

# 基于神经网络的股票价格预测模型

陈巍瑛<sup>1</sup> 张泽星<sup>2</sup> 李文斌<sup>1</sup>

<sup>1</sup>(石家庄经济学院信息工程学院 河北 石家庄 050031)  
<sup>2</sup>(石家庄经济学院网络信息安全实验室 河北 石家庄 050031)

**摘 要** 提出一种基于 BP 神经网络的股票价格预测模型 SPPM (Stock Price Prediction Model)。SPPM 集成了多个神经网络,可预测未来若干天的股价走势。针对 SPPM 中的数据预处理、输出融合、神经网络隐藏层节点数选取等关键问题作了详细讨论。实验结果表明,SPPM 具有一定的实际价值。

**关键词** 神经网络 时间序列 股票

中图分类号 TP183 文献标识码 A DOI:10.3969/j.issn.1000-386x.2014.05.022

## NEURAL NETWORK BASED STOCK PRICE PREDICTION MODEL

Chen Yiyi<sup>1</sup> Zhang Zexing<sup>2</sup> Li Wenbin<sup>1</sup>

<sup>1</sup>(School of Information Engineering, Shijiazhuang University of Economics, Shijiazhuang 050031, Hebei, China)  
<sup>2</sup>(The Network and Information Security Lab., Shijiazhuang University of Economics, Shijiazhuang 050031, Hebei, China)

**Abstract** A BP neural network based stock price prediction model (SPPM) is put forward. SPPM integrates a number of neural networks, so that it can predict stock price trends in a few days. Some key problems of SPPM such as data preprocessing, output fusion, neural network hidden layer node number setting and so on are discussed in detail. Experiment results show that SPPM is of certain practical value.

**Keywords** Neural network Time series Stock

### 0 引 言

股票市场是一个复杂的非线性动力学系统,有显著的非线性、时变性特征,致使对股市指数和股票价格的预测成为一大难题<sup>[1]</sup>。在传统的股票市场预测建模研究中,多元回归分析、时间序列分析、指数平滑等是最常见的方法<sup>[2,3]</sup>。

近年来,众多学者利用神经网络对股票价格进行了大量的预测研究。文献[4]利用不同的神经网络算法产生神经网络集成个体,用遗传算法动态求解集成个体的非负权重系数,进行最优组合集成建模。相对传统的简单平均集成模型,具有预测精度高、稳定性好,易于操作的特点。文献[5]提出了一种基于 RBF 神经网络的股票市场预测模型。文献[6]提出了一种基于主成分的遗传神经网络股票指数预测模型,特点有两个:1) 利用主成份方法选取输入变量;2) 利用遗传算法优化神经网络参数。吴华星对基于神经网络的股票价格预测方法进行了研究<sup>[7]</sup>,他分析了传统的 BP 算法,通过引入调节参数  $\rho$  解决了要同时调整算法中学习速率  $\eta$  和动量参数  $\alpha$  的问题。从基于神经网络的股票预测模型的研究成果来看,关注点主要集中在:1) 对输入数据的处理上;2) 对神经网络的参数学习上;3) 对神经网络的集成方法上。

本文提出了一种基于 BP 神经网络的股票预测模型 (SPMM)。SPMM 的优点及特点为:1) 网络结构简单,易于实现。2) 给出了一种基于窗口技术的样本生成方法。3) 基于前

$N$  天的任意单一指标的运行数据,可以预测未来  $M$  天的趋势。4) 该模型融合了多个神经网络的预测结果,模型的学习能力强,预测精度高。系列实验结果表明,该模型效果良好。

### 1 BP 神经网络

BP 算法用来训练多层前馈神经网络(采用 BP 学习算法的前馈型网络被称为 BP 神经网络),属于有导师的学习算法。BP 网络具有结构清晰,易实现、计算功能强大、性能优越等特点,因而被广泛应用于模式识别、文本分类等多个领域<sup>[8]</sup>。BP 神经网络采用的是并行网络结构,包括输入层、隐含层和输出层。BP 网络已被证明具有强大的非线性映射能力和泛化功能,多层网络能逼近任意非线性函数<sup>[9]</sup>。

