

用于神经网络结构优化的改进遗传算法

赵宏伟 臧雪柏 王立江 凌兴宏

(吉林工业大学信息学院计算机科学与工程系 长春 130025)

摘 要 前馈型神经网络因其结构简单和非线性映射能力而具有广泛的工程应用前景。在研究和应用中,网络的构造是非常重要的。将遗传算法引入神经网络的结构设计,用遗传算法优化神经网络的拓扑结构以及确定 BP 算法的相应参数,对遗传算法的编码方案作了改进,从而避免了无隐层网络结构,提出了一种新的比例变换适应值函数比例系数自适应调整方法,最后给出该算法的实现过程和应用实例。

关键词 人工神经网络,结构优化,遗传算法

中图法分类号 TP18

AN IMPROVED GENETIC ALGORITHM FOR FEEDFORWARD NEURAL NETWORK STRUCTURE OPTIMIZATION

ZHAO Hong-Wei, ZANG Xue-Bai, WANG Li-Jiang, and LING Xing-Hong

(Department of Computer Science and Engineering, College of Information, Jilin University of Technology, Changchun 130025)

Abstract Because of the simple structure and the nonlinear mapping ability, feedforward neural networks (FNN) can be widely used in engineering applications. The structures of FNN are essential. In the paper, a genetic algorithm (GA) is introduced into the design of FNN, to optimize the FNN structures and the parameters of BP. To avoid the situation of no hidden layer structure, an improved GA coding method is developed. A new proportional transform adapter function is introduced in the proportional coefficient adjusting. Finally, the steps of realization and applications are illustrated.

Key words artificial neural networks, structural optimization, genetic algorithms

1 引 言

自 1986 年以 Rumelhart 和 McClelland 为首的 PDP(parallel distributed processing)研究小组发表多层网络学习的误差逆传播算法(BP 算法)以来,人工神经网络的理论和应用研究发展十分迅速,其中,前馈型神经网络因其结构简单和非线性映射能力而具有广泛的工程应用前景,如模式识别、过程控制、故障诊断等,同时,多层前馈神经网络在人工智能研究中又被作为机器学习知识获取的重要工具。

Kolmogorov 已经证明了多层前馈神经网络能够解决可以表示成从某个输入空间到某个输出空间的连续映射的任意问题。然而这只是一个存在性定理,它没有给出构造网络架构的方法和提示。神经网络的拓扑

原稿收到日期:2000-07-17。本课题得到国家自然科学基金(项目编号 59675059)资助。赵宏伟,男,1962 年生,副教授,主要从事计算机仿真与科学计算可视化、人工神经网络、智能制造等方面的研究与教学工作。臧雪柏,女,1963 月生,高级工程师,主要研究方向为计算机仿真与科学计算可视化、人工神经网络。王立江,男,1934 年生,教授,博士生导师,主要研究方向为智能制造、机械加工。凌兴宏,男,1969 年生,硕士,主要研究方向为计算机仿真与科学计算可视化、人工神经网络。

结构对其功能有很大的影响,但是如何构造和优化神经网络的结构,目前尚无理论指导,在实际应用中通常采用凑试法、增长法、修剪法等设计方法。凑试法凭借先验知识构造不同的网络结构,分别进行训练并评价网络性能,从中确定性能最好的网络。凑试法简单易行,但过程复杂且依赖经验,加之初始条件的网络敏感性及存在局部极小,致使不易找到满意的网络结构。增长法在开始时构造一个小规模的网络结构,训练过程中,针对实际问题,根据网络性能要求逐步增加结构的复杂性,直至满足性能要求为止,这种方法有两个主要缺点,①可能的结构空间较大,探测好的结构十分耗时,而且大部分结构将不能被测试;②网络结构依赖于应用要求,既要考虑求解的问题,又要考虑对神经网络解的限制,但目前还没有一个好的技术或方法做到这一点。修剪法与增长法相反,在开始时构造一个含有冗余节点的大规模网络结构,在训练过程中逐步删除那些不必要的节点,降低网络复杂性,直至满足性能要求,但在冗余结构下会产生“过学习”现象,使网络的训练学习了样本模式的确切特征而不是样本集的一般特征,缺乏外推能力而难以实际应用。

研究神经网络的最终目的是解决实际问题,欲实现神经网络,其拓扑结构至关重要,它关系到实现的难易程度、网络可靠性、稳定性、抗干扰性等问题。为了进一步揭示网络拓扑结构与网络性能之间的关系,需要开拓新的研究和分析方法,而遗传算法正具有这种特性。

