

BP 网络隐层神经元数自调节算法的研究

胡 平, 韩 兴, 陈义军

(南京工业大学 电子信息与工程学院, 江苏 南京 211816)

摘 要: 针对传统 BP 算法中隐层神经元数不易选取的问题, 本文提出了 BP 算法的改进模型——DBP(Double BP)网络模型. 该模型可以动态自调节隐层神经元数, 即通过一个给定先验知识的 BP 网络动态调节另一个 BP 网络中隐层神经元数, 并且通过选取合适的权值和阈值使训练误差曲线迅速下降. 解决了 BP 网络拓扑结构中隐层神经元个数以及新增加权值和阈值的选取问题, 还对 BP 网络在陷入假饱和区如何逃逸提出了一种新的方法. 最后通过仿真模拟取得了比较好的效果.

关键词: BP 网络; DBP; 网络; 隐层神经元; 权值; 阈值

中图分类号: TP183

文献标识码: A

文章编号: 1000-7180(2013)02-0108-05

The Research of BP Network Hidden Layer Neuron Self-Regulating Algorithm

HU Ping, HAN Xing, Chen Yi-jun

(College of Electronics and Information Engineering, Nanjing University of Technology, Nanjing, 211816, China)

Abstract: In traditional BP algorithm, the number of hidden layer neurons is uneasy to select. Aiming at sorting out this problem, this article points out the updated BP arithmetic model, which is DBP (Double BP) network model. The new model can adjust hidden layer neurons dynamically and automatically, which means, adjust the hidden layer neurons in another BP network dynamically via a experience and knowledge BP network, make the training error curve decrease rapidly by choosing the proper weight and thresholds. It sorts out the problem in the quantity of hidden layer neurons in BP network topology, as well as the issue of choosing the new added weight and threshold. In addition, it suggests a new approach against how the BP network escapes from the fake saturated area. Finally, it gets a good result through simulation.

Key words: BP network; DBP network; hidden layer neurons; weight; threshold

1 引言

BP 网络系统地解决了多层神经网络中隐层单元连接权的学习问题, 并且在数学上给出了完整的推导, 使得多层前馈网络的研究取得了重大突破^[1]. BP 算法简单、容易实现、计算量小等特点使得它成为目前应用的较为广泛的神经网络算法之一. 但是 BP 网络拓扑结构存在不确定性, 这样就使得在构建网络时, 其隐层层数的确定、隐层神经元个数

的选择、经验性、主观性占据较大部分. 然而正是网络结构的确定, 决定着整个网络的性能. 本文围绕这一问题给出了 BP 网络在三层前馈网络中隐层神经元个数的选取问题, 并且通过一个已经训练好的 BP 网, 根据给定的先验知识, 自适应的根据执行实际任务的 BP 网自身误差的收敛速度, 动态调节该执行实际任务的 BP 网络隐层神经元的个数, 减少了人为干涉的因素, 使得 BP 神经网络的智能化大大提高.

收稿日期: 2012-05-15; **修回日期:** 2012-06-13

基金项目: 国家自然科学基金项目(6073185, 61073197); 江苏省自然科学基金项目(BK2010548); 教育部留学回国人员科研启动基金项目(教外司留[2007]1108 号)

2 BP网络的思想

BP(back-propagation)算法,系统的解决了多层神经网络的学习问题,并且证明了多层神经网络具有很强的学习能力. 它的学习规则是使用最速下降法,通过反向传播来不断调整网络的权值和阈值,使网络的误差平方和最小^[2]. 文献^[3]证明了对任何在闭区间内的连续函数,都可以用一个隐层的BP网络来逼近,即一个三层的BP网络可以模拟任意连续函数. 因此,我们拟用一个隐层来构建BP网络的拓扑结构,BP网络的结构模型如图1所示.