图 1 显示了一个三层结构的前馈神经网络,其中,输入层和输出层各有 2 个节点;隐藏层具有 3 个节点。隐藏层和输出层的每一个节点为一个 sigmoid 单元,它基于一个平滑的可微阈值函数。每一个 sigmoid 单元按下式计算它的输出:  $o = \sigma(\vec{w} \cdot \vec{x})$ ,其中  $\sigma(y) = \frac{1}{1 + e^{-y}}$ ,  $\vec{x}$  为该节点的输入向量,  $\vec{w}$  为它的权向量。 $\sigma$  经常被称为 sigmoid 函数或者也可以称为 logistic 函数。

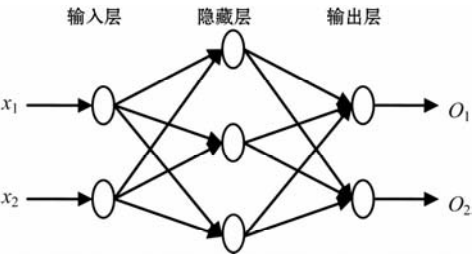


图1 三层结构的前馈神经网络

隐藏层节点的输入来自于输入层,当每一个输入层节点到各隐藏层节点上的权重确定后,隐藏层节点的输出则是确定的。隐藏层的输出作为输出层的输入,同理,当每一个隐藏层节点到各输出层节点的权重确定后,则输出层节点的输出值也是确定的。可见,权向量的学习是关键。

权重的学习问题实质是一个搜索问题,即需要在 $R^{|\vec{w}|}$ 空间中找到一个合理的 $\vec{w}$ ,使相应的网络在训练样本上的误差最小。形式化表示为,找到 $\vec{w}$ ,使下式最小。

$$E(\vec{w}) \equiv \frac{1}{2} \sum_{d \in D} \sum_{k \in outputs} (t_{kd} - o_{kd})^2 \tag{1}$$

其中, $D$ 是训练样本集合; $d$ 是一个训练样本; $outputs$ 是网络输出单元的集合; $t_{kd}$ 是 $d$ 的期望输出向量中第 $k$ 维的值; $o_{kd}$ 是神经网络对 $d$ 的输出向量中第 $k$ 维的值。

显然,遗传算法、粒子群算法等均可以用于寻找 $\vec{w}$ 的近似最优解。本文使用基于随机梯度下降的BP算法<sup>[2]</sup>搜索 $\vec{w}$ 。

2 股票价格预测模型的建立

2.1 预测模型

本文设计的股票价格预测模型(SPPM)如图2所示。该模型分为两个阶段:训练阶段及使用阶段。训练阶段指基于训练样本集,学习神经网络权重的过程。预测阶段指的是使用前一阶段学习到的神经网络,对股价进行预测的过程。为提高SPPM的泛化能力,强化该模型应用于股票价格预测的实效,采用了神经网络集成技术。神经网络集成被证明是一种非常有效的方法,即使仅仅是对一组网络进行简单的投票或平均,也可提高学习系统的处理能力<sup>[10]</sup>。本文对神经网络的集成,体现在两个方面:个体生成方面和结果融合方面。接下去,对这一模型进行详细介绍。

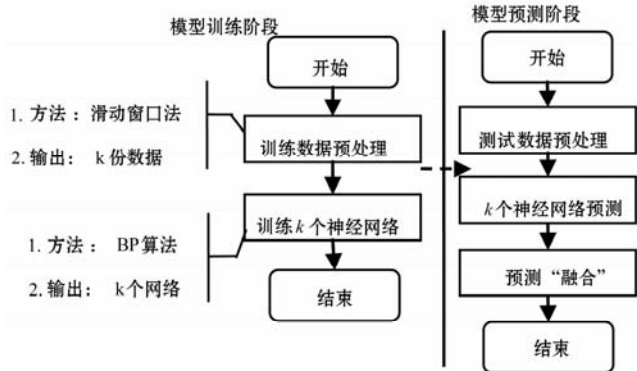


图2 股票价格预测模型

2.2 数据预处理

假设 $t_1, t_2, \dots, t_n (n \geq 2)$ 是一个连续的时间序列,在时刻 $t_i$ 上,可以获得一只股票的任意属性(如:开盘价、最高价、最低价、收盘价、成交量、成交金额等)的值。对于某一属性(以下不妨设为收盘价)在时间序列 $t_1, t_2, \dots, t_n$ 上获取的数据序列,不妨设该数据序列为: $x_1, x_2, \dots, x_n$ 。

