

分类号_____

密级_____

UDC 注1 _____



南京理工大学

NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

硕士学位论文

基于注意力机制和 LSTM 神经网络的

股票价格预测分析

(题名和副题名)

冷天翔

(作者姓名)

指导教师姓名 _____ 安智宇 副教授

学 位 类 别 _____ 经济学硕士

学 科 名 称 _____ 金融学

研 究 方 向 _____ 金融市场投资行为与分析

论文提交时间 _____ 2021 年 3 月

注 1：注明《国际十进分类法 UDC》的类号。

硕士学位论文

**基于注意力机制和 LSTM 神经网络的
股票价格预测分析**

作者：冷天翔

指导教师：安智宇 副教授

南京理工大学

2021 年 3 月

Master Dissertation

**Stock Price Prediction Research Using
Attention-Based LSTM Neural Network**

By

Leng tianxiang

Supervised by Prof. An zhiyu

Nanjing University of Science & Technology

March, 2021

声 明

本学位论文是我在导师的指导下取得的研究成果，尽我所知，在本学位论文中，除了加以标注和致谢的部分外，不包含其他人已经发表或公布过的研究成果，也不包含我为获得任何教育机构的学位或学历而使用过的材料。与我一同工作的同事对本学位论文做出的贡献均已在论文中作了明确的说明。

研究生签名： 冷天翔 2021年3月7日

学位论文使用授权声明

南京理工大学有权保存本学位论文的电子和纸质文档，可以借阅或上网公布本学位论文的部分或全部内容，可以向有关部门或机构送交并授权其保存、借阅或上网公布本学位论文的部分或全部内容。对于保密论文，按保密的有关规定和程序处理。

研究生签名： 冷天翔 2021年3月7日

摘 要

随着社会的进步和科学技术的飞速发展，中国的股票市场越来越受到投资者的关注。为了达成投资目标，投资者通过一些行之有效的方法对于股票价格进行预测显得尤为重要。目前，在金融时间序列的预测方面，比较前沿的方法是 LSTM 神经网络，它能够部分缓解长时依赖问题。而注意力机制作为进一步解决长时依赖问题的利器，具有广阔的应用前景。

本文在 LSTM 神经网络的基础上融入了注意力机制，构建了基于注意力机制和 LSTM 神经网络的股价预测模型（Att-LSTM）。预测标的采用富时中国 A50 指数的最新成分股，这一指数包含了沪深两市市值最大的 50 只股票，预测这一指数成分股的价格具有现实意义。为了能够较大程度上涵盖影响股价的基本信息，引入模型的特征变量不仅包括最基本的开盘价，收盘价，最高价，最低价，成交量以及衍生技术指标，还额外加入了基于皮尔森相关系数所计算的前十只最相关股票的收盘价格。在利用网格寻优法来选择相对较优的超参数的基础上，使用训练集的数据对各模型进行训练，然后使用测试集来测试预测效果。进一步，本文针对在不同的超参数条件下以及极端事件影响下的预测效果差异进行了研究。

结果表明，在不同评价标准下，本文所构建的 Att-LSTM 的三种子模型在大多数预测标的上均优于各对比模型，并且均值及中位数差异通过了显著性检验，在三种子模型中，同时加入三种指标的子模型表现最好。同时，本文还发现极端事件会导致加入衍生技术指标和相关股指标的 Att-LSTM 模型的预测效果减弱，在极端事件情况下，仅加入基础指标的 Att-LSTM 模型表现最好。此外，针对 Att-LSTM 模型，选择合理的超参数组合有助于提升预测效果。

关键词： 股价预测，长时依赖问题，LSTM 神经网络，注意力机制

Abstract

With the progress of society and the rapid development of science and technology, China's stock market has attracted more and more attention from investors. In order to achieve investment goals, it is particularly important for investors to predict stock prices through some effective methods. At present, in the prediction of financial time series, the more cutting-edge method is LSTM neural network, which can partially alleviate the problem of long-term dependence. The attention mechanism, as a sharp tool to further solve the problem of long-term dependence, has broad application prospects.

