

交叉点规模的优化与交叉算子性能的改进*

陈皓¹⁺, 崔杜武¹, 李雪^{1,2}, 韦宏利^{1,3}

¹(西安理工大学 计算机学院, 陕西 西安 710048)

²(陕西师范大学 国际商学院, 陕西 西安 710062)

³(西安工业大学 电子信息工程学院, 陕西 西安 710032)

Optimization Crossover Scale for Improving Performance of Crossover Operator

CHEN Hao¹⁺, CUI Du-Wu¹, LI Xue^{1,2}, WEI Hong-Li^{1,3}

¹(School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China)

²(International Business School of Shaanxi Normal University, Xi'an 710062, China)

³(School of Electronic Information Engineering, Xi'an Technological University, Xi'an 710032, China)

+ Corresponding author: E-mail: chen hao_xaut@yahoo.com.cn

Chen H, Cui DW, Li X, Wei HL. Optimization crossover scale for improving performance of crossover operator. Journal of Software, 2009,20(4):890-901. <http://www.jos.org.cn/1000-9825/3328.htm>

Abstract: Based on the analysis of relationship between the crossover scale and reachable subspace of crossover operator, it can be found that the crossover scale is dynamically adjusted to the population structure. In this paper, three control mechanisms—the well-phased control strategy, the random distribution strategy and the adaptation evolution strategy are built up to adjust the crossover scale. The simulation tests of the classical function show these optimization mechanisms are available and valuable control knowledge of crossover scale for multi-dimension functions have been generated by the adaptation evolution strategy. Furthermore, this research suggests a new method for the operator and parameter optimization of evolution algorithm.

Key words: genetic algorithm; crossover operator; optimization crossover scale

摘要: 通过分析交叉点规模对交叉算子空间搜索性能的影响,可以发现在遗传算法的搜索过程中,其对交叉点规模的需求是随群体状态的演变而动态变化的.为实现对交叉点规模的优化,提出使用分阶段调整策略、随机分配策略以及自适应进化策略 3 种方法来完成对交叉点规模的动态调控.对典型高维函数的优化实验表明,上述方法可以显著提高交叉操作的搜索效率,其中,自适应进化策略利用搜索机制可以发现一类高维函数交叉点规模的控制知识,实验结果证实了此类知识的有效性.此外,该研究也为对进化算法中算子和参数的优化提供了新思路.

关键词: 遗传算法;交叉算子;优化交叉点规模

中图法分类号: TP18 文献标识码: A

在遗传算法(genetic algorithm,简称 GA)中,交叉算子通过模拟生物有性繁殖过程生成新个体投入优胜劣汰

* Supported by the National Natural Science Foundation of China under Grant Nos.60743009, 60501006 (国家自然科学基金); the Natural Science Foundation of Shaanxi Province of China under Grant No.2006F-43 (陕西省自然科学基金)

Received 2007-12-27; Accepted 2008-03-31

的选择竞争,可以说交叉算子在 GA 对待求解问题空间进行搜索并达到全局最优的过程中发挥着至关重要的作用.按编码方式划分,有典型的二进制交叉、实数重组、树形结构染色体位串交叉以及基于自然数编码的部分匹配交叉(PMX)、顺序交叉(OX)和周期交叉(CX)等形式.而就交叉操作所涉及的父辈数量而言,有两父辈交叉和多父辈交叉(multi-parent recombination)等形式.

在交叉算子的理论研究方面,Beyer^[1]针对球函数分析了全局重组算子的收敛速度,章珂等人^[2]针对二进制编码的遗传算法分析了交叉位置的选取对搜索性能的影响,张文修等人^[3]分析了交叉算子的几何性质.任庆生^[4]证明了交叉算子在演化代数增加的时候能够使模式内部各基因趋于独立,并且只要组成模式的各个基因都存在,则该模式一定能够被搜索到.在此基础上,任庆生等人进一步证明了交叉算子的极限一致性^[5],即在演化代数趋于无穷时,交叉算子对染色体的作用结果是一致的.近年来,研究者也从不同角度对交叉算子进行了改进,文献[6]提出通过约束交叉位置和保留全局公共模式来解决标准交叉算子容易破坏高阶、长而好的模式及其在相似个体之间低效的问题,文献[7-9]提出了几种选择交叉父个体和交叉概率的机制,文献[10]提出了一种交叉后代个体对父代个体的替换策略.此外,结合其他技术和方法也产生了一些独特的交叉操作^[11,12].但迄今为止,交叉算子的作用机制和理论分析仍然是进化算法理论研究的薄弱环节之一.

常见的一点交叉、两点交叉、多点交叉和一致交叉几种算子都属于使用二进制编码的两父辈等位交叉操作,此类算子的特性主要由其所具有的交叉点规模以及交叉位间基因段的交换概率所决定.那么,为了提高这类交叉算子的搜索性能,应该如何对交叉点规模进行控制呢?本文分析了交叉点规模对交叉操作搜索效率的影响,并针对交叉点规模的优化问题提出了分阶段调整策略、随机分配策略以及自适应进化策略 3 种调控方法.对典型多维函数的优化实验表明,这些方法是有效的,其中,自适应进化策略可以利用随机搜索机制对交叉点规模的演化规律进行搜索和分析,其统计结果可以作为被优化函数的一类控制知识,实验结果证实此类知识是有价值的,能够有效地辅助交叉操作提高搜索效率.此外,该研究也为对进化算法中算子和控制参数的优化方法提供了新思路.

1 交叉点规模对交叉算子搜索性能的影响

1.1 具有不同交叉点规模的交叉算子对高维函数的优化效果对比

在二进制编码的 GA 中,个体的位串长度为 L ,交叉点等概率选取,可选范围为 $[1, L-1]$.一点交叉的交叉位设置数量为 $L-1$,考虑最右端交叉点可以落在位串之外, k 点交叉的交叉位设置数量为 C_L^k .随着交叉点规模 k 的增加,其设置数量符合以 $L/2$ 为均值的二项分配,而均匀交叉可视为多点交叉的特例.那么,交叉点规模对交叉算子的搜索性能有何影响呢?下面通过对一个高维函数的优化实验来比较不同交叉规模对标准遗传算法(canonical genetic algorithm,简称 CGA)搜索性能的影响.

