

# BP 神经网络在股票预测中的应用

王爱平,陶嗣干,王占凤

(安徽大学 计算智能与信号处理教育部重点实验室,安徽 合肥 230039)

**摘要:**介绍了时间序列和 BP 神经网络算法,并给出信号正向传播和误差反向传播的公式,在此基础上运用 BP 算法对股票时间序列短期的走势进行预测分析。实验结果表明此方法效果良好,识别率高,具有一定的实用价值。

**关键词:**时间序列;BP 算法;股票预测

中图分类号: TP183

文献标识码: B

## New stock forecast based on BP neural network

WANG Ai Ping, TAO Si Gan, WANG Zhan Feng

(Ministry of Education Key Laboratory of Intelligent Computing & Signal Processing, Anhui University, Hefei 230039, China)

**Abstract:** The concept of time series and BP neural network algorithm are introduced in this paper. The signal forward propagation and error back propagation formula are proposed through studying the BP neural network algorithm. According to this, the short-term trend prediction and analysis of the stock time series are completed by using BP algorithm. Experimental results show that the result is excellent, the recognition rate is high.

**Key words:** time series; BP algorithm; stock prediction

股票是证券市场的重要组成部分,股票的发行和交易促进了市场经济的发展。而股票市场是一个高度复杂的非线性动态系统,其变化规律由于受经济、政治、公司经营况况和市场人气等多方面的影响,使股票价格走势呈现不规律性。为了能够预测其变化趋势,人们研究出了一些股票分析方法<sup>[1-2]</sup>,如:技术分析(移动平均法 MA、成交量图法 VOL 等)、基本面分析(宏观经济、国际政治、公司经营况况等)和心理分析(股票市场中的羊群效应)等。而这些方法大部分是线性模型,对股票短期走势预测效果不是很理想。而 BP 神经网络<sup>[3]</sup>能够把输入输出问题转化为非线性映射问题,即完成由  $n$  维输入空间到  $m$  维输出空间的非线性映射,能很好地解决非线性问题。本文利用时间序列对股票价格的原始数据进行预处理,然后利用 BP 神经网络对上证 A 股皖维高新近期股价涨跌走势进行了预测。

### 1 时间序列

对某一个或一组变量  $x(t)$  进行观察测量,将在一系列时刻  $t_1, t_2, \dots, t_n$  ( $t$  为自变量且  $t_1 < t_2 < \dots < t_n$ ) 所得到的离散数据组成的序列集合  $\{x(t_1), x(t_2), \dots, x(t_n)\}$  称之为时间序列<sup>[4]</sup>, 记为  $X = \{x(t_1), x(t_2), \dots, x(t_n)\}$ 。一般时间序列由 4

部分构成,即长期或趋势变化、循环变化、季节性变化和随机变化。本文研究时间序列短期趋势变化,来获取所需要的知识。对于样本集  $X = \{x(t_1), x(t_2), \dots, x(t_n)\}$ , 每一个训练样本  $x(t_i)$  由一些重要的特征属性组成,令  $Y = \{y(t_1), y(t_2), \dots, y(t_n)\}$  为目标期望,即由特征属性所决定的决策属性值。例如在股票涨跌预测中,一只股票每天的开盘价、最高价、最低价和收盘价等特征属性影响次日股票的走势。由此可以按日时间序列来构造样本向量,但这种方式没有考虑到时间序列数据前后相互影响,割裂了时序特性。为了更好地反映时间序列数据之间的关系,将前  $N$  天的股票特征属性对当日股票的影响也考虑进来,从而可以得到  $N$  天时间序列,这时得到的样本中特征属性增加了许多,给 BP 网络训练增加了难度。由于股票的涨跌是依据昨日与今日收盘价的高低比较得到的,这时可以约去前  $N-1$  天的开盘价、最高价、最低价,得到前  $N-1$  天的收盘价和第  $N$  天开盘价、最高价、最低价和收盘价样本,从而使网络得到进一步简化。

