广州 510450)

**摘 要:** 针对量子粒子群算法存在的不足, 将变异算子引入其中, 提出一种高斯变异量子粒子群算法 (GM-QPSO), 并将其应用于数据库查询优化中. 首先建立数据库查询优化数学模型, 然后采用量子粒子代表一个可行的数据库查询方案, 然后通过量子粒子之间的信息交流, 找到数据库查询最优解, 最后在 Matlab 2012 上进行了仿真实验. 仿真结果表明, GM-QPSO 克服了量子粒子群算法存在的不足, 不仅提高了数据库查询速度, 而且获得了更加理想的查询优化方案.

**关键词:** 数据库查询; 变异算子; 遗传算法; 粒子群算法

## Query Optimization of Database Based on Gauss Mutation Quantum Behaved Particle Swarm of Optimization Algorithm

LI Guo-Fang, LI Jing

(Health Vocational College, Guangzhou Medical University, Guangzhou 510450, China)

**Abstract:** In order to solve the defect of quantum particle swarm algorithm, mutation operator of the genetic algorithm is introduced into quantum particle swarm optimization algorithm. It produces a novel query optimization method of database (GM-QPSO). Firstly, the mathematic model is established for database query optimization problems. And then the optimal scheme of database query optimization problems is found by the sharing message of quantum particle. Finally, the simulation experiments is carried out on Matlab 2012. The results show that the proposed algorithm has solved the defect of quantum particle swarm algorithm, and improved query speed of database and can obtain better query scheme.

**Key words:** database query; mutation operator; genetic algorithm; quantum behaved particle swarm optimization algorithm

随着数据库规模日益增大, 数据查询频率增加, 因此数据查询效率至关重要, 如何快速找到满足用户查询条件的记录, 是现代大规模数据研究的热点问题<sup>[1]</sup>.

针对数据库查询优化问题, 国内外学者投入了大量时间和精力进行了研究<sup>[2]</sup>. 数据库查询优化是一个典型的多约束的组合优化问题, 传统穷举搜索算法存在效率低, 效果差等缺陷. 为了解决该难题, 一些学者将启发式算法引入到数据库查询中<sup>[3]</sup>. 如: 出现了基于遗传算法、蚁群算法、粒子群算法、模拟退火、蛙跳算法以及它们的组合算法<sup>[4-8]</sup>, 一定程度上提高了查询优化效果, 获得了比较理想的查询效果, 但

是随着数据库规模的不断扩大, 这些算法的问题开始暴露出来, 出现了收敛速度慢、易陷入局部最优、早熟现象, 难以得到全局最优的数据库查询方案<sup>[9]</sup>. 近年来, 随着量子计算的发展, Sun 等人在粒子群算法的基础上, 引入量子行为, 提出了量子粒子群算法 (quantum-behaved particle swarm optimization, QPSO) 算法<sup>[10]</sup>. 量子粒子群是粒子在进化过程中加入了量子行为, 因此, 相比标准的粒子群算法具有较好的全局收敛性. 但是, 该算法仍旧不可避免早熟<sup>[11]</sup>.

为了提高数据库查询的优化效率, 提出一种高斯变异量子粒子群算法的数据库查询优化方法 (Gaussian mutation quantum behaved particle swarm optimization,

<sup>①</sup>收稿时间:2014-02-19;收到修改稿时间:2014-03-28

GM-QPSO)的数据库查询优化算法,并通过仿真实验对 GM-QPSO 算法在数据库查询优化问题求解中的性能.

