

引用格式:孔盼盼,崔英花.多重自适应性遗传算法的研究[J].微电子学与计算机,2019,36(3):53-57. [KONG Pan-pan,CUI Ying-hua. Research on multiple adaptive genetic algorithm [J]. Microelectronics & Computer,2019,36(3):53-57.]

多重自适应性遗传算法的研究

孔盼盼,崔英花

(北京信息科技大学 信息与通信工程学院,北京 100101)

摘 要: 为了改善遗传算法“早熟”的问题,增强算法的全局收敛能力并加快收敛速度,提出了一种新的多重自适应遗传算法。由于种群在不同的遗传阶段具有不同的特征,所以将遗传代数分为三个阶段,每一阶段采用不同的选择策略,加快算法的收敛,并且设计了一种基于种群多样性和个体适应度的多重自适应交叉概率和变异概率,增强算法的全局搜索能力。通过分别对两处改进的仿真对比,验证了改进的有效性,对交叉和变异概率的改进加快了收敛速度和精度,改善了“早熟”问题,对选择策略的改进在不影响收敛精确度的前提下,进一步加快了收敛速度。新算法在搜索全局最优解和收敛速度上具有较好的综合表现。

关键词: 自适应;遗传算法;种群多样性;收敛速度;“早熟”问题

中图分类号: TP301.6 **文献标识码:** A **文章编号:** 1000-7180(2019)03-0053-05

Research on multiple adaptive genetic algorithm

KONG Pan-pan,CUI Ying-hua

(Institute of Information and Communication, Beijing Information Science and Technology University, Beijing 100101, China)

Abstract: An improved multiple adaptive genetic algorithm is proposed to solve the "premature" problem of genetic algorithm, enhance the global convergence ability of the algorithm and accelerate the convergence speed. For the population has different characteristics in different genetic stages, the generation number is divided into three stages, and each stage adopts different selection strategies to accelerate the convergence of the algorithm. A new kind of population diversity and individual fitness based adaptive crossover probability and mutation probability are also designed to enhance the global search ability of the algorithm. Through the comparison of two improved simulations, the effectiveness of the improvement is verified. The improvement of the crossover and mutation probabilities accelerates the convergence speed and accuracy and solves the "premature" problem. By improving the selection strategy, the speed of convergence is further accelerated without affecting the convergence accuracy. The new algorithm has a good overall performance in the search for global optimal solution and convergence speed.

Key words: adaptive; genetic algorithm; population diversity; convergence speed; prematurity

1 引言

遗传算法(Genetic Algorithm,GA)是一种模仿生物群体进化的随机搜索和优化算法,以适者生存为基础,在潜在的解决方案中逐渐求得一个近似最

优解^[1]。遗传算法有着鲜明的优势:它不依赖于问题的具体领域,具有良好的并行性和可操作性。因此,遗传算法已在函数优化、机器学习、图像识别、网络构造等许多领域得到了广泛应用^[2]。

遗传算法的优越性得到学术界的一致认可,然

收稿日期:2018-07-09;修回日期:2018-08-14

基金项目:国家自然科学基金(61340005);北京市自然科学基金面上项目(4132012);北京市教委科技发展计划资助项目(KM201411232011);北京市优秀人才培养资助 D 类项目(2013D005007000006)

而,基本遗传算法也存在诸如局部搜索能力差、收敛速度慢和“早熟”等缺点^[3].很多专家学者在群体大小、交叉概率和变异概率等方面提出了各种改进的自适应遗传算法(Adaptive GA, AGA)^[5-9].Srinivas^[4]等第一次提出了自适应遗传算法(Adaptive GA, AGA),交叉和变异概率可随适应度值的变化而变化,改善了由固定交叉和变异概率造成的“早熟”问题.文献[5-7]做了进一步改进,提出了不同的交叉和变异概率公式,改善了传统 AGA 算法中群体中最大适应度个体的交叉和变异概率为零的缺陷,避免了优良个体进入进化停滞状态.文献[7]提出了一种可根据种群相异度调节交叉和变异概率的遗传算法,进一步加快了收敛.但这些算法在交叉概率和变异概率公式的设计上,要么是只根据个体适应度来调整交叉和变异概率,没有从遗传算法的全局角度进行分析,要么是根据种群相异度来调整交叉和变异概率,没有考虑个体的好坏.

本文在综合考虑了不同进化代数种群遗传特征的不同和个体优良差异之后,提出了一种多重自适应遗传算法(Multiple Adaptive GA, MAGA).

