

# 改进小生境遗传算法在元搜索引擎 调度优化中的研究\*

刘双印,徐龙琴,沈玉利

(广东海洋大学 信息学院,广东 湛江 524088)

**摘要** 针对多独立搜索引擎组合调度时查询精度、查询完全度和响应时间不理想等问题,结合元搜索引擎调度特点对多独立搜索引擎组合调度进行动态优化。文中借鉴小生境思想,将小生境技术与遗传算法相结合,提出了一种多目标组合优化调度的改进小生境遗传算法。该算法使每个个体在其小生境内进行局部寻优操作,保证了群体的多样性,增强了局部搜索能力,抑制了种群的早熟现象。在多个子目标不能同时达到最优时,采用个体综合适应度对各个目标函数的适应度进行加权,来协调优化各搜索引擎的组合,找到搜索引擎组合调度序列的非劣解。仿真实验结果表明该算法提高了元搜索引擎的调度效率,在查询精度和计算速度上均优于常用的查询优化技术。

**关键词** 多目标优化;最优解;小生境;遗传算法

中图分类号: TP311

文献标识码: A

文章编号: 1672-6693(2008)03-0046-05

面对众多用户的搜索请求,元搜索引擎本身不提供搜索信息的数据库,而是将其请求分发到实际的独立搜索引擎上,对独立搜索引擎返回的结果进行分析处理后再将搜索结果返回给用户。元搜索将整个搜索空间分为若干不同的领域,对于不同领域,各独立搜索引擎的性能有所不同,元搜索引擎根据用户的查询请求,调度若干个独立搜索引擎的组合来响应用户请求。对于任一领域的查询请求,元搜索引擎的目标是使得搜索该领域的信息性能最佳,这就要求所分派调度的独立搜索引擎组合是一个优化组合,最简单的调度方法就是通过人工模拟的方法来测试不同搜索引擎的组合,然后指定某一固定搜索引擎组合为某一领域的请求服务。但独立搜索引擎的数量、服务功能、性能也都在不断地变化,这种简单的搜索引擎调度策略不能反映系统的变化,特别是在查询精度、查询完全度和响应时间等综合方面查询效果不理想。

近年来,广泛应用于各种优化技术的全局优化搜索算法——遗传算法(Genetic Algorithm, GA)<sup>[1]</sup>被引入到信息检索领域,从而为信息查询优化技术提供了新的解决途径。尽管GA在信息检索中的应用取得了一定的效果<sup>[2]</sup>,但存在局部搜索能力不强、

易陷入局部最优和早熟等缺陷,使得传统的GA在进行查询优化时效果不理想。许多文献针对遗传算法的缺陷提出了各种改进方法,如引入爬山法、模拟退火法<sup>[2]</sup>、自适应遗传算法、实数编码遗传算法<sup>[3-4]</sup>,并获得了性能上的一定改善。与其它优化的遗传算法不同,本文结合元搜索引擎查询调度的特点,引入小生境的思想,提出了一种基于小生境的改进遗传算法,通过构造遗传算法的操作算子,有效提高局部搜索能力、克服早熟缺陷、加快收敛速度,以解决动态优化多独立搜索引擎组合调度问题。经仿真验证,该算法在查询精度和计算速度上均优于常用的查询优化技术。

## 1 小生境技术原理

在自然界中,“物以类聚,人以群分”的小生境现象普遍存在。生物总是倾向于与自己特性、形状相类似的生物生活在一起,交配繁殖后代。在生物学中,把某种特定环境及其在此环境中生存的组织称为小生境<sup>[5]</sup>。小生境的形成在生物学上是有积极意义的,它为新物种的形成提供了可能性,在小生境形成初期,小生境中的物种的基因常常是不同的,由于多个小生境间相对隔离,缺少必要的基因交流,使

\* 收稿日期 2007-11-09 修回日期 2008-02-20

资助项目 广东省粤港关键领域重点突破项目( No. 2006A25007002 ) 广东省自然科学基金项目( No. 7010116 )

