

族群进化算法的经验学习机制

陈皓¹, 崔杜武¹, 崔颖安^{1,2}, 陶永芹^{1,2}

CHEN Hao¹, CUI Du-wu¹, CUI YIN-an^{1,2}, TAO Yong-qin^{1,2}

1.西安理工大学 计算机学院, 西安 710048

2.西安交通大学 电子信息工程学院, 西安 710049

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

E-mail: chen hao_xaut@yahoo.com.cn

CHEN Hao, CUI Du-wu, CUI YIN-an, et al. Experience-learning mechanism of ethnic group evolution algorithm. Computer Engineering and Applications, 2008, 44(29): 27-30.

Abstract: Based on the classifying process according to the coding similarity among typical individuals, a kind of population organization mechanism, ethnic group, has been created in chaotic population, and a novel evolution model, ethnic group evolution algorithm (EGEA), is developed. Based on ethnic group evolution mechanism, some typical individuals can be sifted out from population conveniently and a kind of experiential knowledge can be discovered in it. In this paper, an experience-learning mechanism is used to improve EGEA and the simulations show the experience-learning mechanism is feasible and valid, which improves the searching efficiency of EGEA greatly.

Key words: genetic algorithm; ethnic group evolution algorithm; ethnic group experience learning mechanism

摘要: 族群是依据个体编码特征的相似性对群体进行分类后形成的一种群体结构化组织, 基于该机制形成了一种新的进化模型—族群进化算法(ethnic group evolution algorithm, EGEA)。族群机制可有效调控群体结构, 协调算法的全局搜索和局部搜索时间, 同时利用其所具有的分类能力也可方便地获取群体中的典型个体。设计了族群的经验学习机制来挖掘蕴含于群体中的进化经验知识, 并利用这些知识来引导群体的搜索, 提高 EGEA 的收敛速度。仿真实验表明族群的经验学习机制不仅是可行的而且是有效的, 它显著提高了 EGEA 的搜索效率。

关键词: 遗传算法; 族群进化算法; 族群经验学习机制

DOI: 10.3778/j.issn.1002-8331.2008.29.007 **文章编号:** 1002-8331(2008)29-0027-04 **文献标识码:** A **中图分类号:** TP18

1 前言

在遗传算法(Genetic Algorithm, GA)“生成+检测”的迭代机制中, 群体既是信息搜索的途径也是信息传递的载体, 其结构直接影响着 GA 的搜索效率。传统的 GA 群体是随机而无序的, 但为了使群体结构更适合于问题求解, 一些进化算法通过调节群体结构来提高自身的搜索性能, 如小生境技术^[1]以及结合聚类方法的小生境技术^[2]。这类策略主要是通过选择操作等被动的过程来间接地调整群体结构。而组织进化算法^[3]则明确地在群体中建立了清晰的组织形式, 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。这类研究表明, 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)作为社会群体的基本单元, 并以族群为观点来分析社会群体的结构和演变趋势以及族群间的相互竞争对社会发展的影响。受该思想启发, 文献[4]提出从编码相似性的角度对群体进行结构分析以掌握群体中蕴含的进化趋势, 并以各进化趋势为单元来控制群体的搜索过程以达到调控群体结构的目的。该文将对群体分类而形成的组织形式称为族群, 并将基于族群结构的进化模型称为族群进化算法(Ethnic Group Evolution Algorithm, EGEA)。

根据 NFL 理论^[5], 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, 自学

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60743009, No.60501006); 陕西省自然科学基金(the Natural Science Foundation of Shaanxi Province of China under Grant No.2006F-43); 陕西省教育厅科学技术研究计划项目(the Science and Technology Research Project of Shaanxi Province Department of Education No.07JK327)。

作者简介: 陈皓(1978-), 男, 博士研究生, 主要研究领域为进化计算与数据挖掘; 崔杜武(1945-), 男, 教授, 博士生导师, 主要研究领域为进化计算、计算机网络和多媒体技术; 崔颖安(1975-), 男, 博士研究生, 主要研究领域为智能计算与分布式计算; 陶永芹(1960-), 女, 博士研究生, 主要研究领域为智能计算与知识进化。

收稿日期: 2008-05-15 **修回日期:** 2008-07-23

习能力已成为进化计算改进的一个重要方向,而在此类研究中如何获取有效的知识则成为了一个关键技术。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^[6]。但 GA 的一个主要优势在于其搜索机制不需对问题本身的特征有过多的了解,从这个角度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在于缺乏对有价值典型个体的筛选机制,文化算法^[7]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 可以方便地从群体中筛选出典型个体,基于这种能力提出了族群的经验学习机制。该机制通过挖掘典型个体中蕴含的进化经验知识来引导群体的进化方向,促进信息交换、提高 EGEA 的收敛速度。将这种具有经验学习能力的族群进化算法(L-EGEA)应用于多维函数优化问题的求解,发现族群的经验学习过程不仅是可行的而且是有效的,它显著提高了 EGEA 的搜索效率。