遗传算法是一种模拟生物界自然选择和自然遗传机制的高度并行、随机、自适应最优化搜索算法,具有隐含的并行性和对全局信息的有效利用能力。它只需搜索少数结构就能反映搜索空间的大量区域,利用群体

的适应值信息,通过简单的复制、杂交和变异算子,就能以很大的概率找到全局最优解。利用遗传算法可以找出网络的规模、结构和学习参数,自动搜索一个最优的网络设计,部分学者对此做了一些研究工作。本文在前期工作的基础上,对遗传算法的编码方案作了改进,从而避免了无隐层网络结构,提出了一种新的比例变换适应值函数比例系数自适应调整方法,最后给出该算法的实现过程和应用实例。

2 用遗传算法设计神经网络

用遗传算法设计神经网络的流程如图1所示。本文主要讨论前馈神经网络的结构设计。

染色体串可以编码前馈多层神经网络的结构、连接形式和BP算法的控制参数(如学习率、动量项、初始权值范围等)。一种染色体串二进制编码方案是:设网络最多有两个隐含层,每一个隐含层6位编码,1位表示有/无隐含层,5位表示该隐含层中神经元节点数(神经元的最大数目为32)。学习速率、动量系数、初始权值范围分别用2位编码。这样一个染色体串用18位二进制编码。

在遗传算法中,适应值用来区分群体中个体的好坏,适应值越大的个体越好,适应值越小的个体越差,以此保证适应性好的个体有机会在下一代中产生更多的子个体。

3 遗传算法的改进

遗传算法是一种参数搜索算法,采用适应度比例选择、单点杂交和单点变异的遗传算法称为简单遗传算法,它的缺陷是收敛速度慢、易陷入局部极小等。在应用遗传算法中,要首先给定一组控制参数,如群体规模、杂交率和变异率等,控制参数的不同选取对遗传算法的性能产生很大的影响;各种算法是针对实际问题的,都有各自的特点,采用不同的选择、杂交和变异策略,对算法的性能会有很大的影响。

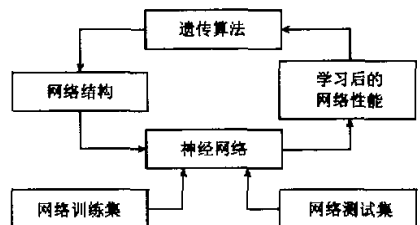


图1 遗传算法设计神经网络

存在如下缺点:

(1)缺乏全局性和稀疏性.单隐层的网络结构多于双隐层的结构,且单隐层的神经元数相对集中在1~32中间的一些数上,有可能重复;

(2)出现无隐层的现象.

采用二进制编码,虽然比较容易操作,但如果要表示较大的范围,则串的长度比较大.从理论上已经证明:具有短定义长度、低阶并且适应值高于群体平均值的模式,在遗传算法迭代中将按指数率增长,对群体中 N 个串的处理实际上隐含地并行处理了大约 $O(N^3)$ 个模式^[1].从短定义长度、低阶模式的角度出发,对上述传统二进制编码方案作如表1的改进.改进的编码方案不仅位串少了一位,而且表示的范围较传统的编码范围大,表示的模式也较多,保证了遗传算法搜索的全局性和稀疏性,同时也避免了出现无隐层的现象.

表1 改进遗传算法染色体串编码结构

序号	域名称	二进制位数	取值说明
1	学习速率	2	{0.5, 0.25, 0.125, 0.0625}
2	动量系数	2	{0.9, 0.8, 0.7, 0.6}
3	初始权值范围	2	{ ± 1 , ± 0.5 , ± 0.25 , ± 0.125 }
4	多层结构	1	0 表示单隐层,神经元数为序号5、6之和;2~64 1 表示双隐层
5	第1隐层节点数	5	1~32
6	第2隐层节点数	5	1~32

3.2 适应值的比例变换

适应值的度量问题是一个很重要的问题.在简单遗传算法中通常采用指数比例变换

$$E = \text{训练样本集平均相对误差} + \text{测试样本集平均相对误差},$$

适应值函数定义为

$$Syz = \exp(\min(E)/E) \quad (1)$$

适应值范围在1~2.718之间,这种指数比例变换在搜索初期可限制优势个体的复制数量,防止算法早熟收敛,但到搜索后期对于提高相似串之间的竞争性不利,因为到搜索后期按此式计算出的适应值差别很小.

对式(1)加以修改,将指数比例变换为下面的关系式:

$$Syz = \exp(\beta \min(E)/E) \quad (2)$$

指数比例可以使适应值高的串保持多的复制机会,同时又限制了它的复制数目以避免其很快控制整个群体.系数 β 决定了选择的强制性, β 越大,选择越趋向于那些具有高适应值的串.

