

# BP神经网络预测算法的改进及应用

王晓敏<sup>1</sup>, 刘希玉<sup>2</sup>, 戴芬<sup>1</sup>

(1. 山东师范大学 信息科学与工程学院, 山东 济南 250014;

2. 山东师范大学 管理与经济学院, 山东 济南 250014)

**摘 要:** BP算法是应用广泛的神经网络算法, 具有较强的非线性拟合能力, 可以用来预测非线性时间序列数据的发展趋势。但在实际应用和仿真过程中, 由于算法本身的限制和不足, 对于仿真和计算都会带来很多问题, 比如网络训练过程中程序异常中止、训练时间过长、仿真精度不高等。针对这样的情况, 通过分析算法本身和训练仿真过程, 找到了相应的原因和解决方法, 研究了传统BP神经网络模型的缺陷并提出附加动量的方法来改进BP神经网络, 最后通过在Matlab仿真环境下的实际仿真过程, 验证了改善效果。

**关键词:** BP算法; 数据预测; 仿真; 神经网络

**中图分类号:** TP301

**文献标识码:** A

**文章编号:** 1673-629X(2009)11-0064-04

## Improvement and Application of BP Neural Network Forecasting Algorithm

WANG Xiao-min<sup>1</sup>, LIU Xi-yu<sup>2</sup>, DAI Fen<sup>1</sup>

(1. Department of Information Science and Engineering, Shandong Normal University, Jinan 250014, China;

2. Department of Management and Economics, Shandong Normal University, Jinan 250014, China)

**Abstract:** BP(back propagation) algorithm which is one of the most widely used neural network algorithms, has very high nonlinear fitting ability, and it can be used to predict the developing trend of time series data in practical application and simulation. But some kinds of problems and exceptions may happen because of the limitation and deficiency of the algorithm itself, such as abnormal termination, long training time and low accuracy. Aiming at improving the performance, through analyzing the algorithm and simulation, corresponding causes and problem-solving way are found. In this paper some limitations of BP neural network have been analyzed and optimized methods have been supposed. Finally, through practical simulating experiments in Matlab, the effect is certificated.

**Key words:** BP algorithm; data forecasting; simulation; neural network

## 0 引言

神经网络是一门发展十分迅速的交叉学科, 它是由大量的处理单元组成的非线性大规模自适应动力系统。目前已经提出了多种训练算法和网络模型, 其中应用最广泛的是前馈型神经网络中的误差反向传播(BP)学习算法<sup>[1]</sup>。BP算法系统地解决了多层网络中隐含单元连接权的学习问题, 使BP网络成为应用最为广泛的一种神经网络模型。但随着使用的广泛, 人们发现BP网络存在收敛速度缓慢、易陷入局部极小

等缺陷, 极大地影响了神经网络的进一步应用<sup>[2]</sup>。文中结合BP网络的训练过程, 分析其产生问题的原因, 提出一种改进的BP网络训练方法, 避免了传统的BP算法容易陷入局部极小的问题, 同时可以提高BP算法的训练速度, 降低收敛时间。

## 1 BP神经网络

### 1.1 传统BP网络算法

BP神经网络的学习过程分为信息的正向传播和误差的反向传播过程两个阶段。外部输入的信号经输入层、隐含层的神经元逐层处理向前传播到输出层给出结果。如果在输出层得不到期望输出, 则转入逆向传播过程, 将实际值与网络输出之间误差沿原来连接的通路返回, 通过修改各层神经元的联系权值, 使误差减少, 然后再转入正向传播过程, 反复迭代, 直到误差

收稿日期: 2009-03-02; 修回日期: 2009-06-12

基金项目: 国家自然科学基金重大项目(60873058; 60743010); 山东省自然科学基金重大项目(Z2007G03)

作者简介: 王晓敏(1983-), 女, 硕士研究生, 研究方向为数据挖掘与智能算法; 刘希玉, “泰山学者”, 教授, 博士生导师, 研究方向为数据挖掘与人工智能。

小于给定的值为止<sup>[3]</sup>。以一个三层网络为例,网络由  $N$  个输入神经元,  $K$  个隐层神经元,  $M$  个输出神经元组成(见图 1)。  $O2_{pm}$  和  $O1_{pk}$  分别为输出层和隐层的输出值,  $w2_{km}$  和  $w1_{nk}$  分别为隐层到输出层和输入层到隐层的连接权值, 设输入学习样本为  $x_{pm}$ , 其对应的希望输出值为  $t_{pm}$ 。

