

动态变异遗传算法

张宇^{*1} 郭晶² 周激流²

(1. 成都航空职业技术学院电子工程系 成都 610061; 2. 四川大学电子信息学院 成都 610064)

【摘要】遗传算法是根据达尔文生物进化理论而提出的一种优化算法。该文提出了一种新的遗传算法，理论分析显示，它不仅能保持遗传种群的多样性，而且能快速收敛。计算机仿真实验证明了改进后的遗传算法能够有效地克服不成熟收敛、进而搜索到全局最优解，并将这种新遗传算法用于BP网络的拓扑结构的优化和连接权值的训练，实例表明了该算法的有效性和可行性。

关键词 遗传算法; 变异算子; BP网络; 优化算法

中图分类号 TP302.6

Dynamic Mutation Genetic Algorithm

Zhang Yu¹ Guo Jing² Zhou Jiliu²

(1. Dept. of Electronic Engineering, Chengdu Aeronautic Vocational & Technical College Chengdu 610061;

2. College of Electronic and Information, Sichuan University Chengdu 610064)

Abstract Genetic Algorithms is Optimal Algorithm, which employ a search technique based on ideas from Darwin's natural evolution theory. A new genetic algorithm (NGA) is proposed in this paper, which not only can keep the population diversity but also has quicker convergence speed. The experiment results show that the improved genetic algorithm can efficiently find global optimal beyond premature convergence. Finally, using the NGA, optimizing the topology and training the weights for BP neural network are done. The results of the applications show that the new genetic algorithm is practical and efficient.

Key words genetic algorithm; mutation; BP neural network; optimal algorithm

遗传算法(GA: Genetic Algorithm)是一种以达尔文自然进化理论和孟德尔遗传变异理论为基础的全局随机搜索优化计算技术，它同时考察多个候选解、模拟生物遗传进化过程、淘汰劣质解、鼓励发展优质解、逐步提高解群体的质量，直至收敛，并获得(准)最优解。同时，由于GA仅需知道如何根据解求得其相应的适应度，而不要求适应度函数满足连续可微等条件，因而应用广泛。但是，典型遗传算法(CGA: Canonical Genetic Algorithm)在不断搜索更优解的过程中将不可避免地使种群的多样性渐趋于零，常使种群进化收敛于局部极值，造成不成熟收敛现象，这是长期困扰GA，进而影响GA进一步应用的主要原因。

本文根据文献[1]分析的造成CGA不成熟收敛的原因，提出了一种以改进变异算子为核心的遗传算法，该算法在发生不成熟收敛时能从当前局部解中跳出来并向其附近的另一个具有更高适应度的局部/全局解收敛，因而具有能迅速地搜索到全局最优解的优点。

1 不成熟收敛的原因分析

典型遗传操作由选择、交叉、变异三算子构成。由于通常变异概率较小，故变异使种群发生显

2002年1月19日收稿

* 女 40岁 大学 讲师

著改变的概率极小, 故遗传操作中选择和交叉算子起主要作用。

若当前种群各个体具有的共同模式记 H , 则交叉算子有能力搜索且仅能搜索 H 包含的所有个体^[1]。当种群多样性越大时, H 的阶就越小, 从而包含更多的潜在解, 交叉算子的搜索能力就越强, 但 H 以概率1单调递减到单点型早熟集(从而种群多样性以概率1递减到0), 选择和交叉算子在起搜索作用的同时, 不可避免地使种群多样性渐趋于零, 从而不断缩小自己的搜索范围, 最终导致不成熟收敛。因此, 选择和交叉算子具有二重性, 既可以强迫算法收敛, 使其达到问题的全局最优, 又可能在搜索过程中不成熟收敛而使算法终止于非(全局或局部)最优状态, 并且成熟化效应是造成遗传算法不成熟收敛的起因。

2 对典型遗传算法CGA的改进

从理论上分析, 只要有变异存在, 进化必然不收敛, 从而也不会发生不成熟收敛。但如果在进化过程中一直使用较大的变异概率, 则迭代过程趋于纯随机搜索, 使算法丧失模拟进化的仿生实质^[2]。因此, 先不考虑不成熟收敛的发生, 而是设法加速收敛过程, 以缩短总的搜索时间, 当不成熟收敛发生时, 加大变异概率, 使用特殊的变异算子, 使种群从局部极值中跳跃到新的搜索空间, 特殊的变异算子将保证变异产生的个体有相当大的概率落入全局最优解附近, 并能向该全局最优解收敛。