### 2 BP 算法

BP 神经网络是人工神经网络的重要模型之一,也

是目前应用最广泛的一种多层前馈神经网络,可以解决大多数神经网络面临的问题。BP 算法<sup>[3,5]</sup>的思想是学习过程由信号的正向传播与误差的反向传播两个过程组成,若输出层的实际输出与目标期望不符,则转入误差的反向传播,调整各层神经元的权值,此过程一直进行到训练网络预先设定好的学习次数或小于网络输出的误差值。目前使用最多的是三层 BP 网络,包括输入层、隐层和输出层。设三层 BP 网络,输入向量为  $X=(x_1, \dots, x_i, \dots, x_n)^T$ , 隐层输出向量为  $Y=(y_1, \dots, y_j, \dots, y_m)^T$ , 输出层输出向量为  $O=(o_1, o_2, \dots, o_k, \dots, o_l)^T$ , 期望输出向量为  $d=(d_1, d_2, \dots, d_k, \dots, d_l)^T$ , 输入层到隐层之间的权值为  $V=(V_1, V_2, \dots, V_j, \dots, V_m)$ , 隐层到输出层之间的权值为  $W=(W_1, W_2, \dots, W_k, \dots, W_l)$ 。

(1) 正向传播过程中隐层的输出可表示为:

$$y_j = f\left(\sum_{i=0}^n v_{ij} x_i\right) \quad j=1, 2, \dots, m$$

其中  $x_0 = -1$  是隐层神经元引入阈值而设定的。

输出层的输出可表示为:

$$o_k = f\left(\sum_{j=0}^m w_{jk} y_j\right) \quad k=1, 2, \dots, l$$

其中  $y_0 = -1$  是输出层神经元引入阈值而设定的。

(2) 反向传播过程中输出层的权值修改可表示为:

$$\Delta w_{jk} = \eta (d_k - o_k) o_k (1 - o_k) y_j \quad k=1, 2, \dots, l; j=0, 1, \dots, m$$

隐层的权值修改可表示为:

$$\Delta v_{jk} = \eta \left( \sum_{k=1}^l \delta_k^0 w_{jk} \right) y_j (1 - y_j) x_i$$

其中  $\delta_k^0 = (d_k - o_k) o_k (1 - o_k)$   $k=1, 2, \dots, l; i=0, 1, \dots, n; \eta$  为学习速率。

### 3 使用 BP 神经网络对股票进行预测

#### 3.1 数据处理

文中皖维高新股票数据是从华林证券软件上证 A 股日线报表中整理而得,每个样本数据包括开盘价、最高价、最低价和收盘价 4 个特征属性以及第  $m$  日的收盘价和第  $(m+1)$  日的收盘价比较结果的样本,即目标期望值。

定义:若第  $m$  日的收盘价小于第  $(m+1)$  日的收盘价,令期望值为 0,表示股价下跌;反之若第  $m$  日的收盘价大于第  $(m+1)$  日的收盘价,令期望值为 1,表示股价上涨。部分样本信息如表 1 所示。

为了更好地反映时间序列数据内部之间的关联性,引入  $m$  日交叉时间序列。样本集  $X=\{x(t_1), x(t_2), \dots, x(t_n)\}$ , 其中  $x(t_i)$  为第  $i$  个交易日样本数据,令  $x(t_i)=(v_{i1}, v_{i2}, v_{i3}, v_{i4})$   $i=1, 2, \dots, n$ , 其中  $v_{i1}, v_{i2}, v_{i3}, v_{i4}$  分别代表第  $i$  天的开盘价、最高价、最低价和收盘价。于是可得  $m$  日交叉时间序列样本为:

$$y_1 = \{x(t_1), x(t_2), \dots, x(t_m), o_{m+1}\}$$

$$y_2 = \{x(t_2), x(t_3), \dots, x(t_{m+1}), o_{m+2}\} \dots$$

$$y_{n-m+1} = \{x(t_{n-m+1}), x(t_{n-m+2}), \dots, x(t_n), o_{n+1}\}$$

表 1 皖维高新股票部分训练样本和期望值

时间	开盘价	最高价	最低价	收盘价	期望值
2008-10-09	5.36	5.38	5.15	5.18	0
2008-10-10	4.81	5.03	4.78	4.79	1
2008-10-13	4.80	5.27	4.62	5.27	0
2008-10-14	5.42	5.50	4.80	4.91	0
2008-10-15	4.88	4.94	4.75	4.87	0
2008-10-16	4.53	4.78	4.49	4.55	1
2008-10-17	4.61	4.72	4.58	4.67	1
2008-10-20	4.61	5.03	4.53	4.96	0
2008-10-21	5.00	5.04	4.80	4.84	0