首先对该数据序列中的每一维进行归一化处理:

$$x_i = \frac{x_i - \min}{\max - \min} \tag{2}$$

其中, $1 \leq i \leq n$ , $\max$ 和 $\min$ 分别表示 $x_1, x_2, \dots, x_n$ 的最大值和最小值。

其次,按式(3)构造 $n - N - M + 1$ 个样本,其中, $N$ 和 $M$ 表示用前 $N(N \geq 1)$ 天的收盘价预测未来 $M(M \geq 1)$ 天的收盘价; $n > N + M \geq 2$ 。

$$\begin{cases} X_1 = \langle x_1, x_2, \dots, x_N \rangle, Y_1 = \langle x_{N+1}, x_{N+2}, \dots, x_{N+M} \rangle \\ X_2 = \langle x_2, x_3, \dots, x_{N+1} \rangle, Y_2 = \langle x_{N+2}, x_{N+3}, \dots, x_{N+M+1} \rangle \\ \vdots \\ X_{n-N-M+1} = \langle x_{n-N-M+1}, x_{n-N-M+2}, \dots, x_{n-N-M+M} \rangle, Y_{n-N-M+1} = \langle x_{n-N-M+M+1}, x_{n-N-M+M+2}, \dots, x_n \rangle \end{cases} \tag{3}$$

所得到的样本集用 $D$ 表示,显然, $|D|$ 为 $n - N - M + 1$ 。 $D$ 中第 $i(1 \leq i \leq n - N - M + 1)$ 个样本,表示为 $\langle X_i, Y_i \rangle$ ,其中 $|X_i|$ 为 $N, |Y_i|$ 为 $M$ 。对于一个时间序列,如图3所示,当窗口大小发生调整时,则可以得不同的样本集。使用不同的样本集,则可以训练得到不同的神经网络个体。以下, $D_1, D_2, \dots, D_k$ 分别表示 $k$ 个样本集, $k$ 表示窗口个数或神经网络的总数。

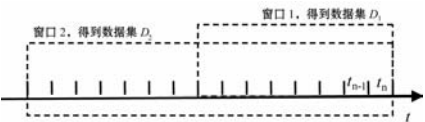


图3 基于窗口技术的时间序列

2.3 网络结构的确立

Hornik等早已证明:若输入层和输出层采用线性转换函数、隐藏层采用sigmoid转换函数,则含一个隐藏层的多层神经网络能够以任意精度逼近任何有理函数<sup>[8]</sup>。因此,本文使用的前馈网络具有三层结构:输入层、隐藏层和输出层。该网络输入层的节点数由输入向量的维数决定,由于输入向量的维数是 $N$ ,故输入层节点数确定为 $N$ 个。输出层节点数由输出向量的维数大,而且是训练时出现“过拟合”的直接原因。然而,遗憾的是,目前理论上还没有一种科学的和普遍的确定方法。为尽可能避免训练时出现“过拟合”现象,保证足够高的网络性能和泛化能力,确定隐藏层节点数的最基本原则是:在满足精度要求的前提下取尽可能紧凑的结构,即取尽可能少的隐藏层节点数。在确定隐藏层节点数时必须满足下列条件:1) 隐藏层节点数必须小于 $|D| - 1$ ;2) 训练样本数必须多于网络模型的连接权数,一般为 $2 \sim 10$ 倍。

不妨设隐藏层节点数为 $h$ ,连接权数则为: $N \times h + M \times h$ 。

第2个条件可表示为: $2 \leq \frac{|D|}{N \times h + M \times h} \leq 10$ ,得 $\frac{|D|}{10 \times (N + M)} \leq h \leq \frac{|D|}{2 \times (N + M)}$ 。从而,上述两个条件则可表示为:

$$\frac{|D|}{10 \times (N + M)} \leq h \leq \frac{|D|}{2 \times (N + M)} \quad \text{且} \quad h \leq N - 1 \tag{4}$$

