

# 基于贝叶斯正则化 BP 神经网络的股票指数预测

杨海深 傅红卓

(华南理工大学理学院数学系, 广州 510640)

**摘 要** 提出了利用贝叶斯正则化 BP 神经网络对股票指数进行预测。通过对比实验表明, 贝叶斯正则化的 BP 神经网络比相同条件下采用其他改进算法有较好的泛化能力, 对股票指数预测有很好的效果。

**关键词** 神经网络 贝叶斯正则化 上证指数

**中图法分类号** TP183 **文献标志码** A

股票市场的建模和预测研究对我国的经济发展和金融建设具有重要的意义。在当前的证券分析中, 普遍采用的方法有两种: 基本分析法和技术分析。技术分析立足于股票交易中历史事件重现的假设上, 以人为判断和将过去的股价图相比较得到最有可能的结果, 带有较大的主观性; 另外, 技术分析存在着一定的时滞现象<sup>[3]</sup>。由于股票价格序列可以看作是含有白噪声的时间序列, 因而可以通过建立股价序列的适应或最优统计模型对股价的变化做出分析和预测<sup>[4]</sup>。近年来, 基于人工神经网络的分析预测方法日益成熟, 由于其良好的非线性映射能力及自适应、自学习等优良特性, 因而神经网络已经被广泛地应用于对金融领域股票价格、收益、汇率及风险等的分析和预测中<sup>[5]</sup>。

BP 神经网络具有优良的非线性逼近能力, 在处理缺失值和非线性问题时有明显的优越性。然而标准 BP 算法存在着局部收敛、收敛速度过慢和过度拟合等问题, 为克服其弱点, 人们在标准 BP 算法的基础上进行了许多有益的改进, 如加入动量法, L-M 优化算法以及共轭梯度学习算法。但是这些算法的 BP 神经网络的泛化能力在股票市场预测中没有达到所期望的效果, 而泛化能力是衡量神经网络性能

好坏的重要标志。本文针对上述 BP 算法的不足, 提出了利用贝叶斯正则化 BP 神经网络对股票指数进行预测。通过对比实验表明, 贝叶斯正则化 BP 神经网络比相同条件下采用其他改进算法有较好的泛化能力, 对股票指数预测有很好的效果。

## 1 贝叶斯正则化方法的 BP 神经网络

### 1.1 基本的 BP 网络

BP 神经网络通常由输入层、输出层和若干隐含层构成, 每层有若干个节点组成, 每个节点表示一个神经元, 上层节点与下层节点通过权连接, 同一层节点之间没有联系。<sup>[1]</sup>

BP 神经网络的学习过程有两部分组成: 信息的正向传播和误差的反向传播。当正向传播时, 输入信息经隐含层和输出层的神经元逐层处理, 向前传播到输出层, 给出结果。如果输出层得不到期望输出, 则转入逆向传播过程, 将误差信号沿原来的神经元连接通路返回。在返回的过程中, 修改各层神经元间的连接权值, 然后再转入正向传播过程, 反复迭代, 使得误差不断减少, 直到误差小于给定的值为止。归结起来: “模式顺传播” → “误差逆传播” → “记忆训练” → “学习收敛”过程。具体的算法参见文献[6]。

### 1.2 贝叶斯正则化 BP 神经网络

影响神经网络的泛化能力主要依赖于网络结

2009 年 3 月 18 日收到

第一作者简介: 杨海深 (1982—), 男, 河南商丘人, 硕士研究生, 研究方向: 统计及其应用, 神经网络。E-mail haishen\_y@mail.scut.edu.cn

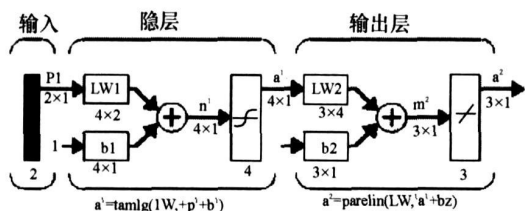


图 1 一个隐层的 BP神经网络

构和训练样本的特性,因此可以选择合适的训练策略和优化网络结构来提高其泛化能力。选择合适的训练策略包括在训练样本中加入噪声、提前停止法等方法;网络结构优化方法主要有修剪法、正则化法、进化法等;本文采用正则化的方法。

