Advances in Applied Mathematics 应用数学进展, 2022, 11(1), 65-70
Published Online January 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.111010

Hans汉斯

基于LSTM模型的股价分析及预测

李　悦，李　俏

辽宁师范大学数学学院，辽宁　大连

收稿日期：2021年12月11日；录用日期：2022年1月1日；发布日期：2022年1月14日

摘　要

股价预测一直都是股票投资者重点关注和重点研究的方向。针对股价具有高度非线性、高噪声、动态性等问题，本文采用长短期记忆网络(LSTM)模型对股价进行预测。数据取自半导体行业公司股票价格，采用python深度学习框架构造长短期记忆网络模型，分别对每一组股票的开盘价进行预测，再通过均方误差和决定系数对预测结果进行评价。实验结果表明将LSTM神经网络用于股票价格预测具有较好的效果，可以为投资者提供一定的参考。

关键词

Python，长短期记忆网络，股票价格，半导体行业

Analysis and Forecast of Stock Price Based on LSTM Model

Yue Li, Qiao Li

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: Dec. 11th, 2021; accepted: Jan. 1st, 2022; published: Jan. 14th, 2022

Abstract

Stock price forecasting has always been the focus and research direction of stock investors. Aiming at the problems of high non-linearity, high noise and dynamics in stock prices, this paper uses a long short-term memory network (LSTM) model to predict stock prices. The data is taken from the stock prices of companies in the semiconductor industry, and a long- and short-term memory network model is constructed using the python deep learning framework to predict the opening price of each group of stocks, and then the prediction results are evaluated through the mean square error and the coefficient of determination. The experimental results show that the use of LSTM neural network for stock price prediction has a good effect, and it can provide a certain reference for investors.

文章引用: 李悦, 李俏. 基于 LSTM 模型的股价分析及预测[J]. 应用数学进展, 2022, 11(1): 65-70.
DOI: 10.12677/aam.2022.111010

Keywords

Python, Long Short-Term Memory Network, Stock Price, Semiconductor Industry

Copyright © 2022 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 前言

从国内来看，半导体行业的景气度处在持续攀升过程中。虽然一直面临着发达国家的技术封锁，但是，在国家政策的扶持下，我国半导体行业发展的步伐加快。国内半导体企业在芯片国产化的大潮之下，景气度亦在持续攀升之中，有望迎来发展的黄金期，实现跨越式赶超。考虑到各个半导体及芯片制造公司股票价格瞬息万变，较难预测，投资者只有准确掌握股票价格的涨落与趋势，才能做出正确的决策。

目前，越来越多的学者采用人工智能技术来实现股价的预测，从传统的人工神经网络模型到深度学习模型，再到现在的多种模型相互结合进行预测。张如梦[1]等人采用两种最热方法——分别是 BP 神经网络模型和 ARMA-GARCH 模型分析并预测股票价格，对比得出实验结果；韩山杰[2]等人采用了很经典的多层感知机模型预测了苹果公司股票价格的收盘价；杨琦[3]等人同样采取了创新后的 ARMA-GARCH 模型对不同股票的价格进行了分析和预测；黄颖[4]等人则是采用了 XGBoost 和 LSTM 模型进一步对比分析预测了时间金融序列包括不同股票的价格；张康林[5]是用了不同的软件 pytorch 来构建 LSTM 模型对不同的股票价格先进行分类再进行分析预测；宋刚[6]等人采用了自适应粒子群优化后的 LSTM 模型对沪、深、港的不同股票价格进行了预测，具有更高的适用性；李雄英[7]等人采用了三个模型分别分析和预测了四大银行的股票价格，得出不同模型有不同的特点，其中 LSTM 是比较实用的。基于以上分析，得出 LSTM 模型在股票价格的分析与预测上展现了很大的优势，对业绩比较优良的股票来说，采用 LSTM 模型预测精度更高。因此本文采用 LSTM 模型对国内多个半导体行业龙头企业的股票价格进行预测。

2. 神经网络模型介绍

2.1. RNN 神经网络模型

循环神经网络是一种对序列数据进行建模的神经网络，它像一个循环动态系统，在该结构中当前的输出会流入下一步的输入中，为下一次输出做出贡献。其主要形式是该结构有个循环结构会保留前一次循环的输出结果并作为下一次循环输人的一部分输入。图 1 为循环神经网络结构图。

