

族群进化算法*

陈皓¹⁺, 崔杜武¹, 崔颖安^{1,2}, 陶永芹^{1,2}, 梁琨¹

¹(西安理工大学 计算机科学与工程学院, 陕西 西安 710048)

²(西安交通大学 电子信息工程学院, 陕西 西安 710049)

Ethnic Group Evolution Algorithm

CHEN Hao¹⁺, CUI Du-Wu¹, CUI Ying-An^{1,2}, TAO Yong-Qin^{1,2}, LIANG Kun¹

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

²(School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

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

Chen H, Cui DW, Cui YA, Tao YQ, Liang K. Ethnic group evolution algorithm. Journal of Software, 2010, 21(5):978-990. <http://www.jos.org.cn/1000-9825/3484.htm>

Abstract: Enlightened by the conception of ethnic group in social science and a perspective to analyze the structure and evolutionary tendency of population in terms of ethnic group, a population structured technology, ethnic group mechanism, is proposed. Meanwhile, an ethnic group evolution algorithm (EGEA) with a dual track co-evolution process and special ethnic group operators is designed for binary coding. The simulation tests of the classical function and challenging composition test function show that the EGEA can restrain premature convergence effectively during the evolutionary process while improving the search efficiency greatly. The comparisons between EGEA and other typical algorithm show EGEA is a competent algorithm for solving numerical optimization problems.

Key words: genetic algorithm; ethnic group mechanism; ethnic group dual track co-evolution

摘要: 借鉴人类学研究中族群的概念以及以族群为视角来分析群体的结构及其演变趋势的方法,提出了一种简单、有效的群体结构调控技术——族群机制.设计了针对二进制编码方式的族群分类方法,并基于该族群结构形成了具有双轨协同进化特征的族群进化算法以及相应的族群算子.针对高维函数和复杂混合函数的数值优化实验表明,族群进化机制可以显著提高群体的抗早熟能力和搜索效率,与其他典型算法的对比也表明,族群进化算法是一种具有竞争力的函数优化算法.

关键词: 遗传算法;族群机制;族群双轨协同进化

中图法分类号: TP18 **文献标识码:** A

遗传算法(genetic algorithm,简称 GA)是一种利用“生成+检测”的迭代机制进行搜索的算法,交叉和变异两

* Supported by the National Natural Science Foundation of China under Grant Nos.60743009, 60873035 (国家自然科学基金); the Shaanxi Provincial Natural Science Foundation of China under Grant No.2006F-43 (陕西省自然科学基金); the Science and Technology Research Project of Shaanxi Province Department of Education of China under Grant No.07JK327 (陕西省教育厅科学技术研究计划)

Received 2008-05-04; Revised 2008-07-08; Accepted 2008-10-09; Published online 2009-09-11

种主要的遗传算子为群体提供了搜索动力,同时依靠选择算子来控制群体的搜索方向.在保留上代最佳个体的前提下,GA 的迭代过程是全局收敛的^[1].但由于标准遗传算子都是在一定发生概率控制下的随机操作,随机性导致搜索过程在提供个体进化机会的同时也可能促使个体衰退,由此引起了群体进化的不确定性以及搜索过程的不稳定性.近年来,结合自然科学中的各种思想和方法来改进已有的进化模型成为进化计算研究的一个重要方向,如借鉴生命科学中免疫的概念和理论提出的免疫遗传算法^[2]、通过应用基因库理论而产生的遗传算法^[3]、将经济学中“组织”的概念引入进化算法形成的组织进化算法^[4]、利用模糊逻辑来自适应调整交叉和变异概率的遗传算法^[5]以及基于水平集和拉丁方方法的进化算法^[6]等等.

现代人类学研究中使用依据文化特征进行分类而形成的“族群(ethnic group)”作为社会群体的组织单元,并以族群为视点来分析人类社会的组织结构和演变趋势以及各种进化趋势间相互学习和竞争对社会发展的影响.本文力图将这种对复杂大群体进行单元划分以便于对群体结构进行分析以掌握群体发展动向的研究方法引入遗传算法,形成了一种新的进化模型——族群进化算法(ethnic group evolution algorithm,简称 EGEA).

本文提出了族群的基本定义以及在二进制编码空间中族群的分类方法和族群进化的评估指标,并针对多维数值优化问题设计了具有双轨协同进化特征的 EGEA,同时使用 6 个典型多维函数和 6 个复杂的混合函数对 EGEA 的性能进行测试.与多种同类算法的对比结果显示,EGEA 能够显著抑制“早熟”现象的产生,同时有效提高群体的搜索速度以及收敛的稳定性.

1 族群进化基础

1.1 一种针对群体结构的调控机制——族群

传统的进化群体是随机而无序的,为了使群体更适合于多目标搜索,小生境技术^[7]利用排挤策略和适应值共享策略来抑制群体中个体的相似程度以及降低高适应度地区的个体密度.排挤策略是在遗传选择时仅使新的字串替代与之相似的父代个体,而适应值共享策略则是通过定义群体中个体的共享度来调整个体的适应度值.因此,小生境技术主要是通过改进常规选择操作来间接地调整群体的结构.聚类^[8]是一个无监督的分类,其目的是发现样本点之间最本质的“抱团”性质.文献[9,10]将聚类方法与适应值共享策略和排挤策略相结合,通过聚类过程来分析群体结构以动态获取小生境半径,并根据群体状态有针对性地产生小生境以降低遗传漂移的影响.这种改进机制已使群体具有了某种模糊的组织形式.组织进化算法^[4]则明确地在群体中建立了清晰的组织结构,其进化的核心单位不是个体而是组织,即一组个体的集合.针对数值优化问题,这种群体结构取得了较好的效果.上述研究表明,针对问题对群体结构进行适当的调整将有助于提高进化算法的搜索效率.因此,我们试图寻找一种简单而有效的群体结构调控机制.