## 1 高斯变异的量子粒子群优化算法

### 1.1 量子粒子群优化算法

粒子群优化算法是 1995 年由美国学者 J.Kennedy 和 R.C.Eberhart 受鸟群觅食行为的启发而提出的<sup>[7]</sup>,是一种基于迭代的优化算法,但是,标准粒子群算法还存在一些有待于改进的问题,主要有搜索空间有限、容易出现早熟、收敛精度不高、容易陷入局部优化等<sup>[12]</sup>. J.Sun 等人 2004 年在标准的 PSO 基础上引入量子行为,提出了量子粒子群(QPSO),使得粒子可以在整个可行解的空间中进行搜索,从而寻求全局最优解, QPSO 算法比标准 PSO 算法具有更好的全局收敛性和搜索能力. QPSO 是以 DELTA 势阱为基础,认为粒子具有量子的行为,该算法的迭代公式为:

$$mbest = \frac{1}{M} \sum_{i=1}^D p_i \quad (1)$$

$$p_d = \frac{R_{1d} p_{id} + R_{2d} P_{gd}}{R_{1d} + R_{2d}} \quad (2)$$

$$X_{id}(t+1) = P \pm \beta \times |mbest_d - X_{id}(t)| \times \ln\left(\frac{1}{u}\right) \quad (3)$$

其中,  $mbest$  为平均最优位置;  $M$  为粒子的数目;  $p_i$  第  $i$  个粒子的个体最优位置;  $X_i$  为第  $i$  个粒子的位置,  $u$ 、 $R_{1d}$ 、 $R_{2d}$  是介于(0,1)之间的

### 1.2 高斯变异量子粒子群优化算法(GM-QPSO)

为了增加粒子的多样性,避免陷入局部极值,把高斯分布融合到量子粒子群算法中,用高斯分布变异代替随机数变异操作,具体步骤如下:

- 1) 初始化粒子群中粒子的位置,并把每个粒子位置的设置为个体最优位置.
- 2) 计算每个粒子的适应度值,如果  $f(X_i) < f(P_i)$ , 则更新粒子的个体最优位置  $P_i = X_i$ .
- 3) 更新粒子群的全局最优位置  $P_g = \min(P_i)$ .
- 4) 利用公式(3)计算平均最优位置  $mBest$ ;
- 5) 利用公式(8)和(9)分别对  $mBest$  和  $P_g$  进行扰动.

$$mutate(mbest) = mbest + \phi\delta \quad (4)$$

$$mutate(P_g) = P_g + \phi\delta \quad (5)$$

式中,  $\phi$  是变异大小,  $\delta$  是一个柯西分布的随机变量,

6) 对每个粒子利用式(3)更新粒子的位置;

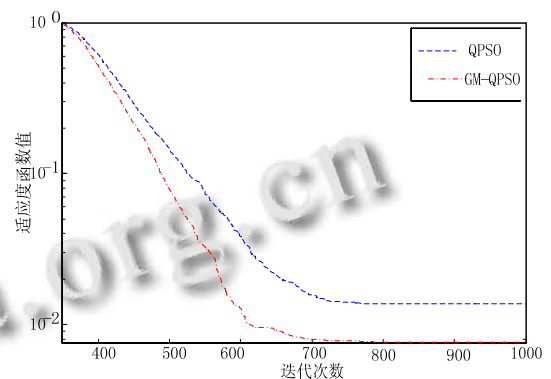
7) 如果满足结束条件,则算法结束;否则,返回步骤 2).

采用 4 个准测试函数: Sphere、Griewank、Rosenbrock 和 Rastrigin 对 QPSO 算法和 GM-QPSO 算法的性能进行测试,所有算法的运行结果如图 1 所示.从图 1 可知,对所有函数,相对于 QPSO 算法,GM-QPSO 获得更优的全局搜索能力、收敛精度和收敛速度,这主要是由于当粒子因为陷入局部最优而出现早熟现象的时候,高斯变异使得粒子有机会在小范围内进行搜索,获得了更优的结果.