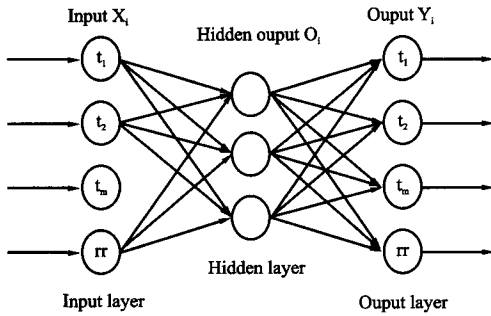


图1 BP网络的结构模型

整个网络由一个输入层,一个隐含层,一个输出层构成,各层之间用相应的权进行连接,而相同层之间的神经元不进行连接. 输入、输出层的神经元个数是根据具体的实际需要确定. BP算法由数据流的前向计算和误差信号的反向传播两部分构成,具体算法步骤如下^[2]:

(1)假设有 p 个训练样本,即 p 个输入-输出对 $(x(k), y(k)), k = 1, 2, 3, \dots, p$, 通过输入层结点传至隐层结点,最后由隐层传至输出层结点,最终将实际输出值传出. 每一层的输出成为下一层的输入,此操作如式(1)所示:

$$O_{kj} = f_j(\sum_i w_{ji} O_{ki} + \theta_j) \quad (1)$$

其中 w_{ji} 为前一层第 i 个神经元输入到后一层第 j 个神经单元的权重. θ_j 为每个神经元的阈值, O_k 为网络的实际输出向量,即 $O_k = O_{k1}, O_{k2}, \dots, O_{kn}$, n 为输出向量的维数. $f_j(x)$ 为神经单元的激发函数,一般选用 Sigmoid 函数.

(2)然后根据隐层单元和输出层单元的实际输出计算训练误差,此操作的公式如式(2)、(3)所示:

$$\delta_{kj} = O_{kj}(1 - O_{kj})(t_{kj} - O_{kj}) \quad (2)$$

$$\delta_{kj} = O_{kj}(1 - O_{kj}) \sum_m \delta_{km} w_{mj} \quad (3)$$

其中 $t_k = (t_{k1}, t_{k2}, t_{k3}, \dots, t_{kn})$ 为第 k 个样本的期望输出, n 为输出向量的维数.

(3)再进行误差信号的反向传播,即各神经元之间的连接权值,以及神经元的阈值,根据所得到的训练误差反向进行修正,公式如式(4)、(5)所示:

$$w_{ji}(t+1) = w_{ji}(t) + \eta \delta_j O_{ki} + \alpha [w_{ji}(t) - w_{ji}(t-1)] \quad (4)$$

$$\theta_j(t+1) = \theta_j(t) + \eta \delta_j + \alpha [\theta_j(t) - \theta_j(t-1)] \quad (5)$$

其中 η 为按梯度搜索的学习步长, α 为势态项,均为给定的常数.

(4)直到 p 个训练样本训练一次后,计算一次误差,这里采用均方根误差计算,如式(6)所示:

$$e = \sqrt{\frac{\sum_{i=1}^m \sum_{j=1}^n (t_{ij} - y_{ij})^2}{m * n}} \quad (6)$$

其中 y 为输出层的输出结果, m 为训练集的样本数, n 为网络输出的单元个数.

(5)每 p 个样本学习一次就会产生一个 e , 如果得到的 e 小于事先给定的期望误差,就会继续重新学习 p 个样本,直到满足最终给定的期望误差条件才会停止学习. 这就是 BP 算法的基本思想^[4].

在整个算法实现的过程中,隐层神经元的个数自始至终都是固定的,导致在事先确定隐层神经元个数的过程中,可能由于给定的隐层神经元个数过少,不能最终收敛到给定的误差期望值,或者给定的神经元个数过多,不仅训练时间过长,而且有可能出现所谓的“过拟合”(Overfitting)问题,即过于精细的逼近训练的样本从而导致泛化能力下降. 在实际应用中,已经有研究者通过给定一个较小的隐层神经元个数,然后根据经验和实际收敛情况手动添加神经元个数,或者给定一个足够多的隐层神经元个数,根据权值的灵敏度来减掉那些灵敏度较小的权值. 但是这些无疑给神经网络智能型大打折扣,变成了数据在人为的干涉中进行处理的过程^[5]. 本文提出的思想是再重新构造一个BP网来控制隐层神经元个数的增减,最终达到满足实际需要的目的.