人类社会是个典型的进化群体,全球化过程使各种成员交织混杂在一起,形成了更为复杂、多元的结构.在现代人类学研究中,“族群”常被作为群体的基本单元,用来分析群体内存在的冲突和矛盾以及发展和演变的趋势.虽然对族群的概念还存在一定的争议,但多数人类学家和社会学家认为,族群是指在一个较大的社会体系中具有自身文化特质的一种群体.可见,划分族群的基本思想是首先选定一些特征,然后区分出这些特征在群体中所具有的不同类型,最后根据个体属性与特征类型的匹配程度进行分类.受该思想的启发,我们设计了一种新的群体结构分析与调控机制,并借用“族群”来为其命名.下面是对族群的一个基本定义:

定义 1. 族群 E 是依据一定规则对个体 A 进行分类后在群体 N 中建立的某种组织结构的映射关系.

换言之,族群就是从特定的角度利用分类过程来分析群体结构,并对该结构进行描述的一种组织机制.通过控制族群的进化过程可实现对群体结构演变过程的调控.基于此定义,族群结构具有以下特点:

(1) 确定族群分类规则的过程即是对群体进化状态观察角度的选择过程.角度不同,分类规则也不同,从而会产生不同的族群结构.编码方式、适应度等都可成为观察群体的一种角度,而二进制编码模式则是我们首先尝试的族群分类角度.

(2) 族群是一种组织关系的结构映射,它仅记录了个体与族群间的逻辑从属关系,并未将个体从群体中抽取出来组成实质的子群体.在基于小生境技术的群体中,各小生境是隐性存在于群体中的,个体与小生境之间没

有直接的从属关系.而组织进化算法中的基本进化单位是明确的组织,个体依赖于组织而存在.与上述两种机制不同之处在于,族群结构具有清晰的组织形式,但仅是一种逻辑从属关系,个体依然是群体进化的主体,并分散在群体中.这种组织结构既可以使族群方便地进行群体的结构化调整,又降低了管理过程的计算开销.

(3) 族群依附于群体而存在,其结构随群体的演变而变化.

1.2 基于二进制编码的族群进化评估指标

在 EGEA 中,我们设计了一些新的评估指标,用于综合评价个体和族群的质量.设群体 N 的规模为 n, L 为个体编码长度, m 维个体 $A_i = a_1, \dots, a_L (a \in \{0, 1\})$ 的适应度为 $F_i > 0, F_{\max}$ 和 F_{avg} 分别表示当前群体的最优适应度和平均适应度.

定义 2. 个体 A_i 的相对适应度比为

$$\bar{R}_i = (F_{\max} - F_i) / (F_{\max} - F_{\text{avg}}) \tag{1}$$

显然,当 $F_i > F_{\text{avg}}$ 时, $0 \leq \bar{R}_i \leq 1$; 当 $F_i \leq F_{\text{avg}}$ 时, $\bar{R}_i \geq 1$.

定义 3. 个体 A_i 与 A_j 之间的加权编码距离为

$$\delta_{ij} = \sum_{w=1}^m \sum_{k=1}^{l_w} (b_{w,k} \times \eta_{w,k}) \tag{2}$$

其中, $b_{w,k} = \begin{cases} 1, & a_{i,w,k} \neq a_{j,w,k} \\ 0, & a_{i,w,k} = a_{j,w,k} \end{cases}$, $b_{w,k}$ 的权值为 $\eta_{w,k}$.

定义 4. 个体 A_i 与 A_j 之间的加权编码差异度为

$$D(A_i, A_j) = D_{ij} = \frac{\delta_{ij}}{\sum_{w=1}^m \sum_{k=1}^{l_w} \eta_{w,k}} \tag{3}$$

实际应用中主要关心的是个体与群体中最优个体 A_{best} 的差异率,所以常用的公式为 $D(A_i, A_{\text{best}}) = D_i$.

定义 5. 个体 A_i 的竞争指数为

$$Q(A_i, A_{\text{best}}) = Q_i = D_i^{\zeta \bar{R}_i} \tag{4}$$

其中, ζ 为调节参数.规定若 $D_i = 0$, 则 $Q_i = 0$, 且 A_{best} 的竞争指数为 1, 则 $Q \in [0, 1]$.

定义 6. 第 k 代群体的平均编码差异度为

$$\gamma_k = \frac{1}{n-1} \sum_{i=1}^{n-1} \left(\frac{1}{n-i} \sum_{j=i+1}^n D_{ij} \right) \tag{5}$$

其中, $\eta_{w,k} = 1$ 则 $\gamma \in (0, 1)$.

通常认为,适应度大于群体平均适应度的部分个体对群体进化更为有价值,相对适应度比能量化地描述个体的这种性质,因此该参数被作为从适应度的角度评估个体价值的指标.若个体每一维编码的基因位对个体精度的影响都相同,则可令 $\eta_{w,k} = 1$; 而对于数值优化一类问题,高位基因对个体精度的影响总是大于低位基因,则可通过权值来体现基因位对个体精度影响力的差异.本文令 $\eta_{w,k} = l_w - k_w + 1$, 其中 l_w 为个体第 w 维的编码长度, $k=1, \dots, l_w$. 因此,可用加权编码距离和加权编码差异度来衡量个体之间的编码差异性.基于个体对群体多样性保持以及适应度增长两方面作用的综合评价,竞争指数被用来作为个体竞争能力的评价指标.图 1 为竞争指数与相对适应度比和加权编码差异度的关系曲线.基于竞争指数,高于群体平均适应度且与群体中最优个体存在较大差异的个体将获得更强的竞争能力,反之则相反.群体平均编码差异度 γ 越接近 1, 说明个体间相异基因位越多,在族群进化中将其作为群体多样性的评价指标.