This paper incorporates the attention mechanism on the basis of the LSTM neural network, and builds a stock price prediction model (Att-LSTM) based on the attention mechanism and the LSTM. The forecast target uses the latest constituent stocks of the FTSE China A50 Index, which contains the 50 largest stocks in Shanghai and Shenzhen stock markets by market capitalization. It is of practical significance to predict the price of the constituent stocks of this index. In order to cover the basic information that affects stock prices to a greater extent, the characteristic variables introduced into the model include not only the most basic opening price, closing price, highest price, lowest price, trading volume, and derivative technical indicators, but also additional Pearson-based correlations. The closing prices of the top ten most relevant stocks calculated by the coefficient. On the basis of using the grid optimization method to select relatively optimal hyperparameters, the data of the training set is used to train each model, and then the test set is used to test the prediction effect. Furthermore, this paper studies the difference in prediction effects under different hyperparameter conditions and under the influence of extreme events.

The results show that under different evaluation criteria, the three sub-models of Att-LSTM constructed in this paper are superior to the comparison models in most of the prediction targets, and the mean and median differences have passed the significance test. In the three sub-models Among them, the sub-model that adds three indicators at the same time performs best. At the same time, this article also finds that extreme events will weaken the prediction effect of the Att-LSTM model that adds derivative technical indicators and related stock indicators. In the case of extreme events, the Att-LSTM model that only adds basic indicators performs best. In addition, for the Att-LSTM model, selecting a reasonable combination of hyperparameters can help improve the prediction effect.

Keywords: Stock price prediction, Long-term dependence problem, LSTM neural network, attention mechanism

目 录

摘 要	I
Abstract	II
1 绪论	1
1.1 研究背景	1
1.2 研究意义	2
1.3 研究内容	3
1.4 研究的创新点	4
2 文献综述	5
2.1 神经网络及其变体在金融时序预测上的应用	5
2.2 注意力机制在金融时序预测中的应用	8
2.3 文献述评	9
3 研究方法概述	11
3.1 循环神经网络	11
3.2 LSTM 神经网络	15
3.3 注意力机制	18
3.4 基于注意力机制和 LSTM 神经网络的时序预测模型	19
3.5 LSTM 神经网络的优化策略	23
3.5.1 自适应动量估计法	23
3.5.2 批次标准化层	24
3.5.3 早停策略	25
4 研究方案设计	26
4.1 特征变量的构建	26
4.1.1 基础指标及衍生技术指标的选取	27
4.1.2 相关股指标的选取	29
4.2 样本选择与数据说明	31
4.3 特征变量的无量纲化	32
4.4 超参数寻优及模型训练	32
4.5 预测结果的评价与检验	33
4.5.1 预测结果的评价标准	33
4.5.2 模型间预测效果差异的显著性检验	34

5 基于 Att-LSTM 的股价预测研究.....	35
5.1 描述性统计分析	35
5.2 相关性分析	36
5.3 超参数寻优	39
5.4 模型预测对比分析	39
5.5 进一步研究	42
5.5.1 极端事件期间预测效果的扩展研究	42
5.5.2 超参数对于预测效果的影响的扩展研究	44
6 结论与展望	46
6.1 结论	46
6.2 展望	46
参考文献.....	47
致 谢	51
附 录 A	52
附 录 B	53

1 绪论

1.1 研究背景

股票市场目前已经成为了整个经济社会不可或缺的一部分，它作为直接融资的典型代表，自从诞生以来，为千百万投资者提供了投资的选择和渠道，让他们的个人资产能够得到合理的保值和增值，更让他们能够和优秀而伟大的公司共同成长。上面是从个人角度说的，从企业角度来说：同时它又为企业谋求发展和扩大规模提供了宝贵的资金的来源，由于股票的无需偿还性，企业得以能够抓住稍纵即逝的发展机遇。而从更加宏观的角度来说，股票市场起到了融通资金的作用，逐渐成为了经济社会高速发展的基石。中国股市发轫于 1989 年，发展至今也已经有三十年的历史。从当初的筚路蓝缕到如今的光辉闪耀，如今，沪深两市的日交易额已经合计约 5000 亿元，上市公司的总数目突破了四千家，所有上市公司的总市值达到了近 80 万亿。可以这么说，中国股市已然成为了中国经济不可或缺的一部分。投资者们投资着中国股市就相当于在投资着中国，股市的变动很大程度上体现出了一个国家经济社会的发展情况，因此，我国股票市场的不断壮大正体现出了我国不断发展的经济实力，国际地位乃至综合国力。