对于系数 β 的确定,本文提出了一种新的自适应调整方法:

$$\beta = \text{avg}(E) / (\max(E) - \text{avg}(E)) \quad (3)$$

式中 $\text{avg}()$ 表示取平均值; $\max()$ 取最大值.并且对 β 做如下限制:如果 β 大于4,则 $\beta=4$;如果小于1,则 $\beta=1$.这样在搜索初期, $\text{avg}(E) < (\max(E) - \text{avg}(E))$,也就是说进化群体在解空间中较分散时,则给予较小的 β 值,限制性能好的群体的复制数目,以免其很快控制整个群体;当 $\text{avg}(E) > (\max(E) - \text{avg}(E))$,也就是说在搜索后期,整个群体的性能值都比较接近的情况下,我们采用较大的比例系数,提高相似串间的竞争性.

3.3 自适应杂交率和变异率的设计

在简单遗传算法(SGA)或标准遗传算法(CGA)中,变异率是个常数.通常,对于变异率是一常量的情况,经过多次迭代后,群体的素质会趋于一致,这样就形成了“近亲繁殖”,近亲繁殖对后代的质量不利.同样,如果双亲的基因码链非常接近,杂交以后其后代对于双亲,素质提高也极少.因此,群体基因的多样性变差不仅会减慢进化历程,而且可能导致进化停滞,过早收敛于非最优解.目前许多学者都认识到变异率需要随着遗传进程而自适应变化,这种有组织性性能的遗传算法具有更高的鲁棒性、全局最优性和效率.文献[2]提出一

种杂交率 P_c 和变异率 P_m 。随基因操作的在线性能自适应变化的有效方法,性能提高则 P_c 增加,反之则 P_m 增加;文献[1]研究了 P_m 随迭代次数变化的效果,实例证明 P_m 随指数下降有较好性能;文献[3]提出一种自适应变异方式, P_m 与一对父串间的海明距离成反比,结果显示能有效保持基因的多样性;可见,在今后的研究中,遗传算法结构、基因操作和参数都会向自组织的形式发展并将进行系统的综合。

借鉴上述文献思想,提出一种自适应杂交和变异概率,其公式如下:

$$p_c = \begin{cases} 0.8(f_{\max} - f_{\text{avg}})/f_{\text{avg}}, & (f_{\max} - f_{\text{avg}}) < f_{\text{avg}} \\ 0.8, & (f_{\max} - f_{\text{avg}}) > f_{\text{avg}} \end{cases} \quad (4)$$

$$p_m = \begin{cases} 0.4f_{\text{avg}}/(f_{\max} - f_{\text{avg}}), & f_{\text{avg}} < (f_{\max} - f_{\text{avg}}) \\ 0.4, & f_{\text{avg}} > (f_{\max} - f_{\text{avg}}) \end{cases} \quad (5)$$

式中 p_c 为杂交概率, p_m 为变异概率, f_{\max} 为每代中个体最高适应值, f_{avg} 为每代平均适应值。且对变异率作如下的重新定义:每代中个体发生变异的概率,也就是每代中有 $p_m \times N$ 个个体发生变异。

4 算法性能分析和试验结果

遗传算法中,群体规模、杂交和变异算子的概率等控制参数的选取是非常困难的。而对于算法性能的评价也是个复杂的问题,同样也存在收敛速度和收敛精度的矛盾。特别是通过 BP 算法评价结构又带有一定的随机性,因为网络的训练效能依赖于随机的初始化条件;如果定义停止准则为某代个体的最优性能达到规定的要求或两代之间最好性能值之差为某个值,由于初始群体的随机性,以算法的收敛速度去评价性能就失去意义。为了能够综合评价算法的性能,我们定义停止准则为规定的代数、比较算法的离线性能(度量算法的收敛性能)和在线性能(度量算法的在线性能),并对最后的收敛结果加以分析和比较。

简单遗传算法与综合改进遗传算法的性能比较如图 2 和图 3 所示。简单遗传算法也采用改进的编码结构,其杂交率为 0.6,变异率为 0.02。图 3 中也给出了传统编码方案的简单遗传算法(杂交率为 0.6,变异率为 0.01)离线性能和优化结果,而在线性能由于单隐层多于双隐层,且有无隐层的现象,虽然呈下降趋势,但总体性能比较差,图 2 中没有画出。由图 2 和图 3 可看出:改进遗传算法的离线性能明显好于简单遗传算法。由于对网络结构性能的评价非常费时,对上述算法运行了大约同样的时间,而没有迭代到规定的代数。采用新编码的简单算法和改进遗传算法迭代了 18 次,而传统编码的简单遗传算法才迭代了 7 次,这是由于简单遗传算法中,初始群体缺乏多样性,初始参数值都比较低,网络训练速度慢。从实验结果可以看出如果变异率较小,简单遗传算法也存在局部最优解问题,也就是早熟收敛问题。