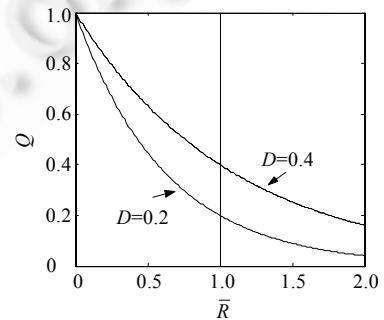


Fig.1 Race exponent curves

图 1 竞争指数曲线

1.3 个体的角色分工——大配子

在群体的演化过程中,保持个体的稳定性有利于遗传信息的传递,但是个体也需要通过自身的改变来实现对新信息的探索 and 发现.当二者无法达到一个有效的平衡时,就会对群体的进化过程产生负面的影响.生物的有性繁殖过程为群体协调信息的传递和探索两者的关系提供了新思路:不同的配子具有截然相反的特点,雌性配子数量少却体积较大,雄性配子数量众多但体积较小.其中的小配子更加灵活,而大配子则保证了初生后代所需的能量.通过这种分工机制,有性繁殖过程既保持了遗传的稳定性也增加了变异量,从而使生物具有了更快的适应性.本文引用大配子的概念来对个体进行角色分工,并用大配子系数 $\lambda \in (0,1)$ 作为大配子选择的阈值参数.

定义 7. 若 A_i 的竞争指数 $Q_i > \lambda$, 则个体 A_i 为大配子, 即 $A_i \in M$.

大配子是群体中部分具有较高竞争指数的典型个体,这部分个体承载了群体所拥有的主要进化信息.

1.4 族群的代表——族群中心个体

从不同的角度对大配子进行分类可以形成不同的族群结构,而每个族群都代表了这种分类规则下的一类进化趋势,其中最典型的个体将被用来作为该族群的代表.

定义 8. 族群 $E(i)$ 中具有最大竞争指数的个体被称为族群中心个体 $A_c(i)$.

可见,族群中心个体是一个族群中最典型的个体,作为此个体所在族群的标识,其所具有的竞争指数将代表该族群的竞争能力.

2 族群进化算法(EGEA)

2.1 数值优化问题

本文的主要目的是研究 EGEA 的搜索性能,因此仅针对无约束数值优化问题进行求解.一个全局无约束数值优化问题可以被描述为下述数学模型:

$$\text{Min } f(\bar{x}), \bar{x} = (x_1, \dots, x_m) \in S \quad (6)$$

其中, $S \subseteq R^m$ 为由 m 维变量组成的问题空间,其每一维变量的搜索区间为 $x_{low,j} \leq x_j \leq x_{up,j}, j=1, \dots, m$.

2.2 基于二进制编码方式的族群分类

以大配子为对象进行族群分类,既可以降低计算量又可以保证所产生族群的有效性和代表性.在二进制编码空间中,个体的编码特征可作为一种搜索模式的分类依据.族群半径系数 $\theta \in (0,1)$ 被用来作为进化单元间差异性的判断标准.因此,以加权编码差异度为分类指标,可在大配子中建立一种基于二进制编码的族群结构.以下为族群分类算子的执行流程:

步骤 1. 筛选大配子:计算群体的竞争指数并从中挑选大配子,大配子系数 λ 可由下式获得:

$$\lambda = (\bar{Q} + v) / 2 \quad (7)$$

其中, \bar{Q} 为当前群体的平均竞争指数,常数 $v \in (0,1)$.

步骤 2. 确定族群半径:族群半径系数 θ 可由下式获得:

$$\theta = (\gamma + \varepsilon) / 2 \quad (8)$$

其中, γ 为当前群体的多样性指标,常数 $\varepsilon \in (0,1)$.

步骤 3. 筛选族群中心个体:对大配子按其竞争指数排序,以当前群体中最优个体 A_{best} 为第 1 个族群中心个体 $A_{best} \in A_c$,并按队列顺序依次判断,若 A_i 对于任意的 $A_c(j)$ 都有 $D(A_i, A_c(j)) > \theta$, 则 $A_i \in A_c$.

步骤 4. 大配子分类:以族群中心个体为每个族群的核心,对大配子进行分类并形成族群组织.对于非族群中心个体 $A_i \in M$,判断 A_i 与所有 A_c 的加权编码差异度,若 A_i 与 $A_c(j)$ 具有最小值,则 $A_i \in E(j)$.

步骤 5. 分配繁殖空间:族群 $E(k)$ 可繁殖的后代规模为

$$S_k = n \frac{Q_c(j)}{\sum Q_c} \quad (9)$$

其中, $Q_c(i)$ 为中心个体 $A_c(i)$ 的竞争指数, $\sum Q_c$ 为群体中族群竞争指数之和.

大配子系数 λ 决定了大配子的数量, 而族群半径系数 θ 则决定了族群分类的粒度. 这两个指标将决定族群分类的质量, 从而影响群体进化的效率.

2.3 族群双轨协同进化机制

族群分类过程使群体能够按照进化特征的差异形成若干个平行的进化单元, 其中每个单元代表了群体中蕴含的一种进化趋势. EGEA 以族群为单位来控制群体的繁殖过程, 同时, 族群的分类能力也使 EGEA 可以方便地从群体中筛选出典型个体, 并从这些个体所承载的信息中发现有价值的经验性知识. 利用这类后验性知识可以修正个体的编码, 从而影响群体的结构, 起到加速信息交换、提高收敛速度的作用. 结合族群繁殖过程以及基于族群进化经验知识的挖掘和应用的自学习过程这两类个体搜索途径, 形成了一种独特的进化模型, 族群双轨协同进化 (ethnic group dual track co-evolution, 简称 EGDE). 图 2 为该进化机制的主要流程.

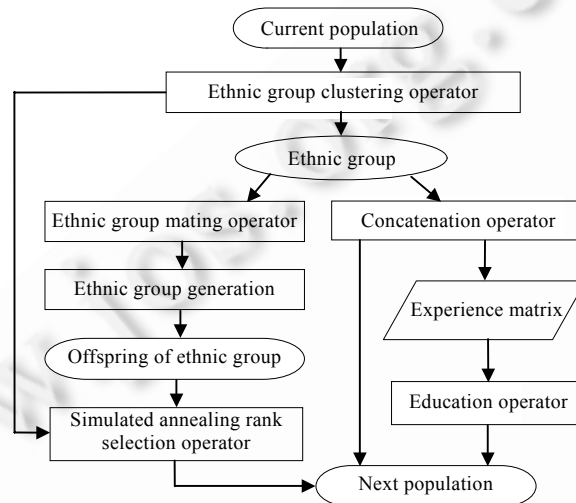


Fig.2 Flowchart of EGDE

图 2 族群双轨协同进化流程

2.3.1 由大配子主导的族群择偶繁殖

在以族群为单位的群体繁殖过程中, 择偶算子负责父个体的挑选, 族群交叉和变异算子负责族群后代的生产, 最后由模拟退火排序选择算子完成群体的更替.