其中  $o_i$  为第  $(m+1)$  日股价的期望值。

#### 3.2 网络参数的确定

由于股价涨跌期望值是由 1 或 0 组成,故网络的传递函数选择单极性 Sigmoid 函数,即  $f(x) = \frac{1}{1+e^{-x}}$ , 用测试样本预测时,设网络输出向量为  $O$ ,若分量值大于 0.5,判为 1;反之,判为 0。使用 Matlab 工具箱提供的自适应 lr 的 BP 算法训练函数,网络训练要求输出误差  $E \leq 1e^{-3}$ 。

#### 3.3 实验结果

以皖维高新为例,对该股票信息按照 1 日、3 日和 5 日交叉时间序列进行数据整理(时间从 2008 年 10 月 9 日~2009 年 1 月 5 日共 60 天数据,2009 年 1 月 6 日~2009 年 1 月 21 日共 12 天数据),预测结果如表 2 所示。

表 2 1 日、3 日和 5 日交叉时间序列预测结果

$m$ 日交叉时间序列	样本数/个	测试样本数/个	准确率/%
1	60	12	66.7
3	58	10	80
5	56	8	50

由表 2 可知,3 日交叉时间序列样本,预测准确率最高,达到 80%。而 5 日交叉时间序列样本预测准确率最低,只有 50%。对于股票短期的走势可能受到国家政策因素(比如降息、经济振兴规划等),外围股票市场多方面的影响,5 日交叉时间序列数据可能存在一定的噪声,使最终预测的结果偏低。比较 3 日交叉时间序列和 1 日时间序列预测的结果,发现 3 日预测的正确率远高于 1 日的正确率,从中可以看出,第  $(m+1)$  天的股价不仅受第  $m$  天股价的影响还受  $(m-1)$  天股价的影响,只是影响股价的程度不一样而已。由此可以看出时间序列数据之间存在相互依赖性,选择好时间周期对预测结果的准确率起到关键作用。

股票数据样本按交叉时间序列处理,更能体现数据之间的关联性,在处理非线性问题上,BP 算法发挥了很大的优势。对于短期股票走势的预测,效果较好,具有一定的实用价值。但在实验中,也存在一些问题:(1)BP 算法收敛速度慢,易形成局部极小而得不到全局最优,网

络训练要经过几次或十几次才能达到期望输出,当然这也是BP网络自身的缺点,不易克服;(2)样本数据的选取,时间序列数据不能太短也不能太长,如果选取的时间序列数据小于股票变化的周期,预测的结果不能很好地表现股票价格的变化趋势;但是如果选取的时间序列数据远大于股票变化的周期,网络训练加重,而且数据中可能含有噪声,影响预测的准确度。所以,选择多长时间周期的股票数据,适合大部分股票的预测,是一个值得研讨的问题。

#### 参考文献

- [1] 张杨武.基于时间序列和神经网络相结合的股票预测研究[J].电脑知识与技术,2009,5(1):208-209.
- [2] 胡书晓.基于神经网络的股票K线特征图形预测研究[D].上海:同济大学学位论文,2008.

[3] 韩力群.神经网络[M].北京:北京邮电大学出版社,2006.

[4] 张善文,雷英杰,冯有前.MATLAB在时间序列分析中的应用[M].西安:西安电子科技大学出版社,2007.

[5] 姚培福,许大丹.BP神经网络在股票预测中的应用研究[J].广东自动化与信息工程,2006(1):7-9.

(收稿日期:2009-10-22)

#### 作者简介:

王爱平,女,1956年生,教授,主要研究方向:数据挖掘、人工智能、编译技术、计算机仿真以及滤波算法收敛性等领域。

陶嗣干,男,1981年生,硕士研究生,主要研究方向:数据挖掘。

王占凤,女,1984年生,硕士研究生,主要研究方向:智能计算。

电子技术应用  
APPLICATION OF ELECTRONIC TECHNIQUE  
www.chinaAET.com