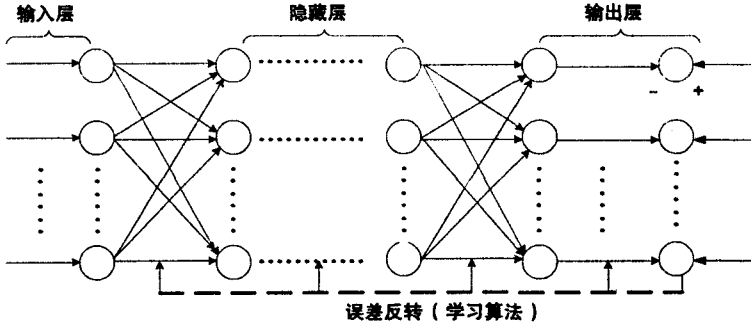


图 1 BP 神经网络示意图

标准算法步骤如下:

(1) 初始化权值, 设定学习率  $\mu$ , 允许误差  $\epsilon$ , 最大迭代次数, 置循环步数  $i = 0$ 。

(2) 正向计算: 将第  $p$  个样本  $\{X_p = \{X_{p1} \cdots X_{pm}\}\}$  顺序输入到网络中, 按下式分别计算  $O1_{pk}$  和  $O2_{pm}$ :

$$O1_{pk}(i) = f\left(\sum_{n=1}^N w1_{nk}(i)x_{pn}\right)$$

$$O2_{pm}(i) = f\left(\sum_{k=1}^K w2_{km}(i)O1_{pk}(i)\right)$$

激活函数常采用 S 型 sigmoid 函数:

$$f(x) = 1/(1 + e^{-x})$$

(3) 计算均方误差  $E$ , 若  $E \leq \epsilon$ , 则停止迭代, 否则执行下一步。

$$E = \frac{1}{M} \sum_{m=1}^M (t_{pm} - O2_{pm})^2$$

(4) 反向计算: 计算权值的改变量, 公式如下:

$$\Delta w1_{nk}(i+1) = \mu \sum_{p=1}^P \delta_{pk}(i)x_{pn}$$

$$\Delta w2_{km}(i+1) = \mu \sum_{p=1}^P \bar{\delta}_{pm}(i)O1_{pk}(i)$$

而

$$\bar{\delta}_{pm}(i) = (t_{pm} - O2_{pm}(i))O2_{pm}(i)(1 - O2_{pm}(i))$$

$$\delta_{pk}(i) = O1_{pk}(i)(1 - O1_{pk}(i)) \sum_{m=1}^M \bar{\delta}_{pm}(i)w_{km}(i)$$

更改权值

$$w1_{nk}(i+1) = w1_{nk}(i) + \Delta w1_{nk}(i+1)$$

$$w2_{km}(i+1) = w2_{km}(i) + \Delta w2_{km}(i+1)$$

(5) 置  $i = i + 1$ , 返回(2)。

## 1.2 BP 算法的不足

虽然 BP 算法得到广泛的应用, 但其也存在不足, 由于算法实质上是非线性优化问题的梯度算法, 所以

它存在收敛性的问题<sup>[4]</sup>。由于网络误差是在高维空间中具有复杂形状的曲面, 对这样复杂的误差曲面, 沿负梯度方向调整连接权值, 可能出现以下几种情况:

(1) 收敛到局部极小点。由于算法采用的是梯度下降法, 训练是从某一起始点沿误差函数的斜面逐渐达到误差的最小值, 而网络误差曲面是高维的凹凸不平的误差曲面, 因此, 在学习过程中可能陷入某个局部极小点。

(2) 在平坦区域内连接权值调整缓慢。误差曲面往往存在一些平坦区, 由于激活函数的导数趋于零, 即使误差较大, 但梯度调整过程几乎处于停顿状态。

(3) 在陡峭的沟壑中振荡。误差曲面通常也存在陡峭的沟壑, 这就使网络的权值收敛不到极小点上, 而在极小点周围来回振荡。

(4) 跳过较好的极小点。网络在学习过程中, 也可能在网络权值的调整过程中, 跳过较好的极小点, 这主要是由于误差曲面的凹处太窄, 而梯度值又较大, 致使网络权值从一个较好的极小点附近跳到另一个次好的极小点区域。