1) 族群择偶算子

通常, GA 从群体中随机挑选用于繁殖的父个体, 并通过固定概率来控制后代个体的产生. 文献[11]利用一种虚拟父母机制来促使后代个体出现在最有潜力的区域, 文献[12]则通过对个体的性别分类机制来控制产生后代个体的区域范围. 这类研究说明, 合适的父母个体能够有效提高群体繁殖的效率和后代个体的质量. 在基于族群机制的群体中存在两类个体, 大配子和普通个体. 作为群体中最典型的个体, 大配子将主导族群的繁殖过程. 族群择偶算子的工作流程如下:

步骤 1. 挑选大配子: 依据竞争指数, 以轮盘赌选择方式从当前族群所属的大配子中挑选个体 M_i .

步骤 2. 挑选普通配子: 依据竞争指数, 以轮盘赌选择方式从整个群体中挑选普通配子个体 A_j .

步骤 3. 般配判断: 双方的般配率为

$$\omega_{ij} = D_{ij}(Q_i + Q_j) \quad (10)$$

如果 $\omega_{ij} > \text{Random}(0, 1)$, 则择偶成功; 否则, 返回步骤 1 重新配对.

择偶机制可自适应地调节族群繁殖的概率, 此概率不仅取决于父个体自身的质量, 同时也取决于父个体间

的特征差异,从而使有较高质量和存在一定差异性的父个体有更多的繁殖机会.

2) 族群交叉算子

在基于二进制编码的多维优化问题中通常采用多点交叉算子,其中交叉点规模极大地影响着该算子的搜索效率^[13].由于进化群体结构演变的随机性,导致合适的交叉点规模很难估计,而且随着群体的演变,该参数也需要动态地进行调整.在族群交叉过程中,我们采用一种具有对交叉点规模进行自我优化能力的多点交叉算子:

步骤 1. 设置初始交叉点规模 X_0 .

步骤 2. 若第 i 代交叉点规模初始值为 X_i ,则此代中一次交叉操作的交叉点规模可由下式获得.

$$x = f(\alpha) = X_i + r - e, \quad \text{if } \frac{r}{2e+1} < \alpha \leq \frac{r+1}{2e+1} \quad (11)$$

其中, $\alpha = \text{random}(0,1)$; $e=1,2,\dots$ 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 α 在区间 $r=0, 1, \dots, 2e$ 内,则 $r-e$ 将决定 X 的调整方向和幅度.

步骤 3. 若第 i 代个体繁殖过程中父个体 A_1 与 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

$$v^k = \max[F(A_{1,2}^1), F(A_{1,2}^2)] / F_{best}^{i-1} \quad (12)$$

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,并据此选择下一代的交叉点规模期望值:

$$X_{i+1} = \begin{cases} g\left(\max(\bar{v}^{X_i+1}, \dots, \bar{v}^{X_i+e})\right), & \text{if } \max(\bar{v}^{X_i}, \bar{v}^+, \bar{v}^-) = \bar{v}^+ \\ g\left(\max(\bar{v}^{X_i-e}, \dots, \bar{v}^{X_i-1})\right), & \text{if } \max(\bar{v}^{X_i}, \bar{v}^+, \bar{v}^-) = \bar{v}^- \\ X_i, & \text{else} \end{cases} \quad (13)$$

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

步骤 4. 若群体进入下一个迭代过程,则返回步骤 2.

X_i 表示了第 i 代群体中使用交叉点规模的期望均值,利用公式(11)可对交叉点规模进行随机调整,随机数 α 的分布特征将决定不同交叉点规模在交叉操作中的使用数量.例如,若 α 是均匀分布的,则 x 在 $[X_i-e, X_i+e]$ 区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.本文使用的 α 都服从正态分布.

3) 族群变异算子

由族群交叉操作产生的两个个体经族群变异操作后将作为该族群后代个体.变异概率由下式确定:

$$p_m = \begin{cases} \text{random}(0.001, 0.01), & \max(Q_i, Q_j) = 1 \\ 1 - \max(Q_i, Q_j), & \text{otherwise} \end{cases} \quad (14)$$

族群变异根据父代个体的竞争能力自适应地控制后代个体的变异概率,若父代个体的竞争指数较高,则其后代个体将具有较高的稳定性,从而保证优良基因模式的生存概率;反之,将以不同的概率去改变其后代个体的结构,从而促使竞争能力较弱的模式以不同幅度去搜索新的模式.

4) 模拟退火排序选择算子

在模拟退火排序选择算子中,竞争指数被用来作为个体的绝对价值指标,依据其对当前群体及其后代个体进行排序.利用模拟退火过程产生的排序队列的位置参数,并以其作为个体的相对价值指标.综合队列中个体的两类价值指标,可通过下式来确定个体的生存概率:

$$P_s(k) = \sigma Q_k + (1 - \sigma) f_k \quad (15)$$

其中, Q_k 为队列中第 k 位个体的竞争指数, $\lambda(k)$ 为其位置参数, 权值 $\delta \in (0, 1)$ 用于控制个体的竞争指数和模拟退火参数对选择概率的影响程度.

规定个体排序队列中 $\lambda(k)=1$, 其余 $k=2, \dots, 2n$, 位置的个体位置参数由以下模拟退火函数产生:

$$\lambda(k) = \exp(-\gamma/T_k) \quad (16)$$

其中, γ 为群体多样性指标, 队列中第 k 位个体的退火温度为 $T_k = T_0 \beta^k$, 一般 $T_0 = 2n$, β 是一个略小于 1 的数.