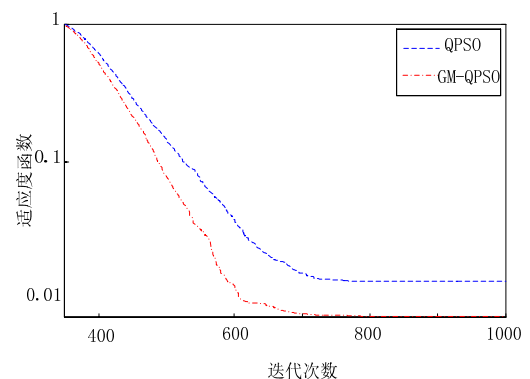
## 2 GM-QPSO的数据查询优化问题求解

### 2.1 数据库查询优化的数学模型

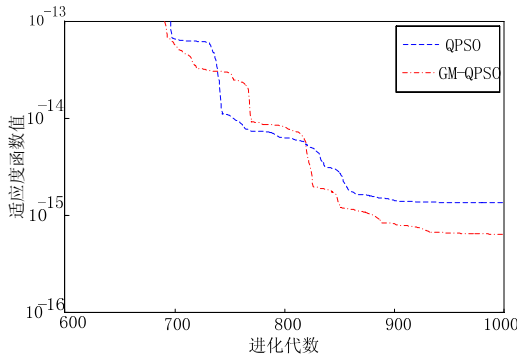
对于一个含有  $n$  个关系的连接树,  $(j_1, j_2, \dots, j_i)$  为连接树内部节点(即非叶子节点),其代价评估模型为:查询优化方案可以描述为 4 种情况,具体如图 5 所示,此处选用(a)“左深树”搜索策略空间为例进行分析.



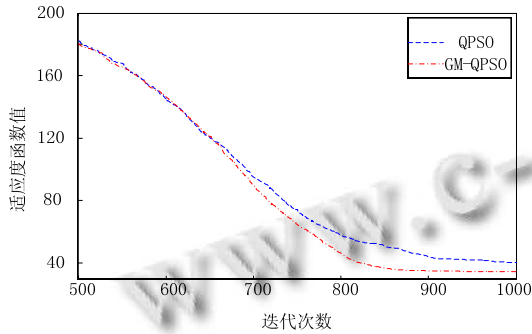
(a) Sphere 函数适应度变化曲线



(b) Griewank 函数适应度变化曲线



(c) Rosenbrock 函数适应度变化曲线



(d) Rastrigin 函数适应度变化曲线

图 1 QPSO 改进前后的收敛性能比较

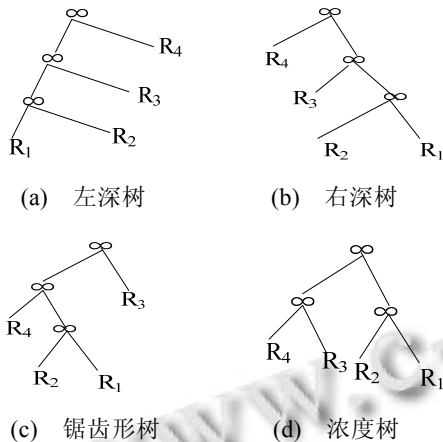


图 2 四种连接情况图示

$$COST = \sum_{i=1}^{n-1} cost(j_i) \quad (6)$$

对于一个连接  $J=R \text{ join } S$ , 其计算代价  $cost(J)$  如式(13)所示:

$$cost(J) = \frac{|R| \times |S|}{\prod_{c \in C} \max(V(c, R), V(c, S))} \quad (7)$$

(7)中  $V(c,R)$  的计算方式如公式(8):

$$V(c, J) = \begin{cases} V(c, J) & c \in R-S \\ V(c, S) & c \in S-R \\ \min(V(c, R), V(c, S)) & c \in R \cap S \\ 1 \leq i_1, i_2, \dots, i_m < n \end{cases} \quad (8)$$

式中,  $|R|$  表示关系  $R$  的元组数;  $C$  表示关系  $R$  和关系  $S$  的公共属性集合;  $V(c, R)$  表示关系  $R$  中属性  $c$  的不同取值的个数.