2 族群进化算法(EGEA)简介

文献[4]首先从二进制编码的角度设计了基本的族群进化框架以及相应的评估指标,其过程主要包括族群分类、族群繁殖以及下代群体选择三个步骤。设 EGEA 的群体 N 由 n 个 m 维个体组成 A_i , 其适应度为 $F_i > 0$, F_{best} 为当前群体中的最优适应度值,第 k 代群体中的族群 E_k 的数量为 ξ , 族群后代个体为 C_k 。

ALGORITHM 1 EGEA

```

01: begin
02:  $k=0$ ;
03: initialize population  $N_0$  randomly;
04: evaluate the race exponent( $N_0$ );
05: while(未达到终止条件)do
06: begin
07:  $E_k$ =ethnic group clustering( $N_k$ );
08: for  $i=1$  to  $\xi$  do
09: while(族群  $E_k(i)$  的后代数量未达到)do
10:  $C_k(i)$ =mutation(crossover(mate( $E_k(i)$ ),  $N_k$ )));
11: evaluate the race exponent( $N_k$ ,  $C_k$ );
12:  $N_{k+1}$ =select next population( $N_k$ ,  $C_k$ );
13:  $k=k+1$ ;
14: end;
15: 搜索  $A_{best}$  的邻域输出优化结果;
16: end.

```

为了更全面地评价个体的质量,EGEA 使用竞争指数(race exponent) Q_i 作为个体 A_i 的评价指标

$$Q_i = D_i^{\bar{R}_i}$$

其中 \bar{R}_i 为个体 A_i 相对当前群体中的最优个体 A_{best} 的相对适应度比; D_i 为二者的加权编码差异率。基于竞争指数,高于群体平均适应度且与群体中最优个体具有较大差异的个体将获得更强的竞争能力。

为协调群体在信息搜索和传递间的关系,EGEA 引用生物的有性繁殖过程中大配子的概念来对个体进行角色分工。大配

子是依据个体的竞争指数从群体中筛选出的部分典型个体,其承载了群体所拥有的主要进化信息。以大配子为对象进行族群分类,既可降低计算量也可保证所产生族群的有效性和代表性。在二进制编码空间中,依据个体编码的差异性,可在大配子中建立一类族群结构。这种族群结构具有清晰的组织形式,但仅是大配子与族群间从属关系的一种记录,所以具有结构简单易于管理的特点。

族群是从特定角度对群体所蕴含进化趋势的分类,EGEA 以族群为单位来控制群体的繁殖,协调进化机会在各种进化趋势类型间的分配。族群繁殖过程使用具有择偶功能的多点交叉算子以及自适应变异算子产生后代个体。择偶多点交叉算子首先通过择偶过程(mating)分别从当前族群所属的大配子和整个群体中依据竞争指数以轮盘赌选择的方式挑选两个父个体 M_i 和 A_j , 并判断二者的般配率

$$\bar{\omega}_{ij} = D_{ij} (Q_i + Q_j)$$

其中 D_{ij} 为 M_i 与 A_j 的编码差异率。若 $\bar{\omega}_{ij} > \text{Random}(0, 1)$ 则择偶成功,否则需要重新择偶。接着通过交叉和变异过程产生后代个体,其变异概率为

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

当一代中的族群繁殖过程完成后,将由选择算子挑选产生下代群体,并完成群体迭代过程。当群体迭代的终止条件被满足时,一个针对当前群体中的最优个体 A_{best} 进行的邻域搜索用来提高最终解的精度。

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,并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。

3 具有经验学习能力的族群进化算法(L-EGEA)

族群中最具竞争力的一个个体可作为一个族群的代表,在族群的经验学习机制中可利用族群进化经验矩阵(Experience Matrix, EM)来存储若干个此类个体,以 EM 为基础通过拼接实验来重组个体,以分离出蕴含于典型个体中有价值的基因片段作为进化的经验知识,并用这类知识影响下代群体的结构,以实现引导和加速群体进化的目的。

3.1 族群进化经验矩阵

以维为单位,可以将一个族群中的典型个体切割为 m 个信息单元(Information Unit, IU),该信息单元的基本结构如下

$$IU = (BC, RC, WV)$$

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。如果群体中存在 e 个族群时,则族群经验矩阵可以表示为

$$EM = \begin{Bmatrix} IU_{11} & \cdots & IU_{1j} & \cdots & IU_{1m} \\ \vdots & & \vdots & & \vdots \\ IU_{e1} & \cdots & IU_{ej} & \cdots & IU_{em} \end{Bmatrix}$$