模拟退火排序选择算子具有精英个体保留的特点, 这保证了群体的收敛性^[1]. 另外, 它以竞争指数作为个体的绝对价值指标保证了被选择个体的代表性, 同时以模拟退火过程产生的位置参数作为个体的相对价值指标则保证了被选择个体的多样性. 参数 δ 用来调节竞争指数和位置参数对选择概率的影响率, 为使该过程适应于群体状态的演变, 我们通过下式来动态地调整权值参数:

$$\delta = \gamma \quad (17)$$

在实施族群繁殖过程中, 一方面会不断在问题的解空间中发现新的有价值区域, 在这些区域建立族群并繁殖后代, 而竞争中处于劣势的族群将萎缩直至淘汰, 这种群体中族群的新陈代谢过程保证了算法全局搜索的效率; 另一方面, 围绕大配子进行的族群繁殖过程保证了典型模式的生存概率与重组概率, 起到了局部求精的作用, 同时也促进了个体间的信息交换, 提高了群体的搜索效率. 因此, 族群繁殖过程可有效协调群体全局搜索和局部搜索的关系.

2.3.2 针对族群进化经验知识的自学习过程

根据 NFL(no free lunch)理论^[14], 相关知识的指导可以使进化算法与问题更为贴近, 从而可以有效提高进化算法的搜索效率. 因此, 自适应、自学习能力已成为改进进化计算的一个重要机制. 而在此类研究中, 如何获取有效的知识则成为一项关键技术. 与问题相关的先验性信息是一类重要的知识来源, 如免疫算法中使用的疫苗^[2]. 但 GA 的一个主要优势在于其搜索机制无须对问题本身的特征有过多的了解, 从这个角度来看, 后验性的经验知识则是更为重要的一类知识. 传统的 GA 无法充分利用积累于群体中的经验性信息的一个重要原因是缺乏筛选有价值的典型个体的机制, 文化算法^[15]对 GA 的最大改进之处就在于设立了一个信仰空间来提取、存储群体中的典型个体, 从而使群体进化过程具备了自我学习的基础.

族群机制具有的分类能力使 EGEA 可以方便、有效地从群体中筛选出典型个体, 基于这种能力, EGEA 形成了一种独特的自学习机制. 该机制与族群繁殖过程并行, 利用族群进化经验矩阵(experience matrix, 简称 EM)来存储族群中的典型个体, 并以这些信息为基础通过拼接实验重组个体, 以分离出蕴含于典型个体中有价值的基因片段, 同时利用这些经验信息来影响下一代群体的结构, 以实现引导和加速群体进化的目的.

1) 族群进化经验矩阵

每个族群中的中心个体是最有代表性的个体, 以维为单位, 可以将一个族群中心个体切割为 m 个信息单元(information unit, 简称 IU). 该信息单元的基本结构如下:

$$IU = (BC, RC, WV) \quad (18)$$

其中, BC 为 IU 的二进制编码信息, RC 为 IU 的实数编码信息, 而 WV 则表示该信息单元的权值.

基于上述结构的信息单元, 可利用族群经验矩阵 EM 来存储多个族群中心个体的信息. 如果族群存在 e 个中心个体时, 一个族群经验矩阵可以表示为

$$EM = \begin{Bmatrix} IU_{11} & \dots & IU_{1j} & \dots & IU_{1m} \\ \dots & & \dots & & \dots \\ IU_{e1} & \dots & IU_{ej} & \dots & IU_{em} \end{Bmatrix} \quad (19)$$

2) 拼接实验算子

基于 EM , 拼接实验算子以维为单位, 使用不同的 IU 尝试进行个体重组, 以发现更优秀的个体, 同时验证被采用的 IU 的价值, 从而实现从这些典型个体中分离出有价值基因片段的目标.

步骤 1. 初始化 $EM: IU_{ij}(i=1, \dots, e, j=1, \dots, m)$ 的权值为

$$W_{ij} = F_i / F_{best} \quad (20)$$

步骤 2. 重组实验个体:以维为单位,依次从各维的 e 个 IU 中依据 WV 通过轮盘赌选择方式挑选出 m 个 IU , 并组合产生实验个体 A_{test} ,则在此过程中, IU_{ij} 被选择的概率为 $WV_{ij} / \sum_{i=1}^e WV_{ij}$.

步骤 3. 评估实验效果:计算 A_{test} 的适应度值 F_{test} ,此次实验的效果为

$$\rho = F_{test} / F_{best} \quad (21)$$

如果 $\rho > 1$,暂令 $Q_{test} = 1$,并用 A_{test} 随机替换下一代群体中的一个个体,实验中被选中 IU 的 WV 值将被修改为 $MV' = MV + t\rho$;如果 $\rho < 1$,则 A_{test} 将被放弃,被选中 IU 的 WV 值将被修改为 $MV' = MV + \rho/t$.一般设置 $t = 2$.

步骤 4. 若实验次数限制 $e \times m$ 已达到,则终止此算子;否则返回步骤 2.

通过该算子的处理, EM 中各 IU 的权值发生了显著的变化,依据该统计结果,可以从中 EM 筛选出蕴含于典型个体中最有价值的一些基因段,并将这些信息作为族群进化过程中积累下来的一类经验性的知识.

3) 教育算子

教育算子以一定的概率选择经验矩阵中较有竞争力的信息单元,用于修正下一代群体中个体的结构,以促进信息交换并增加具有进化潜力的模式在下一代群体中的浓度.

步骤 1. 从下一代群体中随机选择一个个体 A_i ,则其接受教育的概率为

$$P_{education} = 1 - Q_i \quad (22)$$

步骤 2. 对 A_i 的每一维依次进行判断,如果 $P_{education} > \text{Random}(0,1)$,则依据 IU 的 MV 值通过轮盘赌选择方式从 EM 中对应维的 e 个 IU 中选择一个单元替代 A_i 现有的编码信息;否则保留该维的编码信息.

步骤 3. 评估被教育个体的适应度值,并以其替换原有个体.

2.3.3 邻域搜索算子

邻域搜索算子通过下降迭代搜索过程对在二进制编码空间中发现的最优个体进行邻域搜索,以提高解的精度.其主要流程如下:

步骤 1. 以二进制编码空间中搜索到的最优结果 $A_{best} = (x_1, \dots, x_m)$ 为初始点,设其适应度为 F^0 , x_i 实数编码的搜索区域为 $[o_i - \eta_i, o_i + \eta_i]$, $\eta_i = \frac{|x_i - \underline{x}_i|}{(2^{l_i} - 2)}$, \bar{x}_i 和 \underline{x}_i 分别表示 x_i 的上、下限, l_i 为 x_i 的二进制编码长度.若设搜索步数 $\alpha = 1, 2, \dots$,则 x_i 的搜索步长为 $g = \eta_i / \alpha$.