综上,本文采用的是: $N - h - M$ 的神经网络,隐藏层节点为sigmoid单元,输出层节点为线性单元。

2.4 预测融合

$k$  个数据集可以训练得到  $k$  个神经网络。假设对于一样未知样本  $\langle X, ? \rangle$ , 设 SPPM 分别输出  $Y_1, Y_2, \dots, Y_k$ , 接下去, 将讨论“融合”这些输出并最终给出未来  $M$  天的股价走势的具体方法。

一种最简单的方法就是取各输出的均值, 如下式:

$$Y = \frac{\sum_{i=1}^k Y_i}{k}$$

(5)

其中,  $Y$  的大小为  $M$ , 每一维代表对未来某一天的预测值。

在许多情况下, 我们并不需要知道未来的具体预测值, 而仅需要知道未来是涨还是跌。此时, 仅需要观察  $Y$  中, 当前这一天相比于前一天的值即可获知。

3 实验设计与结果

本文实验的硬件环境为: CPU 为 Intel Core 2 Duo Processor T5500; 内存为 1G。编程环境为: Matlab 7.1。实验数据取自“同花顺”, 总共取了五个时间段 (即  $k=5$ ) 的收盘价数据, 如表 1 所

表 1 五个时间段的数据集

数据集	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$
时间段	2011-03-15 至 2012-03-16	2011-06-15 至 2012-03-16	2011-09-15 至 2012-03-16	2011-12-15 至 2012-03-16	2011-10-14 至 2012-03-16
交易天数( $n$ )	246	184	119	59	103
数据集大小 ( $n-N-M+1$ )	235	173	108	48	92
建议隐藏层节点数范围( $h$ )	$1.96 \leq h \leq 9$	$1.44 \leq h \leq 7.2$	$0.9 \leq h \leq 4.5$	$0.4 \leq h \leq 2$	$0.77 \leq h \leq 3.83$

表 2 中, SPPM 预测 2012-03-19 日上证指数的收盘价为 2 374.64, 相比于前一交易日 (2012-03-16) 的实际收盘价 (即 2 404.74) 而言, 预测趋势为下跌。这与当天的真实运行结果有所背离。

表 2 中, SPPM 预测 2012-03-20 日上证指数的收盘价为 2 373.84, 相比于前一交易日 (2012-03-19) 的实际收盘价 (即 2 374.64) 而言, SPPM 预测趋势为下跌。这与当天的真实运行结果完全符合, 且预测值与实际收盘价非常接近。

总体上看, SPPM 预测未来两天相对于当前日期 (即 2012-

示。由于篇幅关系, 本文仅讨论  $N=10, M=2$  的情况, 即根据前 10 天的数据预测未来 2 天的走势。值得说明的是: 表 1 中的交易天数, 对于不同的股票来说, 可能略有不同 (比如: 有的股票可能会有停牌现象)。在同一时间段内, 表 1 中的  $n, n-N-M+1$  及  $h$  的值对于不同的股票可能略有差别。

表 1 中的五个数据集, 分别训练五个神经网络, 神经网络权重学习的迭代次数为 800, 隐藏层节点的转换函数取 tansig, 输出层节点的转换函数取 purelin, 其他参数取 newff 函数的默认值。这五个神经网络分别记作: NN1、NN2、NN3、NN4 及 NN5。表 2 是 SPPM 对上证指数的预测结果, 预测的指标值为: 收盘价。值得指出的是, 每一个神经网络的训练和预测都独立运行 10 次, 神经网络对某一天收盘价的预测均取这 10 次独立实验的均值。比如: NN1 被独立训练了 10 次, 每一个独立训练后, NN1 都将对 2012-03-19 上证指数的收盘价得出一个预测值, 这 10 次预测的平均值为 2 335.8。这样做的好处是: 从一定程度上有效避免了随机的乐观或悲观。本文所有的实验数据都是 10 次独立实验的平均值。

03-16) 而言, 将处于下行趋势, 这与真实的运行情况是相符合的。

为了进一步验证 SPPM 的性能, 我们还对贵州茅台 2012-03-19 及 2012-03-20 的收盘价进行了预测。表 3 给出了 SPPM 对贵州茅台的预测结果。2012-03-16 日, 贵州茅台的实际收盘价为 207.59, 2012-03-19 及 2012-03-20 日, 贵州茅台分别有小幅上涨。SPPM 预测 2012-03-19 及 2012-03-20 日的收盘价分别为 207.16 和 207.22, 这与实际观测值非常接近。

