World Economic Research 世界经济探索, 2020, 9(2), 24-32
Published Online June 2020 in Hans. http://www.hanspub.org/journal/wer
https://doi.org/10.12677/wer.2020.92004

Hans汉斯

Forecast of CSI 300 Index Based on Artificial Neural Network Model

Shun Xiao, Leqi Yang, Si Yu, Huimin Xin, Huqin Yan

Xiamen National Accounting Institute, Xiamen Fujian
Email: yanglq_1997@163.com

Received: May 12th, 2020; accepted: May 28th, 2020; published: Jun. 4th, 2020

Abstract

Based on the background of the current epidemic impact on the economy and the suppression of demand, frequent foreign super loose monetary policies, this paper uses a two-layer artificial neural network model and selects a total of 1021 Shanghai and Shenzhen 300 index data from July 6, 2017 to April 24, 2020 as a sample for stock market forecasting. The prediction results prove that the artificial neural network model predicts the stock price error rate to be controllable, and can provide some reference and guidance for stock price prediction in the short term.

Keywords

Artificial Neural Network, Stock Forecast, CSI 300 Index

基于人工神经网络模型的沪深300指数预测

肖　顺，杨乐祺，余　偲，辛慧敏，阎虎勤

厦门国家会计学院，福建 厦门
Email: yanglq_1997@163.com

收稿日期：2020年5月12日；录用日期：2020年5月28日；发布日期：2020年6月4日

摘　要

基于当前疫情冲击经济、抑制需求，国外超级宽松货币政策频出的背景下，本文采用二层人工神经网络模型，选取了2017年7月6日至2020年4月24日共1021个沪深300指数数据为样本，进行股票市场预测。预测结果证明人工神经网络模型预测股价误差率可控，可以在短期内为股价预测提供一定借鉴和指导。

文章引用：肖顺，杨乐祺，余偲，辛慧敏，阎虎勤. 基于人工神经网络模型的沪深 300 指数预测[J]. 世界经济探索, 2020, 9(2): 24-32. DOI: 10.12677/wer.2020.92004

关键词

人工神经网络，股票预测，沪深300指数

1. 引言

突如其来的新冠肺炎疫情，引发了一场全球性危机，国际货币基金组织预测 2020 年全球经济将萎缩 3%。世界卫生组织称，新冠肺炎疫情将与人类长期共存，而且北美、欧洲各国疫情形势依然严峻，复工复产政策尚不明朗，失业率不断上升。虽然国内疫情已经基本扑灭，但内需不足，而且在全球产业链深度融合的背景下，中国企业的原料供应与产品销路受阻，出口停摆。全球经济发展充满了不确定性。加之原油期货价格暴跌至负数，欧佩克提前减产，股市的走向扑朔迷离。借助模型预测股市走向，准确率最高可达 80%，能够帮助投资者在重重迷雾中制定正确的投资策略，避免不理智的追涨杀跌。

国内外学者提出了多种股票预测方法，例如时间序列预测、灰色预测、组合预测。但是由于历史股票数据规模庞大，多噪声和高度模糊非线性的特点，而且常受到黑天鹅事件影响，上述方法的适应性不佳。人工智能算法 BP (Back Propagation)神经网络所具有的良好的自适性能力、自学习能力，能够以任意精度逼近复杂的非线性关系，弥补线性模型的缺陷，构建股票指数预测系统[1]。

BP 神经网络由 Rumelhart 和 McClelland 等科学家于 1986 年提出，是目前应用最广泛的神经网络模型之一。本文采用二层人工神经网络，以沪深 300 指数为例进行股票数据的预测和分析。二层人工神经网络由输入层和输出层组成，输入层神经元对应自变量，为沪深 300 指数的历史数据[2]，经过激活函数，输出因变量，然后通过调整自变量缩小输出值与样本的误差，提高准确性[3]。