### 1.2.1 正则化方法

正则化方法是通过修正神经网络的训练性能函数来提高其泛化能力。一般情况下,神经网络的性能函数采用均方误差函数。记神经网络模型训练样本  $D = (x_i, t_i)$ ,  $i = 1, 2, \dots, n$ ,  $n$  为训练样本总数,  $W$  为网络参数向量。给定网络结构  $H$  和网络参数  $W$  的条件下,网络的误差函数  $E_D$  取误差的平方和:

$$E_D = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^k (f(x_i, W, H) - t_i)^2 \quad (1)$$

(1)式中  $f(\cdot)$  为网络的实际输出,  $k$  为神经网络的输出量。

常用的正则化方法是在误差函数后加上权重衰减项  $E_W$ :

$$E_W = \frac{1}{2} \|W\|^2 = \frac{1}{2} \sum_{i=1}^m w_i^2 \quad (2)$$

(2)式中  $m$  为网络参数总数,于是总误差函数可定义为:

$$F(W) = E_D + \alpha E_W \quad (3)$$

(3)式中  $\alpha, \beta$  是超参数,控制着其他参数(权及阈值)的分布形式。超参数的大小决定着神经网络的训练目标,若  $\alpha \ll \beta$  则侧重于减小训练误差,但可能过拟合;若  $\alpha \gg \beta$  则侧重于限制网络权值规模,但可能误差较大。在实际应用中,需要折中考虑,极小化目标函数是为了减少网络训练误差的同时,降低网络结构的复杂性。对于正则化方法而言,难点在于超参数的确定,该文采用贝叶斯方法实现超参数

的选择。

### 1.2.2 神经网络的贝叶斯学习

David Mackay<sup>[1]</sup>将贝叶斯方法用于神经网络建模过程中,之后又对网络训练过程中后验概率计算问题提出了基于高斯逼近方法来计算 Hessian 矩阵。在贝叶斯理论的框架下,网络的参数被认为是随机变量,给定样本数据下,由贝叶斯规则,参数的分布函数为:

$$p(W|D, \alpha, \beta, H) = \frac{p(D|W, \beta, H)p(W|\alpha, \beta, H)}{p(D|\alpha, \beta, H)} \quad (4)$$

(4)式中  $p(D|W, \beta, H)$  为似然函数,  $p(D|\alpha, \beta, H)$  是归一化因子,  $p(W|\alpha, \beta, H)$  是先验密度,表示在没有数据样本下的参数  $W$  (权向量)的先验知识。由于对参数的分布只有很少的知识,因此先验分布是一个很宽的分布,一旦有了数据,可转化为后验分布,后验分布较为紧凑,即只有在很小范围中的权值才可能与网络的映射一致。

假设训练样本的总体分布是正态分布,权参数的先验分布也是正态的。从而似然函数和先验分布函数分别为

$$p(D|W, \beta, H) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \quad (5)$$

$$p(W|\alpha, \beta, H) = \frac{1}{Z_W(\alpha)} \exp(-\alpha E_W) \quad (6)$$

式(5)、式(6)中  $Z_D(\beta) = (\frac{\pi}{\beta})^{\frac{n}{2}}$ ,  $Z_W(\alpha) = (\frac{\pi}{\alpha})^{\frac{m}{2}}$ 。

把式(5)、式(6)代入式(4),得到:

$$p(W|D, \alpha, \beta, H) = \frac{\frac{1}{Z_W(\alpha)} \frac{1}{Z_D(\beta)} \exp(-(\beta E_D + \alpha E_W))}{\text{归一化因子}} = \frac{1}{Z_F(\alpha, \beta)} \exp(-F(W)) \quad (7)$$

贝叶斯学派认为,最优的权参数应该极大化后验概率  $p(W|D, \alpha, \beta, H)$ 。从式(7)可以看出,极大化后验概率等价于极小化总误差函数  $F = \beta E_D + \alpha E_W$ 。

下面仍然使用贝叶斯规则来优化超参数  $\alpha$  和  $\beta$  还是需要求后验分布:

$$p(\alpha, \beta|D, H) = \frac{p(D|\alpha, \beta, H)p(\alpha, \beta|H)}{p(D|H)} \quad (8)$$