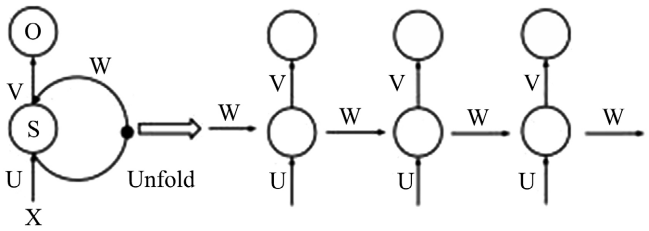

Figure 1. Cyclic network unfolding diagram
图 1. 循环网络展开图

2.2. LSTM 神经网络模型

LSTM 算法全称为 Long short-term memory，最早由 Sepp Hochreiter 和 Jürgen Schmidhuber 于 1997 年提出，是一种特定形式的 RNN (Recurrent neural network，循环神经网络)，而 RNN 是一系列能够处理序列数据的神经网络的总称。LSTM 神经网络结构图如图 2 所示。

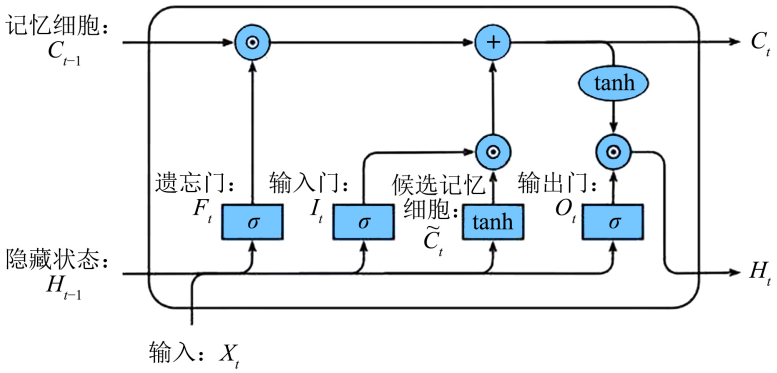

Figure 2. LSTM neural network structure
图 2. LSTM 神经网络结构

LSTM 模型在 RNN 的基础上添加了 3 个门控结构，分别为“遗忘门”“输入门”和“输出门”。

遗忘门：在我们 LSTM 中的第一步是决定我们会从细胞状态中丢弃什么信息。这个决定通过一个称为忘记门层完成。该门会读取 h_{t-1} 和 x_t，输出一个在 0 到 1 之间的数值给每个细胞状态 C_{t-1} 中的数字。1 表示“完全保留”，0 表示“完全舍弃”。

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right)$$

其中 h_{t-1} 表示的是上一个 cell 的输出，x_t 表示的是当前细胞的输入。σ 表示 sigmod 函数。

输入门：下一步是决定让多少新的信息加入到 cell 状态中来。实现这个需要包括两个步骤：首先，一个叫做“input gate layer”的 sigmoid 层决定哪些信息需要更新；一个 tanh 层生成一个向量，也就是备选的用来更新的内容。在下一步，我们把这两部分联合起来，对 cell 的状态进行一个更新。

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right)$$

$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right)$$

输出门：最终，我们需要确定输出什么值。这个输出将会基于我们的细胞状态，但是也是一个过滤后的版本。首先，我们运行一个 sigmoid 层来确定细胞状态的哪个部分将输出出去。接着，我们把细胞状态通过 tanh 进行处理(得到一个在−1 到 1 之间的值)并将它和 sigmoid 门的输出相乘，最终我们仅仅会输出我们确定输出的那部分。

$$o_t = \sigma\left(W_0\left[h_{t-1}, x_t\right] + b_0\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

3. 基于 LSTM 模型的股票价格预测

3.1. 数据选取

随着信息时代的来临，科技的发展离不开半导体，关于我国半导体的发展一向被大家关注，本文数

据取自半导体行业的几个龙头股票价格。

长电科技(600584)：半导体龙头股。公司主营半导体，电子原件，专用电子电气装置的研制、开发、生产及销售，公司拥有 IGBT 封装技术，同时其控股子公司正在积极研发 IGBT 产品，为 IGBT 设计公司。

闻泰科技(600745)：半导体龙头股。旗下安世集团拥有生产氮化镓相关的技术，安世半导体生产 GaN 产品，车载 GaN 已经量产，全球最优质的氮化镓供应商之一。

三安光电(600703)：半导体龙头股。全资子公司厦门市三安集成电路有限公司主要从事化合物半导体集成电路业务，涵盖 PA 射频、电力电子、光通讯和滤波器等芯片，主要应用于大数据、云计算、物联网、电动汽车、智能移动终端、通讯基站、导航等领域的信息技术产品，应用广泛。

在 2011-01-01 到 2021-10-31 期间两千多个交易日的各个指标(日收盘价、日开盘价、日最低价、日最高价、日交易量、日交易额)为样本数据(数据来源 163 网易股票市场)。