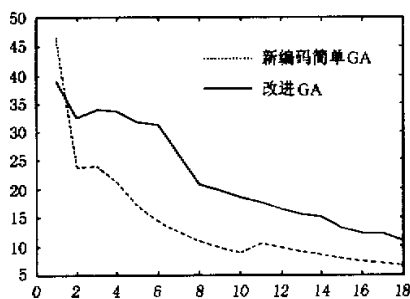


图2 简单GA与综合改进GA在线性能比较

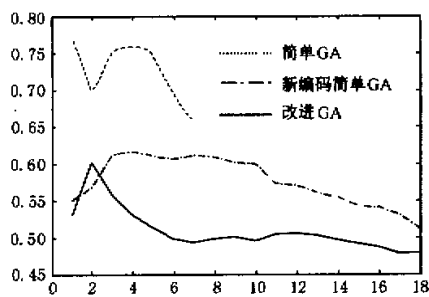


图3 简单GA与综合改进算法离线性能比较

对网络性能的评价,只考虑了收敛精度,而对于网络的收敛速度未进行评价。由于单隐层网络收敛速度快于双隐层结构的网络,因此,通过遗传算法的优化实验可以得出如下结论:在设计神经网络结构时,尽量采用单隐层的结构,尽管单隐层结构的隐层单元数多于两个隐层,但收敛速度仍快于双隐层结构;如果需要两个隐层时,把神经元尽量放在第1隐层。简单遗传算法与改进遗传算法的优化结果分别如表2和表3所示。

表 2 简单 GA 算法结果

网络结构	初始 η	动量系数	初始 W	训练误差	测试误差
3-20-16-1	0.125	0.6	0.125	0.2701	0.7180
3-20-25-1	0.125	0.6	0.125	0.2560	0.2308
3-20-17-1	0.125	0.6	0.125	0.2282	0.6596
3-20-25-1	0.125	0.6	0.125	0.2305	0.8641
3-13-21-1	0.25	0.8	0.125	0.2550	0.8628
3-17-1	0.0625	0.7	0.125	48.1938	63.8624
3-17-1	0.5	0.8	0.125	48.1882	63.9191
3-20-21-1	0.125	0.6	0.125	48.2585	67.5008
3-20-5-1	0.125	0.6	0.125	0.2358	0.4405
3-20-21-1	0.125	0.6	0.125	0.2890	0.3974

表 3 改进 GA 算法结果

网络结构	初始 η	动量因子	初始 W	训练误差	测试误差
3-27-5-1	0.25	0.9	0.5	0.2066	0.2222
3-27-5-1	0.25	0.9	0.5	0.2622	0.1798
3-20-12-1	0.5	0.9	0.25	0.2424	0.5060
3-27-5-1	0.25	0.9	1.0	0.2482	0.5100
3-27-5-1	0.25	0.9	1.0	0.2860	0.6213
3-18-9-1	0.125	0.6	0.5	0.2779	1.3053
3-30-1	0.25	0.7	0.5	0.2408	0.6707
3-45-1	0.25	0.9	1.0	0.2697	2.7511
3-27-5-1	0.5	0.9	0.5	0.2688	0.3135
3-15-24-1	0.25	0.6	0.25	0.2900	0.7959

5 结 论

对传统的网络结构编码进行了改进,扩大了编码范围,增加了表示模式,保证了遗传算法搜索的全局性和稀疏性,同时也避免无隐层现象的出现;提出一种适应值指数比例系数自适应调整方法,能使进化群体在解空间较分散时,限制性能好的群体的复制数目,以免其很快控制整个群体,而在群体性能值比较接近的情况下,又能提高相似串间的竞争性。改进的遗传算法对于神经网络结构的设计具有优良的性能,该方法的工程应用是作者的进一步研究课题。

参 考 文 献

- 1 Fogarty T C. Varying the probability of mutation in genetic algorithms. In: Proc of the 3rd Int'l Conf on Genetic Algorithms. Bos. 1989. 104~109
- 2 Davis L. Adapting operator probabilities in genetic algorithms. In: Proc of the 3rd Int'l Conf on Genetic Algorithms. Bos. 1989. 61~69
- 3 Whitley D. Genitor II: A distributed genetic algorithm. Journal Expt Ther Intelligence. 1990. (2): 189~214