假设先验分布  $p(\alpha, \beta|H)$  是一种很宽的分布函

数,也即均匀分布。因为式(8)中的归一化因子 $p(D|H)$ 与 $\alpha$ 和 $\beta$ 无关,因此求取最大后验分布的问题就转化为求解最大似然函数 $p(D|\alpha, \beta, H)$ 。注意到似然函数 $p(D|\alpha, \beta, H)$ 是式(4)的归一化因子,因此可以得到:

$$p(D|\alpha, \beta, H) = \frac{p(D|W, \beta, H)p(W|\alpha, H)}{p(W|D, \alpha, \beta, H)} \quad (9)$$

将式(5)、式(6)、式(7)代入式(9)得到:

$$p(D|\alpha, \beta, H) = \frac{Z_F(\alpha, \beta)}{Z_w(\alpha)Z_D(\beta)} \quad (10)$$

当样本数据量较多时,后验分布趋于正态分布,若后验分布曲线足够窄,峰值足够尖锐,则可利用泰勒展开式对问题作进一步的简化,以求出 $Z_F(\alpha, \beta)$ 。

设 $F(W)$ 取最小值时所对应的权值(包含阈值)为 $W_{MP}$ ,将 $F(W)$ 在 $W_{MP}$ 附近泰勒展开,忽略高次项,得到:

$$Z_F \approx (2\pi)^{\frac{m}{2}} (\det(\dot{\dot{F}}(W_{MP}))^{-1})^{\frac{1}{2}} \times \exp(-F(W_{MP})) \quad (11)$$

(11)式中 $\dot{\dot{F}}(W_{MP}) = \beta \dot{\dot{F}}_{ED} + \alpha \dot{\dot{F}}_{EW}$ 是 $F(W)$ 在 $W_{MP}$ 点的Hessian矩阵。

将式(11)代入式(10),并利用最大似然原理,求出满足似然函数 $p(D|\alpha, \beta, H)$ 最大的 $\alpha$ 和 $\beta$ 即得到最优的超参数。

$$\alpha_{MP} = \frac{\gamma}{2E_W(W_{MP})} \text{ 和 } \beta_{MP} = \frac{n-\gamma}{2E_D(W_{MP})} \quad (12)$$

(12)式中 $\gamma = m - 2\alpha_{MP} \text{tr}(\dot{\dot{F}}(W_{MP}))^{-1}$ 称为有效参数个数。 $m$ 是网络参数总数。 $\gamma$ 表示网络中有多少参数在减少总误差函数方面起作用,它的取值范围为 $[0, m]$ 。

在进行优化求解时需要计算 $F(W)$ 在其最小点 $W_{MP}$ 处的Hessian矩阵,也即需要计算 $\dot{\dot{F}}(W_{MP})$ ,计算量比较大。为了提高计算速度,可利用高斯-牛顿逼近法对Hessian阵作进一步简化,得到: $\dot{\dot{F}}(W_{MP}) \approx 2\beta J^T J + 2\alpha I_m$ ,其中 $J$ 是 $E_D$ 的在点 $W_{MP}$ 雅可比矩阵<sup>[2]</sup>。

### 1.2.3 贝叶斯正则化BP神经网络训练步骤

(1)确定网络结构,初始化超参数 $\alpha$ 和 $\beta$ ,一般设定为 $\alpha=0$ 和 $\beta=1$ 。根据先验分布对网络参数赋初值;

(2)用BP算法训练网络使总误差 $F(W)$  =

$\beta E_D + \alpha E_W$ 达到最小;

(3)利用高斯-牛顿逼近法求Hessian阵 $\dot{\dot{F}}(W_{MP}) \approx 2\beta J^T J + 2\alpha I_m$ ,并计算有效参数个数 $\gamma$ ;

(4)由式(12)计算超参数 $\alpha$ 和 $\beta$ 的新的估计值;

(5)重复执行(2)、(3)、(4)直到达到所需精度。

贝叶斯方法正则化神经网络是个迭代过程,每个迭代过程总误差函数随着超参数的变化而变化,最小值点也在变化,网络的参数也在不断修正,最终达到总误差函数在迭代过程中没有较大改变,也称此时收敛。因为目前在网络结构的选择方面还没有理想的方法,在实际工作中常常需要用试验的方法确定最佳的网络结构,因此可采用不同的网络结构进行网络训练,然后比较这些网络模型的显著度,选择显著度较大的网络作为模型。