作者简介 刘双印(1978-)男,讲师,硕士,研究方向为人工智能、分布式计算、智能信息系统等。

得这种基因差异得以保留。各小生境中生物的变异是随机发生的,因而通常有着不同的变异方向,这种变异的差异使物种间的基因差异不断扩大。由于地理位置和自然环境的不同,自然选择的方向和压力也各不相同,这一差异导致了物种间基因组成产生更大的差异,于是各物种向各自的方向发展进化,正因为如此自然界中生物保持近乎于无限多样性,使其充满了生机和活力<sup>[6]</sup>。

在具体的工程应用,小生境技术演变为:将每一代个体划分为若干类,每个类中选出若干适应度较大的个体作为一个类的优秀代表组成一个种群,再在种群以及不同种群之间通过杂交、变异产生新一代个体群,同时采用预选择机制、排挤机制或共享机制完成选择操作。也就是说让个体在一个特定的生存环境中进化,形成多个小生境,最终达到小生境内的峰值,从而找到全局最优解。受此启发,近年来人们将小生境现象引入到遗传算法中,实践证明,这一技术对于改善遗传算法全局收敛性能具有良好的效果<sup>[7]</sup>。

## 2 基于小生境的改进遗传算法

基本遗传算法在求解多独立搜索引擎组合调度动态优化问题时,往往只能找到几个局部最优解,而无法收敛到全局最优解。这是因为在标准的遗传算法的初期,群体保持了多样性,但是到了算法后期,群体的多样性遭到了破坏,大量个体集中于某一个极值点附近,它们的后代造成了近亲繁殖,这样就易造成收敛于一个局部最优解,而无法跳出该局部搜索<sup>[8]</sup>。

借鉴小生境思想,来改进传统遗传算法。生境是一个适应度函数值的集合,  $\{f_{ij}\}$  为一小生境  $i = 1, 2, \dots, n$   $j = 1, 2, \dots, size$   $n$  为适应度函数个数,  $size$  为种群规模。改进遗传算法是多个目标的预先寻优过程,因此它的设计关键是在进化过程中,要对各个目标的优良个体进行保留,如果有  $n$  个可能的目标函数  $f_1, \dots, f_n$ , 则只要一个个体对某一个函数较优,就可保留。假定有  $n$  个候选的目标函数,可以根据经验只是选择一个作为主目标函数,例如  $f_1$  为主目标函数,而  $f_2, \dots, f_n$  作为隐含目标函数。

当前种群中的个体对每一个目标函数都有相应的目标函数值,在  $f_1$  进化时,通过正常的选择、交叉和变异操作形成新个体集合。新个体对每一个目标函数也有不同的适应度和函数值,在进行用新个体

替换原来种群中的部分个体时,就不能仅仅按照对  $f_1$  的适应度来考虑,要考虑其所有目标函数的适应度。要替换掉那些对所有函数都不优的个体,而不仅仅是替换掉所有对  $f_1$  不优的个体。对此,设置一个综合适应度,这个综合适应度是个体对各个目标函数的适应度的加权。

基于小生境改进的遗传算法描述如下。

Step1 定义遗传算法运行需要的参数,如种群规模、染色体长度、选择概率、交叉概率、变异概率、目标函数个数等;

Step2 确定一个主目标函数  $f_1$ ;

Step3 确定编码方式并初始化种群  $G$ ;

Step4 计算初始种群  $G$  中个体对所有目标函数的目标函数值  $obj_1, obj_2, \dots, obj_n$ ;

Step5 计算初始种群  $G$  中个体的综合加权适应度  $OBJ$ ;

Step6 While( 不满足迭代结束条件 )

1) 计算个体对主目标函数的适应度;

2) 应用选择算子根据个体对主目标函数适应度选择若干个体构成一个新的种群  $G'$ ;

3) 对新的种群  $G'$  进行交叉操作;

4) 对新的种群  $G'$  进行变异操作;