随着计算机网络技术的迅猛发展，各种投资致富的新闻案例不绝于耳，因此，投资理财的理念也逐渐被普通人所广泛接受。但是想在股票市场取得预想中的收益是非常困难的，股票市场每时每刻都在产生巨量的数据，全市场每天的分笔数据就有数十 GB，这还仅仅是股票市场本身的数据信息，如果算上外部环境，数据量是天文数字级别。因此，越来越多的投资者开始尝试运用计算机技术针对海量的股票数据进行处理，以期获得预期中的高额回报。然而，对于股票价格有影响的特征变量非常多，一只股票的价格不仅和公司本身的经营状况，市场投资者短期的供求关系有关，尤其还容易受到外部宏观经济环境，整体市场的情绪冷热等等高噪声因素的干扰，所以，能在纷繁复杂，噪声密布的股价数据中寻找出有价值的信息，并且对于这些信息进行再处理，最终提炼出投资决策意见的能力显得非常重要。

在过去，学者们对于股票市场研究的目光往往聚焦于波动性预测（韦艳华和张世英, 2004）以及重大金融风险管理领域（巴曙松和蒋峰, 2019）。但目前，由于投资者对于准确预测股价，指导投资决策的需求日益高涨，投资者们的首要目标已经变成了最大限度地减亏增盈，提高投资收益率，因此，建立各种模型对股票价格进行准确的预测成为了专业投资者最重要的研究内容之一。在研究方法上，以往的学者们所利用的主要是各种基于统计学和概率论的方法，并将其应用于金融时间序列的预测。多元线性回归模型、自回归滑动平均模型、自回归条件异方差模型（Ching 等, 2013）的出现极大地推

动了金融时间序列预测领域的研究发展。但是由前述可知，由于股票市场是高度动态的，复杂的，非线性的，混沌的和非参数的（Abu-Mostafa 等，1996），所以这些基于统计学和概率论的传统模型对于股价的非线性变化规律的反应能力比较低，因此导致预测的效果也不好，用它们的结果来辅助进行投资决策的实用性没有达到预期的效果。

近年来，“摩尔定律”使得计算机的算力不断地提高，在此基础上，以神经网络为代表的人工智能技术也在飞速发展。越来越多的学者们开始运用人工智能算法对于以股价为代表的金融时间序列进行预测，他们运用的方法主要有改进的神经网络（Hansen 等，1997），改进的聚类算法（Zakaria 等，2012）和改进的支持向量机回归（王卫红等，2016）等等，都达成了较好的预测效果。而对于金融时间序列这种前后可能存在某种关联的特殊形式的数据来说，目前学术界应用最多的人工智能模型为循环神经网络（RNN）模型（Rumelhart 等，1986），循环神经网络提出的目的就是为了处理时间序列数据，它可以通过递归形式的网络结构提取出时间序列数据更深层次的特征，使得获取时间序列数据的固有非线性信息成为可能。然而，传统的循环神经网络模型的结构具有很大局限性，一旦遇到所处理的时间序列过长，或者特征变量数目过多，就极易容易面临梯度爆炸或梯度消失的问题（Bengio 等，1994），这个问题的存在导致模型的拟合效果变得非常低，从而使得模型无法学习到时间序列中所固有的长时依赖信息。Hochreiter 和 Schmidhuber（1997）针对长时依赖问题专门提出了长短时记忆神经网络（LSTM）结构，LSTM 神经网络可以部分解决传统的循环神经网络模型的梯度爆炸和梯度消失的问题，从而使得整个模型不仅能够提取出时间序列数据更深层次的特征，还能学习到时间序列中所固有的长时依赖信息，从而大大提升了模型的预测效果（Heaton 等，2017）。