3. DBP神经网络模型

DBP神经网络模型是本文的核心. 本文的算法思想是通过构造一个可以自己学习的BP网络在误差进入平坦区的时候去控制执行实际应用的BP网的隐层神经元个数.

3.1 DBP神经网络定义

在DBP神经网络模型中,全部的神经网络由主网和副网两部分组成. 其中主网是用于执行实际任

务的神经网络,副网是用于完成配置和训练其它神经网络的神经网络

3.2 DBP 神经网络的功能

主网:工作网,由副网监控训练,执行具体的实际任务^[6-7].

副网:配置和训练主网,包括配置和调整主网的结构、参数,且该网具有自身的配置和学习的功能.该网一般应具有学习、记忆、识别、推理的能力.学习样本主要来自人工干预的参数、一些算法的参数值及对主网配置和训练的成功例子(神经网络库).其主要目的是引入先验知识,必要时控制主网的工作.

3.3 神经网络(副网)控制主网的神经元个数的

模型引入

引入神经网络(副网)控制主网的神经元个数的模型具体步骤如下:

- (1)选取副网样本并构造训练该网;
- (2)主网开始训练;
- (3)在主网误差 e 达到一个比较容易进入的固定值 m 时,副网开始实施监控;
- (4)判断主网训练过程若总误差达到要求的精度,且没有增加隐层神经元个数,则转到结束;
- (5)记录主网本次与上一次样本的总误差的差 ΔE ,即误差曲线的斜率 k ;
- (6)若 $k > \Delta E_1$ (ΔE_1 为误差曲线陷入平坦区时的斜率值),则副网监控主网,输入增加隐层神经元个数为 0;
- (7)若 $k < \Delta E_1$,则说明输出单元落入平坦区,此时开始记录满足 $k < \Delta E_1$ 的迭代次数 count ,若 $\text{count} = n$ 时,副网监控主网输入增加隐层神经元数为 1,直至主网最终满足精度要求转至结束;
- (8)结束.

4 实验测试

我们利用 DBP 网络模型模拟在 $[0, 2\pi]$ 区间内的 $\sin$ 函数,并给出较多的测试样本对其进行测试.

4.1 DBP 网的构造以及样本的选取

副网的拓扑结构为一个三层 BP 网络结构,输入层取两个结点,分别表示为误差曲线图的斜率和达到此斜率所迭代的次数,隐层取十个结点,输出层取一个结点.误差精度规定为 0.001.本实验在增加隐层神经元个数的基础上对增加的隐层神经元个数实施了简化,通过大量的实验表明,只要处理单元进入了饱和区内,就可以有效的通过增加神经元的个数来使处理单元逃离饱和区,即使误差快速下降.于

是在本实验中我们规定:

$$y = \begin{cases} 1 & k < 1.0 \times 10^{-8}, p = 1.0 \times 10^5 \\ 0 & \text{else} \end{cases} \quad (7)$$

式中 y 为副网的输出(即主网隐层神经元增加的个数), k 为主网误差曲线的斜率,即当 $k < 1.0 \times 10^{-8}$ 且 $p = 1.0 \times 10^5$ 时,副网输出 1,不满足此条件主网输出 0.此处的斜率 k 的选择具有一定的通用性,也即斜率 $k = 1.0 \times 10^{-8}$ 时,误差曲线基本进入饱和区内,而 p 的选择本例实验中选的是 1.0×10^5 ,也可以根据不同的情况作出适当的调整.副网部分样本选取如表 1、表 2 所示.