5) 计算新种群中个体对所有目标函数的函数值  $obj_1', obj_2', \dots, obj_n'$ ;

6) 计算新种群中个体的综合加权适应度  $OBJ'$ ;

7) 根据原种群  $G$  对所有目标函数的综合适应度  $OBJ$  和新种群  $G'$  对所有函数的综合适应度  $OBJ'$ , 用新种群中的个体替换原种群中综合适应度低的个体,形成下一代种群  $G_1$ ;

8) 目标函数的滚动,从隐函数中选择一个目标函数为当前的主目标函数;

9) 计算7)形成的下一代种群  $G_1$  中个体对所有目标函数的适应度值和综合适应度值,并令  $G = G_1$ 。如不满足迭代结束条件则继续迭代;

Step7 End While;

Step8 染色体解码。

基于小生境改进的遗传算法是为了目标函数寻优算法能很快收敛,因此要保证保留所有目标函数的优良个体,并且保持群体的多样性。

## 3 改进遗传算法在元搜索引擎调度中的应用

元搜索引擎的调度问题是一个多目标优化问

题(Multi-objective Optimization Problem, MOP)。对于某一领域的查询请求,元搜索引擎调度的独立搜索引擎组合有3个子目标:搜索精度、搜索完全度、平均响应时间。这3个子目标存在着冲突,例如提高搜索完全度,可能会引起不相关文档的附带率上升,则搜索精度下降,若提高搜索精度和搜索完全度,平均响应时间必然增加。提高其中一个子目标函数的结果,则必然会降低另外一个或几个子目标函数的结果,也就是说要同时使这多个子目标都一起达到最优是不可能的,而是只能在它们中间进行协调和折中处理。

### 3.1 编码方法及初始化种群

传统的二进制编码存在解码技术问题,更为重要的是二进制编码还存在计算精度与计算速度之间两难选择的问题。十进制浮点数编码可以避免数制的转换、提高计算精度和速度。在多参数优化的问题中为了计算精度和使用方便,通常采用十进制浮点数编码。假设有  $n$  个独立搜索引擎,则  $n$  个搜索引擎组合的十进制浮点数排列在一起成为一个个体,如  $x = x_1 \ x_2 \ \dots \ x_n$ 。

### 3.2 适应度函数的选择

在元搜索引擎中,每次用户提交查询串后,检索接口调用若干独立搜索引擎组成一个搜索引擎序列,其性能指标包含多个方面,如搜索精度  $p$ ,搜索完全度  $c$ ,平均响应时间  $t$ 。

搜索引擎的搜索空间  $\mathbf{R}$  表示为  $\{r_1 \ r_2 \ \dots \ r_n\}$ ,其中  $r_1 \ r_2 \ \dots \ r_n$  分别表示每个领域。

定义1 在搜索空间  $\mathbf{R}$  中,领域  $r_i$  在搜索引擎  $E$  上所占的权重为  $w_i$ ,可表示为二元组  $(r_i \ w_i)$ ,所有二元组的集合叫做搜索空间  $\mathbf{R}$  在搜索引擎  $E$  上的领域分布集,用  $f(E)$  表示。即  $f(E) = \{(r_i \ w_i) | r_i \in \mathbf{R} \ 0 \leq w_i \leq 1 \ j = 1 \ 2 \ \dots \ n, \sum_{i=1}^n w_i = 1\}$ 。

定义2 在搜索空间  $\mathbf{R}$  中,领域  $r$  在搜索引擎  $E$  上搜索精度  $precision$  定义为在  $r$  上结果集中的相关文档数与结果集总文档数之比,用二元组  $(r, precision)$  表示,所有这种二元组的集合叫做搜索空间  $T$  在搜索引擎  $E$  上的搜索精度,用  $\mu(E)$  表示。即  $\mu(E) = \{(r \ precision) | r \in \mathbf{R} \ 0 \leq precision \leq 1\}$ 。