3.2. 实验设计

三组数据分别为 2536，2406，2594 个交易日指标，为预测开盘价，实验步骤如下。

第一步，本文只预测开盘价，所以对每组数据的开盘价进行归一化处理。

第二步，输入前两千多工作日的开盘价为训练集，最后 300 天的开盘价为测试集。设置时间戳为 40，训练轮数为 20。

第三步，构建 LSTM 模型，分别有三种，单层 LSTM，多层 LSTM 和双向 LSTM，对于每组数据用三种模型分别进行实验，选取误差最小的双向 LSTM 模型，得到实验结果。

第四步，测试集输入模型进行预测，对预测数据和真实数据反归一化到原始数值，并绘制真实数据和预测数据的对比曲线，求出均方误差，均方根误差，平均绝对误差和决定系数。

3.3. LSTM 模型拟合和预测

利用 python 里的 spyder 软件，两个安装包分别是 TensorFlow 和 keras，通过设置不同训练集得出拟合优度，并比较实验结果。三组数据分别进行拟合得出预测结果图如下。

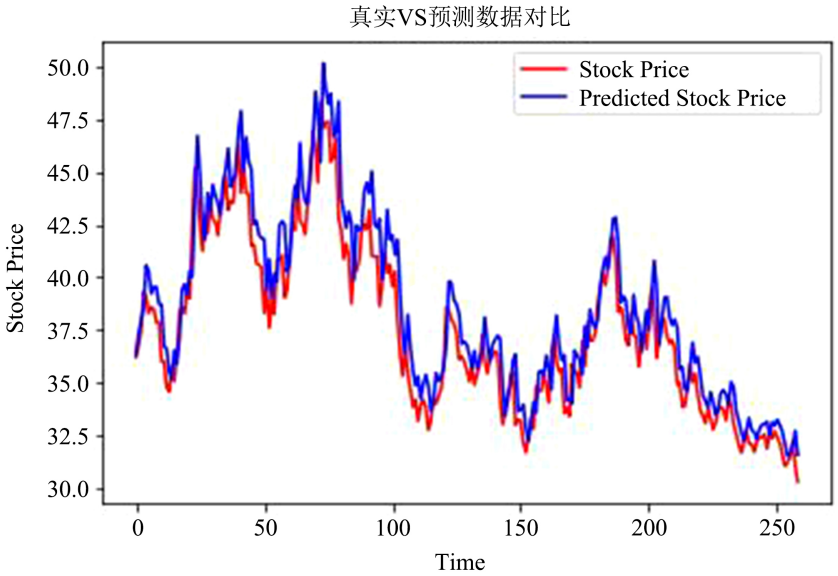

Figure 3. Changdian Technology (600684) experimental results
图 3. 长电科技(600684)实验结果图

上图 3 是股票代码长电科技(600584)，得出实验结果，均方误差是：2.52943。均方根误差：1.59042。平均绝对误差：1.26634。其中 R2 表示是决定系数，可以简单理解为反映模型拟合优度的重要的统计量，R2 的值越大表示模型拟合程度越好。

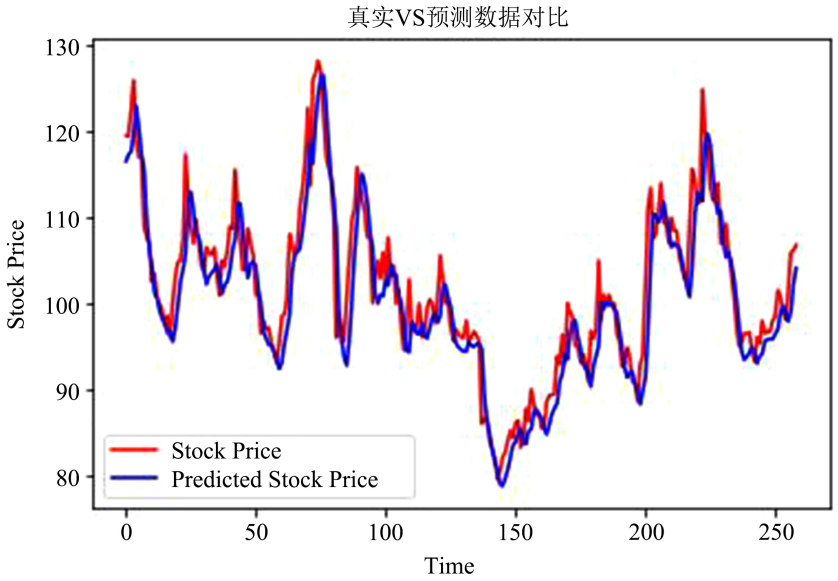

Figure 4. Wentai Technology (600745) experimental results
图 4. 闻泰科技(600745)实验结果图

上图 4 是股票代码闻泰科技(600745)，得出实验结果，均方误差是：14.96051 均方根误差：3.86788。平均绝对误差：2.86100。这组数据的拟合程度没有第一组数据结果好。由于这组数据比较大，级别是 100 多，而其它两组数据都是几十，所以误差比较大，但是通过比较 R2 的值可以知道拟合程度还是比较好的。