2.1 加快收敛速度的措施

为了加快收敛速度, 本文对适应度进行了标准化处理, 以突出较优解的优势, 首先, 求得种群适应度值的平均值 \bar{f} 和方差 s

$$\bar{f} = \sum_{i=1}^N f_i / N \quad (1)$$

$$s = \sqrt{\sum_{i=1}^N (f_i - \bar{f})^2 / N} \quad (2)$$

再将各适应度值标准化

$$f'_i = \frac{f_i - \bar{f}}{s} \quad (3)$$

若标准化后的适应度最小值小于0, 则将所有适应度值都加上适应度最小值的绝对值, 以保证适应度函数值非负。设 $f_{\min} = \min\{f'_i\}$, $i=1, 2, \dots, N$; 若 $f_{\min} < 0$, 则 $f'_i = f'_i + |f_{\min}|$, $i=1, 2, \dots, N$ 。

2.2 新变异算子

通常二进制编码的变异算子是随机选中一个变异位, 将该位上的二进制码取反^[3, 4], 这种借鉴纯随机搜索, 在实践中对帮助种群跳出局部解的作用不大。

新变异算子基于这样一个事实: 当进行函数优化时, 对于多峰函数, 在某局部解 S_A 的周围, 与 S_A 距离为 r 处必有一个适应度更高的局部解 S_B , 当陷入局部解 S_A 时, 种群中绝大多数个体落在 S_A 的极小邻域内。若使这些个体均发生变异, 且变异的范围为一个半径为 $r_{\min} \sim r \sim r_{\max}$ 的圆环, 则有相当大的概率使某个个体变异后会落入局部解 S_B 的邻域, 从而促使种群向 S_B 收敛。

以两变量问题为例, 其具体的变异方法是待变异个体 S_i 解码得到 (x, y) , 然后随机产生 r 和 q , 使 $r_{\min} \sim r \sim r_{\max}$, $0 \leq q < 2\pi$, 则 $\Delta x = r \cos q$, $\Delta y = r \sin q$, 再将 $(x + \Delta x, y + \Delta y)$ 编码生成 S'_i , 即为变异结果。

2.3 新变异算子的物理及数学解释

以Schaffer函数极值点求解为例, 其函数表达式为

$$f(x, y) = 0.5 - \frac{\sin^2 \sqrt{x^2 + y^2} - 0.5}{[1 + 0.001(x^2 + y^2)]^2} \quad (4)$$

其自变量搜索范围为 $-10 \leq x, y \leq 10$ ，该函数在最大值峰(1.0)周围有一圈脊，取值均为0.990 283，本文采用如下的适应度变换函数

$$F = \frac{a}{a + (1 - f)} \quad a = 0.1 \quad (5)$$

将适应度取值限定在 $[0, 1]$ 范围内，且最大值(1.00)与次最大值(0.91)之间有较明显的区别。变换后的适应度函数 F 的图像在 $y=0$ 处的剖面图如图1a所示， F 的俯视示意图如图1b所示。在图1b中，矩形标示出函数的搜索范围(记作 S_T)，正中的实心圆点表示全局最大值的适应度值大于0.91的邻域(记作 S_O)，围绕正中实心圆点的圆表示全局最大值附近的一圈脊(该圆记作 C_B ，其上各点的适应度值均为0.91)。