## 1.3 改进的 BP 算法

上面分析的几种情况, 在利用网络预测时间序列数据的训练过程中都有可能发生。在应用中网络的初始权值、隐节点数、动量因子、学习率、目标误差、训练步数等参数在相应 BP 网络训练方法中都必须注意选择。初始权值和隐节点数选取缺乏具体的理论指导, 通常选取不同的小伪随机数作为初始权值, 但由于随机性需要训练和仿真多次取平均值<sup>[5]</sup>。隐节点太少网络容错性差, 降低非线性映射能力; 隐节点太多训练易陷入局部极小点、训练收敛慢甚至发散, 并且降低网络泛化能力, 目前虽然存在许多数目选择经验公式但极不通用。具体问题中最优数目往往通过试验确定<sup>[6]</sup>。目标误差太大或步数太少, 训练不充分; 反之, 一则使训练时间太长, 二则造成对训练样本的过度拟和而降低对新样本的泛化能力。一般学习率选为 0.1~0.9, 动量因子选为 0.8~0.9。

网络算法对网络性能起决定性影响, 近年来, 许多研究人员对其作了深入的研究, 提出了许多改进的方法, 如加入动量项, 自适应学习率调节和弹性算法以及遗传算法和模拟退火算法等全局优化思想的引入<sup>[5~7]</sup>。在基于梯度下降方式的改进中最佳代表是加入动量因子和学习率自适应相结合的动量项自适应学习率方法, 此方法以加入动量因子作为阻尼项, 学习率也自适应调节, 因此既减少了震荡又加快收敛速

度。

基于数值优化的数学思想与 BP 算法相结合,训练速度极快而且测试性能很好,由于 BP 网络运算实质就是一个非线性函数优化问题,考虑将训练算法与非线性优化算法相结合作为改进的途径。将高斯-牛顿法加入因子  $\mu_k$  得 L-M (Levenberg-Marquardt) 优化方法:

$$x_{k+1} = x_k - [J^T(x_k)J(x_k) + \mu_k^I]^{-1}J^T(x_k)V(x_k)$$

进而得到 LM-BP 权值调节公式为:

$$w_{k+1} = w_k - [J^T(w_k)J(w_k) + \mu_k^I]^{-1}J^T(w_k)e(w_k)$$

其中,  $w_k$  为权值,  $e(w_k)$  为误差函数。这样通过自适应调节参数  $\mu_k$ , LM-BP 算法就综合了梯度下降法和高斯牛顿法的优点,网络性能极好。BP 网络改进方式很多,需要选择的网络参数又如此繁杂特殊,因此在具体实验时必须考虑改进方式以及网络参数的选择。

## 2 用改进的 BP 算法对数据进行预测

神经网络对时间序列进行预测,是用过去的  $N$  个时刻数据预测未来  $M$  个时刻数据,把  $L$  个样本数据分成  $K$  段长度为  $N+M$  的重叠数据段,其中,输入为  $X_k, X_{k+1}, \dots, X_{k+N-1}$  输出为  $X_{N+k}, X_{N+k+1}, \dots, X_{N+M+k-1}, k=1, 2, \dots, K$ 。共得到  $L-(N+M)-1$  个样本,实现一个 NR-MR 输入输出非线性映射,进而用网络进行训练和预测。

下面用 BP 神经网络改进方式并确定网络参数,对某省 2006 年 9 月份缺测数据补缺,然后对 2007 年 11 月份城镇居民人均可支配收入进行预测。采用单隐层 BP 网络,即由输入、一个隐含层、一个输出层组成,传递函数采用 Sigmoid 函数,由于各指标数据太大,指标之间又有不同的量级,输入太大会使网络传递函数提前达到饱和值,导致训练失败,而且大量级输入对小量级输入会产生信息覆盖,需要对网络的输入数据进行归一化预处理<sup>[8]</sup>。用最简单的归一化公式对数据进行归一化处理,  $x_i(i) = (x(i) - x_{\min}) / (x_{\max} -$

$x_{\min}), i=1, 2, \dots, p$ , 其中  $x(i)$  归一化到  $[0, 1]$ ,  $x(i)$  为城镇居民家庭人均可支配收入实际样本数据,  $x_{\max}, x_{\min}$  为输入实际样本数据的最大、最小值。归一化后大大加快了收敛速度,同时也不影响数据间的联系。网络的初始权值和阈值由网络自动选取。误差函数和训练次数自定,隐含层元数亦自定。经过不断的训练来最终确定合适的隐含层元数。模型见图 2。