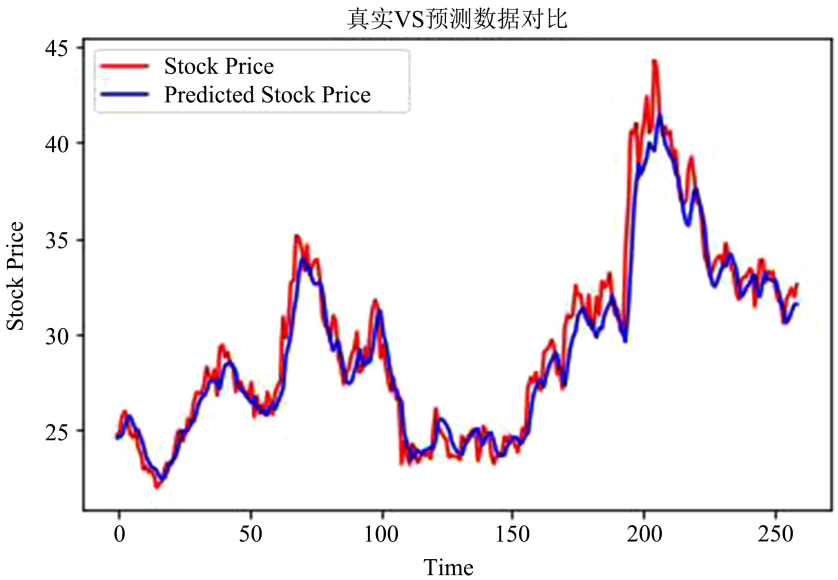

Figure 5. Sanan Optoelectronics (600703) experimental results
图 5. 三安光电(600703)实验结果图

上图 5 是股票代码三安光电(600703)，得出实验结果，均方误差是：1.74834。均方根误差：1.32225。平均绝对误差：0.93061。对于三个实验结果进行误差比较，见表 1。

Table 1. Comparison table of three experimental parameters
表 1. 三次实验参数对比表

股票代码	长电科技(600584)	闻泰科技(600745)	三安光电(600703)
MSE	2.5293	14.96051	1.74834
RMSE	1.59042	3.86788	1.32225
MAE	1.26634	2.86100	0.93061
R2	0.86248	0.83625	0.91408

4. 结论

股票预测对于政府决策起着非常重要的作用，本文根据长电科技、闻泰科技、三安光电 2011 到 2021 的股票真实数据，利用 LSTM 模型预测未来股票具体收盘价，可进一步判断股票上涨下跌情况，预测效果良好。根据三次实验结果，可以得到误差较小的预测价格。其中长电科技和三安光电的股票价格在 100 以内变化，闻泰科技的股票价格在 100 以上。因为使用相同模型，得到结果显示长安科技和三安光电的误差更小。但是可以看到三个公司的股价都将呈上升趋势。表明了近期半导体行业的发展较好，虽然过了风口之后最高的下降期，但是在缓慢上升。本文对这三支代表股票数据进行分析，在实际股票中，每天的数据波动性都非常大，因此后期应对更多半导体股票数据做进一步分析。

本文对股价的预测方法是比较简单基础的，通过 LSTM 模型得到比较精确的实验结果。但是仍存在处理数据上的不足。此模型在未来金融时间序列研究中也比较广泛，可以为投资者提供一定的参考信息，也能为后续的股价研究者提供相应的参考。

参考文献

[1] 张如梦, 张华美. BP 神经网络与 ARMA-GARCH 模型在股票预测中的对比分析[J]. 高师理科学刊, 2021, 41(4): 14-20.

[2] 胡衍坤. 基于 TensorFlow 进行股票预测的深度学习模型的设计与实现[J]. 计算机应用与软件, 2018, 35(6): 267-271.

[3] 杨琦, 曹显兵. 基于 ARMA-GARCH 模型的股票价格分析与预测[J]. 数学的实践与认识, 2016, 46(6): 80-86.

[4] 黄颖, 杨会杰. 基于 XGBoost 和 LSTM 模型的金融时间序列预测[J]. 科技和产业, 2021, 21(8): 158-162.

[5] 张康林. 基于 pytorch 的 LSTM 模型对股价的分析与预测[J]. 计算机技术与发展, 2021, 31(1): 161-167.

[6] 宋刚, 张云峰, 包芳勋. 基于粒子群优化 LSTM 的股票预测模型[J]. 北京航空航天大学学报, 2019, 45(12): 2533-2542.

[7] 李雄英, 陈小玲, 曾凯华. 基于三类模型的四大银行股票收益率预测研究[J]. 经济数学, 2018, 35(4): 21-27.