2. 数据来源及模型介绍

2.1. 数据来源

为了通过人工神经网络对股指进行预测，需要使用股指历史真实数据，本文通过 Pyhthon 财经数据库接口包 Tushare，获取沪深 300 指数数据。不同于最高价是大多数人认为的好的卖出价格，也不同于最低价是大多数人认为适合买进的价格，收盘价是不再进行交易的价格。因此研判收盘价有着重要意义，无论当天股价如何振荡，最终将定格在收盘价上。收盘价是市场参与者们所共同认可的价格，因此本文选择收盘价作为样本数据，作为对沪深 300 指数进行后续拟合预测的基础。为保证本文的预测及分析能够基于充足的样本量进行，本文选择 2017.07.06~2020.04.24 期间所有交易日的沪深 300 指数收盘价，共 1021 个数据来进行后续的预测分析[4]。

2.2. 模型原理及步骤

2.2.1. 二层神经网络算法原理

二层神经网络算法实际上是将样本数据代入时间序列的向量自回归(Vector AutoRegression, VAR)模型进行拟合，构成一个自循环矩阵方程，使得输入值和输出值来源于同一个样本数据序列。将矩阵代入激活函数，同时样本数据在循环过程中实现自学习，然后求解矩阵方程。最后，通过反向传播机制调整

矩阵方程的解，使得样本值与预测值的误差达到最小[5]。

2.2.2. 二层神经网络算法步骤

① 分割样本数据

假设样本总数为 N_1，被区分为模拟窗口 $\mathrm{NN}-N_1=n_0$ 和预测窗口 $\mathrm{NN}=N_1-n_0$ (如图 1)。对于模拟窗口 n_0，如果输入矩阵列数为 $n-1$，则行数为 $m=n_0-n+1$；对于全样本窗口 N_1，如果输入矩阵列数为 $n-1$，则行数为 $M=N_1-n+1$ [6]。样本窗口分割如图 1：

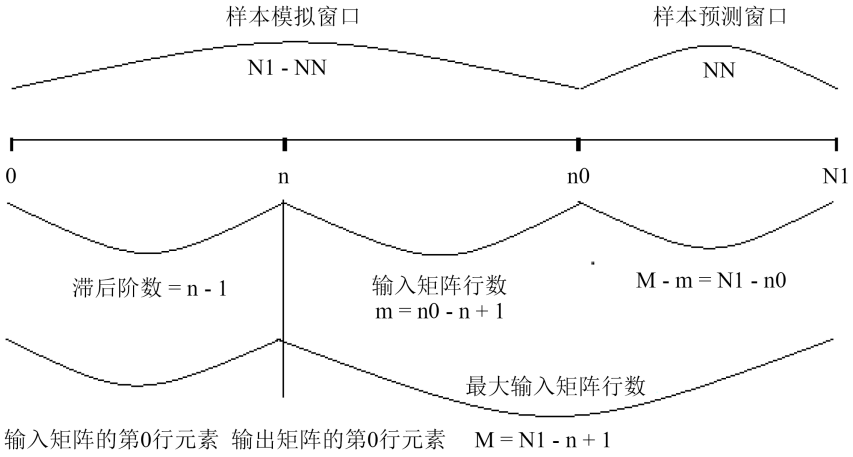

Figure 1. Sample data segmentation
图 1. 样本窗口分割示意图

② 正向求解逻辑

假设股票指数样本序列为：

$$S=\left\{y_1,y_2,\cdots,y_{n-1},y_n,y_{n+1},\cdots,y_{n_0},y_{n_0+1},\cdots,y_{N_1}\right\} \tag{1}$$

