

基于遗传神经网络成绩预测的研究与实现

陈 勇

(四川工商学院, 四川 成都 611745)

摘 要: 针对BP神经网络固有的局限性和在应用于成绩预测时出现的问题,运用遗传算法对BP神经网络的权值和阈值进行优化,通过详细设计遗传算法的编码方式、适应度函数,遗传算子使二者结合后的遗传神经网络模型具有更快的学习训练收敛速度,为了提高优化效果,设计了自适应的遗传算法交叉算子和变异算子,并通过与基本BP神经网络和自适应BP神经网络的对比,显示了优化的有效性和可行性。运用Matlab实现了遗传神经网络模型,并完成了模型的训练,运用Java语言完成了模型的调用和成绩预测系统的实现。分析结果表明,该遗传神经网络模型在成绩预测方面具有较高的准确性,具有一定的实用价值。

关键词: 成绩预测; BP神经网络; 遗传算法; Matlab; Java

中图分类号: TN711-34; TP183

文献标识码: A

文章编号: 1004-373X(2016)05-0096-05

Research and implementation of result prediction based on genetic neural network

CHEN Yong

(Sichuan Technology and Business University, Chengdu 611745, China)

Abstract: Since the BP neural network has inherent limitations and some problems existing in result prediction, the genetic algorithm is used to optimize the weighting and threshold of BP neural network, and its coding scheme, fitness function and genetic operator are designed detailed. The genetic neural network model has fast convergence speed of learning and training. The adaptive crossover operator and mutation operator of the genetic algorithm were designed to improve the optimization effect. The validity and feasibility of the optimization are showed by the comparison of basic BP neural network and adaptive BP neural network. The genetic neural network model was realized with Matlab, and its training was completed. The model call and result prediction system were implemented by means of Java language. The analysis results indicate that the genetic neural network model has high accuracy in the aspect of result prediction, and certain practical value.

Keywords: result prediction; BP neural network; genetic algorithm; Matlab; Java

0 引 言

近年来,神经网络的研究被广泛应用于如生物、医学、经济等诸多领域,基于神经网络进行成绩预测的研究也逐渐开始受到人们的关注^[1]。有效的学生成绩统计分析和预测在指导学校合理分配教学资源,全面提高教学质量方面都发挥着至关重要的作用。因此,将神经网络技术应用于学生课程成绩预测具有很强的现实意义与研究价值^[2]。目前,在为数不多的课程成绩预测实践中,绝大部分的学者采取应用最为广泛的BP神经网络模型进行实现,这在取得一定成果的同时暴露了BP神经网络算法收敛速度慢,效率低下等弊端和对课程成绩预测领域的不适应性。本文通过与基本BP神经网络和

自适应BP神经网络的对比,验证模型的收敛速度,体现出算法改进的效果^[3]。运用Matlab实现了设计好的遗传神经网络模型,并完成了模型的训练,运用Java语言完成了模型的调用和成绩预测系统的实现。研究结果能有效地指导学校合理分配教学资源,预防可能发生的重大教学事故,对于全面提高教学质量发挥着至关重要的作用。

1 遗传算法结合BP神经网络的设计

遗传算法具有较强的全局搜索能力和很强的鲁棒性,非常适合对BP神经网络进行优化^[4]。将二者结合起来,可大大减少网络陷入局部极小的概率,同时进一步提高网络的收敛速度,能较快的获得所求问题的全局最优解。遗传算法优化BP神经网络的基本原理就是将BP神经网络的权值交给遗传算法来控制,即将BP神经网络各隐层的节点权值和阈值作为遗传算法的输入,将

它们进行编码,生成初始种群,利用遗传算法的选择交叉和变异来产生新的后代,即新的神经网络权值和阈值,再交还给BP神经网络,由神经网络进行后续的求解过程。

从算法的角度上讲就是先通过遗传算法在目标问题的解空间中进行搜索^[5],当搜索到一个较优的网络形式时,再利用BP算法进行定位,确切地找到这个较优的解空间中的最优解或满意解。遗传算法结合BP算法的主要流程如图1所示。