Generalized Schwefel's Problem 2.26:

$$f_1(x) = \sum_{i=1}^m (-x_i \sin \sqrt{|x_i|}), x_i \in [-500, 500]^m.$$

在 CGA 中分别使用 8,15,20 点交叉算子和均匀交叉算子以及变异算子和轮盘赌选择算子对 20 维的 f_1 函数连续进行 20 次运算.CGA 的群体规模为 40,交叉概率为 0.7,变异概率为 0.02,优化终止代数为 2 000 代,每维的编码长度为 17.表 1 为实验统计结果,图 1 为使用不同交叉算子的 CGA 在 20 次运算中最接近平均结果的一次运算的函数值优化曲线.在这一系列实验中,15 点交叉算子取得了最佳的优化效果,函数值优化曲线也显示,15 点交叉算子的搜索速度要明显快于 8 点和 20 点交叉算子以及均匀交叉算子.此实验说明,交叉点规模直接影响着交叉算子的搜索效果,但很难直接

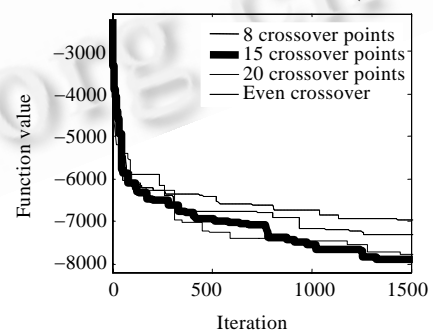


Fig.1 Optimization curves of f_1 by CGA with four fixed crossover

图 1 使用 4 种固定规模交叉算子的 CGA 对 f_1 的优化曲线

确定最佳的交叉点规模.为了寻找有效的交叉点规模的选择方法,有必要对交叉点规模与交叉操作搜索性能之间的关系加以分析.

Table 1 A comparison of optimization statistic results of f_1 by CGA with four fixed crossover operator

表 1 使用 4 种固定规模交叉算子的 CGA 对 f_1 的优化统计结果对比

Crossover scale	8 crossover points	15 crossover points	20 crossover points	Even crossover
Mean function value	-7 049.7	-7 902.2	-7 720.8	-7 669.7
Standard deviations	1 696.1	1 646.4	1 702.2	1 363.2

1.2 任意 k 点交叉算子的空间搜索能力分析

规模为 2^L 的二进制编码的解空间是一个 L 维超立方体 M , 其中任一顶点 $v \in M$ 都是解空间中的一个个体, 顶点 v_1, v_2 间的编码差异值可用海明距离 $H(v_1, v_2)$ 表示.

定义 1. 设 $H(v_1, v_2) = d$, 顶点 v_1, v_2 通过基因位的交换所能达到的所有顶点构成了一个待搜索空间, 其规模为 $2^{d[13]}$; 经任意 k 点交叉运算可能搜索到的所有顶点构成了该算子的可搜索空间 Φ , 其规模表示为 $|\Phi(k)|$.

Φ 反映了交叉算子的空间搜索范围, 其规模越大, 说明交叉算子具有越广的空间搜索能力. 已有工作表明, 一点交叉的可搜索空间规模 $|\Phi(1)| = 2d^{[13, 14]}$, 两点交叉的可搜索空间规模 $|\Phi(2)| = d^2 - d + 2^{[13]}$, 一致交叉的可搜索空间规模为 $2^{d[14]}$, 下文将对任意 k 点交叉算子的空间搜索能力进行分析.

显然, v_1, v_2 间相同基因位的交换对新顶点的发现没有实际作用, 有效的交叉操作依靠的是相异基因位的交换. 由于相异基因位数限制了可能用于交换的基因段的数量, 而交叉点的规模决定了实际可产生交换的基因段数量, 因此, 这两个因素直接影响着可产生的相异基因位交换组合类型的数量.

定理 1. 设 $H(v_1, v_2) = d$, 则 v_1, v_2 间进行一次有效交叉, 最大可交换的基因段规模为

$$\Theta(v_1, v_2) = \lfloor d/2 \rfloor \quad (\lfloor \cdot \rfloor \text{表示取整操作}) \quad (1)$$

证明: 显然, 一个基因段的最小单位是一个基因位, 任意连续基因位的交换都可等价为一个基因段的交换, 只有相隔至少一个相异位的交叉操作需要至少两次基因段交换. 当 d 为偶数时, 以一个基因位为交换单位且只相隔一个相异位的基因段最多有 $d/2$ 个; 当 d 为奇数时, 交换 $(d+1)/2$ 个基因段的效果可等效于交换 $\lfloor d/2 \rfloor$ 个基因段. 因此, $\Theta(v_1, v_2) = \lfloor d/2 \rfloor$. 证毕. \square

定理 2. 设 $H(v_1, v_2) = d$, $\Theta(v_1, v_2) = t$, 则对于任意 k 点交叉算子, 当 $\lfloor k/2 \rfloor \geq t$ 时, k 点交叉的可搜索空间规模为 $|\Phi(k)| = 2^d$; 当 $\lfloor k/2 \rfloor < t$ 时, 若 k'_{even} 是小于 k 的最大偶数且 $|\Phi(k'_{\text{even}})| = 2m$, 即由该算子可达到 m 对子顶点且达到第 j 对子顶点时, 最后一个交叉点在 d 个相异基因位中的位置为 $q_j \in [1, d]$, 则 k 点交叉的可搜索空间规模为

$$|\Phi(k)| = |\Phi(k'_{\text{even}})| + 2 \sum_{j=1}^m \psi_j \quad (2)$$

其中, $2 \sum_{j=1}^m \psi_j$ 表示可新达到的顶点数量. 若 k 为偶数, 当 $q_j < d-1$ 时, $\psi_j = C_{d-q_j}^2$, 其余情况下不会产生新顶点 $\psi_j = 0$; 若 k 为奇数, 当 $q_j < d-1$ 时, $\psi_j = d - q_j - 1$, 其余情况下, $\psi_j = 0$.

证明: 令 $i = \lfloor k/2 \rfloor$, 显然, 当 $i \geq t$ 时, d 个相异基因位间的任意基因交换组合都是可实现的, 因此 v_1, v_2 构成的待搜索空间都是可达的, 所以, $|\Phi(k)| = 2^d$; 当 $i < t$ 时, 进行 $i+1$ 个以上基因段的交换组合是不可能实现的, 因此交叉操作的可搜索空间是待搜索空间的一个子集.

考虑只有部分交叉点在相异基因位间以及交换基因段可能只包含相同基因位的情况, 则任意较小规模交叉操作的可搜索空间都是较大规模交叉操作可搜索空间的一个子集. 若 k 为偶数, 则交叉操作最多可完成 i 个基因段的交换, 新增一次基因段交换达到的新顶点都是以上个基因段交换达到的顶点为出发点. 因此, i 个基因段的交换与连续 i 次交换基因位不产生相交的两点交叉操作等价. 若 $q_j \geq d-1$, 则再新增交叉点不会产生相异基因位的交换, 因此可新达到的顶点数 $\psi_j = 0$; 反之, 通过 $[q_j+1, d]$ 间的两点交叉操作能够新达到的顶点数量为 $\psi_j = C_{d-q_j}^2$.