1.2 研究意义

股票市场常被称作经济市场的晴雨表，这说的是通过股票市场的走势可以预知未来经济社会的发展变动趋势（Mabrouk，2011），及时对于未来可能出现的风险和机遇进行准备，由于经济市场的良好运行对于整个国家的良性发展十分重要，所以股票市场便成为了整个国家经济体最重要的组成部分。及时有效地预测股票市场的价格，不仅可以为管窥经济社会的运行趋势提供帮助，为国家领导层进行经济领域的决策提供参考，从而确保经济社会的稳定发展，而且对于股票市场进行预测可以及时地发现个股的异动，从而能够敏锐地捕捉投资机会，规避投资风险。现如今，股票市场的规模在不断地扩大，面对海量的上市公司，能通过某种方法对于股价变动进行预测显得尤为重要，这可以帮助及时发现某些风险情况的发生，防范风险进一步扩大，造成更大的损失。最后，对于普通投资者来说，股票价格预测能够直接指导微观的投资决策，帮助普通投资者获取预

期的收益。因此，对股票价格进行预测研究，在经济社会层面和普通投资者层面都意义非凡。

随着金融科技的不断发展和管理层的大力支持，最近几年，将以 LSTM 神经网络模型为代表的深度学习技术应用于金融领域的风潮持续盛行。各大券商的金融工程团队都将大量的人力和精力投入了这一交叉领域的研究，并且推出了很多种借助深度学习技术进行选股甚至投资决策的金融产品，取得了非常良好的反响。

因此本文的研究意义在于通过将经过注意力机制改进的 LSTM 神经网络模型运用在金融时间序列的预测上，给投资者的投资决策带来一定的辅助指导。这与目前股票市场的大趋势相符合，具有较大的研究价值和实际意义。

1.3 研究内容

在现有研究的基础之上，本文的研究目的在于验证利用注意力机制和 LSTM 神经网络预测股票价格的有效性。在广泛阅读文献，设计好完善的研究方案的基础上，本文研究了所构建的基于注意力机制和 LSTM 神经网络的时间序列预测模型（Att-LSTM）预测股票收盘价的预测效果，同时与多种基准模型进行对比，并进行了统计学显著性检验，保证研究结果的可靠性。本文还研究了在不同的超参数条件下的预测效果差异，并且针对极端事件时期的预测效果差异也进行了研究。本文的主要内容以及结构安排如下：

第一章：绪论，该部分论述本文的研究背景及意义，指出了传统的 LSTM 神经网络模型在时间序列预测领域存在一些固有缺陷，引出本文的主要研究目的是构建基于注意力机制和 LSTM 神经网络的时间序列预测模型（Att-LSTM）对于股票收盘价进行预测，验证所构建的模型的有效性。最后说明了本文的研究内容以及可能的创新之处。

第二章：文献综述，该部分对于国内外利用 LSTM 神经网络以及注意力机制预测股票价格的相关研究进行回顾与综述。

第三章：研究方法概述，该部分介绍了 LSTM 神经网络的相关研究方法，提出了基于注意力机制和 LSTM 神经网络的时间序列预测模型（Att-LSTM），并且详细分析了模型的具体结构。最后给出在实际训练模型的过程中采用的一些关键的优化方法。

第四章：研究方案设计。该部分主要包含五个部分：一是特征变量的构建，本文构建了三类可能对于股票价格产生影响的特征变量，包括基础指标，衍生技术指标以及相关股指标；二是样本选择与数据说明；三是特征变量的无量纲化；四是神经网络的超参数寻优以及模型训练，本文将采用网格寻优法确定各个模型中的最优超参数，在训练策略上使用早停策略来进一步防止过拟合的发生。五是预测结果的评价方法，本文使用了三种评价指标对于预测结果进行评价，同时还将对不同模型之间的预测差异进行参数和非参数检验，保证结果的可靠性。