表 2 对“上证指数”的预测结果 (参数设置:  $k=5, N=10, M=2$ )

日期	真实值 (N/A)	真实涨跌	NN1 ( $h=8$ )	NN2 ( $h=6$ )	NN3 ( $h=4$ )	NN4 ( $h=2$ )	NN5 ( $h=3$ )	预测值 (N/A)	预测 涨跌
2012-03-16	2 404.74								
2012-03-19	2 410.18	涨	2 355.0	2 363.0	2 345.6	2 400.3	2 409.3	2 374.64	跌
2012-03-20	2 376.84	跌	2 352.7	2 358.7	2 346.0	2 402.1	2 409.7	2 373.84	跌

表 3 对“贵州茅台”的预测结果 (参数设置:  $k=5, N=10, M=2$ )

日期	真实值 (元)	真实 涨跌	NN1 ( $h=8$ )	NN2 ( $h=6$ )	NN3 ( $h=4$ )	NN4 ( $h=2$ )	NN5 ( $h=3$ )	平均 值(元)	预测 涨跌
2012-03-16	207.59								
2012-03-19	207.64	涨	207.35	206.67	207.41	207.13	207.24	207.16	跌
2012-03-20	207.75	涨	207.15	206.85	207.56	207.71	206.84	207.22	涨

图 4 分别给出了 SPPM 在上证指数和贵州茅台  $D_1$  数据集上的预测结果与实际结果的对比曲线。总体上来看, SPPM 对上证指数的拟合情况要好于贵州茅台。可能的原因在于: 相比于个股而言, 上证指数受少数机构操控的可能性较小, 即上证指数更反映了市场的整体规律。这也就是说, 上证指数的运行规律性更容易被 SPPM 捕捉和学习到。而由于炒作的原因, 个股的规律即使被捕捉到了, 预测的值也可能与实际的值有差距。越是拟合效果差的股票, 越反映出其市场规律性的成份少, SPPM 则越难预测准确。

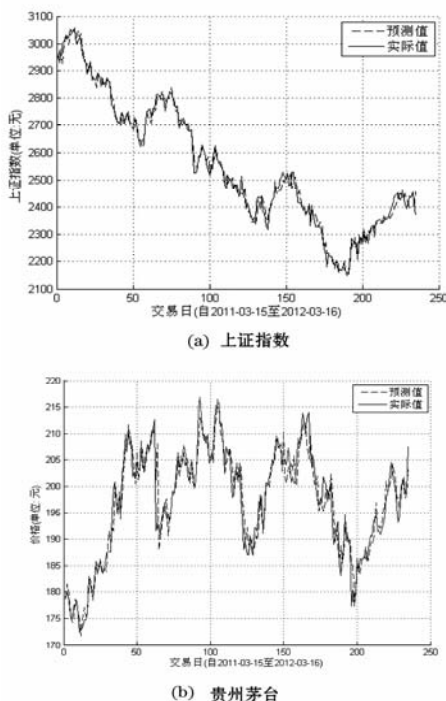


图 4 SPPM 预测值与实际值的对比曲线

## 4 结 语

本文提出了一种股票价格预测模型 SPPM, 它可以预测未来若干天的股票价格。由于集成了多个神经网络, 所预测的价格具有较高程度的准确性。实验结果证明了这一点。未来的工作包括: 1) 研究  $N$  与  $M$  对预测准确性的关系; 2) 研究样本数与预测准确性的关系等。