2 自适应遗传算法及其改进

自适应遗传算法是对基本遗传算法的改进,通过自适应调整算法的参数,包括交叉概率和变异概率等,改善了由于固定交叉和变异概率而造成的“早熟”问题,加快了算法的收敛速度,提高了收敛精确度.

自适应遗传算法的详细步骤为:

- (1)编码.将表示问题可行解的个体进行编码.
- (2)初始化群体.根据取值范围,随机产生一定规模的初始种群.
- (3)计算适应度.用适应度来评价个体的优劣.
- (4)遗传操作.循环进行以下操作,直到达到最大遗传代数:
 - ①选择.通过选择操作来形成新的种群.
 - ②交叉和变异.根据概率公式自适应调整交叉和变异概率来实现交叉和变异操作.
- (5)输出适应度最大的个体.

自适应算法中,交叉概率和变异概率的设计会直接影响到算法的收敛,交叉概率的大小影响新个体的产生速度,交叉概率越小,旧的种群越不容易破坏,这不会延缓新个体的产生,更会导致算法的局部收敛,造成“早熟”的问题,但是如果交叉概率过大,就会使原本比较优良的个体遭到破坏,使算法收敛变慢.

变异概率是决定算法跳出局部最优解的一个关键因素.变异概率过小,不容易产生新的个体模式,反之过大,对最优解的破坏比较大,会使 GA 算法变成纯粹的随机搜索算法^[4].所以,找到合适的交叉和变异概率对整个遗传算法的性能是至关重要的.

自从 Srinivas 等提出了 AGA 算法之后,一些专家学者提出了一些改进的自适应遗传算法.张京钊等^[7]提出了一种新型自适应遗传算法(New Adaptive GA, NAGA),改进的交叉概率和变异概率公式如下:

$$P_c = \begin{cases} \frac{P_{c1}}{(P_{c1} - P_{c2}) + \exp(\frac{f' - f_{avg}}{f_{max} - f_{avg}})} & f' \geq f_{avg} \\ P_{c1} & f' < f_{avg} \end{cases} \quad (1)$$

$$P_m = \begin{cases} \frac{P_{m1}}{(P_{m1} - P_{m2}) + \exp(\frac{f' - f_{avg}}{f_{max} - f_{avg}})} & f' \geq f_{avg} \\ P_{m1} & f' < f_{avg} \end{cases} \quad (2)$$

式中, f_{avg} 是种群中所有个体的平均适应度值; f_{max} 是种群中最大的个体适应度值; f' 是两个要交叉的个体中适应度值较大的个体的适应度值; f 是要变异的个体的适应度值; P_{c1}, P_{c2} 取 $(0, 1)$ 之间的值; P_{m1}, P_{m2} 取 $(0, 0.1)$ 之间的值.可以看出,当个体适应度值大于等于 f_{avg} 时,交叉和变异概率随适应度的变化而变化,缺点是当个体适应度值小于 f_{avg} 时,交叉和变异概率不变.

文献[8]提出了一种新的根据种群相异度来调节交叉和变异概率的改进方法.相异度的具体定义如下:

假设种群中有 M 个个体,每个个体有 N 个基因,这样就可以用一个 $M \times N$ 的矩阵 $\mathbf{X}$ 来表示该种群,首先计算第 j 列元素的标准差 d_j 表示第 j 列基因的变异度,矩阵 $\mathbf{X}$ 第 j 列的标准差为:

$$d_j = \left[\frac{1}{M} \sum_{i=1}^M (x_{ij} - \overline{x_j})^2 \right]^{1/2} \quad (3)$$

式中, $\overline{x_j}$ 为第 j 列元素的平均值.

然后将 d_j 映射到 $[0, 1]$ 内,得到变换后的衡量各基因差异度的参数 d_j' ,映射关系如下:

$$d_j' = \frac{2}{\pi} \arctan(2d_j) \quad (4)$$

最后用 $\text{diff} = \frac{1}{N} \sum_{j=1}^N d_j'$ 来表示种群的相异度, diff 值越大,说明个体之间越不同,也就是种群的多

样性越好. 交叉和变异概率公式如下:

$$p_c = \begin{cases} 0.55 + 0.45\text{diff} & 0.55 + 0.45\text{diff} \leq 0.95 \\ 0.95 & 0.55 + 0.45\text{diff} > 0.95 \end{cases} \quad (5)$$

$$p_m = 0.01k_m(1 - d')\left(\frac{f}{f_{\text{avg}}}\right) \quad (6)$$

式中, $k_m > 0$. 从公式可以看出, 交叉概率与种群相异度有关, 变异概率与个体适应度有关. 这些算法概率公式的设计没有综合考虑个体适应度与种群相异度对算法进化的影响, 只关注了其中一种因素的影响.