定义3 在搜索空间  $\mathbf{R}$  中,领域  $r$  在搜索引擎  $E$  上搜索完全度  $completeness$  定义为在  $r$  上搜索到的信息条数与实际相关信息条数之比,用二元组  $(r, completeness)$  表示,所有二元组的集合叫做搜索空间  $\mathbf{R}$  在搜索引擎  $E$  上的搜索完全度,用  $\rho(E)$

表示。即  $\rho(E) = \{(r \ completeness) | r \in \mathbf{R} \ 0 \leq completeness \leq 1\}$ 。

改进遗传算法中一个搜索引擎调用序列可以看作是一个个体,记作  $A_i$ ,一个搜索引擎序列  $A_i = (E_{i1} \ E_{i2} \ \dots \ E_{in})$ ,每个独立搜索引擎可以看作个体的基因。元搜索引擎在各领域的评价指标包括:搜索精度、搜索完全度、平均响应时间。分别表示如下。

$$A_i.p = \sum_{E \in agent_i} p(E)$$

$$A_i.c = \sum_{E \in agent_i} \rho(E)$$

$$A_i.t = \bar{T}(E_{i1} \ E_{i2} \ \dots \ E_{in})$$

其中  $\bar{T}(E_{i1} \ E_{i2} \ \dots \ E_{in})$  为各独立搜索引擎响应时间的平均值。以上3式可作为搜索精度、搜索完全度、平均响应时间3个目标函数的适应度函数。

### 3.3 元搜索引擎调度协调和折中处理

元搜索引擎评价参数值并不是始终不变,而是随着系统的运行、搜索事务的处理而动态变化的。在系统中对每次用户的搜索请求、搜索请求所在领域、返回的结果集以及用户对结果的选择都会以日志的形式保存下来。从日志中就可以统计各个搜索引擎中各种领域的分布,搜索精度,搜索完全度,以及平均响应时间。

元搜索引擎智能化调度中由  $A_i$  协调和折中处理的搜索引擎组成,它的性能参数可以用四元组  $(f, p, c, t)$  来表示。 $A_i$  对于所有领域的分布集  $f$  可表示成构成该个体的搜索引擎的相应各领域权重的平均值组成的集合,即

$$f(A_i) = \{(r_i \ w_i) | w_i = AVG_{0 \leq i \leq \lambda}(w_i) \ r_i \in \mathbf{R}$$

$$0 \leq w_i \leq 1 \ j = 1 \ 2 \ \dots \ n, \sum_{i=1}^n w_i = 1\}$$

个体的搜索精度  $p$  可表示成构成该个体的搜索引擎的相应各领域搜索精度的平均值组成的集合,即

$$\mu(A_i) = \{(r_i \ p_i) | p_i = AVG_{0 \leq i \leq \lambda}(p_i) \ r_i \in \mathbf{R}$$

$$0 \leq p_i \leq 1 \ j = 1 \ 2 \ \dots \ n\}$$

个体的搜索完全度  $c$  可表示成构成该个体的搜索引擎的相应各领域搜索完全度的平均值组成的集合,即

$$\rho(A_i) = \{(r_i \ c_i) | c_i = AVG_{0 \leq i \leq \lambda}(c_i) \ r_i \in \mathbf{R}$$

$$0 \leq c_i \leq 1 \ j = 1 \ 2 \ \dots \ n\}$$

由于个体需要从每个搜索引擎得到搜索结果并进行处理后才返回给用户,因此

$$f(A_i) = \max(\mu(E_{i1}), \dots, \mu(E_{in})) + \bar{T}(E_{i1} \ E_{i2} \ \dots \ E_{in})$$

其中  $T(E_{i1}, E_{i2}, \dots, E_{in})$  为个体处理搜索引擎  $E_{i1}, E_{i2}, \dots, E_{in}$  返回结果所用的平均时间。

将个体  $A_i$  的综合适应度函数设置为  $A_i.obj = K(f, p, c, t)$  其中  $K$  为比例变换函数  $f, p, c, t$  为个体  $A_i$  上的性能指标。改进的遗传算法通过对初始化种群的重组、变异,产生新一代个体种群,最后笔者选择  $A_i.obj$  最大值的  $A_i^*$  作为该领域  $r_i$  的个体,负责该领域的搜索请求。