因此, 当 k'_{even} 新增两个交叉点时, 由 m 对子顶点可新达到的所有顶点的数量为 $2 \sum_{j=1}^m \psi_j$. 当 k 为奇数时, 其交叉过程

等价于交换基因位不产生相交的 i 次两点交叉与一次单点交叉.与偶数点交叉不同之处在于,通过 $[q_j+1,d]$ 间一点交叉操作可新达到的顶点数量为 $\psi_j=d-q_j-1$.综上所述,定理 2 成立.证毕. \square

进行交叉操作是为了在可搜索空间中发现更有价值的顶点.定理 2 表明,交叉点规模越大,产生的相异基因位交换组合类型越多,则其可搜索的范围也越广.但是,随着 k 的增大, $d-q_j$ 的值必然减小,这意味着在增加交叉点规模的同时,可搜索范围获得的增长反而在减小.当 $\lfloor k/2 \rfloor \geq i$ 时,再增加交叉点规模将不会扩大交叉操作的搜索范围.而且交叉操作可搜索空间范围越大,也就意味着此空间中任一顶点被搜索到的概率越小.所以,过大的交叉点规模反而会影响有价值顶点被搜索到的概率.

在实际运算中,相异基因位总是相邻了一定的相同基因位,此时并非所有的交叉点都能产生新个体,交叉操作实际搜索的空间仅是其可搜索空间的一个子集 $A \subseteq \Phi$.显然,当交叉点规模不同时, A 接近 Φ 的概率是不一样的,其极端情况是交叉操作没有搜索到任何新个体,即 A 中只有两个父个体.

定理 3. 设 $H(v_1, v_2)=d, v_1, v_2$ 的交叉概率为 P_c , 第 k 个交叉点在位串中的位置为 l_k , 且 k 个交叉点都不重合, 则任意 k 点交叉可搜索到新顶点的概率下限为

$$P_{new} \geq \begin{cases} \left[1 - \frac{(L-d-1)!}{(L-1)^k (L-d-k-1)!} \right] P_c, & k = 2i \\ \left[1 - \frac{(L-d-1)! l_k!}{(L-1)^{k+d-1} (L-d-k)! (l_k-d)!} \right] P_c, & k = 2i-1 \end{cases}, i = 1, 2, \dots \quad (3)$$

证明:当交叉操作产生交换的基因位都是相同位,或所有的相异位同时都被交换时,交叉操作无法达到新的顶点,相异位同时都被交换的概率可忽略.当 k 为偶数时,交换的基因位数至少为 k , 则被交换的基因位都是相同位的上限概率为 $\prod_{j=0}^{k-1} \frac{L-d-1-j}{L-1} = \frac{(L-d-1)!}{(L-1)^k (L-d-k-1)!}$; 当 k 为奇数时,其结果等效为 $k-1$ 点偶数交叉与一个一

点交叉的叠加,第 k 个交叉点产生的交换基因段 $L-l_k$ 中不存在相异位的概率为 $\prod_{j=0}^{d-1} \frac{l_k-j}{L-1} = \frac{l_k!}{(L-1)^d (l_k-d)!}$, 则此基因段和其他被交换的 $k-1$ 个基因位都是相同位的上限概率为

$$\left(\prod_{j=0}^{k-2} \frac{L-d-1-j}{L-1} \right) \left(\prod_{j=0}^{d-1} \frac{l_k-j}{L-1} \right) = \frac{(L-d-1)! l_k!}{(L-1)^{k+d-1} (L-d-k)! (l_k-d)!}$$

综上所述,定理 3 成立.证毕. \square

定理 3 表明, d 和 k 是影响相异基因位被交换概率的两个主要因素.当 d 较大时,即使 k 较小,交叉操作也依然可以交换较多的相异基因位,此时, A 将以较大的概率接近于 Φ ; 但当 d 随群体收敛过程逐渐降低时, A 接近 Φ 的概率也随之降低, k 的增大可在一定程度上提高 A 的规模.

定理 2 和定理 3 从交叉算子对静态编码空间搜索范围的角度分析了交叉点规模与交叉算子搜索性能的关系.基于以上分析可以得出在群体迭代过程中交叉点规模所呈现的大致调控趋势:在进化初期可采用较小的交叉点规模,随着群体的收敛可适当增大交叉点规模,基本保持一个逐渐递增的过程.另外,文献[15]从模式重组与存活概率的角度分析了交叉点规模与交叉算子模式搜索性能的关系,对文献的分析结果进行归纳可以得出以下结论:若某类交叉算子的模式生存概率较大,则其模式重组能力较差;反之亦然.综上所述,在遗传算法搜索过程中,交叉算子对交叉点规模的需求会随群体状态的演变而动态地发生变化.为保持交叉操作的搜索效率,需要在群体进化过程中对交叉点规模进行动态调控.但是,由于遗传算法搜索过程的随机性导致群体状态演变的不确定性,因此很难通过简单的方法直接获得理想的交叉点规模.

2 对交叉点规模的优化

鉴于前述分析,可提出 3 个层次的优化策略.

2.1 交叉点规模分阶段调整策略

将遗传算法搜索过程划分为若干个使用逐渐增大的交叉点规模的进化阶段,是一种最直接、最简单的调控

方法.下面分别采用三阶段、四阶段和五阶段这 3 种阶段划分方式对 f_1 函数进行优化实验,来验证这种方法的有效性,实验条件与第 1.1 节相同.

图 2(a)为 3 种阶段划分方式中交叉点规模的调整过程,表 2 显示出采用四阶段和五阶段调整策略的 CGA 其优化效果要好于 15 点交叉算子.图 2(b)为 3 种交叉点规模阶段调整实验中优化结果最接近平均优化结果的一次运算的函数值优化曲线,由图可见,交叉点规模的阶段调整过程可明显提高算法的搜索效率.实验结果说明,分阶段对交叉点规模进行调控是可行的,其中,五阶段调整方式是 3 种模式中最接近 CGA 对 20 维的 f_1 函数实际交叉点规模调整需求的演变过程.