步骤 2. 分别计算 o_i 在递增与递减两个方向上进行单步调整后函数适应度的变化,选择增量最大的方向作为 x_i 的搜索方向.

步骤 3. 为每个搜索步数产生一个新的个体,其各维变量为

$$o_i^{k+1} = o_i^k + hg^k d_i^k \quad (23)$$

其中, $d_i^k = -1, 1$ 为 x_i 的搜索方向, $h = 1, \dots, \alpha$. 评估新产生的 α 个个体的适应度值,设其中具有最大适应度 F^{k+1} 的个体为 $A_{best}^{k+1} = (o_1^{k+1}, \dots, o_m^{k+1})$.

步骤 4. 如果 $F^{k+1} < F^k$ 则停止搜索,输出最终搜索到的解个体 A_{best}^k ; 否则,以 $A_{best}^{k+1} = (o_1^{k+1}, \dots, o_m^{k+1})$ 为新的初始点,步长 $g^{k+1} = g^k / \alpha$,返回步骤 2,在区间 $[o_i^{k+1} - g^{k+1}, o_i^{k+1} + g^{k+1}]$ 内进行搜索.

2.4 EGEA 的执行过程

以下为 EGEA 的执行过程,其中 ξ 表示当前群体中族群的数量, C_k 表示第 k 代的族群后代个体,接受教育比例 $\psi \in (0,1)$,表示下一代群体中有 ψm 个个体将接受教育操作.

Algorithm EGEA.

- 01: **begin**
- 02: $k = 0$;
- 03: initialize population N_0 randomly;
- 04: evaluate the race exponent (N_0);
- 05: **while** (the termination criteria are not reached) **do**


```

06:  begin
07:       $E_k$ =ethnic group clustering ( $N_k$ );
08:      for  $i=1$  to  $\xi$  do
09:          while (children number of  $E_k(i)$  are not reached) do
10:               $C_k(i)$ =mutation (crossover (mate ( $E_k(i), N_k$ )));
11:          evaluate the race exponent ( $N_k, C_k$ );
12:           $N_{k+1}$ =selection next population ( $N_k, C_k$ );
13:           $EM_k$ =initialize experience matrix ( $E_k$ );
14:           $N_{k+1}$ =education (splice experiment ( $EM_k, \psi$ ));
15:          evaluate the race exponent ( $N_{k+1}$ );
16:           $k=k+1$ ;
17:      end;
18:      adjutant-searching ( $A_{best}$ );
19:  end.

```

3 算法性能测试

3.1 测试函数

6个典型多维函数和6个混合函数被选择用来测试EGEA的性能.测试函数 $f_1 \sim f_4$ 和 f_5, f_6 分别选自文献[16]中的多模态函数 $f_1 \sim f_4$ 和单模态函数 f_7, f_8 .上述典型测试函数具有以下特征^[17]:全局最优解中各维变量相同,全局最优解常常在原点而且全局最优解中各变量一般处于搜索区域的中心或边界位置.如果算法针对以上特点设计搜索算子,如邻域竞争算子^[16],那么算法的搜索效率将大幅度提高.为了更真实地模拟现实问题,文献[17]通过对函数进行坐标变换、旋转变换以及空间叠加等一系列措施生成更具复杂性和挑战性的混合测试函数 $Cf_1 \sim Cf_6$.实验中,群体规模 $n=100$,族群分类参数 $v=0.6, \epsilon=0.3$,教育比例 $\psi=0.2$,图3为12个测试函数各维的编码长度.

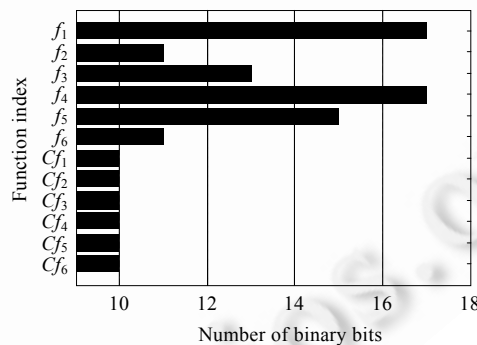


Fig.3 Code length of functions

图3 函数的编码长度

3.2 实验结果

3.2.1 典型测试函数优化实验

在30维的 $f_1 \sim f_6$ 函数优化实验中,EGEA将与OGA/Q^[18],HTGA^[19],MAGA^[16],StGA^[20],LEA^[6]和OEA^[4]这6种在函数优化方面表现出良好性能的算法进行比较.OGA/Q采用的终止条件是在1000代迭代后连续50代解的质量没有提高^[18],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^[19],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^[20],LEA的终

止条件与 OGA/Q 相同^[6],OEA 的终止条件是函数评估次数大于 300 000^[4].EGEA 采用的终止条件是对每个函数都搜索 200 代.表 1 为对每个函数连续 50 次运算后的优化结果.

Table 1 Experimental statistic results of EGEA for $f_1 \sim f_6$ based on 30 dimensions

表 1 EGEA 对 30 维函数 $f_1 \sim f_6$ 的实验统计结果

Function		f_1	f_2	f_3	f_4	f_5	f_6
Generations	Maximum	197	162	153	199	168	62
	Minimum	79	18	29	70	33	9
	Mean	148	70	77	166	108	35
	Standard deviations	24.2	27.6	28.1	31.9	24.1	10.3
Function value	Maximum	-12 569.469 8	0	0	1.4e-4	0	0
	Minimum	-12 569.486 6	0	0	0	0	0
	Mean	-12 569.485 7	0	0	1.25e-5	0	0
	Standard deviations	1.15e-3	0	0	8.67e-4	0	0
Function evaluations	Maximum	35 779	11 695	22 001	33 189	20 881	12 836
	Minimum	14 790	4 005	5 914	15 787	9 324	2 725
	Mean	27 186	8 914	13 409	21 858	18 392	8 106
	Standard deviations	5 494	2 354	3 598	4 782	3 623	1 264

EGEA 在对 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 函数的 50 次优化过程中有 3 次没有在 200 代内收敛到全局最优解.对 f_1 函数搜索的最优结果为-12 569.486 6.