### 3.4 交叉算子

在自然界生物进化过程中起核心作用的是生物遗传基因的重组(交叉)和变异。同样,遗传算法中起核心作用的是遗传操作的交叉算子。所谓交叉是指把两个父代个体的部分结构加以替换重组而生成新个体的操作。通过交叉,遗传算法的搜索能力得以飞跃提高。

交叉以概率  $P_c$  (通常  $P_c$  取值范围[0.6, 1.0]) 发生。交叉算子采用单点交叉。具体操作是:在个体串中随机设定一个交叉点,实行交叉时,该点前或后的两个个体的部分结构进行互换,并生成两个新个体。交叉点时随机设定的,当染色体长为  $l$  时,可能有  $l-1$  个交叉点设置,所以,单点交叉可能实现  $l-1$  个不同的交叉结果。

### 3.5 变异算子

遗传算法引入变异的目的是有两个:①使遗传算法具有局部的随机搜索能力。当遗传算法通过交叉算子已接近最优解邻域时,利用变异算子的这种局部随机搜索能力可以加速向最优解收敛。显然,此种情况下的变异概率应取较小值,否则接近最优解的模式串会因变异而遭到破坏。②使遗传算法可维持群体多样性,以防止出现未成熟收敛现象。此时收敛概率应取较大值。

遗传算法中,交叉算子因其全局搜索能力而作为主要算子,变异算子因其局部搜索能力而作为辅助算子。遗传算法通过交叉和变异这一对相互配合又相互竞争的操作而使其具备兼顾全局和局部的均衡搜索能力。所谓相互配合,是指当群体在进化中陷于搜索空间中某个超平面而仅靠交叉不能摆脱时,通过小生境变异操作可有助于形成所期望的模式串,变异操作有可能破坏原有的模式串。

变异操作发生的概率  $P_m$  取得比较小,通常为 1% 左右,但足以保持种群内个体的多样化。变异算子的基本内容是对群体中的个体串的某些基因座上的基因值作变动。本文所采用的变异算子基本步骤

如下。

1) 在群体中所有个体的编码串范围内随机地确定基因座。

2) 以事先设定的变异概率  $P_m$  来对这些基因座的基因值进行变异。

### 3.6 算法描述

Step1 设置种群规模  $size$ , 交叉概率  $P_c$ , 变异概率  $P_m$ , 目标函数集  $f = \{A_i.p, A_i.c, A_i.t\}$ , 目标函数标记  $n = 1$ , 迭代次数为  $epoch$ , 精度为  $\varepsilon$ ;

Step2 设置搜索精度  $f(n)$  为当前目标函数;

Step3 初始化种群  $G$ ;

Step4 While( 迭代次数  $m \leq epoch$  or  $d < \varepsilon$  )

1) 计算初始种群  $G$  中各个个体的目标函数值  $A_i.p, A_i.c, A_i.t, i = 1, 2, \dots, size$ ;

2) 计算初始种群  $G$  中各个个体的综合加权适应度  $A_i.obj, i = 1, 2, \dots, size$ ;

3) 根据“主目标函数适应度”选择若干个体构成一个新的种群  $G'$ ;

4) Crossover(  $G'$  );

// 利用单点交叉函数对新种群进行交叉操作

5) Mutation(  $G'$  );

// 利用变异函数对新种群进行变异操作

6) 计算新种群  $G'$  中各个个体的目标函数值  $A_i.p', A_i.c', \dots, A_i.t', i = 1, 2, \dots, size$ ;

7) 计算新种群  $G'$  中个体的综合加权适应度  $A_i.obj, i = 1, 2, \dots, size$ ;

8) 根据原种群  $G$  对所有目标函数的综合适应度  $A_i.obj$  和新种群  $G'$  对所有函数的综合适应度  $A_i'.obj$ , 用新种群  $G'$  中的个体替换原种群中综合适应度低的个体。形成下一代种群  $G_1$ ;

9) 计算  $d_o, d$  为种群  $G$  的最优个体的综合适应度与种群  $G'$  的最优个体的综合适应度之差;

10) 目标函数的改变  $n = n + 1$ , 设置  $f(n)$  为当前目标函数;

11) Set  $G = G_1$ ;

Step5 End While.