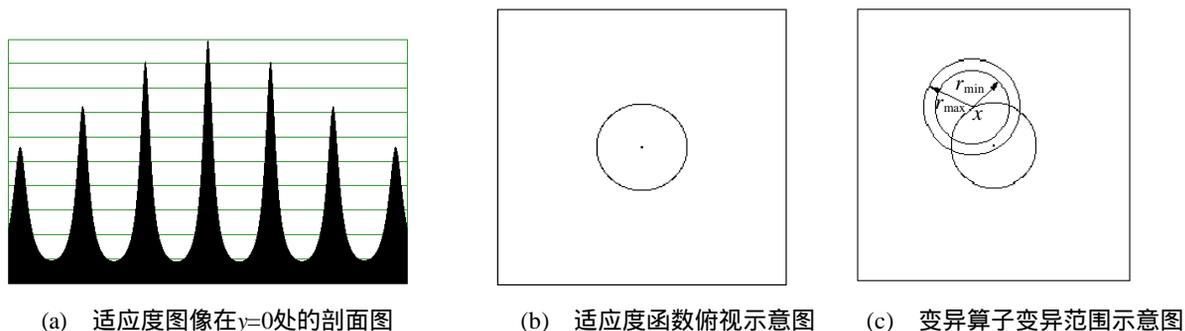


图1 Schaffer函数示意图

在选择算子和最佳个体保存法的作用下，若种群中有一个以上的个体落入 S_O 范围内，则因为其适应度值高于搜索空间的其他区域，种群就可以向全局最大值逐渐收敛。因 S_O 与 S_T 相比太小，故单靠初始种群的随机生成及选择、交叉算子作用，很难有个体落入其中，所以，进化较易收敛于 C_B 上，陷入不成熟收敛。又因常规的变异算子基于纯随机搜索方法， C_B 上某点的变异结果可能为 S_T 中任一点，其落入 S_O 中的概率(S_O/S_T)相当小，导致变异后落入非 S_O 区域的个体又逐渐收敛于 C_B 上，因而采用常规的变异算子不能使进化有效地跳出不成熟收敛。而本文提出的变异算子把变异范围由全搜索空间 S_T 缩小到一个由内半径 r_{\min} 和外半径 r_{\max} 界定的圆环(此圆环面积记作 S_C)，如图1c所示。图中 X 点表示不成熟收敛种群中的被变异的一个个体。该圆环包含 S_O 区域，对位于该圆环圆心的个体 X 采用本文提出的变异算子进行变异，其结果必然落入 S_C 中，因此，其进一步落入 S_O 的概率为 S_O/S_C 。因为 S_C 远小于 S_T ，所以概率 S_O/S_C 远大于概率 S_O/S_T ，故在新变异算子的作用下，从不成熟收敛区域跳跃到全局最优解邻域的概率比采用常规变异算子时的概率大得多。

2.4 新变异算子中半径 r 的动态调整策略

确定 r_{\min} 和 r_{\max} 本身是一个优化问题。因为 r_{\min} 和 r_{\max} 与适应度函数的数学特性有关，在不知其特性的情况下，很难给出具体的推导规则。但可借鉴Grefenstette利用GA来优化GA的控制参数的方法^[5]，即利用GA来优化 r_{\min} 和 r_{\max} ，但计算量很大。为此，本文提出如下策略，可以有效地动态调整 r ，使改进后的遗传算法能快速跳出不成熟收敛，进而找到全局最优解。

首先，由用户指定搜索范围，以此定出变异半径的最大允许取值 R_{\max} ，而变异半径的最小允许取值 $R_{\min} = 0$ 。再取 $r_{\min} = 0$ ， $r_{\max} = r_{\min} + r_{\text{step}}$ ，其中 r_{step} 是由用户指定值，变异在此范围内进行。若种群已收敛于某局部解 S_A ，而在其附近具有更大适应度的另一个局部解 S_B 已包含在变异圆环内，则经过变异后有一个以上个体落入 S_B 邻域的期望迭代次数可近似表示为

$$n = \left\lceil 1 - \left(1 - \frac{S_o}{\pi r_{\max}^2 - \pi r_{\min}^2}\right)^N \right\rceil a \quad (6)$$

式中 S_o 表示 S_B 邻域中适应度值大于 S_A 的适应度值的区域的面积; r_{\max} 为变异圆环外半径; r_{\min} 为变异圆环内半径; N 为种群规模; a 为容错系数。容错系数 a 的设置主要是因为变异的随机性, 增大 a 可使迭代在不调整 $[r_{\min}, r_{\max}]$ 的情况下多变异几代, 但过大的 a 会降低搜索速度。

在得到期望变异迭代次数 n 后, 若在 n 代内通过变异都未使种群最佳适应度上升, 则说明当前的 $[r_{\min}, r_{\max}]$ 不合适, 应调整为 $r'_{\min} = r_{\min} + r_{\text{step}}$, $r'_{\max} = r_{\max} + r_{\text{step}}$, 然后重新计算 n 值并再试。当 r_{\max} 达到 R_{\max} 时, 则 $[r_{\min}, r_{\max}]$ 的调整方向应反向, 即 $r'_{\min} = r_{\min} - r_{\text{step}}$, $r'_{\max} = r_{\max} - r_{\text{step}}$, 直至 r_{\min} 达到 0 后又再反向。

2.5 新遗传算法(NGA)全局寻优能力实验

在实验中, 本文采用二进制编码, 种群规模取为500。

Schaffer函数采用常规变异算子和新变异算子的迭代过程如图2所示。图中水平方向表示遗传迭代次数, 垂直方向表示适应度值, 粗曲线为种群平均适应度, 细曲线为种群当前最佳适应度。