表 2 为 EGEA 与 OGA/Q,HTGA,MAGA,StGA,LEA 和 OEA 这 6 种算法对 30 维 $f_1 \sim f_6$ 函数优化结果的对比.由于编码方式和收敛条件不同,很难比较上述几种算法的优劣,但根据其收敛精度以及花费的函数评估次数,可以从一个侧面反映各种算法针对对不同函数优化能力的差异.

Table 2 Comparison of experimental results between EGEA and 6 algorithms for $f_1 \sim f_6$ based on 30 dimensions

表 2 EGEA 与 6 种算法对 30 维函数 $f_1 \sim f_6$ 实验结果的对比

Function		f_1	f_2	f_3	f_4	f_5	f_6
EGEA	MNFE	27 186	8 914	13 409	21 858	18 392	8 106
	MFV	-12 569.485 7	0	0	1.25e-5	0	0
	(Std)	(1.15e-3)	(0)	(0)	(8.67e-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)
LEA	MNFE	287 365	223 803	105 926	130 498	110 674	110 031
	MFV	-12 569.4542	2.103e-18	3.274e-14	6.104e-16	4.727e-16	4.247e-19
	(Std)	(4.831e-4)	(3.359e-18)	(3.001e-14)	(2.513e-17)	(6.218e-16)	(4.236e-19)
OEA	MNFE	300 019	300 019	300 018	300 020	300 017	300 014
	MFV	-12 569.4866	5.430e-17	5.336e-14	1.317e-2	2.481e-30	2.068e-13
	(Std)	(5.555e-12)	(1.683e-16)	(2.954e-13)	(1.561e-2)	(1.128e-29)	(1.440e-12)

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

在对 f_2 和 f_6 函数优化中,EGEA 花费了最少的函数评估次数找到了全局最优解;

在对 f_1 函数优化中,StGA 是唯一找到最优值的算法,EGEA 花费的函数评估次数和收敛精度要好于 OGA/Q,HTGA 和 LEA;

在对 f_3 函数优化中,MAGA 和 StGA 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 OGA/Q,LEA 和 OEA;

在 f_4 和 f_5 函数优化中,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,StGA,LEA 和 OEA.

综合来说,在对 $f_1\sim f_6$ 函数优化中,MAGA 的表现最佳,EGEA 与其他 5 种算法针对不同函数的优化效果各有优势.

3.2.2 混合函数优化实验

文献[17]中记录了 PSO,CPSO,CLPSO,CMA-ES,G3-PCX 和 DE 这 6 种算法对混合函数的优化结果.每种算法对各函数都进行了 20 次运算,每个函数的最大评估次数被限制为 50 000 次.EGEA 也对每个函数进行了 20 次运算,最大迭代次数限制为 300 代.表 3 为 EGEA 运算后的统计结果.表 4 为 EGEA 与上述 6 种算法优化结果的对比.

Table 3 Experimental statistic results of EGEA for $Cf_1\sim Cf_6$

表 3 EGEA 对函数 $Cf_1\sim Cf_6$ 的实验统计结果

Function		Cf_1	Cf_2	Cf_3	Cf_4	Cf_5	Cf_6
Generations	Maximum	287	289	299	298	296	297
	Minimum	102	155	110	175	122	198
	Mean	221	239	245	257	225	272
	Std	56.7	52.2	54.4	35.6	54.8	32.7
Function value	Maximum	4.90e-2	28.1	75.3	120.0	21.4	375
	Minimum	2.03e-4	1.25e-6	1.2e-4	90.5	1.9	304
	Mean	1.03e-3	4.54	13.2	100.7	2.3	310
	Std	1.36e-3	4.01	37.3	23.6	7.6	5.6
Function evaluations	Maximum	51 079	54 729	52 334	54 807	56 247	55 253
	Minimum	18 527	28 164	19 443	30 867	22 389	35 195
	Mean	39 425	44 374	42 912	48 877	41 660	48 870
	Std	6 975.9	7 128.7	7 563.7	7 418.5	7 103.8	6 390.7

Table 4 Comparison of experimental results between EGEA and 6 algorithms for $Cf_1\sim Cf_6$

表 4 EGEA 与 6 种算法对函数 $Cf_1\sim Cf_6$ 的实验结果的对比

Function		Cf_1	Cf_2	Cf_3	Cf_4	Cf_5	Cf_6
EGEA	MFV	1.03e-3	4.54	13.2	100.7	2.3	310
	Std	1.36e-3	4.01	37.3	23.6	7.6	5.6
PSO	MFV	100.00	155.91	172.03	314.30	83.45	861.42
	Std	81.650	131.76	32.87	20.07	101.11	125.81
CPSO	Mean	156.26	242.29	362.64	522.37	255.56	853.14
	Std	134.27	148.95	196.31	122.09	175.63	127.98
CLPSO	MFV	5.74e-8	19.16	132.81	322.32	5.371	501.16
	Std	1.04e-7	14.75	20.03	27.46	2.61	0.78
CMA-ES	MFV	100.00	161.99	214.06	616.40	358.53	900.26
	Std	188.58	151.00	74.18	671.92	168.26	0.0831
G3-PCX	MFV	60.00	92.699	319.80	492.96	26.02	772.08
	Std	69.921	99.067	125.19	142.49	41.58	189.39
DE	MFV	0.07	28.76	144.41	324.86	10.79	490.94
	Std	0.11	8.63	19.40	14.75	2.60	39.64

在 Cf_1 函数优化实验中,EGEA 达到的最高精度是 10^{-4} ,20 次实验中达到 10^{-3} 精度的有 17 次,其他 6 种算法中只有 CLPSO 的优化结果优于 EGEA.

在 Cf_2 函数优化实验中,EGEA 达到的最高精度是 10^{-6} ,20 次实验中达到 10^{-3} 精度的有 14 次,其他 6 种算法都未能达到这个精度.