### 3.7 实验结果分析

从搜索引擎超市(<http://searchenginewatch.com/>)获得 10 个独立搜索引擎针对“数码电子”领域, 查询串为“GPRS(General Packet Radio Service)”、“军事”领域, 查询串为“terrorism”时各独立搜索引擎的搜索精度  $P$ 、搜索完全度  $C$ 、平均响应时间  $T$  的数据, 如表 1、表 2 所示。

表 1 “数码电子”领域各独立搜索引擎的性能指标

搜索引擎 编号	1	2	3	4	5	6	7	8	9	10
<i>P</i>	0.4	0.4	0.3	0.4	0.2	0.1	0.5	0.3	0.4	0.2
<i>C</i>	0.5	0.6	0.7	0.7	0.6	0.5	0.8	0.7	0.9	0.6
<i>T</i>	5	4	3	2	4	6	3	2	4	3

表 2 “军事”领域各独立搜索引擎的性能指标

搜索引擎 编号	1	2	3	4	5	6	7	8	9	10
<i>P</i>	0.4	0.5	0.3	0.2	0.2	0.7	0.2	0.1	0.5	0.3
<i>C</i>	0.5	0.7	0.5	0.6	0.6	0.9	0.4	0.3	0.8	0.7
<i>T</i>	4	4	2	3	4	8	3	4	7	3

采用基于小生境的改进遗传算法和标准遗传算法分别实现对这两个领域的搜索引擎调度,在独立搜索引擎个数为 10 个时,可通过穷举法求出所有非劣解。对于改进遗传算法和标准遗传算法,种群规模 *size* 均为 100,考察每个解,如果它属于当前数据的 Pareto 最优解集,则为非劣解。非劣解个数越多说明该搜索引擎组合精度和搜索完全度越好,进化代数越少该搜索引擎组合平均响应时间越短。实验结果如图 1、2 所示。

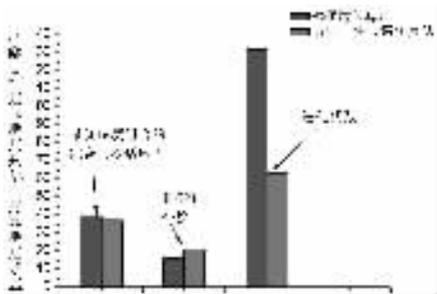


图 1 “数码电子”标准遗传算法与改进遗传算法调度结果对比

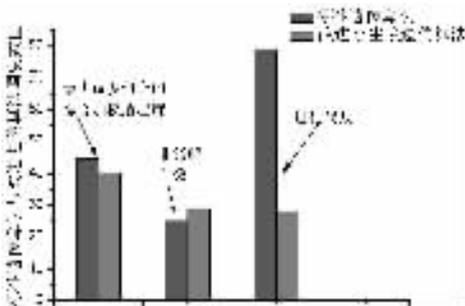


图 2 “军事领域”标准遗传算法与改进遗传算法调度结果对比

代数远远少于标准遗传算法,调度所用的时间大大减少,非劣解个数多于标准遗传算法,其搜索精度和完全度也优于标准遗传算法的原搜索引擎的组合调度,基本能满足实际的元搜索引擎调度时间需求。在调度结果上不一定能找到最优的调度组合,但是能在较短时间内找到搜索精度、搜索完全度、平均响应时间 3 方面均较优的独立搜索引擎组合。综合来看,改进遗传算法提高了元搜索引擎的调度效率。