第五章：基于 Att-LSTM 的股价预测研究：该部分首先对于数据集进行基本的描述性统计，观察是否存在明显的离群值可能对于预测效果产生影响，然后呈现本文构建的各个指标和股票收盘价之间的相关性分析结果。在利用网格寻优法确定了每个预测标的使用的具体超参数之后，利用训练集进行训练，并将测试集数据代入训练好的模型，得出本文构建的 Att-LSTM 模型和其他基准模型的预测效果以及差异，并进行显著性检验。最后，本文进一步研究了各超参数组合下和极端事件时期的预测标的的预测效果的差异。

第六章：结论与展望，该部分对本文所得出的研究结论进行简要总结，指出了研究的不足以及未来的改进方向。

1.4 研究的创新点

在将注意力机制引入金融时间序列分析的领域，学者们的研究相对较为简单，本文在 LSTM 神经网络中融入了注意力机制，并在此基础上加入了批次标准化层和 Dropout 方法，构建了基于注意力机制和 LSTM 神经网络的时间序列预测模型对股票收盘价进行预测。在模型的训练方面应用了早停策略，来提高预测结果的可靠性。同时，本文计算了预测标的股票与其同行业股票的皮尔森相关系数，将其绝对值从高到低排序，取前十名作为此预测标的股票的相关股，通过加入前十大相关股的收盘价这一外生特征变量指标，以期进一步改善模型的预测效果。模型都有各自的适用范围，不同超参数组合以及极端事件的发生，均有可能影响模型的预测效果。本文也将针对这个问题进行进一步研究。

2 文献综述

针对金融领域时间序列的预测研究一直以来都是众多学者关注的热点问题。鉴于金融时间序列的不确定性及高噪声特征，精准预测仍然非常困难，加之自变量和因变量之间的关系通常会随时间发生动态变化，致使传统时间序列模型很难有效进行金融市场预测。此外，仅适用于平稳序列建模也大幅限制了时间序列模型的应用与拓展。

计算机的算力的增加以及越来越多的神经网络变体模型的推出，使得越来越多的学者开始将神经网络方法应用于金融时序预测上。在近几年，注意力机制被提出用来改进 LSTM 神经网络的预测效果，有一些学者也开始探索其在金融时序预测领域的应用。

2.1 神经网络及其变体在金融时序预测上的应用

目前，神经网络及其变体模型主要包括基于人工神经网络（Artificial Neural Network, ANN）的各种模型，应用于金融时序预测上的主要有多层感知机（MLP），反向传播神经网络（BPNN），循环神经网络（RNN），长短期记忆神经网络（LSTM），门控循环单元（GRU）。

MLP, BPNN 是较为传统的方法，输入的特征变量数据没有先后顺序，导致无法反映时间序列的本质特征，同时容易陷入局部极值问题，即无法找到全局最优解。

Hammd 等（2007）利用三层 BP 神经网络对于约旦股票市场的股票收盘价进行了预测，结果表明，和所使用的传统统计学基准模型相对比，三层 BP 神经网络的预测效果要更加好，相对来说精度也要更高。Dhar 等（2010）使用 MLP 多跟感知机模型来预测印度证交所的几种综合股指的收盘价，其通过实验结果发现，预测实验取得了良好的效果，同时预测结果对于模型的超参数变化显得十分敏感，因此导致预测结果的可变性非常强。许兴军和颜钢（2011）使用多层 BP 神经网络预测在上交所上市的浦发银行在金融危机后的收盘价格，他们选取了各周期均线，量价以及各种技术指标作为特征变量输入，结果表明实现了良好的预测效果。朱永明和邵庚云（2013）则将目光聚焦于上市公司的基本面，他们使用财务数据作为输入的特征变量，由于财务指标的数量很多，因此先使用主成分分析方法对于财务指标进行降维，而后输入 BP 神经网络进行预测，与基准模型的对比结果表明，BP 神经网络的预测效果明显较好。Ticknor（2013）提出一种人工神经网络，利用贝叶斯模型进行修正，使用量价关系以及财务数据作为输入特征变量进行单步预测，结果表明，在美国股票市场的蓝筹股上该模型的预测效果有了质的提升。