在 Cf_3 函数优化实验中,EGEA 达到的最高精度是 10^{-4} ,20 次实验中达到 10^{-3} 精度的有 6 次,其他 6 种算法都未能达到这个精度.

在 Cf_4,Cf_5 和 Cf_6 这 3 个混合函数由多种不同函数组成,使得结构更为复杂,但 EGEA 对这 3 个函数的优化结果都明显要好于上述 6 种算法.

4 结 论

族群是在进化群体之上建立的个体间逻辑层面的组织形式,这种机制可以方便地调控群体结构,有效地协调进化算法全局搜索和局部搜索之间的关系,从而提高进化算法的搜索效率和抗早熟能力.本文设计了针对二进制编码方式的族群分类方法,并基于该族群结构形成了具有双轨协同进化特征的族群进化算法以及相应的族群算子.利用族群的分类能力,双轨协同进化机制融合了基本的“生成+检测”迭代过程和自学习过程,并使二者之间形成了相互利用、相互促进的关系.在对典型多维函数的优化实验中,EGEA 通过与在函数优化方面具有优异性能的 6 种算法的对比,说明族群进化算法是一种有竞争力的多维函数优化算法.为测试族群进化算法解决复杂问题的能力,6 个更具挑战性的混合测试函数被用于测试,EGEA 的优化结果明显要好于文献[17]中记录的几种算法的优化结果.上述结果表明,基于族群结构的进化模型不仅是可行的,而且是有效的,同时,这种进化机制也为群体智能优化提供了一个良好的框架,因此,下一步我们将用该算法解决其他优化问题.

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老师和同学表示衷心的感谢!

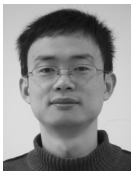
References:

- [1] Rudolph G. Convergence analysis of canonical genetic algorithms. *IEEE Trans. on Neural Networks*, 1994,5(1):96–101.
- [2] Jiao LC, Wang L. A novel genetic algorithm based on immunity. *IEEE Trans. on System, Man, and Cybernetics—Part A: Systems and Humans*, 2000,30(5):552–561.
- [3] Wang CX, Cui DW, Wan DS, Wang L. A novel genetic algorithm based on gene therapy theory. *Trans. of the Institute of Measurement and Control*, 2006,28(3):253–262.
- [4] Liu J, Zhong WC, Jiao LC. An organizational evolutionary algorithm for numerical optimization. *IEEE Trans. on Systems, Man, and Cybernetics—Part B*, 2007,37(4):1052–1064.
- [5] Zhang J, Chung HSH, Lo WL. Clustering-Based adaptive crossover and mutation probabilities for genetic algorithms. *IEEE Trans. on Evolutionary Computation*, 2007,11(3):326–335.
- [6] Wang YP, Dang CY. An evolutionary algorithm for global optimization based on level-set evolution and Latin squares. *IEEE Trans. on Evolutionary Computation*, 2007,11(5):579–595.
- [7] Xie T, Chen HW, Kang LS. Evolutionary algorithms of multi-objective optimization problems. *Chinese Journal of Computers*, 2003,26(8):997–1003 (in Chinese with English abstract).
- [8] Sun JG, Liu J, Zhao LY. Clustering algorithms research. *Journal of Software*, 2008,19(1):48–61 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/48.htm>
- [9] Gan J, Warwick K. Dynamic niche clustering: A fuzzy variable radius niching technique for multimodal optimization in GAs. In: Kim JH, ed. *Proc. of the IEEE Congress on Evolutionary Computation*. Seoul: IEEE Press, 2001. 215–222.
- [10] Ling Q, Wu G, Yang ZY, Wang QP. Crowding clustering genetic algorithm for multimodal function optimization. *Applied Soft Computing*, 2006,8(1):88–95.
- [11] Ortiz-Boyer D, Hervás-Martínez C, García-Pedrajas N. Improving crossover operator for real-coded genetic algorithms using virtual parents. *Journal of Heuristics*, 2007,13(3):265–314.
- [12] García-Martínez C, Lozano M, Herrera F, Molina D, Sánchez AM. Global and local real-coded genetic algorithms based on parent-centric crossover operators. *European Journal of Operational Research*, 2008,185(3):1088–1113.
- [13] 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.
- [14] Wolpert DH, Macready WG. No free lunch theorems for optimization. *IEEE Trans. on Evolutionary Computation*, 1997,1(1):67–82.
- [15] Reynolds RG, Zhu S. Knowledge-Based function optimization using fuzzy cultural algorithms with evolutionary programming. *IEEE Trans. on Systems, Man and Cybernetics—Part B: Cybernetics*, 2001,31(1):1–18.

- [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] Liang JJ, Suganthan PN, Deb K. Novel composition test functions for numerical global optimization. In: Kennedy J, ed. Proc. of the IEEE Swarm Intelligence Symp. Pasadena: IEEE Press, 2005. 68–75.
- [18] Leung YW, Wang Y. An orthogonal genetic algorithm with quantization for global numerical optimization. IEEE Trans. on Evolutionary Computation, 2001,5(1):41–53.
- [19] Tsai JT, Liu TK, Chou JH. Hybrid Taguchi-genetic algorithm for global numerical optimization. IEEE Trans. on Evolutionary Computation, 2004,8(4):365–377.
- [20] Tu ZG, Lu Y. A robust stochastic genetic algorithm (StGA) for global numerical optimization. IEEE Trans. on Evolutionary Computation, 2004,8(5):456–470.

附中文参考文献:

- [7] 谢涛,陈火旺,康立山.多目标优化的演化算法.计算机学报,2003,26(8):997–1003.
- [8] 孙吉贵,刘杰,赵连宇.聚类算法研究.软件学报,2008,19(1):48–61. <http://www.jos.org.cn/1000-9825/19/48.htm>



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.



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.



崔颖安(1975—),男,博士生,讲师,主要研究领域为智能计算,分布式计算.



陶永芹(1960—),女,博士生,讲师,主要研究领域为智能计算,知识进化.



梁琨(1976—),女,工程师,主要研究领域为分布式计算.