## 2 实证分析

### 2.1 研究样本的获取

股票市场是一个很不稳定的动态变化过程,不仅受国内外经济因素的影响,而且受到庄家的炒作、政府的调控等因素的影响。因此,研究所选取的是在股价波动较为平稳的市场下,即没有或少有暴涨和暴跌等不稳定现象的样本数据。否则,如果样本很“特殊”,就只能抽取到某些特定的规律,而降低了网络的泛化能力。我们采用的数据是2008年3月28日—2008年12月30日的上证指数(数据来源于锐思金融研究数据库)共188个数据作为例子说明本方法,其中158个数据作为训练样本,30个测试样本。采用上证指数有三个原因:一是个股容易被人为炒作,上证指数综合了很多个股的各种因素,即使有个别个股被人为炒作,但只要大部分个股的走势反映了股市的内在规律,从总体上来说,这时的上证指数仍然能反映股市的内在规律,即上证指数受外界干扰的程度远不如个股那样明显;二是这个指数可以最大程度地反映市场整体价格水平,产生的信号对投

投资者的影响最为强烈;三是指数能够比较准备的反映国内股市行情动态,具有较高的预测价值和较好的可预测性。

## 2.2 预测指标的选取

本文预测指标的选取立足于原始数据,即上证指数的开盘价,收盘价,最高价,最低价,交易量与交易金额。选择了过多的指标和收集过多的数据,这些信息中可能有冗余信息。而这些冗余信息不仅增加了问题的复杂性,而且还会影响预测的准确性。在不损失信息的情况下应充分利用各种方法,如相关性分析、逐步回归分析、主成分分析法、单变量显著性检验等方法对初选的指标进行筛选。通过反复比较最终确定了 11 输入变量。而输出变量为明日收盘价。

表 1 预测指标表

$X_1$ : 今日 开盘价	$X_2$ : 今日 最高价	$X_3$ : 今日 最低价	$X_4$ : 今日 收盘价	$X_5$ : 今日 成交 金额	$X_6$ : 今日 成交 量
$X_7$ : 昨日 收盘价	$X_8$ : 前日 收盘价	$X_9$ : 前五 日平均 收盘价	$X_{10}$ : 前 十日平 均收 盘价	$X_{11}$ : 前 二十日 平均收 盘价	

## 2.3 构建贝叶斯 BP网络模型

### 2.3.1 数据的归一化处理

为了计算方便、消除量纲影响、提高预测精度,需要对输入数据进行预处理。应用公式  $p_n = 2(p - p_{\min}) / (p_{\max} - p_{\min}) - 1$  可以把原始数据变化到  $[-1, 1]$  之间。本文采用预测领域常用的近期放大,远期缩小的数据处理办法。将分量  $X_8$  限定在  $[0, 0.3]$  之间,分量  $X_9$  限定在  $[0, 0.24]$ ,分量  $X_{10}$  限定在  $[0, 0.23]$ ,分量  $X_{11}$  限定在  $[0, 0.2]$ ,其余各分量在区间  $[-1, 1]$ 。在试验数值表明,输入向量各分量经这样处理后,预测结果确实比所有分量毫无差别地统一转化为区间  $[-1, 1]$  要好。

### 2.3.2 网络结构的确定

BP神经网络的结构难以确定,存在一个结构优化的问题。网络的精确性和训练效率很大程度上取决于网络结构和控制参数。网络结构包括网络层数、各层神经元个数等控制参数是指输入变量、

允许误差、学习速率和初始权值等参数。

### (1) BP网络层数和激活函数的确定

模型层数的确定也就是隐含层的确定,本文选择三层 BP神经网络来建立预测模型;即一个输入层,一个隐含层,一个输出层。隐含层节点数的选择有一些基本的原则(参见文献[6])。本文用试验试凑法来确定隐含层节点数  $J$ ,分别测试了  $J=4, 5, 6, 7, 8, 9$ ,比较误差效果及其收敛速度,最终选择了  $J=7$ 。也即网络结构为  $11-7-1$ ,隐含层的激活函数是双极性 S 型函数,输出层是线性函数。