代入向量自回归模型构造矩阵，所在行数 $i=0,1,\cdots,m-1$，则通项公式为：

$$y_{1+i}x_1+y_{2+i}x_2+\cdots+y_{n-1+i}x_{n-1}+h=y_{n+i} \tag{2}$$

该公式共 m 个，将输入值中的 y 用 A 表示，则构成的矩阵表示为：

$$A_{mn}X_{n1}+I_{m1}h=Y_{m1} \tag{3}$$

其中 X_{n1} 是方程的解，h 为常数项。采用激活函数 $f(x)$ 作用于矩阵方程，激活函数通过变换，把方程组转换成了一个新的线性方程组：

$$f\left(A_{mn}X_{n1}+I_{m1}h\right)=Y_{m1} \tag{4}$$

其中激活函数：

$$f(x)=\frac{1}{1+\mathrm{e}^{-x}}=\begin{cases}0, & x\to-\infty\\ \dfrac{1}{2}, & x\to 0\\ 1, & x\to+\infty\end{cases} \tag{5}$$

该函数的定义域为 $x\in(-\infty,\infty)$，而值域为 $y\in(0,1)$。所以，当作为因变量的样本值超过 1 时，就不能直接使用该函数作为激活函数。因此非负序列 $\{y_1,y_2,\cdots,y_i,\cdots,y_n\}$，设定一个上限值 $K>0$，使得：

$$M = \max_{i=1,2,\cdots,n}\{y_1, y_2, \cdots, y_i, \cdots, y_n, K\} \tag{6}$$

对原样本值 y_i 重新进行处理，使得新序列 Y_i 符合激活函数值域要求：

$$Y_i = \frac{y_i}{M},\ (i = 1, 2, \cdots, n) \tag{7}$$

但是在取得最优解之后，必须还原预测值：

$$y_i = MY_i,\ (i = 1, 2, \cdots, n) \tag{8}$$

最后进行正向求解，正向求解即为直接求得使得方程(8)成立的解 X_{n1} 及常数项 h。

③ 反向传播调解

假设矩阵 U_{m1} 定义如下：

$$U_{m1} = A_{mn}X_{n1} + I_{m1}h \tag{9}$$

那么，经过激活函数 f 作用之后，方程式可以表示为：

$$f(U_{m1}) = Y_{m1} \tag{10}$$

假设 E_{m1} 为误差项，满足关系式：

$$E_{m1} = Y_{m1} - f(U_{m1}) \tag{11}$$

当 $E_{m1} \neq 0$ 时进行反向传播调解，那么通过对方程解 X_{n1} 和 h 增加调节量：

$$X_{n1} = X_{n1} + \delta_{n1} \tag{12}$$

$$h = h + \tau \tag{13}$$

那么，经过激活函数调解之后能够使 $E_{m1} = 0$，则满足条件：

$$f\left(A_{mn}\left(X_{n1} + \delta_{n1}\right) + I_{m1}\left(h + \tau\right)\right) = f\left(U_{m1} + A_{mn}\delta_{n1} + I_{m1}\tau\right) = Y_{m1} \tag{14}$$

对激活函数按照一阶 Taylor 展式展开，则有：

$$f\left(U_{m1} + A_{mn}\delta_{n1} + I_{m1}\tau\right) = f\left(U_{m1}\right) + f'\left(U_{m1}\right)\left[A_{mn}\delta_{n1} + I_{m1}\tau\right] = Y_{m1} \tag{15}$$

那么就有：

$$f'\left(U_{m1}\right)\left[A_{mn}\delta_{n1} + I_{m1}\tau\right] = Y_{m1} - f\left(U_{m1}\right) = E_{m1} \tag{16}$$

由于有关系式：

$$A_{mn}\delta_{n1} + I_{m1}\tau = \frac{E_{m1}}{f'\left(U_{m1}\right)} \tag{17}$$

假如 $\tau = 0$，则有：

$$A_{mn}\delta_{n1} = \frac{E_{m1}}{f'\left(U_{m1}\right)} \tag{18}$$

$$A_{mn}^{\mathrm{T}}A_{mn}\delta_{n1} = A_{mn}^{\mathrm{T}}\left[\frac{E_{m1}}{f'\left(U_{m1}\right)}\right] \tag{19}$$

$$\left(A_{mn}^{\mathrm{T}}A_{mn}\right)^{-1}A_{mn}^{\mathrm{T}}A_{mn}\delta_{n1} = \left(A_{mn}^{\mathrm{T}}A_{mn}\right)^{-1}A_{mn}^{\mathrm{T}}\left[\frac{E_{m1}}{f'\left(U_{m1}\right)}\right] \tag{20}$$

$$\delta_{n1} = \left(A_{mn}^{\mathrm{T}}A_{mn}\right)^{-1}A_{mn}^{\mathrm{T}}\left[\frac{E_{m1}}{f'\left(U_{m1}\right)}\right] \tag{21}$$