为了克服传统神经网络所面临的种种问题，Hinton 教授在 2006 年首次提出深度学习的概念（Hinton 等，2006），利用深度学习理念建立的神经网络模型能够根据数据特征进行一般化学习，尤其擅长学习不同维度的信息，在学习过程中摒弃无关因素，将重点聚焦于有用的因素上，因此可以获得更好的预测效果。Hinton 在其论文中使用预训练结合微调的方法成功训练出了多层的深度置信网络，使模型能够从原始数据中抽象出具有代表性的特征，开启了深度学习的时代。在此之后，Hinton 和 Bengio（2007）又提出了 Dropout 方法，这个方法可以在神经网络训练过程中，随机弃用某些比例的神经元，这种方法可以有效地解决神经网络的过度拟合问题，他们还证明了随着神经网络的深度的增大，会导致神经网络的泛化能力也变得越来越好。局部极值问题长久以来一直是困扰神经网络预测效果的障碍，而 LeCun 等（2015）从技术角度论证了深层网络可以使局部极值问题带来的问题的影响降到最低，因而深度学习技术成为目前最前沿的人工智能技术。Heaton 等（2016）则总结了将深度学习技术用于金融市场研究的三大优势：第一是相对于以往的传统统计学模型对于输入变量的限制，深度学习对于输入变量的具体形式没有限制，模型的输入可以是与金融时间序列预测相关的任何特征变量；第二是深度学习技术可以有效学习到输入变量和预测变量之间的非线性的复杂关系，提高模型的学习能力；第三是金融时间序列具有鲜明的时序性特征，而深度学习技术所构建的神经网络可以有效捕捉时序数据之间的前后关联性，很好地利用这一特征。

目前，基于深度学习理念的神经网络模型主要包括循环神经网络（Recurrent Neural Network, RNN）及其衍生模型 GRU、LSTM，卷积神经网络（Convolutional Neural Network, CNN）和深度信念网络（Deep Belief Network, DBN）等。在金融时间序列预测领域，所使用的主要是 RNN, LSTM, GRU。

Xiong 等（2015）利用 LSTM 神经网络对标准普尔 500 种股票价格指数进行建模，实验结果是 LSTM 神经网络相对于基准模型针对着有复杂噪声的数据有明显预测优势。Chen 等（2015）把目光聚焦于中国市场，使用 LSTM 神经网络预测股票三天的总收益率，依然是使用主成分分析方法将量价等数据转化为一个十维输入特征变量矩阵，结果表明，与传统的预测方法对比，LSTM 神经网络成功地将预测的准确率提高了 12.9 个百分点。Di Persio 和 Honchar（2016）使用 LSTM 神经网络作为主模型，MLP 和 CNN 作为基准模型，对于标准普尔 500 指数收盘价涨跌的虚拟变量进行预测，结果证明了相较于基准模型，LSTM 神经网络的预测效果是最好的。Di Persio 和 Honchar（2017）又将循环神经网络的三种变体形式 RNN、LSTM 神经网络和 GRU 神经网络用于谷歌公司的收盘价涨跌预测，实验结果再一次证明 LSTM 神经网络在金融时间序列预测方面具有独特优势。Samarawickrama 和 Fernando（2017）使用循环神经网络的变体 GRU 预测斯里兰卡股票指数的收盘价，结果证明 GRU 在金融领域的时间序列预测上也能取得很好的