### 4 结语

本文根据元搜索引擎调度特点,在求解多独立搜索引擎组合调度动态优化问题时借鉴小生境思想,提出了一种改进小生境遗传算法,并给出了算法的实施步骤。该算法使每个个体在其小生境内进行局部寻优操作,保证了群体的多样性,增强了局部搜索能力,抑制了种群的早熟现象,并且在元搜索引擎组合调度应用。实验结果表明该算法在查询精度、查询全面和平均响应时间 3 方面均优于常见独立搜索引擎组合,综合来说它提高了元搜索引擎的调度效率。

### 参考文献:

- [1] 徐宝文, 张卫丰. 搜索引擎与信息获取技术[M]. 北京: 清华大学出版社, 2003.
- [2] 王礼刚, 左源瑞, 李盛瑜. 一种基于改进型遗传算法的关联规则提取算法及其应用[J]. 重庆师范大学学报(自然科学版), 2006, 23(2): 42-45.
- [3] 罗方芳, 陈国龙. 基于遗传算法求解带约束的最小生成树的研究[J]. 计算机科学, 2003, 30(8): 166-168.
- [4] TANG K W, KAK S. Fast Classification Networks for Signal Processing[J]. Circuits Systems Signal Processing, 2002, 21(2): 207-224.
- [5] JELASITY M, DOMBI T. GAS: A Concept on Modeling Species in Genetic Algorithm[J]. Artificial Intelligence, 1998, 99(1): 1-19.
- [6] 陈国良. 遗传算法及应用[M]. 北京: 人民邮电出版社, 1996.
- [7] 王小平, 曹立明. 遗传算法-理论、应用于软件实现[M]. 西安: 西安交通大学出版社, 2002.
- [8] 罗方芳. 智能化搜索引擎关键技术研究[D]. 福州: 福州大学计算机系, 2005. 12.
- [9] 周洪伟, 徐松林, 徐静. 改进的小生境遗传算法[J]. 微计算机信息, 2007, 23(6): 208-209.
- [10] 周宇恒, 王允建. 基于小生境的开放式遗传算法[J]. 计算机应用, 2007, 27(4): 960-965.

分析仿真实验结果得知,改进遗传算法的进化

- [ 11 ] 王淑佩, 林亚平, 易叶青. 一种新的基于小生境的自适应遗传算法[ J ]. 兰州理工大学学报, 2006, 23( 5 ) : 83-87.
- [ 12 ] 赵正文, 康耀红, 方磊坤, 等. 信息检索中的遗传算法应用研究[ J ]. 郑州大学学报( 理学版 ), 2006, 38( 4 ) : 64-68.
- [ 12 ] 凌霄汉, 吉根林. 一种快速选择性神经网络集成方法研究[ J ]. 郑州大学学报( 理学版 ), 2006, 38( 4 ) : 69-73.

## Research into the Application of Improved Genetic Algorithm Based on Niches in Search Engine Optimization

*LIU Shuang-yin , XU Long-qin , SHEN Yu-li*

( Information Institute of Guangdong Ocean University , Zhanjiang Guangdong 524088 , China )

**Abstract** Based on the inquiry accuracy , inquiry completion and the response time of the combination and operation optimization of the multi-independent research engine , combined with the feature of the research engine operation , the multi-independent research engine is put in to dynamic optimization. Borrowed the thought of niches , an improved niche genetic algorithm of the multi-objective assembled optimized operation is put forward by combining the technology of niches and the genetic algorithm. This algorithm enables each individual to have local optimization-oriented operation , ensures the diversity of the group , strengthens the capability of local research and restrains the prematurity. When many subobjects cannot reach optimization at the same time , this algorithm , in use of individual comprehensive sufficiency can enforce weighting each objective function harmonize and optimize each research engine combination in order to fine a good solution to the combination and operation sequence of research engine. Results of the experiments prove that this algorithm can improve the efficiency of research engine operation and it surpasses the common inquiry technique in the inquiry accuracy and calculation speed.

**Key words** multi-objective optimization ; optimization solution ; niches ; genetic algorithm

( 责任编辑 游中胜 )