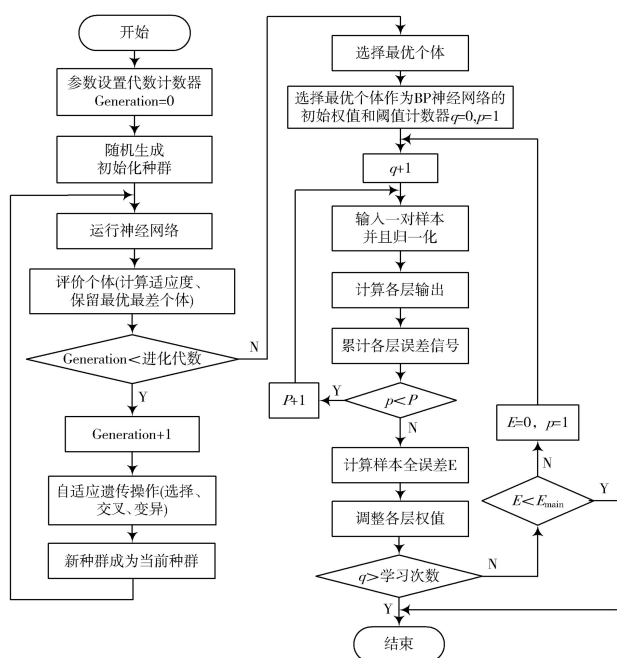


图1 遗传算法结合BP神经网络的流程图

1.1 BP神经网络设计

(1) 输入、输出层节点数量的设计

在文中输入的因子为与待预测课程的5个相关性最高的课程成绩,所以输入层的节点数量为5。输出层为神经网络的输出结果,即目标课程的成绩,所以输出层节点数为1。

(2) 隐层数量和节点数的设计

BP神经网络需要设定隐层层数、每个隐层的单元数。研究表明,双隐层的BP神经网络足够处理大部分的复杂问题,故只需要分析无隐层、单隐层、双隐层的表现即可。最终对BP网络的拓扑结构设置为:输入层+双隐层+输出层。其中每个隐藏层均为15个节点。

(3) 隐层函数的设计

在以tansig为整体的网络结构中,适当的将其中部分隐藏层的传递函数改为purelin可以有效地提高网络整体的预测准确率。将部分隐藏层的传递函数修改为logsig的结果并不尽如人意,可以从某角度说明在本课题中的成绩预测目的下,tansig函数相比logsig函数不论从整体或是个体上都相对更加有效。

(4) 初始学习速率和动量因子的设计

对BP神经网络的优化,采用自适应调整学习速率,在训练初期,选择较大的学习速率,以加快前期的收敛,所以网络的初始速率为0.05,该速率会随着网络的训练而不断发生变化。为了避免网络在训练过程中陷入局部极小值,引入动量因子避免这种现象的出现。根据经验,动量因子一般值为0.85左右。

1.2 遗传算法设计

本文的神经网络模型设计采用实数编码的方式,现神经网络的输入层有5个神经元,第一隐层有15个神经元,第二隐层15个神经元,输出层有1个神经元。将BP神经网络的所有权值和阈值作为待编码的参数,因此有 $(5+1) \times 15 + (15+1) \times 15 + (15+1) \times 1 = 346$ 个参数待优化,这些参数组成一个染色体。染色体中基因的排列顺序为输入层到第一隐层权值、第一隐层到第二隐层权值、第二隐层到输出层权值、第一隐层阈值、第二隐层阈值、输出层阈值。

遗传算法设计了一个能反映染色体适应性的参数,它反映的是每一个染色体与最优解相比较的优良程度,如何做出适应度函数的选择,就是要在压缩适应度和拉开适应度之间寻求一种平衡,同时还应该实现较小的计算量、较强的通用性和一致性等要求。

根据训练样本集,训练神经网络得到总误差E为:

$$E = \sum_{p=1}^p E_p = \frac{1}{2} \sum_{p=1}^p \sum_{l=1}^k (t_{pl} - o_{pl})^2 \quad (1)$$