3 多重自适应遗传算法

本文改进的算法在交叉和变异概率公式设计上采用基于种群相异度和个体适应度的双重自适应交叉和变异概率, 同时为进一步加快收敛, 改进了选择策略, 使算法的全局搜索能力更强, 进化更快更准. 下面分别对两处改进进行分析.

3.1 MAGA 的交叉和变异操作

基于种群相异度和个体适应度的双重自适应交叉和变异概率公式如下:

$$P_c = \begin{cases} P_{c1} \frac{(1 - \text{diff}) \exp \frac{f_{\text{avg}} - f'}{f_{\text{max}} - f_{\text{min}}}}{f_{\text{avg}} - f_{\text{min}}} & f' < f_{\text{avg}} \\ P_{c2} \frac{(1 - \text{diff}) \exp \frac{f_{\text{max}} - f'}{f_{\text{max}} - f_{\text{min}}}}{f_{\text{max}} - f_{\text{avg}}} & f' \geq f_{\text{avg}} \end{cases} \quad (7)$$

$$P_m = \begin{cases} P_{m1} \frac{(1 - \text{diff}) \exp \frac{f_{\text{avg}} - f}{f_{\text{max}} - f_{\text{min}}}}{f_{\text{avg}} - f_{\text{min}}} & f < f_{\text{avg}} \\ P_{m2} \frac{(1 - \text{diff}) \exp \frac{f_{\text{max}} - f}{f_{\text{max}} - f_{\text{min}}}}{f_{\text{max}} - f_{\text{avg}}} & f \geq f_{\text{avg}} \end{cases} \quad (8)$$

式中, 由于个体适应度较大时, 交叉概率和变异概率取较小的值, 所以 $P_{c1} > P_{c2}$, $P_{m1} > P_{m2}$, 都取 $(0, 1)$ 之间的值.

从公式可以看出, 当个体的适应度低于当代种群的平均适应度时, 认为该个体性能不佳, 将对其采用较大的交叉率和变异率. 当个体适应度接近当代种群中的最大适应度时, 认为该个体性能较好, 应尽可能保留其优良模式, 所以对其采用较低的交叉率和变异率. 当种群的相异度较小时, 交叉和变异概率会变大, 这样可以增加种群多样性, 避免算法进入局部收敛, 提高全局搜索能力. 反之, 当种群多样性大时, 采用较小的交叉和变异概率, 有利于算法的收敛.

3.2 MAGA 的选择操作

由于在不同的进化阶段, 种群的进化特征不同, 所以本文提出的 MAGA 算法将总的进化代数 gen 分为三个阶段: $[0, 0.4)\text{gen}$, $[0.4, 0.7)\text{gen}$, $[0.7, 1]\text{gen}$, 对不同遗传阶段进行不同的选择操作, 具体的实现步骤总结如下:

(1) $[0, 0.4)\text{gen}$: 这一阶段属于进化的初始阶段, 主要目的是要保留种群的多样性, 所以采取轮盘赌选择结合最优保留法, 既能保证被选择到的个体的多样性, 又能使优良基因不被丢失, 在后面的进化过程中发挥作用, 防止算法的局部收敛;

(2) $[0.4, 0.7)\text{gen}$: 种群进化到这个阶段, 出现了个体的优劣差距, 因此这一阶段考虑淘汰掉适应度比较低的个体, 保留适应度高的个体. 将个体按适应度从高到低排序, 然后分成三份, 将第一份直接复制成两份, 第三份随机淘汰 $1/2$, 在第二部分中随机选择 $1/2$ 补充第三部分淘汰的部分, 加上第一部分复制的两份, 就形成了新的种群, 很明显, 新种群的适应度要比旧种群的高;

(3) $[0.7, 1]\text{gen}$: 最后这一阶段, 目的是要提高算法的收敛速度, 采取的措施是, 按适应度从高到低排序, 直接复制两份前 $1/2$ 个体, 这样选择到的个体都是适应度排在前面的个体, 可以有效地加快算法的收敛速度.

4 仿真测试

4.1 映射函数的选择

在计算种群相异度时, 对 d_j 的映射函数提出了另一种算法:

$$d_j' = \tanh(d_j) = \frac{e^{d_j} - e^{-d_j}}{e^{d_j} + e^{-d_j}} \quad (9)$$

为了验证第二种算法的优势, 对公式(4)和(9)进行仿真对比. 选择策略采用轮盘赌选择法, 遗传代数为 400 代, 种群规模为 30, 交叉和变异概率公式为公式(7)和(8), $p_{c1} = 0.8$, $p_{c2} = 0.6$, $p_{m1} = 0.06$, $p_{m2} = 0.03$, 测试函数:

$$\max f(x) = x \sin(10\pi x) + 2 \quad x \in [-2, 2] \quad (10)$$

该函数最大值在 3.85 附近. 图 1 是用两种映射函数的每一代的最大适应度值的仿真对比.