图2a表明, 采用常规变异算子的CGA算法经过多次变异均不能超越次最优解而向全局最优解搜索。而图2b表明, 当进化收敛于局部解时, 新的变异算子使种群中的部分个体有能力跳跃到全局最优解附近, 从而促使种群向全局最优解收敛。

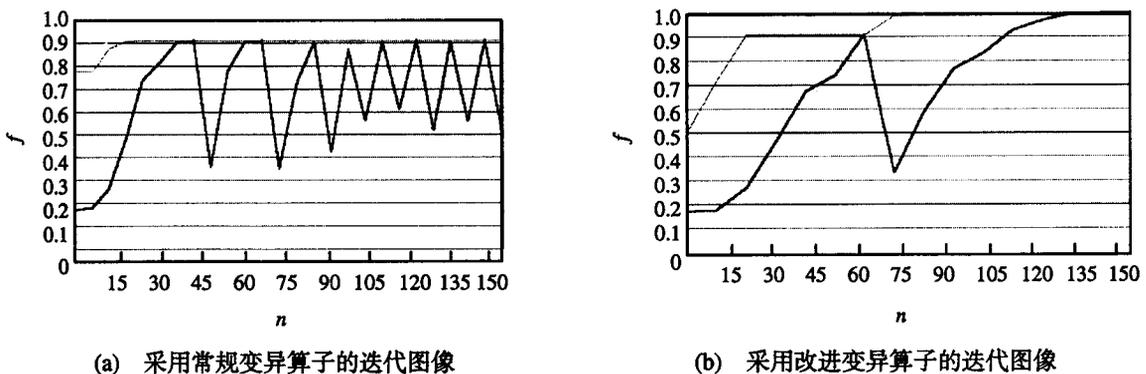


图2 Schaffer函数采用两种变异算子的迭代过程

3 NGA在BP网络学习中的应用

3.1 BP网络的学习能力

BP神经网络的学习包括拓扑结构和权值两个方面。由于Kosmogorov定理证明了三层结构的前馈网络可以逼近任意的联系函数, 所以, 通常情况下BP网络由输入层、隐层、输出层组成, 每层中都包含若干个神经元。神经网络拓扑结构的学习即指确定隐单元的具体个数, 但该问题目前没有一个满意的解决方案, 通常是依赖经验公式; 神经网络权值的学习一般要付出大量的调试时间和高昂的计算代价, 由于BP网络的训练基于梯度下降法, 因而受到了梯度法固有的两个弱点的影响: 1) 学习陷入局部极小后不能跳出来, 2) 学习进入平坦区时优化效率变低且时间变长。而且, 根据实验结果(见表1)还可以看出, 当指定的最大允许均方误差越小(即精度要求越高), 则训练时间大大增加。

表1中, * 表示迭代失败。由于GA具有不要求目标函数连续, 也不要求可微, 仅要求可计算, 且搜索空间始终遍及整个解空间, 容易得到最优解等优点, 因此, 众多学者都把GA应用到对神经网络的拓扑结构设计和连接权重的调整中来^[6,7]。

表1 BP网络学习迭代次数与最大允许均方差关系表

指定的最大允许均方差	第1次实验	第2次实验	第3次实验	第4次实验	第5次实验
0.100	409	419	333	525	615
0.050	758	733	887	831	804
0.010	11 329	10 871	*	10 967	11 143
0.005	41 219	40 402	40 661	*	41 014
0.004	*	63 638	*	64 084	61 110

3.2 NGA与BP网络的学习能力

本文采用NGA算法具体实现了对一个解决XOR问题的三层前馈网络的拓扑结构和连接权重的同时优化,其具体步骤如下。

3.2.1 编码

根据设定隐单元数的经验公式,初始设定网络结构为2-5-1。因此,网络共有15个连接权重、6个阈值。编码采用二值编码,每个权重占用16位,其中最高位为通断位(1表示接通,0表示断开),其余15位用来编码 $[-10,10]$ 范围内的权重,故量化误差为 $\frac{20}{2^{15}} \approx 6.10 \times 10^{-4}$;每个阈值占用16位,编码 $[-10,10]$ 范围内的可能取值,量化误差为 $\frac{20}{2^{16}} \approx 3.05 \times 10^{-4}$ 。