假如$\tau \neq 0$，从而，常数项的调解量就可以是：

$$I_{m1}\tau = \frac{E_{m1}}{f'(U_{m1})} - A_{mn}\delta_{n1} \tag{22}$$

$$I_{m1}^{\mathrm{T}} I_{m1}\tau = I_{m1}^{\mathrm{T}}\left\{\frac{E_{m1}}{f'(U_{m1})} - A_{mn}\delta_{n1}\right\} \tag{23}$$

$$\left(I_{m1}^{\mathrm{T}} I_{m1}\right)^{-1} I_{m1}^{\mathrm{T}} I_{m1}\tau = \left(I_{m1}^{\mathrm{T}} I_{m1}\right)^{-1} I_{m1}^{\mathrm{T}}\left\{\frac{E_{m1}}{f'(U_{m1})} - A_{mn}\delta_{n1}\right\} \tag{24}$$

整理后得到：

$$\tau = \left(I_{m1}^{\mathrm{T}} I_{m1}\right)^{-1} I_{m1}^{\mathrm{T}}\left\{\frac{E_{m1}}{f'(U_{m1})} - A_{mn}\delta_{n1}\right\} \tag{25}$$

只要重复上述过程，则可以逐步迭代，直到取得增加了调节量的X_{n1}和h的值，也就是具有扰动项的方程组的最优解。

2.2.3. 预测结果检验

如果在模拟窗口n_0下，我们可以得到最优解$x_1, x_2, \cdots, x_{n-1}, h$，那么，只要将其代入对于全样本窗口$N_1$下，自然就可以计算出区间$(n, N_1)$下的预测值，而预测区间$(n_0, N_1)$下的预测值，自然也就会得到了。

因为对于股票指数预测来说，对于上涨或者下跌趋势的预测往往比股票指数值接近更有意义[7]，所以，我们需要对于预测效果进行分析。当原值与预测值变化趋势一致时，我们称为 Same Trend；当趋势不同时，我们称为 Diff Trend，二者之和等于 1，Same Trend 值越大，则预测效果越好。

3. 数据处理及结果分析

本文采用的沪深 300 指数收盘价趋势股市预测的总体流程如图 2，大致分为 5 个步骤：1) 处理数据；2) 基于 Python 工具对数据用神经网络进行拟合；3) 利用实际输出与期望输出的差异，调整神经网络参数[8]；4) 检验最优解的效果；5) 利用得到的最优参数建立基于神经网络的股票价格趋势预测模型并得到拟合曲线[9]。

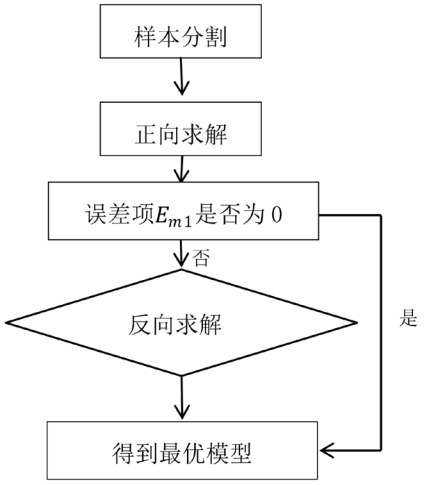

Figure 2. Flowchart: Forecast of CSI 300 Index
图 2. 沪深 300 指数预测流程

3.1. 数据处理

基于前述介绍，本文在 Anaconda 的环境下利用 Python 工具对选取的 1021 个沪深 300 指数数据进行处理。

首先，对样本数据进行了分割，将 $N_1=1021$ 个样本总数区分为模拟窗口 $n_0=1001$ 和预测窗口 $NN=20$ 个。然后，按时间排序得到了其收盘价样本序列 $y_{指数}$ 及每个数据基于基期的增长率样本序列 $y_{指数增长率}$：