从图 1 中可以看出, 使用 $\arctan$ 函数对最大适应度个体的破坏较大, 容易使函数进入局部最优, 产生“早熟”问题, 而 $\tanh$ 函数对最大适应度个体的破坏较小, 有利于算法的收敛, 所以 MAGA 算法的映

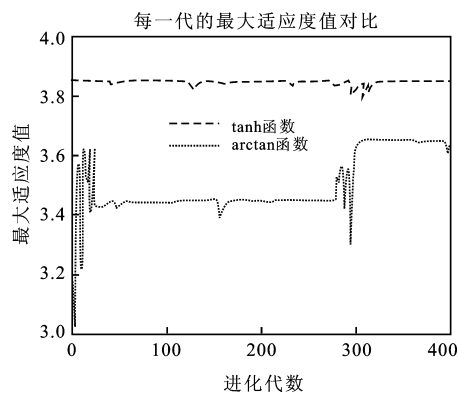


图1 两种映射函数的最大适应度值对比

射函数选择 tanh 函数。

4.2 MAGA 的交叉和变异概率改进的仿真测试

为了验证改进的交叉和变异概率公式的有效性,本文选择了 NAGA。文献[8]的改进方法和本文提出的改进交叉和变异概率公式的算法三种算法,分别对测试函数(8)进行仿真.通过三种算法适应度值的变化情况、收敛速度和最优解的优劣三个方面的对比,来评估算法的性能。

三种算法都采用轮盘赌结合最优保留的选择策略,遗传代数为 400 代,种群规模为 30,NAGA 和本文改进算法: $p_{c1}=0.8$, $p_{c2}=0.6$, $p_{m1}=0.06$, $p_{m2}=0.03$,文献[8]: $k_m=2.5$ 。

三种算法的每一代的最大适应度值的仿真结果对比如图 2,从图中可以看出,NAGA 算法的搜索能力最差,速度慢且最优解精度差,文献[8]的算法的搜索能力改善,改进交叉和变异概率后的算法搜索能力进一步增强,根据种群多样性和个体适应度值双重自适应的交叉和变异概率对最优个体的破坏减小,防止早熟问题的出现,使算法收敛更快。

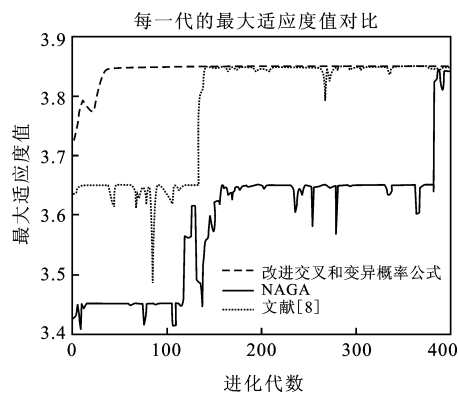


图2 三种算法的最大适应度值变化对比

图 3 是随机抽取的 10 次仿真结果的收敛代数情况图,表 1 是这 10 次的平均收敛情况.可以看出,

在种群规模和迭代次数相同的情况下,改进后的算法的稳定性更好,在收敛代数和最优解的精确度上都要优于另两种算法。

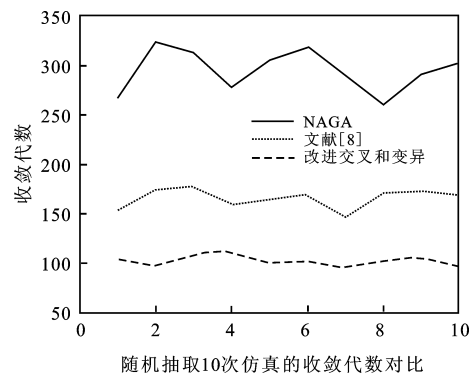


图3 三种算法的收敛代数对比

表 1 平均收敛情况对比

算法	平均收敛代数	平均收敛值
NAGA	295.3	3.753 628
文献[8]	167.9	3.806 692
本文改进	105.2	3.847 253

4.3 MAGA 的选择策略改进的仿真测试

对比的两种算法都采用了本文改进的交叉和变异概率公式,其中第一种算法采用了轮盘赌和最优保留的选择策略,另一种算法采用的是本文改进的选择策略。

图 4 是两种算法平均适应度值的对比情况,表 2 是抽取 10 次仿真结果的平均收敛情况表,从图表可以看出,在保证最优解的精确度不被降低的前提下,通过改进选择策略,可以进一步提高算法的收敛速度。