式中: p 为输入层节点数; k 为输出层节点数; $t_{pl} - o_{pl}$ 表示输出层第 p 个样本的第 l 个节点的实际输出与期望输出的差。

本文的适应度函数为:

$$f = 1/(1 + E) \quad (2)$$

此处遗传算法编码模式选择实数编码,因此交叉算子选择算数交叉,并在算数交叉的基础上采取自适应策略,对个体适应度来说,交叉概率与之成反比,目的是为了很好地保护当前种群中的优秀个体,交叉率的计算公式为:

$$P_c = P_{c1} - \frac{P_{c2}(f' - f_{avg})}{f_{max} - f_{avg}}, f' \geq f_{avg} \quad (3)$$

$$P_c = P_{c1}, f' < f_{avg}$$

式中: f 为 X_1, X_2 适应度的较大者; f_{max}, f_{avg} 为群体的最大适应度和平均适应度。算数交叉的计算方法为:

$$X'_1 = \alpha X_1 + (1 - \alpha) X_2 \quad (4)$$

$$X'_2 = \alpha X_2 + (1 - \alpha) X_1$$

式中: X'_1, X'_2 为交叉后产生的新个体; $\alpha \in (0, 1)$ 。

本文在变异算子的选择上同样采取了自适应策略,

使个体的变异概率与它的适应度成反比,保留优秀解的同时使种群得到进化,变异率的计算公式如式(5)所示。其中, f' 为需变异个体的适应度, f_{\max} , f_{avg} 为群体的最大适应度和平均适应度,在这里均采用均匀变异。

$$P_m = P_{m1} - \frac{P_{m2}(f' - f_{\text{avg}})}{f_{\max} - f_{\text{avg}}}, \quad f' \geq f_{\text{avg}}$$
$$P_m = P_{m1}, \quad f' < f_{\text{avg}}$$

(5)

与此同时,在本文的研究中,遗传神经网络模型的种群规模取 100,遗传神经网络模型的进化代数取 50。

2 成绩预测系统的实现

2.1 数据预处理及相关度分析

本文所进行的成绩预测数据均来自学校网络中心的真实学生成绩数据,获取的数据规格为:每一行为一组输入,内容包括学号、学年、学期、课程名称、课程代码、课程性质、成绩等项,预计输入项在 100 000 以上。针对原始的学生课程成绩数据,首先要做的是数据的预处理,主要内容为数据的清洗,删除与成绩预测无关的课程代码等数据。主要步骤包括:去除空项,去除非正常考试项,去除非专业课成绩,处理成绩数据,处理课程名称等。预处理后的学生课程成绩数据格式如表 1 所示。

表 1 经过预处理学生成绩数据

编号	学年	学期	课程名称	课程性质	成绩	备注
St1	2011—2012	1	计算机软件实践	必修课	79	
St1	2011—2012	1	商务英语	校公选课	70	
St1	2011—2012	1	中国传统文化杂说	校公选课	88	
St1	2010—2011	1	中国近现代史纲要	必修课	83	
St1	2010—2011	2	大学英语国家四级	null	463	
St2	2010—2011	2	微积分 A(I)	必修课	60	
St3	2011—2012	2	计算机软件实践	必修课	60	重修
Std3	2011—2012	1	计算机软件实践	必修课	35	

采取 Apriori 算法对课程进行相关度分析。在横向比较所有学生成绩历史数据的过程中将随目标课程成绩的优秀而同样取得高分、随目标课程成绩过低而同样取得低分的课程称为该课程的相关课程。即便有很多与目标课程成绩相关的课程,也存在相关度的高低,如果将所有的相关课程都作为神经网络的输入项,不但使算法的参数变得复杂,收敛速度缓慢,而且也影响了预测的结果,成为了干扰项。因此,只选择五门相关课程进行成绩的预测。