表 1 副网的部分训练样本

输入斜率	输入次数	期望输出
0.135377	1	0
0.0385268	1	0
0.00660996	1	0
1.69618e-005	1	0
3.09999e-005	1	0
6.70249e-005	1	0
9.59958e-005	1	0
3.60807e-006	1	0
2.67217e-006	1	0
3.54108e-007	1	0
7.29804e-007	1	0

表 2 副网的部分训练样本

输入斜率	输入次数	期望输出
9.8941e-009	100000	1
5.10022e-009	100000	1
6.50216e-009	100000	1
7.5024e-009	100000	1
3.50258e-009	100000	1
2.30001e-009	100000	1
5.93994e-010	100000	1
4.79985e-010	100000	1
7.70001e-011	100000	1
5.41881e-011	100000	1

此样本中当斜率 $k > 1.0 \times 10^{-8}$ 时,我们规定样本输入次数为 1,期望输出均为 0;当斜率 $k < 1.0 \times 10^{-8}$ 时,我们规定样本次数为 100000,期望输出为 1.

主网拓扑结构的选择如下:三层 BP 网络结构,一个输入单元,一个输出单元.隐层神经元为初始个数为 1,误差精度为 0.0009,训练样本数选取 30 个($\sin$ 函数在规定区间内的 30 个点).

4.2 主网增加隐层神经元时副网的输入和输出

在副网的控制下,主网一共增加了两次神经元.

主网在训练过程中当满足 $k < 1.0 \times 10^{-8}$ 时,即开始计算此状态下迭代的次数,反馈给已经训练好的副网,当副网通过主网反馈的数值作为副网的输入计算输出,当输出大于 0.9 时,我们则认为主网的隐层神经元可以增加 1 个. 此实验过程中主网增加隐层神经元时反馈给副网的输入以及副网的输出如表 3 所示.

表 3 主网增加神经元时副网的输入和输出

副网的 输入斜率	副网的 输入次数	副网的 输出
9.89841e-009	100000	1
5.10022e-009	100000	1
6.50216e-009	100000	1

4.3 主网增加的神经元权值以及阈值的选取

当主网增加神经元时,其神经元的权值以及阈值的选择如何进行,我们采用了三种不同方法分别进行模拟,并比较其效果的优劣.

(1)选取 0—1 之间的一个较小的随机数作为新增加神经元的权值和阈值,误差曲线如图 2 所示.

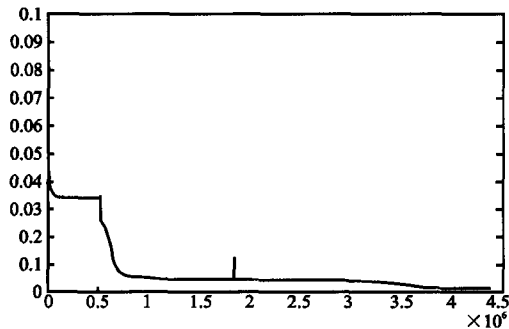


图 2 较小随机数作为新权值阈值的曲线误差

(2)复制任意一个已学习过的隐层神经元的权值和阈值作为新增加神经元的权值和阈值,误差曲线如图 3 所示.

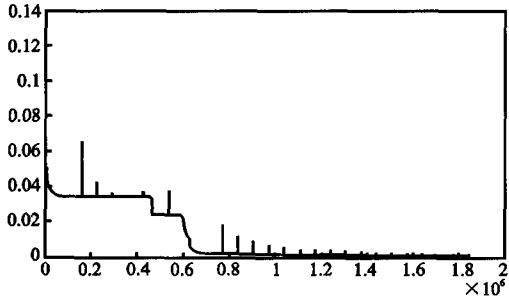


图 3 复制已学过的权值阈值作为新权值阈值的曲线误差

(3)选取已学好的各神经元的权值阈值的平均值作为新增加神经元的权值和阈值,本实验误差曲线如图 4 所示.