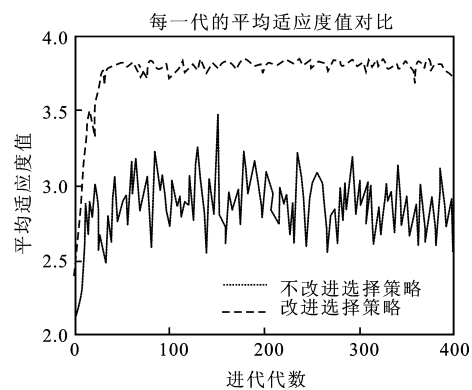


图4 平均适应度值变化对比

表 2 平均收敛情况对比

算法	平均收敛代数	平均收敛值
不改进选择策略	107.3	3.845 374
改进选择策略	72.6	3.846 243

5 结束语

针对遗传算法存在的“早熟”问题并进一步加快算法收敛速度、提高收敛精度^[9-10],本文对遗传算法做了两部分的改进:改进交叉和变异概率公式与改进选择策略,通过分别仿真对比两个改进方法,可以看出,通过改进交叉和变异概率公式能够提高收敛精度和速度,提高算法的全局搜索能力,通过改进选择策略可以进一步加快收敛速度.实验证明,本文提出的 MAGA 算法能够加快收敛速度,有效避免局部收敛和提高算法的稳定性和精确度.

参考文献:

[1] 谭丹丹,曹斌,刘俊,等.一种基于配对的改进遗传算法[J]. 计算机应用,2007,12(27):170-171.

[2] 曲志坚,张先伟,曹雁锋,等.基于自适应机制的遗传算法研究[J]. 计算机应用研究,2015,32(11):3222-3225.

[3] 王小平,曹立明.遗传算法:理论、应用与软件实现[M]. 西安:西安交通大学出版社,2002.

[4] SRINIVAS M, PATNAIK L M. Adaptive probabili-

ties of crossover in genetic[J]. IEEE Trans Sys, Man and Cybem, 1994, 24(4): 656-667.

[5] 卢长娜,王如云,陈耀登.自适应遗传算法[J]. 计算机仿真,2006,23(1):172-175.

[6] 邝航宇,金晶,苏勇.自适应遗传算法交叉变异算子的改进[J]. 计算机工程与应用,2006,42(12):93-96.

[7] 张京钊,江涛.改进的自适应遗传算法[J]. 计算机工程与应用,2010,46(11):53-55.

[8] 王慧妮,彭其渊,张晓梅.基于种群相异度的改进遗传算法及应用[J]. 计算机应用,2006,26(3):668-669.

[9] 沐阿华,周绍磊,于晓丽.一种快速自适应遗传算法及其仿真研究[J]. 系统仿真学报,2004,15(1):122-125.

[10] ROYACHKA K, KAROVA M. High-Performance optimization of genetic algorithms [C]//ISSE' 06. 29thInternational Spring Seminar Electronics Technology. Rome, Italy,2006: 395-400.

作者简介:

孔盼盼 女,(1994-),硕士研究生.研究方向为 RFID 性能测试. E-mail:SunLightKong@163.com.

崔英花 女,(1973-),教授,硕士生导师.研究方向为无线射频识别技术.

(上接第 52 页)

[6] 张宏亭,李学仁,孔韬. BP 神经网络在缺失数据估计中的应用[J]. 计算机工程与设计,2007,28(14):3457-3459.

[7] 刘树洁. 基于神经网络风电场测风数据插补与长年代风速推算方法研究[D]. 湖北:武汉大学,2017.

[8] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J]. 计算机学报,2017,40(6):1229-1251.

[9] 王璨,武新慧,李恋卿,等.卷积神经网络用于近红外光谱预测土壤含水率[J]. 光谱学与光谱分析,2018,38(1):36-41.

作者简介:

张网娟 女,(1992-),硕士研究生.研究方向为大数据、数据管理. E-mail: zhang_wang_juan@foxmail.com.

许国艳 女,(1971-),博士,副教授.研究方向为大数据、数据起源、数据管理.

李敏佳 女,(1994-),硕士研究生.研究方向为社交网络、数据管理.

朱 帅 男,(1992-),硕士研究生.研究方向为社会网络、数据管理.