经过 Apriori 算法对数据进行预处理后,大量原始数据被简化为只有 5 门相关课程和目标课程成绩的数据表,保存为一个文本文档供神经网络模型训练使用,格式如表 2 所示。同时预留另一部分数据作为成绩预测效果的验证数据。

表 2 预处理后的成绩数据

课程 1	课程 2	课程 3	课程 4	课程 5	目标课程
90	85	79	91	76	84
71	80	65	84	76	85
89	75	68	88	71	82
94	90	87	92	83	89
75	63	67	84	71	73
93	80	79	89	84	87
93	80	52	61	64	80
74	67	80	95	83	84

2.2 遗传神经网络模型的实现

根据遗传算法与神经网络相结合的遗传神经网络模型的总体设计,运用 Matlab 实现遗传神经网络模型的过程需要将 BP 神经网络函数与遗传算法的编码解码函数和适应度函数分别进行实现,主函数为 gabpMain.m^[6],实现 BP 神经网络的创建和神经网络参数的设计。同时, gabpMain 函数还需实现神经网络与遗传算法的对接工作,因此还要实现遗传算法种群的初始化和基本参数的设定。

Matlab 实现神经网络模型后,开始用大量的数据对模型进行训练。经过训练的模型才能够被应用于成绩预测系统,通过调用该模型,输入预测课程的 5 门相关课程成绩数据,进行目标课程的成绩预测。通过对模型的训练,进行算法和参数的调整,同时比较各个算法的收敛曲线,验证基于遗传算法对 BP 神经网络进行优化的效果。比较的对象选择一个没有遗传算法进行改进,基本 BP 神经网络参数设置完全相同的神经网络模型,采用同样的数据源进行训练,两个模型分别进行仿真训练各五次,分别选出其中目标进化函数收敛效果最好的结果图进行比对,其中,采用基本 BP 神经网络的模型收敛图如图 2 所示。而本文设计的采用遗传算法进行优化的 BP 神经网络模型收敛图如图 3 所示。

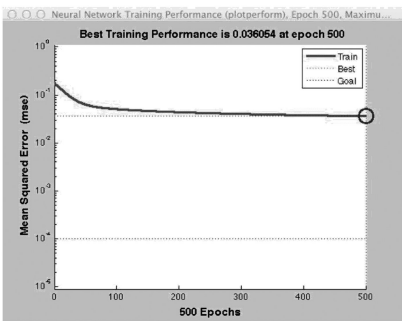


图 2 基本 BP 神经网络模型函数收敛效果图

通过比较不难发现,单一 BP 算法和遗传算法优化神经网络的混合算法在收敛速度方面的差别十分巨大,两者在训练次数的上限均为 500 次,目标误差平方差均为 10⁻³ 的情况下,基本 BP 神经网络模型收敛速度缓

慢,在进行了500次训练后,误差停留在了0.032 915,仍然没有达到目标值,而遗传神经网络模型的收敛速度大大加快,仅在第56次训练就达到了目标值。

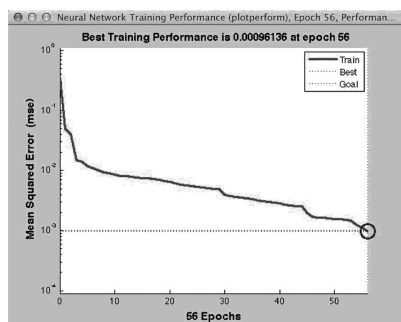


图3 遗传神经网络模型函数收敛效果图

为了进一步加强对比,再次建立基于自适应学习算法改进后的BP神经网络模型,用同样的样本数据进行训练,以验证是否是特殊原因造成二者差距过大的现象,同时验证遗传算法相对其他算法优化BP神经网络的优势,训练的结果如图4所示。