效果。Liu 等（2017）采用了卷积神经网络和 LSTM 神经网络相结合的混合模型对于中国股市的动量收益进行预测，结果表明，这一混合模型预测所构建的投资组合能够取得比一般经典动量模型更高的收益。冯宇旭和李裕梅（2019）利用 LSTM 神经网络来对沪深 300 指数的收盘价格进行预测，他们在传统的量价指标中额外加入了移动平均成交量特征变量，与基准模型 SVR 和 Adaboost 的预测结果进行对比，在 RMSE 评价指标上 LSTM 神经网络要更加低。杨青和王晨蔚（2019）构造了一个深度 LSTM 神经网络针对全球数十个国家三十余种股票价格指数进行预测，结果表明，LSTM 神经网络具有优秀的预测精度，相较于三种基准模型（ARIMA、MLP 和 SVR）而言，三十余种指数的平均预测效果在多种预测期限上都有了较大改善。宋睿（2020）采用了卷积神经网络和 LSTM 神经网络相结合的混合模型对于股票涨跌的虚拟变量进行分类预测，输入特征变量是运用特征提取技术获取的五十种特征，结果显示，这种混合模型在股价的三种趋势虚拟变量的预测上取得了 86.5% 的准确率。王霄鹏（2020）利用基于主成分分析的 LSTM 神经网络对于平安银行的收盘价进行预测，结果证实了基于主成分分析的 LSTM 神经网络有应用前景，相较于基准模型的预测效果有了较大改善。文宝石，颜七笙（2020）在特征工程的基础上，充分考虑的时间序列间的固有联系和特征变量间的自身相关性，构建了基于多维数据处理的 LSTM 神经网络模型，结果证明，预测评价指标均方根误差以及平均绝对误差相对于所使用的基准模型分别降低了 90.81% 和 91.65%，新模型的使用极大提升了预测的精度。乔忠学（2020）针对 LSTM 神经网络的单双向问题做了研究，他选取了深证成指的日收盘价数据进行预测，实验结果表明，双向的 LSTM 神经网络的预测精度和单向的 LSTM 神经网络并没有显著的差别，他的研究证明了股价信息流动具有单向性，而不是向机器翻译等领域存在着信息双向流动的问题。武大硕等（2020）利用遗传算法对于 LSTM 神经网络的超参数进行改进，所选用的基准模型为 BP 神经网络模型与未经遗传算法优化超参数的 LSTM 网络模型，结果表明，在对美国纳斯达克指数收盘价的预测上，不同的波动阶段预测精度差异各有不同，收盘价波动越小准确率也越高，而经过遗传算法优化了超参数之后，即便是在剧烈波动阶段，主模型表现的也比两种基准模型要好的多。刘宁宁和张量（2020）提出了基于主成分分析和多层 GRU 相结合的预测模型。采用上证指数 26 年的历史收盘价数据进行预测，与基准模型 LSTM 神经网络相比，基于主成分分析的多层 GRU 取得了良好的效果，其中两层 GRU 的效果最好，RMSE 值和趋势预测准确率分别达到了 0.025 和 75.9%。包振山等（2020）同样利用遗传算法来寻找 LSTM 神经网络的最优超参数，他们使用了中证 500 的日内分钟数据进行测试验证。实验得出，改进模型的各方面指标均优于未经遗传算法进行超参数优化的单纯 LSTM 神经网络模型。

2.2 注意力机制在金融时序预测中的应用

随着深度学习技术的不断发展，可以输入的模型中的特征数目也越来越多。就股票市场而言，在过去数十年的研究中，学术界宣称发现了数以百计能够提供超额收益的特征因子。然而经过研究发现，这些特征因子中只有少数能够持续提供超额收益（Harvey 等，2016；McLean 和 Pontiff，2016；Green 等，2017；Linnainmaa 和 Roberts，2018；Hou 等，2019）。过多的输入特征会带来过拟合，训练时间过长等诸多问题，更严重的问题是会给模型预测带来不必要的噪声。

注意力机制（Attention Mechanism）作为一种非常有效地处理多特征输入的深度学习机制，虽然在 90 年代中期（Desimone 和 Duncan, 1995）就已经被提出，但是由于当时计算能力的限制，并没有引起广泛关注。注意力机制首先是在图像识别领域被提出的，但是由于当时计算能力的限制，并没有引起广泛关注。它主要是模仿人在看图像时候的视线移动，乃至在看一个复杂场景时的目光聚焦点。因此当神经网络进行图像识别的时候，可以将注意力集中于某些重要的特征上，使得识别能够更加准确。这一过程的原理是人们在观察图像的时候是有着眼点的，如果单次将目光整个施加于图像上，会导致信息的混乱，所以大多是根据需求将注意力集中到图像的特定部分，更为重要的是，人们会根据之前观察的图像学习到未来要观察图像时注意力应该集中的位置。在经历了长期的发展停滞，随着计算机算力的增加，来自 Google DeepMind 团队的 Mnih 等（2014）发表的论文才使学界重新对注意力机制产生了浓厚兴趣。他们认为由于信息处理的瓶颈，人类往往是根据需求关注信息特定的一部分，同时忽略其他可见的信息。根据人类学习的这个特点，他们构建了基于注意力机制的深度学习模型，在图像识别领域取得了很好的效果。