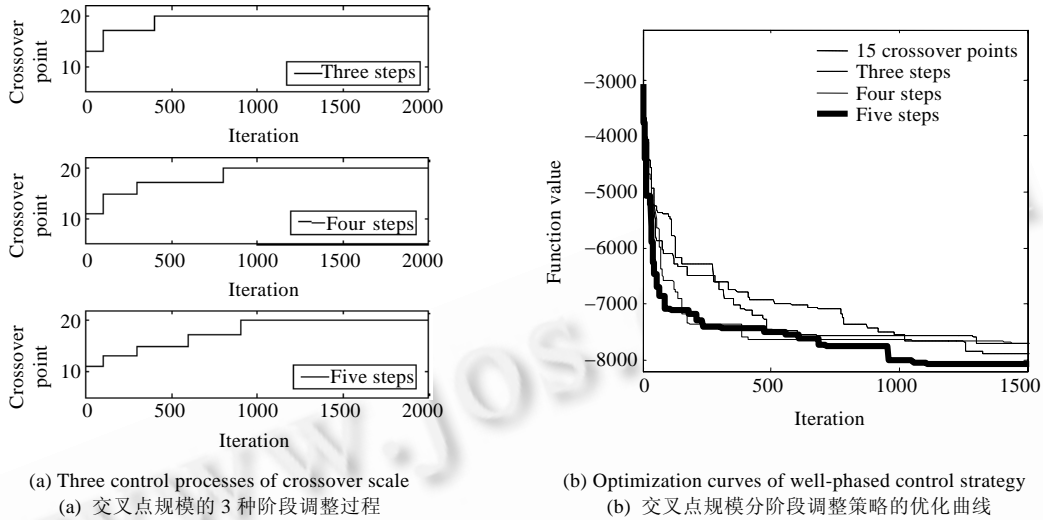


Fig.2 Well-Phased control strategy for crossover scale

图 2 分阶段对交叉点规模进行调整

2.2 基于分阶段调控方法的交叉点规模随机分配策略

一般交叉操作需要同时满足对未知空间的探索和对已发现区域的求精两种类型的搜索,单一的交叉点规模无法达到这一要求,所以,实际应用中往往需要同时使用多种规模的交叉操作.交叉点规模的阶段调整是对其演变过程的一种近似估计,一个阶段内所设置的交叉点规模可视为对此阶段所需各种交叉规模均值的期望.本文提出使用随机分配策略,在有效的阶段调控基础上来实现交叉规模的多样化.

设 α 为一个在 $(0,1)$ 区间的随机数, μ 为当前进化阶段所设定的交叉点规模,则通过下式可确定此阶段内一次交叉操作所使用的交叉点规模:

$$x = f(\alpha) = \mu + r - e, \text{ if } \frac{r}{2e+1} < \alpha \leq \frac{r+1}{2e+1} \quad (4)$$

其中, $e=1,2,\dots$ 为交叉点规模调整步长,它将 $(0,1)$ 区间划分为 $2e+1$ 个等分区间,若获取的随机数在区间 $r=0,1,\dots,2e$,则 $r-e$ 将决定 x 的调整方向和幅度.

公式(4)对交叉点规模的调节类似于一个随机变异过程,显然, μ 是 x 分配结果的均值,而 α 的分布特征将决定不同规模交叉操作的分布数量.例如,若 α 是均匀分布的,则 x 在 $[\mu-e, \mu+e]$ 间的分布数量也是均匀的;若 α 是正态分布的,则 x 的分布结果也接近于正态分布.下文中使用的 α 都服从正态分布,因此也将该策略称为正态分配策略.

下面在基于五阶段交叉点规模调整的 CGA 中加入正态分配策略对 f_1 函数进行优化实验,实验条件与第 1.1 节相同.表 2 为实验统计结果,图 3 为此实验的进化曲线.实验结果显示,正态分配策略使交叉点规模在群体进化过程中的分布更为合理,使 CGA 的搜索过程更为平滑、连续,较少出现进化停滞现象.

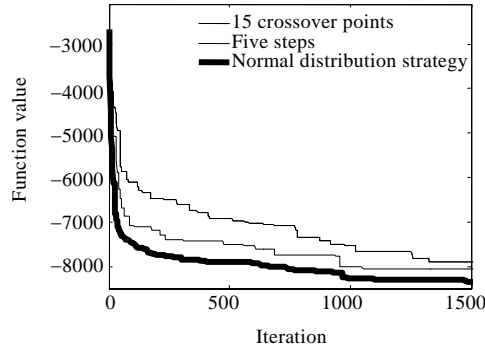


Fig.3 Optimization curves of f_1 by CGA with normal distribution strategy

图 3 采用交叉点规模正态分配策略的 CGA 对 f_1 的优化曲线

Table 2 Comparison of optimization statistic results of f_1 by CGA with different crossover operators

表 2 基于不同类型交叉算子的 CGA 对 f_1 的优化统计结果对比

Crossover strategy	15 crossover points	Well-Phased control strategy			Normal distribution strategy
		Three steps	Four steps	Five steps	
Mean function value	-7 902.2	-7 816.7	-8 088.3	-8 127.3	-8 315.4
Standard deviations	1 646.4	1 667.3	1 168.9	1 080.9	886.1

2.3 交叉点规模的自适应进化策略

下文进一步提出以代为单位的交叉点规模自适应进化策略,利用该方法可对交叉点规模的演化过程进行搜索.交叉算子的进化包括算子初始化过程、利用正态分配策略进行类似于变异操作的随机搜索过程以及对搜索到的算子的效果评估和选择过程.交叉算子的进化过程与群体进化过程通过交叉操作交换信息,实现二者的协同进化.

交叉算子的进化过程描述如下:

- 1) 设置初始交叉点规模 ξ_0 ;
- 2) 若第 i 代交叉点规模期望值为 ξ_i ,则此代中任一交叉操作的规模可由公式(4)产生;
- 3) 若第 i 代个体繁殖过程中,父个体 A_1, A_2 使用 k 点交叉产生的后代个体为 $A_{1,2}^1, A_{1,2}^2$,则一次 k 点交叉的效果为

$$v^k = \frac{\max[F(A_{1,2}^1), F(A_{1,2}^2)]}{F_{best}^{i-1}} \tag{5}$$

其中, F_{best}^{i-1} 为第 $i-1$ 代群体中最优个体的适应度值.该代交叉操作结束后,若 k 点交叉共使用了 t 次,则其平均效果为 $\bar{v}^k = \sum_{j=1}^t v_j^k / t$.统计本代使用的所有规模的交叉算子的平均效果,并据此选择下代的交叉点规模期望值:

$$\xi_{i+1} = \begin{cases} g(\max(\bar{v}^{\xi_i+1}, \dots, \bar{v}^{\xi_i+e})), & \text{if } \max(\bar{v}^{\xi_i}, \bar{v}^+, \bar{v}^-) = \bar{v}^+ \\ g(\max(\bar{v}^{\xi_i-e}, \dots, \bar{v}^{\xi_i-1})), & \text{if } \max(\bar{v}^{\xi_i}, \bar{v}^+, \bar{v}^-) = \bar{v}^- \\ \xi_i, & \text{else} \end{cases} \tag{6}$$

其中, $\bar{v}^+ = \sum_{j=\xi_i+1}^{\xi_i+e} \bar{v}^j / e$ 为交叉点规模增加方向的平均效果,相反方向为 $\bar{v}^- = \sum_{j=\xi_i-e}^{\xi_i-1} \bar{v}^j / e$, $g(v^p) = cp$ 表示获取某个交叉操作的交叉点规模.