### (2) 学习速率的选择

学习速率反映的是网络一次循环训练中权值的变化量,在优化计算中是一个重要的因子,每一个具体网络都存在一个适合的学习速率。在试算过程中采用学习速率为 0.01,收敛速度很快,训练时间很短。

### (3) 初始权值和允许误差的设置

初始权值选取原则是取随机值,且保证权值比较小。一般取一组  $-1 \sim 1$  之间均匀分布的随机数。输出的允许误差小,能得到好的学习精度,但存在对样本的过度拟合,网络的推广能力较差。本文允许误差设为 0.001。

## 2.4 对比实验

将贝叶斯正则化的 BP神经网络算法,加入动量项的 BP算法和基于 LM 算法的 BP网络进行训练与测试。对比效果如图 2~图 4。

从图 2~图 4 很清楚地看到不同 BP神经网络的算法所产生的预测效果,发现应用贝叶斯正则化方法的 BP神经网络有较强的泛化能力,其次是 LM 算法的 BP神经网络。在贝叶斯正则化方法中,迭代 329 次算法收敛,有效参数个数  $\gamma = 13.4$  误差为 0.0018 有较小的权重值 2.62043 可以简化网络,从而可以使神经网络的输出曲线具有一定的光滑度,以避免过分复杂的曲线对训练数据过拟合,提高神经网络的泛化能力。

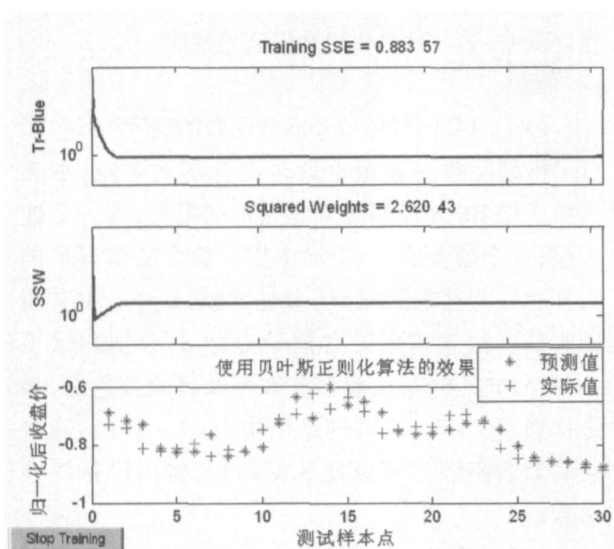


图2 贝叶斯正则化的 BP 神经网络

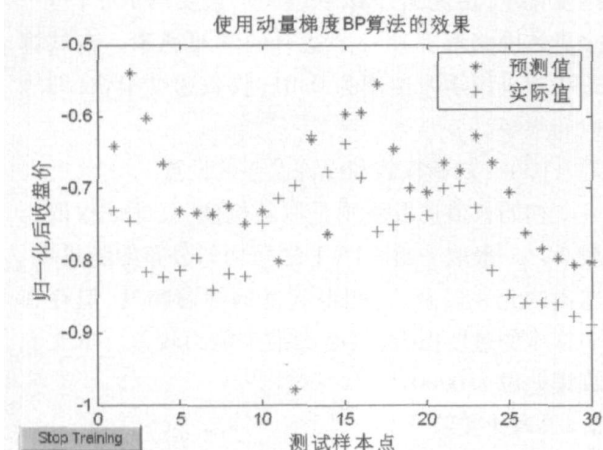


图3 加入动量的 BP 神经网络

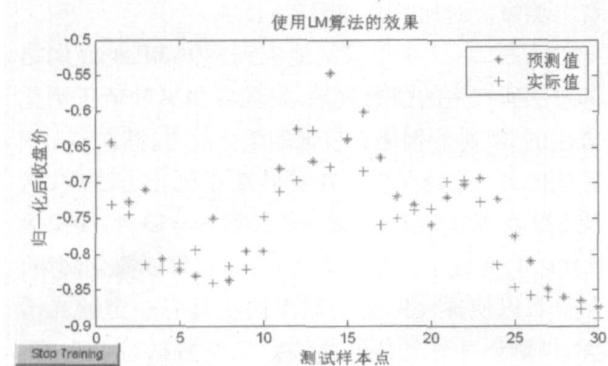


图4 基于 L-M 算法的 BP 神经网络

### 3 结论及展望

本文提出了使用贝叶斯正则化 BP 神经网络对上证指数进行预测, 通过对比实验, 发现其具有较强的泛化能力。贝叶斯正则化神经网络的主要优点是:

(1) 正则化项系数只需通过训练集在训练过程中就能确定, 而无需独立的训练集和测试集;

(2) 能够利用超参数的先验信息, 在应用模拟数据和实例研究中都获得了比较好的效果, 提高神经网络的泛化能力, 并且在样本量有限的情况下, 使得网络结构也具有较好的泛化能力。

同时对检验集进行训练和仿真, 贝叶斯正则化方法的相关系数为 0.832, 加入动量法的相关为 0.682, 基于 LM 算法的相关系数为 0.798, 可见贝叶斯正则化神经网络比传统的神经网络模型有更好的拟合, 预测效果更好。

面对多变的市场环境和经济状况, 对股票指数预测的要求将更为精确、更为快捷。可以预见, 将神经网络等人工智能技术引入到股票价格预测中, 建立准确的预测模型, 将是解决这一类问题的有效途径。

### 参 考 文 献

- 1 Mackay D. J. C. Bayesian interpolation. *Neural Computation*, 1992; 4: 415—447
- 2 Foresee F D, Hagan M T. Gauss-Newton approximation to bayesian learning. In: *Proceedings of the International Conference on Neural Networks*. Houston, Texas, 1997
- 3 陈之大, 贺学会. 证券投资技术分析. 成都: 西南财经大学出版社, 1996
- 4 王 波, 张凤玲. 神经网络与时间序列模型在股票预测中的比较. *武汉理工大学学报·信息与管理工程版*, 2005; 2(6): 69—73
- 5 冯俊秋, 李 挥. 人工神经网络在股票市场预测中的应用. *深圳大学学报(理工版)*, 1998; 15(2—3): 68—75
- 6 阎平凡, 张长水. 人工神经网络与模拟进化计算. 北京: 清华大学出版社, 2000
- 7 杨锦明. 基于贝叶斯正则化 BP 神经网络的上市公司财务困境预警模型. 湖南大学硕士学位论文, 2007
- 8 魏 东, 张明廉, 等. 基于贝叶斯方法的神经网络非线性模型辨识. *计算机工程与应用*, 2005; 11: 5—8

(下转第 3318 页)

## Design of SIP-Based Multimodel Intelligent Measuring and Control Platform

ZHOU Qin-he<sup>1,2</sup>, LIU Guixiong<sup>1</sup>, HONG Xiaobin<sup>1</sup>

(College of Mechanical Engineering, South China University of Technology<sup>1</sup>, Guangzhou 510640, P. R. China)

Guangdong Technical College of Water Resources and Electric Engineering<sup>2</sup>, Guangzhou 510635, P. R. China)

**[Abstract]** Aimed at the difficulty that the traditional measuring and control network is difficult to meet the demand of multi-source data and complex type data in demmanufacturing industry. A SIP-based intelligent measuring and control platform are designed, established the SIP-based intelligent measuring and control network architecture, and planed its functions such as intelligent accessing, self-description ability of data transmission, existence notification mechanism, asynchronous event notification mechanism, online device control mechanism and so on. The practice indicates that this intelligent measuring and control platform has a more perfect function and performance compared with the traditional measuring and control network.

**[Key words]** intelligent measuring and control platform SIP DDP

(上接第 3310页)

## Stock Index Forecast Based on Bayesian Regularization BP Neural Network

YANG Haishen, FU Hong-zhuo

(Department of Mathematics, South China University of Technology, Guangzhou 510640, P. R. China)

**[Abstract]** Using Bayesian regularization BP neural network to predict stock index was proposed. Experimental results show that Bayesian regularization method has stronger generalization capabilities than other improved BP algorithms under the same situation, and is well effective to forecast the stock index.

**[Key words]** BP neural network Bayesian regularization algorithm Shanghai stock index