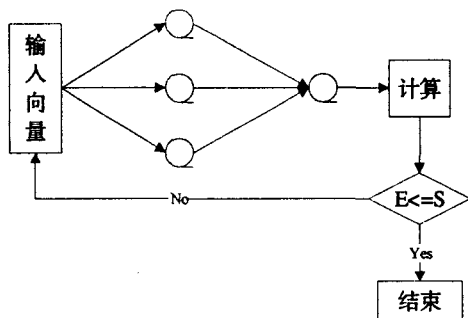


图 2 城镇居民人均可支配收入 BP 网络模型

每个样本有 9 个指标: A1 工业增加值、A2 固定资产投资、A3 地方财政预算收入、A4 地方财政预算支出、A5 对外贸易、A6 社会消费品零售额、A7 居民消费价格指数、A8 城镇居民人均消费性支出、A9 城镇居民人均可支配收入(单位:亿元)。可看出样本各指标之间存在着复杂非线性联系。样本统计数据如表 1 所示。

首先对缺测的 2006 年 9 月份城镇居民人均可支配收入进行补缺。考虑除 2006 年 9 月份以外样本每月份的前 8 个指标 A1、A2、A3、A4、A5、A6、A7、A8, 以及对应月份的城镇居民人均可支配收入 A9, 共组成 27 个样本, 每个样本建立输入输出非线性映射  $f: (A1_i, A2_i, A3_i, A4_i, A5_i, A6_i, A7_i, A8_i) \rightarrow A9_i$ , 其中  $(A1_i, A2_i, A3_i, A4_i, A5_i, A6_i, A7_i, A8_i, A9_i)$  为第  $i$  个样本的相应各指标数据,  $i=1, 2, \dots, 27$ 。采用归一化公式分别对输入输出各指标数据归一化, 然后建立单隐层 BP 网络进行网络训练。由于传统的 BP 算法收敛极慢, 本实验采用基于梯度下降方式的自适应动量学习率法训练, 网络输入节点为 8 个, 输出节点为 1

表 1 某省 3 个年度已知各月份 9 个经济指标的统计数据

(单位:亿元)

年月	A1	A2	A3	A4	A5	A6	A7	A8	A9
2005.06	26.22	46.54	7.12	37.56	0.78	31.20	99.7	2331.20	1768.45
2005.07	34.00	98.66	7.98	46.34	1.09	37.13	99.7	2879.54	2300.22
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
2006.08	77.29	200.00	17.13	68.21	2.56	77.90	112.47	4001.33	4987.48
2006.09	90.54	210.48	15.21	79.86	3.09	86.09	111.11	3987.34	(缺测)
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
2007.10	167.36	334.56	32.56	144.22	4.03	123.56	114.22	5673.67	7345.52
2007.11	187.45	365.67	35.89	176.10	4.15	177.80	100.90	6268.43	8356.45

个,网络隐层节点数过多过少都不好,经过节点数增加试验确定隐节点数为 20 个时网络性能就很好了。训练完成后仿真输出采用  $y_i(i) = (y(i) - y_{\min}) / (y_{\max} - y_{\min})$  函数反归一化。补缺前先随机选取了 2005 年 7 月份的样本数据测试网络性能,测试后反归一化仿真结果为 2331.53,与该月真实值为 2300.22 相比较可见仿真误差非常小。因此可以用 2006 年 9 月份已知前 8 个指标数据作为输入进行补缺。然后取所有反归一化结果的平均值 5483.56 作为该月份城镇居民人均可支配收入补缺数据,可以断定这个补缺值应该非常接近真实值。