- 4) 若群体进入下一个迭代过程,则返回步骤 2).

3 实验

为了验证交叉进化策略的有效性,在 Windows 2000 操作系统的 PC(Pentium4 2.4GHz,256MB Memory)上,对包括函数 f_1 在内的 6 个典型多维函数^[16,17]进行优化实验:

$$\text{Generalized Rastrigin's Function: } f_2(x) = \sum_{i=1}^m [x_i^2 - 10 \cos(2\pi x_i) + 10], x_i = [-5.12, 5.12]^m;$$

$$\text{Ackley's Function: } f_3(x) = -20 \exp \left(-0.2 \sqrt{\frac{1}{m} \sum_{i=1}^m x_i^2} \right) - \exp \left(\frac{1}{m} \sum_{i=1}^m \cos(2\pi x_i) \right) + 20 + e, x_i = [-32, 32]^m;$$

$$\text{Generalized Griewank Function: } f_4(x) = \frac{1}{4000} \sum_{i=1}^m x_i^2 - \prod_{i=1}^m \cos \left(\frac{x_i}{\sqrt{i}} \right) + 1, x_i = [-600, 600]^m;$$

$$\text{Sphere Model: } f_5(x) = \sum_{i=1}^m x_i^2, x_i = [-100, 100]^m;$$

$$\text{Schwefel's Problem 2.22: } f_6(x) = \sum_{i=1}^m |x_i| + \prod_{i=1}^m |x_i|, x_i = [-10, 10]^m.$$

3.1 对高维函数交叉点规模演化规律的发现

在 CGA 环境中分别使用自适应进化策略的交叉算子(evolution crossover,简称 EC)和固定规模交叉算子(fixed crossover,简称 FC)对 30 维函数 $f_1 \sim f_6$ 各进行 50 次运算.CGA 的群体规模为 100,交叉概率为 0.7,变异概率为 0.03,终止代数为 2 000 代,各函数的编码原则是保证搜索步长小于 10^{-3} ,每个函数使用的静态交叉点规模为一个经验值.表 3 为使用两种交叉算子的 CGA 对 6 个函数的优化统计结果,在相同的进化环境中,进化交叉较固定规模交叉的优化效果有明显提高.图 4 为 CGA 使用两种交叉算子的函数值优化曲线和交叉点规模曲线.函数值优化曲线显示,在相似的群体环境中,通过对交叉点规模的优化,CGA 的搜索速度有明显提高.交叉点规模曲线中折线为一次函数优化中交叉点规模的演化过程,直线为固定的交叉点规模,虚线为 50 次优化后交叉点规模演化过程的平均结果.6 个函数的交叉点规模优化曲线显示出了一定的相似特征,即交叉点规模在进化初期在略低于固定交叉点规模处波动,在进化中期在此经验值附近波动,进化后期则会明显增大,而交叉点规模的均值曲线则更明显地表现出了这一变化过程.此结果也基本符合前文对交叉点规模演化规律的分析,因此,均值曲线将被作为相应函数的交叉点规模控制知识.

Table 3 Comparison of optimization statistic results between CGA with FC or EC for $f_1 \sim f_6$ with 30 dimensions

表 3 使用固定规模交叉和进化交叉策略的 CGA 对 30 维函数 $f_1 \sim f_6$ 优化结果比较

Function		f_1	f_2	f_3	f_4	f_5	f_6
EC	Mean function value	-12 272.9	15.5	0.83	0.28	2.8	0.47
	Standard deviations	264.2	5.8	0.6	0.24	14.7	0.17
FC	Mean function value	-10 321.2	142.9	11.2	20.6	47.1	9.4
	Standard deviations	1932.2	16.8	3.6	7.2	20.7	4.7

3.2 对高维函数交叉点规模优化知识的应用实验

基于上述知识,在族群进化算法(ethnic group evolution algorithm,简称 EGEA)^[18]中使用交叉点规模正态分配策略对 6 个多维函数进行优化实验,并将结果与 OGA/Q^[19],HTGA^[20],MAGA^[16]以及 StGA^[17]这 4 种在函数优化方面表现出良好性能的算法进行对比.OGA/Q 采用的终止条件是在 1 000 代迭代后连续 50 代解的质量没有提高^[19],HTGA 的终止条件是群体中最优个体的函数值小于或等于 OGA/Q 提供的平均结果^[20],MAGA 的终止条件是对每个函数都搜索 150 代^[16],StGA 搜索到最优解的代数基本在 200 代以内^[17].EGEA 采用的族群半径参数为 0.4,终止条件是对每个函数都搜索 200 代,其余参数与第 3.1 节相同.通过将前一实验中各函数在 2 000 代进化中产生的交叉点规模均值每 10 代求一平均值作为此实验中一代使用的 μ 值,可将 2 000 代的交叉点规模演化过程映射为 200 代.表 4 为 EGEA 对 6 个函数连续 50 次运算后的统计结果.对 f_2, f_3, f_5 和 f_6 函数的 50 次优化

都在 200 代内找到了最优解.对 f_4 的 50 次优化中有 5 次没有在 200 代内收敛到全局最优解,对 f_1 搜索到的最优解为-12 569.486 6.