$$y_{指数}=\{3053.699,3051.585,\cdots,3829.753,3796.972\} \tag{26}$$

$$y_{指数增长率}=\{1.000000000,0.999307888,\cdots,1.243400905\} \tag{27}$$

为使其符合激活函数值域的要求$(0<y<1)$，本文取两数 K_1 和 K_2，使其分别大于上述两序列中的最大值。此处选取 $K_1=15000$，$K_2=2$，然后用原序列中的每个值除以 K 得到新序列 S：

$$S_{指数}=\{0.203579933,0.203439,\cdots,0.255316867,0.253131467\} \tag{28}$$

$$S_{指数增长率}=\{0.5,0.499653944,\cdots,0.627067779,0.621700452\} \tag{29}$$

将新序列 S 引入向量自回归模型，求解矩阵方程(4)，可以得到 X_n 及 h，将其代回方程(2)，得到基于人工神经网络对股票指数收盘价及该股指收盘价增长率的预测模型。代入相应的样本序列我们可以得到预测值 y_{n+i}，最后将其乘以相应的 K 进行还原，即得到实际的预测值[10]。

另外，为方便检验最优解的效果，我们并通过第二部分对模型介绍进行上涨或者下跌趋势的预测结果的分析，并根据结果理想与否进行参数的优化。本实验中主要改变的参数是方程(2)中的未知数 X 个数 n。

3.2. 实验结果及预测效果

3.2.1. 沪深 300 指数

基于上述数据处理，我们得到本次实验的最佳模型为：

$$y_{n+i}=1.176589y_{1+i}+3.329964y_{2+i}+\cdots+4.863449y_{498+i}+0.00541 \tag{30}$$

沪深 300 指数模型相关参数如表 1：

Table 1. Forecast model of CSI 300 index
表 1. 沪深 300 指数预测模型

参数 n	R_Square	Adjusted R_Square	Same trend	Diff trend
499	0.999522	0.939898	0.6 (12/20)	0.4 (8/20)

由表 1 可知，可决系数与调整后的可决系数均在 0.9 以上，因此该模型的拟合程度与预测效果均较良好；且对指数涨跌的预测正确率达到 60%，以此具有一定的现实指导作用。指数拟合曲线如图 3：

由图3可知，模拟窗口拟合良好，沪深300指数于0~200日逐渐由最初的近4200点跌至3000点左右，而后持续上升至4100点左右，继而呈正常的上下波动状态。预测窗口数值与样本数据数值差距相对较大，由图可知预测值起伏较大，实际样本呈较为平缓的增长态势。相较于股票市场确切的股指数值预测，更为重要的是相对的涨跌情况预测，通过对预测窗口涨跌情况的合理判断，投资者可以更有针对性的选择买入或卖出。