尽管本文的实验数据并不十分丰富, 却提出了一种股价预测的研究思路, 从而使本文具有一定的实际参考价值。

## 参 考 文 献

- [1] 吴建生. 最小一乘回归神经网络集成方法股市建模研究[J]. 计算机工程与设计, 2007, 28(23): 5812-5815, 5818.
- [2] 唐娜. 基于灰色理论和神经网络理论的股票指数预测研究[D]. 武汉理工大学, 2007.
- [3] 陈政, 杨天奇. 基于 RBF 神经网络的股票市场预测[J]. 计算机应用与软件, 2010, 27(6): 108-110.
- [4] 潘晓明, 吴建生. 基于遗传算法神经网络集成股票市场预测研究[J]. 广西师范学院学报: 自然科学版, 2007, 24(1): 77-84.
- [5] 刘磊. 基于遗传神经网络的指数跟踪优化方法[J]. 系统工程理论与实践, 2010, 30(1): 22-29.
- [6] 智晶, 张冬梅, 姜鹏飞. 基于主成分的遗传神经网络股票指数预测

研究[J]. 计算机工程与应用, 2009, 45(26): 210-212.

- [7] 吴华星. 基于神经网络的股票价格预测[D]. 中国科学院研究生院: 计算机技术研究所, 1998.
- [8] Mitchell T M. 机器学习[M]. 曾华军, 译. 北京: 机械工业出版社, 2009.
- [9] 杨行峻. 人工神经网络与盲信号处理[M]. 北京: 清华大学出版社, 2003.
- [10] 周志华, 陈世福. 神经网络集成[J]. 计算机学报, 2002, 25(1): 1-8.

## (上接第 70 页)

从图 7 和图 8 可以看出, 车轮均未出现抱死现象, 轮速曲线变化较为均匀, 并且参考车速能够较好地反映当前时刻的车速, 符合控制模型中所设计的控制算法的控制循环。

从图 9 和图 10 可以看出, 制动初期车速较高时滑移率变化曲线在 15% 左右浮动, 即控制在最佳滑移率控制点附近, 车速较低时为使车辆具有较短的制动时间和控制距离, 控制力度较前期稍猛滑移率变化也相对较大, 符合控制的要求。

## 6 结 语

本文将 ADAMS 中建立的整车模型和在 MATLAB 中建立的 ABS 控制模型以及气压调节控制单元模型进行了联合仿真, 由仿真结果可知 ABS 系统在整车的制动过程中起到了很好的制动防抱死效果。在使用虚拟样机技术后, 可以通过试凑参数, 很快得到制动性能较好的控制门限值, 大大减少了试验次数和缩短了研制周期, 这更能说明虚拟样机技术的优越性。同时可以通过添加能真实模拟汽车行驶状态仿真的驾驶员模型, 选择更匹配的轮胎模型, 开发出更具形象贴近、更复杂的汽车虚拟样机模型, 今后需要进一步研究。

## 参 考 文 献

- [1] Ozdalyan B, Blundell M V. Anti-lock braking system simulation and modeling in ADAMS[C]//IEE Conference Publication, 1998: 140-144.
- [2] 王纪森, 付卫强, 余洋. 防抱制动系统自寻优控制的半物理仿真研究[J]. 系统仿真学报, 2005, 17(12): 2841-2843.
- [3] 李君, 喻凡. 基于道路自动识别的 ABS 模糊控制系统的研究[J]. 农业机械学报, 2001, 32(5): 26-29.
- [4] 张云清, 熊小阳, 陈伟, 等. 基于 ADMAS 与 Matlab 的 ABS 模糊控制仿真研究[J]. 公路交通科技, 2007(11): 148-153, 158.
- [5] Lü H Q, Jia Y M, Du Jun-ping, et al. ABS Composite Control Based On Optimal Slip Ratio[C]//Proceedings of the 2007 American Control Conference Marriott Marquis Hotel at Times Square, New York City, USA, July 11-13, 2007: 5748-5752.
- [6] 程军. 汽车防抱死制动系统的理论与实践[M]. 北京理工大学出版社, 1999: 80-83.
- [7] 唐国元. 基于遗传算法的车辆 ABS 在线整定 PID 控制及仿真[J]. 系统仿真学报, 2007, 19(7): 1554-1557.
- [8] Layne J R. Fuzzy Learning Control for Antiskid Braking System[J]. IEEE Transactions on Control System Technology, 1993, 1(2): 122-129.
- [9] Lin C M, Hsu C F. Self-Learning Fuzzy Sliding-Mode Control for Anti-lock Braking Systems[J]. IEEE Transactions On Control Systems Technology, 2003, 11(2): 273-278.