3.2 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 EM: $IU_{ij} (i=1, \dots, e, j=1 \dots m)$ 的权值为 $W_{ij} = F_i / F_{best}$ 。

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, 并组合产生试验个体 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 $\rho = F_{test} / F_{best}$ 。

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 $MV' = MV + \rho$; 如果 $\rho < 1$, 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 筛选出一些有价值的信息单元。

3.3 教育算子

教育算子以一定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。

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 $P_{education} = 1 - Q_i$ 。

步骤 2 对 A_i 的每一维依次进行判断, 如果 $P_{education} > Rand(0, 1)$, 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; 否则保留该维的信息;

步骤 3 评估被教育个体的适应度值, 并用其替换原有个体。

3.4 L-EGEA 的执行过程

以下为 L-EGEA 的执行过程: 其中教育比例 $\Psi \in (0, 1)$ 用

于设定将接受教育操作的个体数量。

ALGORITHM 2 L-EGEA

```

01: begin
02:   初始化过程;
03:   while(未达到终止条件)do
04:     begin
05:       建立族群  $E_k$  及族群繁殖并选择产生群体  $N_{k+1}$ ;
06:        $EM_k = \text{initialize experience matrix}(E_k)$ ;
07:        $N_{k+1} = \text{education}(\text{splice experiment}(EM_k), \Psi)$ ;
08:       evaluate the race exponent( $N_{k+1}$ );
09:     end;
10:   搜索  $A_{best}$  的邻域输出优化结果;
11: end.
    
```

4 多维函数仿真实验

为验证族群自学习机制的有效性, 本文挑选了 6 个典型的多维函数作为数值仿真验证对象。30 维测试函数 $f_1 \sim f_4$ 和 $f_5 \sim f_6$ 分别为文献[8]中的函数 $f_1 \sim f_4$ 和 $f_7 \sim f_8$ 。函数每维的编码长度保证搜索误差小于 10^{-3} , 群体规模 $n=100$ 。EGEA 和 L-EGEA 分别对 6 个函数进行 200 代优化, 表 1 为连续 50 次实验后两种算法的统计结果。表 2 为 L-EGEA 与 MAGA^[9]、StGA^[9]、OGA/Q^[10]、HTGA^[11] 以及 OEA^[3] 五种具有优异性能算法对 6 个函数优化结果的对比。

表 1 EGEA 与 L-EGEA 对 30 维函数 $f_1 \sim f_6$ 优化统计结果

Function	f_1	f_2	f_3	f_4	f_5	f_6	
EGEA	MNFE	-12 568.256 4	6.8e-2	0.16	0.06	0.45	0
	Std	8.7e-1	5.6e-1	8.3e-2	9.2e-3	6.7e-2	0
	MFV	36 478	18 745	30 258	31 452	20 687	9 654
	Std	6 547	3 562	4 893	6 895	4 165	2 983
L-EGEA	MNFE	-12 569.485 7	0	0	1.25e-5	0	0
	Std	1.15e-3	0	0	8.67e-4	0	0
	MFV	27 186	8 914	13 409	21 858	18 392	8 106
	Std	5 494	2 354	3 598	4 782	3 623	1 264

MNFE: Mean Number of Function Evaluation; MFV: Mean Function Value; Std: Standard deviations

表 2 L-EGEA 与 5 种算法对 30 维函数 $f_1 \sim f_6$ 优化结果对比

Function	f_1	f_2	f_3	f_4	f_5	f_6	
L-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)
OEA	MNFE	300 019	300 019	300 018	300 020	300 017	300 014
	MFV	-12 569.486 6	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)

L-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 次优化过程中有三次没有在 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。EGEA 仅对 f_6 函数的 50 次实验中在 200 代内找到了最优解,对其它函数的优化精度要差于 L-EGEA,而且对所有函数花费的函数评估次数都要多于 EGEA。这说明族群的经验学习机制显著提高了 EGEA 的搜索效率。

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^[10],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^[11],MAGA 的终止条件是对每个函数都搜索 150 代^[8],StGA 搜索到最优解的代数基本在 200 代以前^[9],OEA 的终止条件是函数评估次数大于 300 000^[3]。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应各种算法针对不同函数优化能力的差异。在对 f_2 和 f_6 函数优化中 EGEA 使用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 和 HTGA;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 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 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 与其它 4 种算法针对不同函数的优化效果各有优势。这些比较说明族群的经验学习机制有效提高了 EGEA 的搜索速度和收敛的稳定性,使 EGEA 成为了一种具有竞争力的数值优化算法。