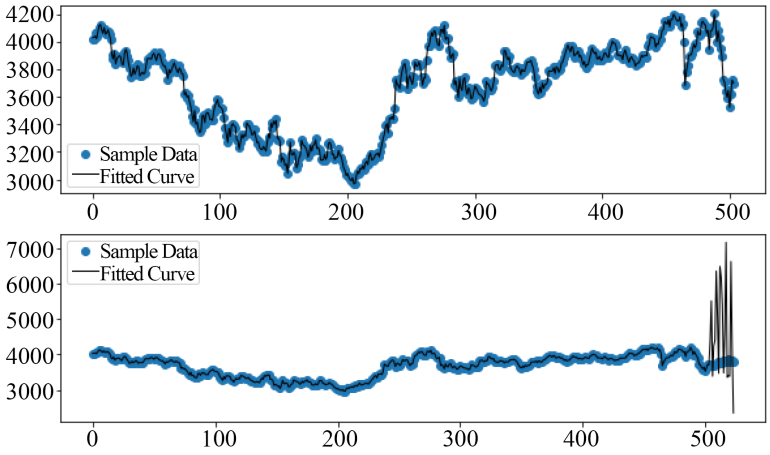

Figure 3. Forecast of the CSI 300 index
图 3. 沪深 300 指数拟合曲线

3.2.2. 沪深 300 指数增长率

基于上述数据处理，我们得到本次实验的最佳模型为：

$$y_{n+i} = 2.489706y_{1+i} - 1.841765y_{2+i} + \cdots + 3.867576y_{489+i} + 0.00919 \tag{31}$$

沪深 300 指数增长率模型相关参数如表 2：

Table 2. Forecast model of CSI 300 index growth rate
表 2. 沪深 300 指数增长率预测模型

参数 n	R_Square	Adjusted R_Square	Same trend	Diff trend
499	0.997098	0.932471	0.6 (12/20)	0.4 (8/20)

由表 2 可知，类似于沪深 300 指数的拟合预测，增长率预测模型的可决系数与调整后的可决系数同样在 0.9 以上，拟合程度与预测效果均较良好；且增长率同方向预测正确率同样达到 60%。增长率拟合曲线如图 4：

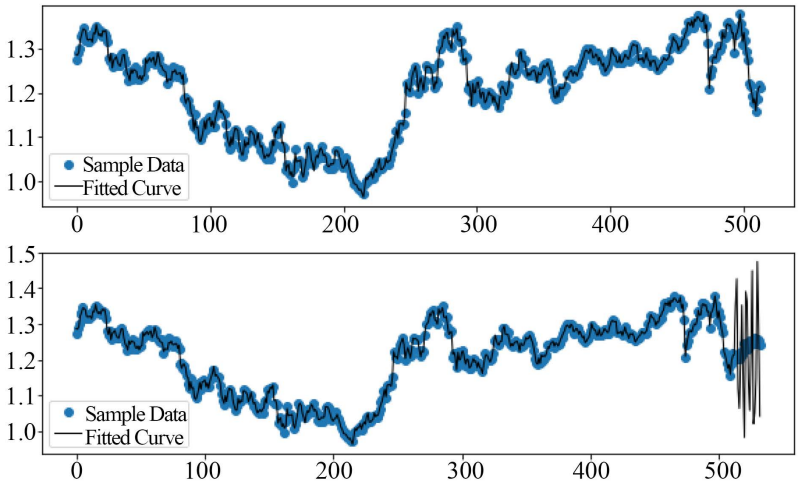

Figure 4. Forecast of the CSI 300 index growth rate
图 4. 沪深 300 指数增长率拟合曲线

沪深300指数增长率拟合图形走势类似于沪深300指数的拟合图形，模拟窗口中于0~220日期间涨幅持续下跌，而后持续走高，于290日左右停止上升进入持续波动阶段。预测窗口中样本数据平缓上升，预测涨跌幅成功率达到60%，结果较为成功。

3.3. 实际预测检验

将二层人工神经网络求解出的方程代入计算，还原得出沪深 300 收盘指数的预测值，与历史实际收盘指数采集结果对比分析，计算相对误差，结果如表 3：

Table 3. Error and accuracy of sample data and prediction results
表 3. 实际值与预测值的相对误差和准确率