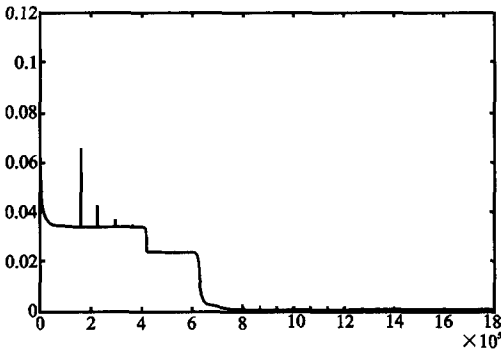


图 4 已学过权值阈值的平均值作为新权值阈值的曲线误差

可以看出,选取 0—1 之间的一个较小的随机数作为新增加神经元的权值和阈值效果明显好于另外两种情况. 出现此种情况的原因如下:在处理单元进入假饱和区的时候,由于触发函数的导函数值很小,所以导致每次的权值阈值的修正都很小,如果此时再复制其中已学好的权值和阈值,无疑不利于逃出饱和区,即不利于权值和阈值的修正,使误差始终难以逃出饱和区. 如果新增加的神经元的权值和阈值选用一个与其他已学好隐层神经元的权值和阈值差异较大,且为较小的数时,可以使整个网络结构的隐层神经元的权值和阈值作出较大的修改,并且帮助逃出饱和区,即使误差在经历短暂的上升之后,迅速的下降收敛,最终达到所规定的误差精度.

(4)通过给定测试样本测试已经学习好的 DBP 网络模型,具体数据如表 4 所示.

表 4 已学好 DBP 网络模型的测试数据

弧度	期望输出	实际输出	预测精确率
2.134	0.845549	0.847547	0.23%
1.345	0.974616	0.965745	0.91%
3.55	-0.397148	-0.487804	2.68%
5.678	-0.568914	-0.574304	0.94%

由表 4 可以看出,在副网控制下已学习好的主网最终预测的数值的精确率始终维持在 5% 以内. 本实验基于 DBP 网络模型,动态增加隐层神经元个数对 sin 函数在 $[0, 2\pi]$ 区间内的模拟取得了较好的效果.

5 结束语

本文介绍了在 DBP 网络模型控制下动态调节隐层神经元个数的思想,并使用该方法对 sin 函数进行了模拟. 在 DBP 网学习过程中,由副网不断学习并记录所有可用的经验,并将其学习经验在主网陷入假饱和区的时候增加主网隐层神经元个数使其快速收敛. 并且在新增加隐层神经元的权值和阈值

的选取上提出了创造性的构想,即选取一个与已学好隐层神经元权值和阈值差别较大的,且数值为较小的值,使新增加的隐层神经元在进入网络学习中带动已有的隐层神经元,使本已进入假饱和状态下的网络的各隐层神经元的权值和阈值由较小的修正变为较大的修正,使误差在经历短暂的上升过后迅速下降,从而整个处理单元快速逃出假饱和区,最终提高整个网络的性能。

通过实验数据可以看出,在副网控制下明显提高了主网学习速度,并且在权值和阈值修正较小的情况下,即误差陷入平坦区的时候,能自动调节隐层神经元个数以达到预期的结果。在遇到其他同类学习问题时,也可以参照此模型,运用副网直接监控并且调节主网隐层神经元个数,选取合适的新增加隐层神经元的权值和阈值来帮助处理单元逃出假饱和区。此方法是对 BP 网络的改进,也对 BP 网络在陷入假饱和区时的处理提出了一种新的方法。

参考文献:

- [1] Rumelhart D E, Hinton G E, Williams R J. Learning representation by back propagation errors [J]. Nature, 1986, 323(9): 533—536.
- [2] 杨建刚. 人工神经网络实用教程[M]. 浙江: 浙江大学出版社, 2001.
- [3] Hecht-Nielsen R. Kolmogorov's mapping neural network existence theorem[J]. Inr Conf on Neural Networks[IEEE], 1987(3): 11—14.
- [4] R. Rojas, editor. Neural Networks—A Systematic Introduction[J]. Springer[IEEE], 1996(9): 25—28.
- [5] 万亿泰. BP 人工智能神经网络拓扑结构及算法[J]. 信息科学, 2009, 28(11): 84—85.
- [6] 纪习尚. 基于最优分类面的神经网络模式分类方法及其应用[D]. 南京: 南京航空航天大学, 2005.

作者简介:



胡平 男, (1962—), 硕士, 副教授, 研究方向为计算机网络、信号处理。



韩兴 男, (1986—), 硕士, 研究方向为计算机网络、信号处理。



陈义军 男, (1986—), 硕士, 研究方向为计算机网络、信号处理。