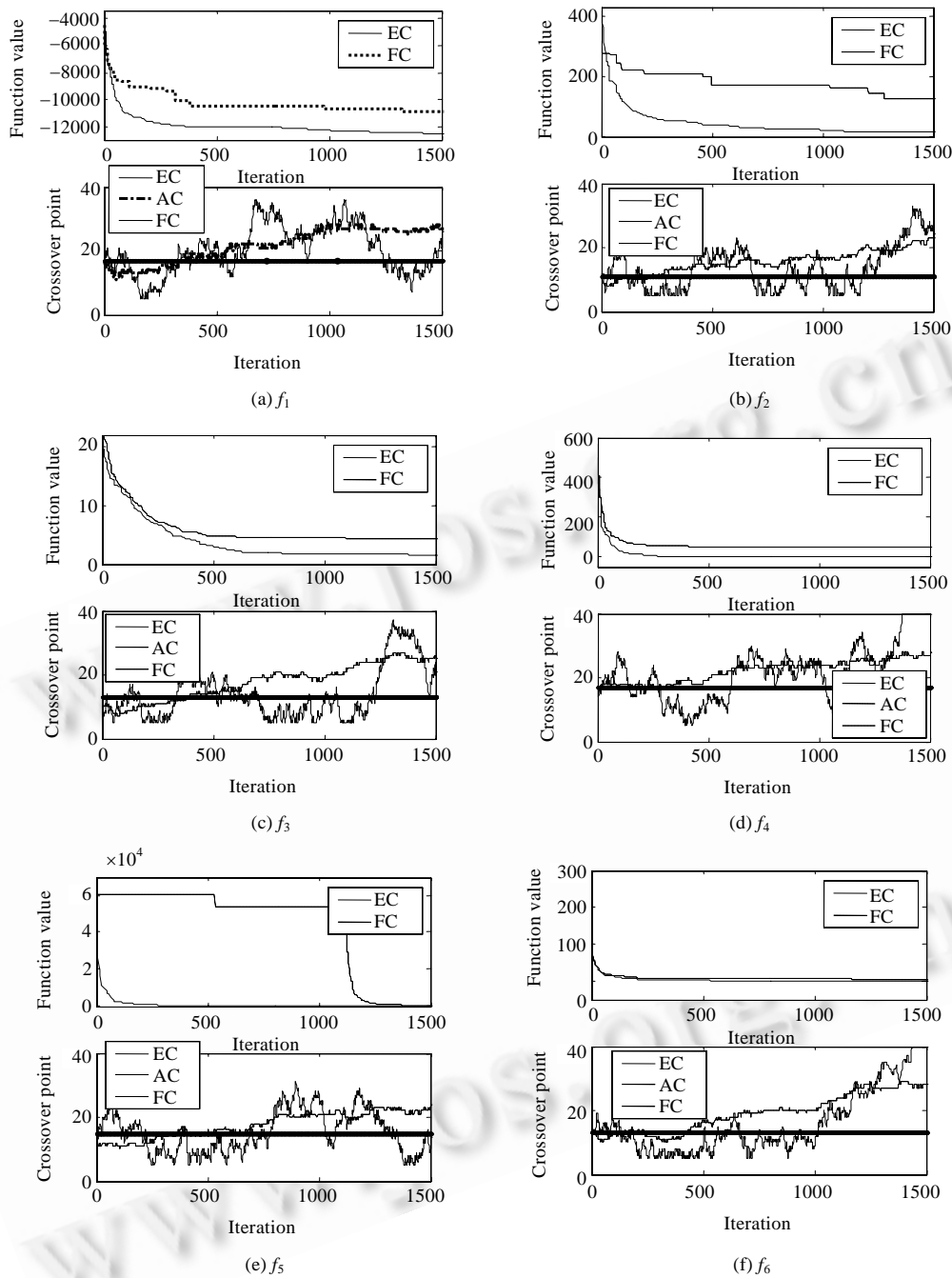


Fig.4 Comparison of the variation of function value and crossover scale between CGA with FC or EC for $f_1 \sim f_6$ with 30 dimensions

图 4 使用固定交叉和进化交叉的 CGA 对 30 维函数 $f_1 \sim f_6$ 的优化中函数值和交叉规模变化过程的比较

Table 4 Optimization statistic results of EGEA for $f_1 \sim f_6$ with 30 dimensions**表 4** EGEA 对 30 维函数 $f_1 \sim f_6$ 的优化统计结果

Function		f_1	f_2	f_3	f_4	f_5	f_6
Iteration	Mean iteration (Std)	176.2 (14.3)	64.3 (32.3)	48.8 (37.1)	137 (71.3)	76 (23.1)	24 (7.7)
Function value	Best value	-12 569.486 6	0	0	0	0	0
	MFV	-12 569.478 8	0	0	1.3e-5	0	0
	Std	8.7e-3	0	0	3.5e-4	0	0
Function evaluation	Minimum value	24 790	6 057	3 245	14 848	10 588	3 665
	MNFE	26 118	8 861	10 190	24 348	17 917	5 685
	Std	2 526	1 794	3 707	3 773	3 941	1 434

表 5 为 EGEA 与 OGA/Q,HTGA,MAGA 以及 StGA 这 4 种算法对 $f_1 \sim f_6$ 函数优化结果的对比.在对 f_2 和 f_6 函数优化过程中,EGEA 取得了最佳结果;在对 f_1 函数优化过程中,StGA 是唯一找到最优值的算法,EGEA 的优化结果要好于 OGA/Q 和 HTGA;在对 f_3 函数优化过程中,MAGA 花费的平均函数评估次数最少,EGEA 与 HTGA 搜索结果的精度最高,且 EGEA 花费的平均函数评估次数要少于 HTGA 和 OGA/Q;在 f_4 和 f_5 函数优化过程中,MAGA 表现出了最优的性能,EGEA 的优化结果与 HTGA 接近,但好于 OGA/Q 和 StGA.综合来说,在对 $f_1 \sim f_6$ 函数优化过程中,MAGA 的表现最佳,EGEA 的优化结果略好于 HTGA,而优于 OGA/Q 和 StGA.

后面的图 5 为 EGEA 对 6 个函数的优化过程曲线,包括函数值优化曲线和进化过程中使用的交叉点规模 μ 值曲线.函数值优化曲线显示,在交叉点规模 μ 值曲线的控制下,EGEA 对 6 个高维函数都有较好的进化速度,说明针对每个函数发现的交叉点规模控制知识都是有效的,在其控制下,采用交叉点规模正态分配策略的交叉操作显著提高了 EGEA 的搜索效率.

Table 5 Comparison of experimental statistic results among five algorithms**表 5** EGEA 与 OGA/Q,HTGA,MAGA 和 StGA 优化结果的对比

Function		f_1	f_2	f_3	f_4	f_5	f_6
EGEA	MNFE	26 118	8 861	10 190	24 348	17 917	5 685
	MFV	-12 569.478 8	0	0	1.3e-5	0	0
	(Std)	(8.7e-3)	(0)	(0)	(3.5e-4)	(0)	(0)
OGA/Q	MNFE	302 166	224 710	112 421	134 000	112 559	112 612
	MFV	-12 569.453 7	0	4.44e-16	0	0	0
	(Std)	(6.447e-4)	(0)	(3.9e-16)	(0)	(0)	(0)
MAGA	MNFE	10 862	11 427	9 656	9 777	9 502	9 591
	MFV	-12 569.486 6	0	4.44e-16	0	0	0
	(Std)	(7.121e-12)	(0)	(0)	(0)	(0)	(0)
HTGA	MNFE	163 468	16 267	16 632	20 999	20 844	14 285
	MFV	-12 569.46	0	0	0	0	0
	(Std)	(0)	(0)	(0)	(0)	(0)	(0)
StGA	MNFE	1 500	28 500	10 000	52 500	30 000	17 600
	MFV	-12 569.5	4.42e-13	3.52e-8	2.44e-17	2.45e-15	2.03e-7
	(Std)	(0)	(1.1e-13)	(3.5e-9)	(4.5e-17)	(5.2e-16)	(2.95e-8)

MNFE: Mean number of function evaluation; MFV: Mean function value; Std: Standard deviations