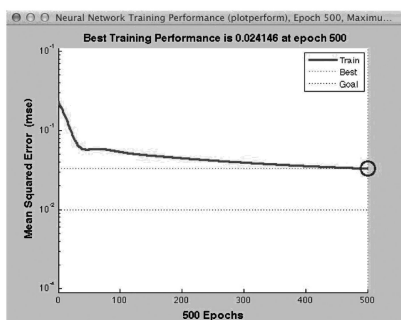


图4 基于自适应学习算法的BP神经网络模型函数收敛效果图

由图中可以看出,在500次训练后,函数的误差均收敛于0.024 146,效果较之于基本的BP神经网络模型大有提高,前期的收敛速度也有所加快,但是仍然没有500次训练以内达到目标。在基于对两种BP算法模型进行了对比之后,可以总结造成这种差距的原因是输入的样本数据数量相对较少,同时课程成绩数据的特点是离散但是重复度很高,训练的效果更加依赖隐层节点和权值的设计,而这正是遗传算法对BP神经网络优化的核心,因此体现出的效果比较显著^[7]。由此可见遗传算法优化BP神经网络确实在应用于成绩预测等类似领域的时候具有一定的优越性。

3 预测结果分析

完成神经网络模型的实现后,此处通过Java语言实现一个调用神经网络模型^[8],允许使用者自行选择目标课程,并直观查看从数据预处理到成绩预测结果的软件

系统。如图5所示,该系统界面简洁明了,进入软件界面后,分别选择年级、学院、科目等3项参数,并连续点击“下一步”,即可直观地看到用数据预处理开始到成绩预测结果的全部中间数据。



图5 基于遗传神经网络的成绩预测系统界面

成绩预测的结果可以采取两种查看方式,即列表方式与图表方式。其中列表方式显示了学生的学号和系统预测该目标课程取得的成绩。能够以红颜色标记成绩预测结果过低的学生,方便使用系统的用户了解学生的成绩和课程情况,如果成绩过低的学生过多,则证明该课程存在发生教学事故的风险,需要采取严格管理等措施来防范。通过这种方式,成绩预测系统达到了通过成绩预测协助进行学生管理和教学事故预防的目的,成绩预测的列表方式显示如图6所示。

图6 遗传神经网络的成绩预测列表方式显示

为了验证成绩预测系统的预测准确度,需要以一些已经有成绩的历史数据进行预测。图表方式以红色的“*”标记目标课程的预测成绩;以圆圈标记目标课程的实际成绩;“*”和圆圈的距离形象地显示了预测成绩与实际成绩之间的差距。成绩预测的图表方式显示如图7所示。

为了通过分析成绩预测结果得知遗传神经网络应用于成绩预测的实际效果,使用系统预测2009级软件学院学生“软件工程实践”这门课程的成绩,并以其中的16条数据进行误差比对,测试的结果如表3所示。

由于预测的分数全部进行取整,所以精度有所影响,因此规定实际输出与预计输出相差小于或者等于3分的情况下,输出的结果为正确,反之为错误。统计得知,使用遗传神经网络的成绩预测系统对学生的学习成绩进行预测,正确率为87.5%,证明遗传神经网络算法在

应用于成绩预测时确实取得了比较好的效果。

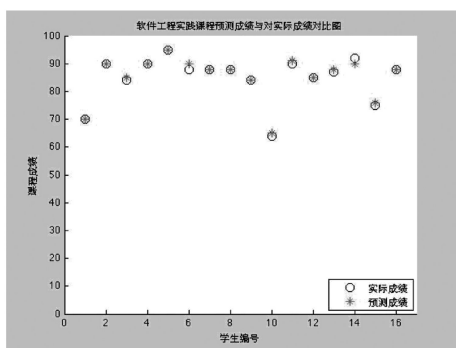


图7 基于遗传神经网络的成绩预测系统的预测结果显示

表3 遗传神经网络的预测结果列表

课程1	课程2	课程3	课程4	课程5	预计输出	实际输出	判断
61	66	73	82	69	70	72	正确
87	90	73	89	90	90	90	正确
46	80	60	76	72	88	86	正确
98	75	80	94	89	88	90	正确
70	81	85	94	85	93	97	正确
81	90	89	89	76	84	80	错误
98	75	80	94	89	88	90	正确
46	80	60	76	72	88	84	正确
76	75	74	92	74	82	81	正确
60	70	61	65	75	70	65	正确
75	85	64	84	75	85	81	正确
81	80	69	93	80	87	82	正确
83	85	87	95	69	92	97	正确
74	85	89	97	85	88	94	错误
60	67	64	82	68	75	74	正确
72	75	75	85	72	88	86	正确