### 2.2 适应度函数设计

量子粒子优劣通过适应度函数进行评价, 数据库查询优化目标是找到一个  $cost(X_i)$  代价最小的查询计划. 由于每一个量子粒子与一个连接查询树对应, 对于粒子  $X_i$ , 采用计算其  $cost(X_i)$ , 然后对  $cost(X_i)$  取倒数作为量子粒子的适应度值, 具体为:

$$f(X_i) = \frac{1}{COST(X_i)} \quad (9)$$

### 2.3 GM-QPSO 的数据库查询优化求解步骤

- 1) 参数初始化: 初始化 GM-QPSO 算法的各个参数, 如最大迭代次数等.
- 2) 初始量子粒子群, 根据式(9)计算量子粒子适应度值并将个体的历史最优位置  $P_{id}$  设为当前位置, 而群体中最优的个体作为当前的  $P_{gd}$ .
- 3) 根据粒子的适应度更新  $P_{id}$  和  $P_{gd}$ .
- 4) 并根据式(4)、(5)对  $mbest$  进行变异, 同时, 根据式(3)更新量子粒子的位置.
- 5) 如果达到最大迭代次数, 输出最优粒子对应的数据库查询方案, 不然, 返回步骤 3) 继续迭代.

## 3 仿真实验

### 3.1 仿真环境

为了测试 GM-QPSO 算法的性能, 在 CPU 为 Intel Core™ 2 Quad 3.0GHz, RAM 为 2.0GB, 操作系统为 Windows 7 的平台下, 采用 Matlab 2012 工具箱进行仿真实验. 采用 QPSO 算法, 遗传算法(GA)、粒子群算法(PSO)进行对比实验. 采用搜索最优查询方案和执行最优查询方案两个方面的代价进行性能评估算法性能, 具体定义如下:

$$\text{搜索代价消耗比} = \frac{\text{当前最优查询方案的搜索时间}}{\text{全局最优查询方案的搜索时间}} \quad (10)$$

$$\text{查询代价消耗比} = \frac{\text{当前最优查询方案的执行时间}}{\text{全局最优查询方案的执行时间}} \quad (11)$$

### 3.2 结果与分析

#### 3.2.1 查询结果比较

经多次实验并求得平均值, QPSO、GM-QPSO、GA、PSO 算法搜索最优查询方案代价比变化曲线如图 3 所示. 执行最优查询方案代价比变化曲线如图 4 所示. 从图 3 和 4 的结果进行分析可以得到如下结论:

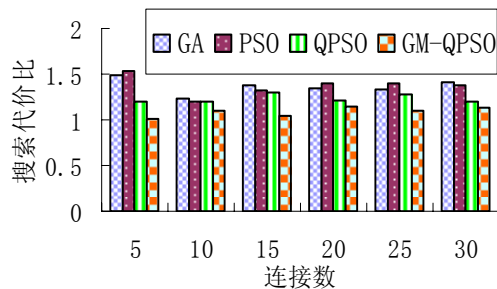


图 3 各种算法的搜索代价消耗比

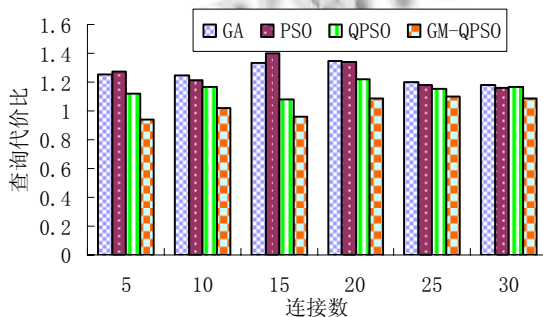


图 4 各种算法的查询消耗代价比

1) 当数据库查询连接数较少时, 所有算法查询和执行代价均较小, 查询效率相当高的, 获得理想的查询结果, 因此, 它们性能几乎没有什么差异, 均可以应用于较小的数据库查询连接数查询。

2) 连接数大于 20 时, 相对于 GA、PSO 算法, QPSO、GM-QPSO 算法优势得到体现, 这主要由于 QPSO、GM-QPSO 具有较好的全局和局部搜索能力, 克服了传统群智能算法的存在不足, 可以算法快速、准确找到数据库查询, 可以满足现代大规模数据库在线查询优化要求。

3) 相对于 QPSO 算法, GM-QPSO 算法降低了数据库查询和执行代价, 这主要因为, GM-QPSO 引入了遗传算法的变异算子, 高斯变异产生微小随机数的几率比较大, 使得粒子的位置在近似最优解的小范围内进行变动, 以提高算法的全局搜索能力, 较好的克服 QPSO 算法存在的局部最优缺陷, 获得了更优