3.2.2 适应度函数

由于每个个体是一套完整的网络的结构和权重设置的编码,将其解码就对应一个具体的网络,然后将4个学习样本(见表2)输入该网络并求得均方差 E ,转换得

$$F = \frac{a}{a + E} \quad (7)$$

式中 E 为网络输出的均方差; a 为灵敏度控制系数,通常 a 应根据应用的实际情况确定,本文取 $a = 0.01$ 。

对于隐层某神经元,连接它的权重有3个,其中2个输入、1个输出,若其中之一通断位为0,则认为此神经元应处于断开状态,于是将其输出置为0。由此可见,断开一个神经元的几率较大,但若导致该神经元断开的通断位经过交叉或变异变为1,则该神经元又可处于接通状态,因此,神经元由断开切换为接通的几率也较大。

3.2.3 遗传算子

选择算子采用赌轮法,交叉算子采用两点交叉,变异算子为正常情况下随机选中个体中一位取反,当不成熟收敛发生时,采用本文提出的变异策略,其进化过程如图3所示,图中横坐标表示迭代次数,纵坐标表示适应度值,细线为种群中的最佳个体适应度值,粗线为种群的平均适应度值。表3~6是整个网络最终的优化结果,从表3~6可以看出,一个2-5-1的网络经GA优化后拓扑结构定为2-2-1,且分类的均方差为 4.09786×10^{-7} ,大大优于用BP算法学习的结果。

表2 XOR网络的学习样本

样本序号	输入值	期望输出
1	(0,0)	0
2	(0,1)	1
3	(1,0)	1
4	(1,1)	0

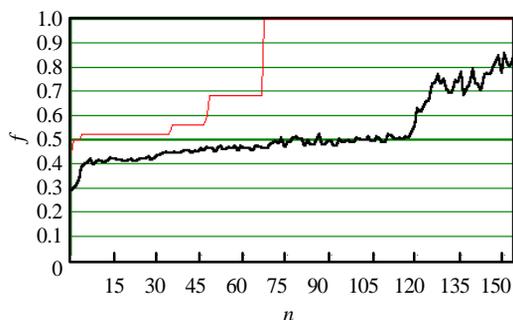


图3 GA训练BP网络的进化过程图像

表3 GA训练BP网第一层权重结果示例

W_{ij}	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
$i=0$	0	0	- 8.064 52	- 3.529 47	0
$i=1$	0	0	4.691	7.076 33	0

表4 GA训练BP网第二层权重结果示例

W_{ik}	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
$k=0$	0	0	- 4.618 37	8.121 28	0

表5 GA训练BP网各层域值结果示例

q_j^l	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
$l=0$	0	0			
$l=1$	0	0	- 5.276 96	5.197	0
$l=2$	- 1.895 5				

表6 训练后BP网络对XOR问题的实验结果

输入模式	分类结果
(0,0)	0.000 114 342
(0,1)	1
(1,0)	0.998 198 000
(1,1)	0.000 131 143
均方误差	4.097 86e - 007

4 结束语

变异是遗传算法的重要操作之一,既可以产生种群中没有的较优基因,也可以恢复在先前迭代中被破坏的较优基因,所以它对于加快搜索速度、确保种群不至于陷入局部解方面具有重要的意义。从实验结果表明:1) 本文提出的新变异算子及其变异半径的动态调整,在函数的优化过程中,能有效地克服不成熟收敛进而搜索到全局最优解;2) 用NGA可同时优化神经网络的结构和权重,且优化得到的网络的均方误差比用BP算法训练的还小,且优化时间较短。

参 考 文 献

- 1 徐宗本,高 勇. 遗传算法过早收敛现象的特征分析及其预防. 中国科学E辑, 1996, 27(4): 364-375
- 2 Dinabandhu B, Pal N R, Pal S K. Directed Mutation in Genetic Algorithms. Information Sciences, 1994, 79(3): 251-270
- 3 Whitley D, Starkeather T, Bogart C. Genetic Algorithms and Neural Networks: Optimizing Connections and Connectivity. Parallel Computing, 1990, 14: 347-361
- 4 Whitley D, Hanson T. Optimizing Neural Networks Using Faster, More Accurate Genetic Search. Proceedings of the Third International Conference on Genetic Algorithms, 1989, 391-396
- 5 Grefenstette J J. Optimization of Control Parameters for Genetic Algorithms. IEEE Trans System Man and Cybernetics, 1986, 16(1): 122-128
- 6 Whitley D, *et al.* Genetic algorithms and neural networks: Optimizing connecticons and connectivity. Parallel Computing, 1990, 14: 347-361
- 7 Kitano H. Designing neural network using Genetic algorithm with graph generation system. Complex System, 1990, 4: 461-476