日期	收盘指数	预测值	相对误差	准确率
2020/4/5	3710.061	4371.73	17.83%	84.86%
2020/4/6	3674.111	5510.16	49.97%	66.68%
2020/4/7	3686.155	3388.96	−8.06%	91.94%
2020/4/8	3675.076	4240.57	15.39%	86.66%
2020/4/9	3734.531	4414.87	18.22%	84.59%
2020/4/10	3713.218	6351.79	71.06%	58.46%
2020/4/11	3798.021	5091.65	34.06%	74.59%
2020/4/12	3780.345	3471.79	−8.16%	91.84%
2020/4/13	3792.811	6482.01	70.90%	58.51%
2020/4/14	3769.178	6166.57	63.61%	61.12%
2020/4/15	3753.257	5404.33	43.99%	69.45%
2020/4/16	3825.699	3480.47	−9.02%	90.98%
2020/4/17	3797.362	4686.89	23.42%	81.02%
2020/4/18	3802.381	7160.52	88.32%	53.10%
2020/4/19	3839.487	3356.01	−12.59%	87.41%
2020/4/20	3853.455	3404.44	−11.65%	88.35%
2020/4/21	3808.047	3387.62	−11.04%	88.96%
2020/4/22	3839.383	6619.92	72.42%	58.00%
2020/4/23	3829.753	4139.70	8.09%	92.51%
2020/4/24	3796.972	2356.27	−37.94%	62.06%

由于 1021 个样本数据覆盖时间较长，因此本文选择了样本最后的 20 个预测窗口数据，通过比较二层人工神经网络模型对沪深 300 指数的预测值与真实指数，得到了二者的相对误差与预测准确率。由对比结果可知，模型预测的误差的较小同时预测准确率较高，因此利用人工神经网络模型对股票价格进行短期预测是可行的，对投资者进行短期的投资具有切实可行的指导价值。

4. 总结

本文采用二层人工神经网络对沪深 300 指数进行分析预测，用实证分析举例证明了模型的可行性。从预测拟合结果来看，所得模型预测值的准确率理论上在 60%以上，结果与实际的吻合程度较高。可见，

采用人工神经网络模型对股价预测的可靠性较强，这为投资者提供了强有力的参考价值和理论基础的同时，为众多股票投资者提升了投资效率，提供了更加有效的信息，也降低了损失发生的概率，增强了收益能力，还能在股票市场趋势预测方面产生较大的效应。

致　谢

在此我们要由衷地感谢所有在论文写作期间帮助过我们的人，特别是厦门国家会计学院信息管理处处长阎虎勤老师，阎老师的《Python 财务数据分析》课程让我们受益匪浅。在我们撰写论文的过程中，得到了阎老师悉心细致的教诲和无私的帮助，无论是在论文的选题、构思和资料的收集方面，还是在论文的模型构造以及成文定稿方面。最后，感谢所有关心、支持、帮助过我们的良师益友。

基金项目

本论文得到了厦门国家会计学院 2019 年“云项课题：Python 财务数据分析”项目的支持。

参考文献

[1] 黄宏运, 吴礼斌, 李诗争. BP 神经网络在股票指数预测中的应用[J]. 通化师范学院学报, 2016, 37(10): 32-34.
[2] 胡振兴, 田大钢. BP 神经网络模型在沪深 300 指数预测中的应用[J]. 现代商业, 2011(29): 43-44.
[3] 石茜子. 基于 BP 神经网络的股价预测模型应用分析[D]: [硕士学位论文]. 深圳: 暨南大学, 2017.
[4] 刘晓敏. 基于 BP 神经网络的股指预测系统[D]: [硕士学位论文]. 大连: 大连理工大学, 2012.
[5] 张建辉. 基于 BP 神经网络的时序预测模型的研究[D]: [硕士学位论文]. 太原: 太原理工大学, 2017.
[6] 阎虎勤, 编著. Python 财务数据分析 2020 春季教材[M]. 厦门: 厦门国家会计学院, 2020.
[7] 李云强, 宋威. 基于 BP 神经网络的股票价格趋势预测[J]. 北方工业大学学报, 2013, 25(1): 11-16+55.
[8] 邢伟琛. 基于 BP 神经网络对上证指数的预测[J]. 企业科技与发展, 2019(12): 124-125.
[9] 王晓东, 薛宏智, 贾雯超. 基于 BP 神经网络的股票涨跌预测模型[J]. 价值工程, 2010, 29(31): 47-49.
[10] 张翱翔. 基于 BP 神经网络股指预测系统的国内外股市预测研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2017.