4 总 结

本文通过分析交叉点规模对交叉算子空间搜索能力的影响发现,随着群体状态的演变,交叉算子对交叉点规模的选择是一个需要动态优化的过程.针对此问题,本文提出使用分阶段调整策略、随机分配策略以及自适应进化策略 3 种方法来对交叉点规模进行动态调控.分阶段调整策略是对交叉点规模演化过程大致的估计,而随机分配策略则实现了交叉规模的多样性,自适应进化策略则能基于进化搜索的方法对交叉点规模的演化过程进行搜索分析.相对来说,自适应进化策略发现的优化规模更为准确,但此过程不可避免地增加了运算开销,因此该方法被主要用于发现交叉点规模控制知识,而将产生的知识应用于随机分配策略中作为实际应用的方法.对高维函数的优化实验表明,此类知识是有价值的,能够有效提高交叉操作的搜索效率.更重要的是,自适应进化策略使遗传算法具有了一种双层次的协同进化结构,即在算法对问题进行优化的同时,交叉算子自身也在

不断地利用进化过程进行调整,以适应问题求解过程的不确定性.此结果启发我们,可利用进化机制对组成遗传算法的各元素的优化规律进行搜索,以实现算法自身的优化.基于此思路寻找进化算法的自我进化机制,是我们今后研究的方向.

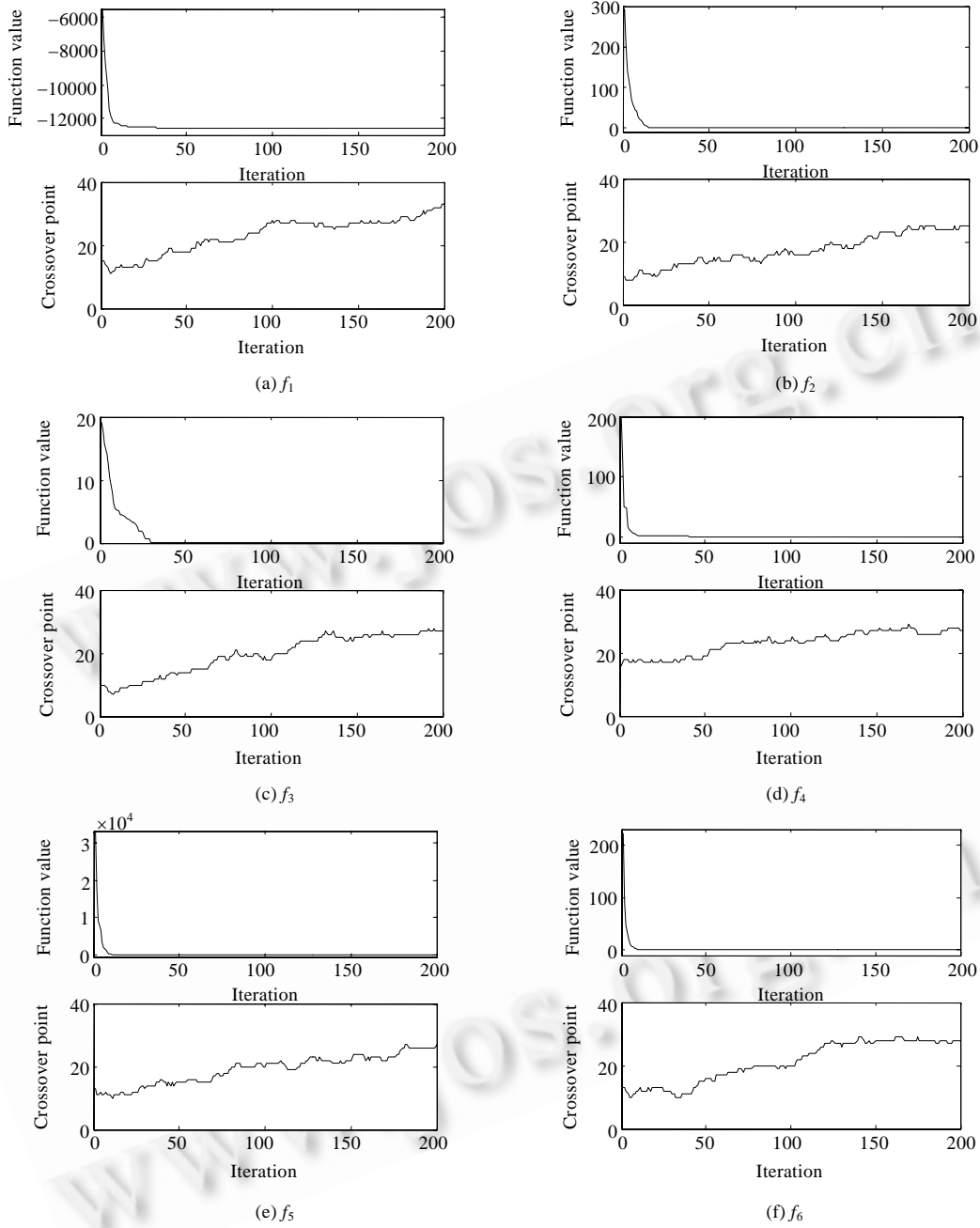


Fig.5 Variation process of function value and μ of EGEA for $f_1 \sim f_6$ with 30 dimensions

图 5 EGEA 对 30 维函数 $f_1 \sim f_6$ 优化过程中函数值和 μ 值的演变过程

致谢 在此,谨向对本文提出宝贵建议的审稿专家以及参与本文内容讨论的所有同学、老师表示衷心的感谢.

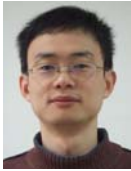
References:

- [1] Beyer HG. Toward a theory of evolution strategies: On the benefits of sex—the $(\mu/\mu, \lambda)$ theory. *Evolutionary Computation*, 1995,3(1): 81–111.
- [2] Zhang K, Liu GZ. Selecting crossover site with unequal probability in genetic algorithms. *Information and Control*, 1997,26(1): 53–60 (in Chinese with English abstract).
- [3] Zhang WX, Leung Y. *Mathematical Foundation of Genetic Algorithms*. 2nd ed., Xi'an: Xi'an Jiaotong University Press, 2000. 37–44 (in Chinese).
- [4] Ren QS, Ye ZX, Zeng J, Qi FH. Search capability of crossover operator. *Journal of Computer Research and Development*, 1999, 36(11):1317–1322 (in Chinese with English abstract).
- [5] Ren QS, Zeng J, Qi FH. Limit consistency of crossover operator. *Chinese Journal of Computers*, 2002,25(12):1405–1410 (in Chinese with English abstract).
- [6] Yao WS, Chen ZQ, Chen SF. CRGA—A genetic algorithm based on preserving global commonality schemata and restricting local of crossover. *Journal of Computer Research and Development*, 2006,43(1):81–88 (in Chinese with English abstract).
- [7] Zhang J, Chung Henry SH, Lo WL. Clustering-Based adaptive crossover and mutation probabilities for genetic algorithms. *IEEE Trans. on Evolutionary Computation*, 2007,11(3):326–335.
- [8] Domingo OB, Cesar HM, Nicolas GP. Improving crossover operator for real-coded genetic algorithms using virtual parents. *Journal of Heuristics*, 2007,13(3):265–314.
- [9] Garcia-Martinez C, Lozano M, Herrera F, Molina D, Sanchez AM. Global and local real-coded genetic algorithms based on parent-centric crossover operators. *European Journal of Operational Research*, 2006,185(3):1088–1113.
- [10] Jin DJ, Zhang JY. A new crossover operator for improving ability of global searching. In: Wang XZ, ed. *Proc. of the 6th Int'l Conf. on Machine Learning and Cybernetics (ICMLC 2007)*. Hong Kong: IEEE SMC Society, 2007. 2328–2332.
- [11] Kaelo P, Ali MM. Integrated crossover rules in real coded genetic algorithms. *European Journal of Operational Research*, 2005, 176(1):60–76.
- [12] Benjamin H, Kevin W. Synapsing variable-length crossover: Meaningful crossover for variable-length genomes. *IEEE Trans. on Evolutionary Computation*, 2007,11(1):118–131.
- [13] Wu SY, Xu ZQ. A heuristic policy for constructing crossover in genetic algorithm. *Chinese Journal of Computers*, 1998,21(11): 1003–1008 (in Chinese with English abstract).
- [14] Zhang JY, Xu J, Bao Z. Attainability of genetic crossover operator. *Acta Automatica Sinica*, 2002,28(1):120–125 (in Chinese with English abstract).
- [15] De Jong KA, Spears WM. A formal analysis of the role of multi-point crossover in genetic algorithms. *Annals of Mathematics and Artificial Intelligence*, 1992,5(1):1–26.
- [16] Zhong WC, Liu J, Xue MZ, Jiao LC. A multiagent genetic algorithm for global numerical optimization. *IEEE Trans. on Systems, Man and Cybernetics—Part B: Cybernetics*, 2004,34(2):1128–1141.
- [17] Tu ZG, Lu Y. A robust stochastic genetic algorithm (StGA) for global numerical optimization. *IEEE Trans. on Evolutionary Computation*, 2004,8(5):456–470.
- [18] Chen H, Cui DW, Li X, Wang ZM. The harmonious evolution of ethnic group algorithm. In: Lei JS, ed. *Proc. of the 3rd Int'l Conf. on Natural Computation (ICNC 2007)*. Haikou: IEEE Computer Society, 2007. 380–384.
- [19] Leung YW, Wang Y. An orthogonal genetic algorithm with quantization for global numerical optimization. *IEEE Trans. on Evolutionary Computation*, 2001,5(1):41–53.
- [20] Tsai JT, Liu TK, Chou JH. Hybrid taguchi-genetic algorithm for global numerical optimization. *IEEE Trans. on Evolutionary Computation*, 2004,8(4):365–377.

附中文参考文献:

- [2] 章珂,刘贵忠.交叉位置非等概率选取的遗传算法. *信息与控制*, 1997,26(1):53–60.
- [3] 张文修,梁怡. *遗传算法的数学基础*. 第2版,西安:西安交通大学出版社, 2000. 37–44.
- [4] 任庆生,叶中行,曾进,戚飞虎.交叉算子的搜索能力. *计算机研究与发展*, 1999,36(11):1317–1322.
- [5] 任庆生,曾进,戚飞虎.交叉算子的极限一致性. *计算机学报*, 2002,25(12):1405–1410.

- [6] 姚望舒,陈兆乾,陈世福.CRGA——一种基于保留全局公共模式和约束交叉位置的遗传算法.计算机研究与发展,2006,43(1):81-88.
- [13] 吴少岩,许卓群.遗传算法中遗传算子的启发式构造策略.计算机学报,1998,21(11):1003-1008.
- [14] 张军英,许进,保铮.遗传交叉运算的可达性研究.自动化学报,2002,28(1):120-125.



陈皓(1978—),男,河北安新人,博士生,主要研究领域为进化计算,数据挖掘.



李雪(1974—),女,博士生,主要研究领域为进化计算,数据挖掘,电子商务管理.



崔杜武(1945—),男,教授,博士生导师,主要研究领域为进化计算,计算机网络,多媒体技术.



韦宏利(1965—),男,博士生,副教授,主要研究领域为进化计算,模糊控制.



全国第4届语义 Web 与本体论学术研讨会(SWON 2009)

征文通知

全国语义 Web 与本体论学术研讨会(SWON)是中国计算机学会暨电子政务与办公自动化专委会主办的系列会议。SWON 2009 会议将于 2009 年 9 月在徐州召开。会议目的是为语义 Web 学术界和工业界提供一个交流平台,反映国际国内关于语义 Web 的最新研究成果和进展。会议录用论文中主要论文初定拟以英文方式由 IEEE Computer Society Press(EI 源刊)正刊出版,其余论文将由核心期刊《计算机科学》专刊、《计算机与数字工程》正刊和清华大学出版社出版(根据录用篇数确定期刊种类)。会议期间除进行会议论文交流外,还将邀请著名专家做特邀报告,并继续评选大会优秀学生论文。

一、征文范围(包括但不限于)

语义 Web 语言与工具	语义 Web 知识表示	语义 Web 知识管理	语义 Web 推理
语义 Web 服务	语义 Web 安全	语义 Web 挖掘	语义信息标注
语义检索和查询	本体学习与元数据生成	本体存储与管理	语义集成和映射

二、来稿要求

1. 本次会议主要通过网上投稿,尽量不要通过 E-mail 投稿,拒收纸质稿件。严禁一稿多投。
2. 中英文稿均可,一般不超过 6000 字,为了便于出版论文集,来稿必须附中英文摘要、关键词、资助基金与主要参考文献,注明作者及主要联系人姓名、工作单位、详细通信地址(包括 E-mail 地址)与作者简介。稿件要求采用 WORD 或 PDF 格式。

三、联系信息

1. 投稿地址: <http://www.easychair.org/conferences/?conf=swon09>.
2. 大会网站: <http://www.neu.edu.cn/wisa2009>
3. 会务情况: 中国矿业大学 姜淑娟 (shjjiang@cumt.edu.cn)

四、重要日期

1. 征文截止日期: 2009 年 4 月 25 日
2. 录用通知发出日期: 2009 年 5 月 20 日
3. 式论文提交日期: 2009 年 6 月 5 日
4. 会议召开日期: 2009 年 9 月 26~28 日