的数据库查询方案。

#### 3.2.2 查询速度比较比较

为进一步分析 GM-QPSO 算法的优越性, 对它们的最优查询方案的迭代次数和运行时间进行仿真对比实验, 结果如图 5 和 6 所示. 从图 5 与图 6 可知, GM-QPSO 算法的性能明显优于对比算法, 迭代 280 次后, MG-QPSO 算法找到了更优的数据库查询方案, 其执行时间为 128.23ms, 远远低于 GA 的 200.24ms, PSO 算法的 195.23ms, QPSO 算法的 175.23ms, MG-QPSO 不仅迭代次数大幅度减少, 而且执行时间急剧下降, 获得更优的数据库查询方案。

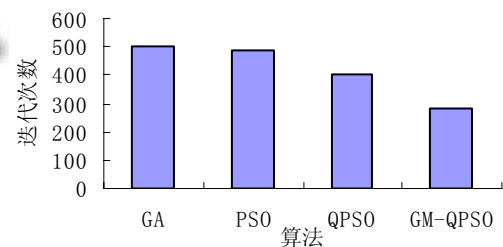


图 5 各种算法的迭代次数对比

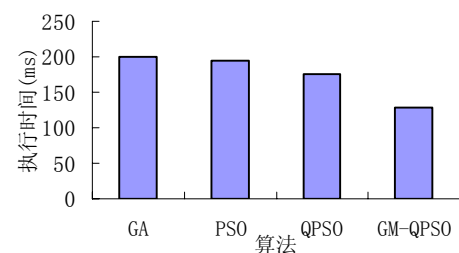


图 6 各算法的执行时间对比

### 4 结语

针对当前数据库查询优化算法存在的效率低, 难以获得最优解的缺陷, 提出一种高斯变异量子粒子群算法的数据库最优查询方法. 实验结果表明, GM-QPSO 引入了遗传算法的变异算子, 高斯变异产生微小随机数的几率比较大, 使得粒子的位置在近似最优解的小范围内进行变动, 以提高算法的全局搜索能力, 较好的克服 QPSO 算法存在的局部最优缺陷, 获得了更优的数据库查询方案, 在数据库中具有广泛的应用前景。

#### 参考文献

- 张敏, 冯登国, 徐震. 多级多版本数据库管理系统全局串行化. 软件学报, 2007, 18(2): 346-347.

- 2 徐增敏,张昆,丁勇,赵汝文.基于动态视图的数据库性能调优.计算机应用与软件,2012,29(12):58-60.
- 3 郭聪莉,朱莉,李向.基于蚁群算法的多连接查询优化方法.计算机工程,2009,35(10):173-176.
- 4 Kumar TVV, Singh V, Verma AK. Distributed query processing plans generation using genetic algorithm. International Journal of Computer Theory and Engineering, 2011, 3(1):38-45.
- 5 Chen PH, Shahandashti SM. Hybrid of genetic algorithm and simulated annealing for multiple project scheduling with multiple resource constraints. Automation in Construction, 2009, 18(4): 434-443.
- 6 林桂亚.基于粒子群算法的数据库查询优化.计算机应用研究,2012,29(3):974-975.
- 7 郑先锋,王丽艳.自适应逃逸动量粒子群算法的数据库多连接查询优化.四川大学学报(自然科学版),2013,50(3):100-103.
- 8 石伟.混沌粒子群算法在数据库查询优化中的应用.科技通报,2012,28(4):116-118.
- 9 Sun J, Fang W, Xu XJ, et al. Quantum-behaved particle swarm optimization: analysis of the individual particle's behavior and parameter selection. Evolutionary Computation, 2012, 20(3): 349-393.
- 10 张兰.量子粒子群算法及其应用[学位论文].西安:西北大学,2010.
- 11 刘俊芳,高岳林.带自适应变异的量子粒子群优化算法.计算机工程与应用,2011,47(3):41-43.
- 12 dos Santos Coelho L. Gaussian quantum-behaved particle swarm optimization approaches for constrained engineering design problems. Expert Systems with Applications, 2010, 37: 1676-1683.