5 结论

本文提出了一种基于族群机制的进化经验学习方法。通过对典型多维函数的优化实验表明这种改进的族群进化算法具有较好的抗早熟能力,与其他典型算法的比较也说明该算法具有性能稳定、搜索效率高等特点,是一种具有竞争力的数值

优化算法。

参考文献:

- [1] Sareni B, Krahenbuhl L. Fitness sharing and niching methods revisited[J]. IEEE Transactions on Evolutionary Computation, 1998, 2(3): 97-106.
 - [2] Qing L. Crowding clustering genetic algorithm for multimodal function optimization[J]. Applied Soft Computing, 2006, 8(1): 88-95.
 - [3] Liu Jing, Zhong Weicai, Jiao Licheng. An organizational evolutionary algorithm for numerical optimization[J]. IEEE Transactions on Systems, Man, and Cybernetics: Part B, 2007, 37(4): 1052-1064.
 - [4] Chen Hao, Cui Du-Wu, Li Xue, et al. The harmonious evolution of ethnic group algorithm[C]//Lei Jingsheng. Third International Conference on Natural Computation (ICNC2007), Haikou, China. IEEE Computer Society, 2007, 4: 380-384.
 - [5] Wolpert D H, Macready W G. No free lunch theorems for optimization[J]. IEEE Transactions on Evolutionary Computation, 1997, 1(1): 67-82.
 - [6] Jiao Licheng, Wang Lei. A novel genetic algorithm based on immunity[J]. IEEE Transactions on System, Man, and Cybernetics-Part A: Systems and Humans, 2000, 30(5): 552-561.
 - [7] Sareni B, Krahenbuhl L. Fitness sharing and niching methods revisited[J]. IEEE Transactions on Evolutionary Computation, 1998, 2(3): 97-106.
 - [8] Zhong W C, Liu J, Xue M Z, et al. A multiagent genetic algorithm for global numerical optimization[J]. IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, 2004, 34(2): 1128-1141.
 - [9] Tu Zhenguo, Lu Yong. A robust stochastic genetic algorithm (StGA) for global numerical optimization[J]. IEEE Transactions on Evolutionary Computation, 2004, 8(5): 456-470.
 - [10] Leung Y W, Wang Y. An orthogonal genetic algorithm with quantization for global numerical optimization[J]. IEEE Transactions on Evolutionary Computation, 2001, 5(1): 41-53.
 - [11] Tsai Jinn-Tsong, Liu Tung-Kuan, Chou Jyh-Horng. Hybrid taguchi-genetic algorithm for global numerical optimization[J]. IEEE Transactions on Evolutionary Computation, 2004, 8(4): 365-377.
-
- (上接 10 页)
- 有助于推动它在国内的广泛应用和发展。本文实现的改进的双字哈希算法经过测试具有良好的分词效果,利用基于它的分词器和 Lucene 工具包,设计并实现了一个视频点播影片搜索引擎,方便用户快速搜索到感兴趣的影片,为提高用户使用点播门户的体验奠定了良好基础。提出的搜索引擎通过缓存技术减轻了网络负载和数据库的访问压力,通过倒排索引技术和双字哈希技术提高了搜索的效率,还支持批量更新索引、模糊查询、多种数据源输入,具有良好的扩展性。
- ## 参考文献:
- [1] Jakarta Lucene home page[EB/OL]. <http://jakarta.apache.org/Lucene>.
 - [2] Vecera R, Rozsnyai S, Roth H. Indexing and search of correlated business events[C]//The Second International Conference on Availability, Reliability and Security, 2007: 1124-1134.
 - [3] Pirrò G, Talia D. An approach to ontology mapping based on the Lucene search engine library [C]//18th International Conference on Database and Expert Systems Applications, DEXA'07, 2007: 407-411.
 - [4] 李庆虎, 陈玉健, 孙家广. 一种中文分词词典新机制—双字哈希机制[J]. 中文信息学报, 2003, 17.
 - [5] 曹红. 林业主题搜索引擎研究[D]. 北京: 北京林业大学, 2005: 25-26.
 - [6] 黄杰. 基于 Lucene 的全文检索系统模型的研究[D]. 广州: 暨南大学, 2007.
 - [7] 车东. 基于 Java 的全文搜索引擎简介[EB/OL]. <http://www.chedong.com/tech/lucene.html>.
 - [8] 王继明, 杨国林. 基于 Lucene 的中文文本分词[J]. 内蒙古工业大学学报, 2007, 26(3): 185-186.
 - [9] 周平. Lucene 全文检索引擎技术及应用[J]. 重庆工学院学报: 自然科学版, 2007, 21.
 - [10] 张校乾, 金玉玲, 侯丽波. 一种基于 Lucene 搜索引擎的全文数据库的研究与实现[J]. 现代图书情报技术, 2005(2).