其次对 2007 年 12 月城镇居民人均可支配收入进行预测。将补缺结果放入样本集,这样就防止样本集出现序列间断而保持了样本集连续性。记  $I_{1i}, I_{2j}, I_{3k}(i, j, k = 1, 2, \dots, 9)$  分别为 2005、2006 和 2007 年已知各月份顺次对应的城镇居民人均可支配收入,建立神经网络的输入输出非线性映射:  $I_{1,i} \rightarrow I_{1,i+1}, I_{2,j} \rightarrow I_{2,j+1}, I_{3,k} \rightarrow I_{3,k+1}$ , 即指标每年度的当前月数据与下一月数据建立输入输出。首先对输入输出进行归一化,然后进行网络训练,训练采用基于梯度下降方式的动量项自适应学习率法,单节点输入输出,隐层和输出层传递函数均用线性函数 purelin。由于网络规模小,网络改进方法和网络参数选取得当,所以训练很快。训练完成以后,采用归一化函数归一化后输入网络,仿真输出再进行反归一化。所有结果取平均值为 7996.75 作为 2006 年 12 月份的预测值,与该月的真实统计数据 8033.90 相比较误差很小。数据的补缺仿真结果有一定误差,但由于神经网络的容错能力,一个样本数据的微小误差基本上不会对样本集造成太大影响。即便有影响,从结果来看预测精度仍很高。由所有样本统计数据可知,2006 年较后月份样本序列与其它月份相比增幅大得多,所以网络训练后外推起来会有一定不足,也是造成预测误差的原因。以上试验完成以后,采用基于数值优化的改进方式 LM-BP 算法重新进

行试验。建立网络,选择合适的网络参数,采用改进的 LM-BP 算法进行权值调整。

### 3 结束语

改进后的 BP 网络在对城镇居民家庭人均可支配收入进行了预测。文中在原有 BP 网络的基础上增加了一个附加动量,它不仅考虑了误差在梯度上的作用,而且也考虑在误差曲面上的变化趋势的影响,无论从速度、仿真效果以及预测误差效果上来看,都优于传统的 BP 网络。但在实际训练过程中,由于隐含层元个数的不确定、学习速率的不断调整,这些都需要花大量时间去训练网络,所以改进的 BP 网络还有待进一步完善。

### 参考文献:

- [1] Pan H. Application of BP neural network based on genetic algorithm[J]. Computer Application, 2005, 25(12): 2777 - 2779.
- [2] He Fangguo, Qi Huan. Back propagation neural network based on modified genetic algorithm and its application[J]. Journal of Huaxhong Normal University: Nat. Sci., 2007, 41(2): 51 - 54.
- [3] Venkatesan R, Balamurugan B. A real - time hardware fault detector using an artificial neural network for distance protection[J]. IEEE Transactions on Power Delivery, 2001, 16(1): 75 - 82.
- [4] 杜华英,赵跃龙. 人工神经网络典型模型比较研究[J]. 计算机技术与发展, 2006, 16(5): 1 - 3.
- [5] 蒋蓉蓉. 一种基于遗传算法的 BP 网络改进方法[J]. 微计算机信息, 2007, 23(11): 234 - 236.
- [6] 王 爽, 张 鹰, 吕瑞霞. BP 神经网络的算法改进及应用[J]. 电脑知识与技术, 2009, 15(4): 933 - 935.
- [7] 王建平, 郭 尚. BP 神经网络预测算法性能的改进策略[J]. 微电子学与计算机, 2007, 24(10): 144 - 145.
- [8] 贾 群. PCA-BP 算法模块化设计的编程实现[J]. 计算机技术与发展, 2008, 18(2): 98 - 101.

(上接第 63 页)

### 参考文献:

- [1] 饶 元, 冯琴博, 李尊朝. 基于 Web Services 的服务合成技术研究综述[J]. 系统工程与电子技术, 2005, 27(8): 1481 - 1489.
- [2] Yang Jiang. Web Service Componentization[J]. ACM, 2003, 46(10): 35 - 40.
- [3] 欧毓毅, 郭荷清. Web 服务动态组合的研究[J]. 计算机应用研究, 2006(4): 22 - 27.
- [4] 邓水光, 俞 镇, 吴朝晖. 面向动态工作流的 Web 服务组合模型研究[J]. 计算机集成制造系统, 2004(6): 601 - 607.
- [5] 李景霞, 侯紫峰. 基于颜色 Petri 网的 Web 服务组合建模及应用[J]. 计算机应用研究, 2006(1): 22 - 27.
- [6] Hamadi R, Benatallah B. A Petri - Net - Based Model for Web Service Composition [C]//Proc. 14th Australasian Database Conf. Database Technologies. [s.l.]: ACM Press, 2003: 191 - 200.
- [7] 岳 昆, 王晓玲, 周傲英. Web 服务核心支撑技术: 研究综述[J]. 软件学报, 2004, 15(3): 428 - 442.
- [8] 周 燕, 姜 浩. 面向动态工作流的 Web 服务组合模型研究: 研究综述[J]. 计算机技术与发展, 2008, 18(1): 63 - 66.