4 结 论

本文在对神经网络和遗传算法的基本理论和优势的研究基础上,设计了遗传算法与BP神经网络相结合的遗传神经网络模型,通过详细设计遗传算法的编码

作者简介:陈 勇(1979—),男,四川遂宁人,讲师。研究方向为数理统计、神经网络。

(上接第95页)

系统,实现系统的选换挡数据监测和显示方面的需求,以飞思卡尔16位单片机为核心搭建硬件平台,利用CAN通信技术,采用VB软件设计了AMT数据采集系统的监控平台。实验结果表明,该系统可以稳定地实现AMT数据采集功能,为提高AMT系统的开发效率提供了一个操作简单而实用的平台,达到了设计要求。

参 考 文 献

- [1] 张志伟.基于LabVIEW的AMT下线检测试验系统[J].湖北汽车工业学院学报,2011,25(3):34-38.
- [2] 马永富,彭忆强,甘海云,等.基于CAN总线与虚拟仪器的AMT

作者简介:刘文秀(1980—),女,内蒙古通辽人,讲师,硕士。从事自动控制方向的教学及研究。

方式、适应度函数、遗传算子使二者结合后的遗传神经网络模型具有更快的学习训练收敛速度,并通过与基本BP神经网络和自适应BP神经网络的对比,显示了优化的有效性和可行性。进一步充分运用Matlab实现了设计好的遗传神经网络模型,并完成了模型的训练,运用Java语言完成了模型的调用和成绩预测系统的实现,以直观的方式显示了预测结果,经过分析,预测结果具有较高的准确率和效率。但是,将RBF神经网络应用于成绩预测领域,是不是会取得比BP神经网络更好的效果,需要通过实验进行进一步的验证。

参 考 文 献

- [1] 周凌翔.改进BP神经网络在模式识别中的应用及研究[D].南京:南京理工大学,2010.
- [2] SOCHA K, BLUM C. An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training [J]. Neural computing and applications, 2007, 16(3): 239-247.
- [3] 黄庆斌.BP算法的改进及其应用研究[D].西安:西安交通大学,2010.
- [4] 尹然,丁晓明,李小亮,等.基于SA-BP神经网络的软件缺陷预测模型的研究[J].西南师范大学学报(自然科学版),2013,38(8):147-152.
- [5] 康琨,孟文俊.一种基于改进神经网络的系统辨识方法[J].计算机与数字工程,2012,40(1):31-33.
- [6] 邹丽娜,丁茜.基于BP算法的成绩预测模型[J].沈阳师范大学学报(自然科学版),2011,29(2):226-229.
- [7] 陈文,庞琳娜.GABP神经网络在交通流预测中的应用研究[J].微计算机信息,2009,25(14):245-247.
- [8] 易安.基于神经网络的模糊推理模型和算法研究[D].重庆:西南大学,2010.
- [9] 选换挡监测系统研制[J].西华大学学报(自然科学版),2012,31(2):26-30.
- [10] 武州,石晓辉,李文礼.AMT试验台控制系统的研发[J].计算机测量与控制,2012(9):2399-2401.
- [11] 范伟成,宗情,朱辰元.基于单片机的CAN-USB通信转换模块的设计与实现[J].计算机测量与控制,2012,20(3):187-189.
- [12] 张晓伟,孙逢春,邢杰.基于CAN总线的混合动力汽车监控系统开发[J].计算机测量与控制,2011,19(10):2427-2429.
- [13] 杜尚丰,曹晓钟.CAN总线测控技术及其应用[M].北京:电子工业出版社,2007.
- [14] 王黎明,夏立,邵英.CAN现场总线系统的设计与应用[M].北京:电子工业出版社,2008.