正是由于注意力机制在图像识别领域的良好效果，对于将基于注意力机制的深度学习模型应用到金融时间序列预测方面的相关研究，近几年不断的涌现。

Cinar 等（2017）用循环神经网络对时间序列进行建模和预测，他们给循环神经网络附加了扩展的注意力机制，实验结果表明将该机制加在循环神经网络上可以弥补时间序列中缺失值的问题。Chen 等（2019）在两层 LSTM 模型的输入层后加入注意力层，应用 softmax 对于输出进行二分类，用以预测涨跌，结果表明：28 只港股在测试集上的涨跌预测准确率有 22 只比传统的 LSTM 模型要高。并且，作者还利用预测出的结果构建了一个简单的买入并持有策略，结果该策略在验证集上明显跑赢了市场。因此，基于注意力机制的 LSTM 能够显著提升模型的预测效果。王理同等（2019）利用加入注意力机制的 LSTM 神经网络模型，使用上证 50 成分股的历史收益率作为输入变量，预测上证 50 指数的收益率，结果证实，基于注意力机制的 LSTM 神经网络模型比传统 LSTM 神经网络模型提升了 3 倍以上。乔若羽（2019）针对股票市场的特征提取困难、预测精度

较低等问题, 给 RNN、LSTM 和 GRU 引入注意力机制来赋予各时间维度的信息不同权重, 区分不同信息对预测的重要程度, 从而提升各模型的性能。实验结果表明, 在加入注意力机制以后, 模型的预测效果都有了明显的改善。林杰和康慧琳(2020)针对中国股票市场, 提出了一种基于注意力机制的 LSTM 股价趋势预测模型。他们选取最新的上证 50 指数成分股作为研究对象, 针对各种技术指标所发出的买入卖出信号进行趋势化处理得出了基于技术指标的趋势预测数据, 然后和传统的交易数据进行组合作为预测模型的新的输入, 然后使用基于注意力机制的 LSTM 模型提取股价趋势特征进行预测。实验结果表明: 引入基于技术指标的趋势预测数据到预测模型中, 能够在单一量价输入特征变量的基础上提升预测精确度, 与传统的机器学习模型 SVM 和单一的 LSTM 模型相比, 基于注意力机制的 LSTM 模型具有更好的预测效果。谷丽琼等(2020)将注意力机制融合进 GRU 预测模型之中, 在科大讯飞股票收盘价上的预测实验结果证明了, 注意力机制的融入可以使得模型能够聚焦在更重要的时间点以及更重要的特征变量信息上, 能够使用这些重要的信息解决传统模型对于时间序列的特征变量不敏感的问题, 提升了预测精度。实验结果表明基于注意力机制的 GRU 模型在 MAPE, RMSE, R2score 三个评价指标上均优于其他模型, 注意力机制能够捕捉重要时间点局部特征, 对预测模型的优化是可行和有效的。

2.3 文献述评

根据文献综述可以看出, 在将神经网络应用于金融时间序列预测的领域, 传统的神经网络模型(MLP, BPNN 等)已经大量的实践证明预测效果不如循环神经网络模型及其变体(RNN, LSTM, GRU 等)。越来越多的学者的关注重点放在如何处理日益复杂的输入特征变量, 大部分学者选择采用简单的主成分分析方法进行处理, 有些学者简单地应用了注意力机制, 做了一些研究尝试, 结果表明, 注意力机制的应用的确能够提升循环神经网络模型的预测效果。但是, 这些学者的模型结构相对较为简单, 优化策略应用较少, 给模型的预测效果带来了不利影响。从特征变量选取的角度来看, 学者们主要考虑的还是基础的交易数据以及衍生的技术指标, 这一些变量属于内生特征变量。他们对于影响股票价格的外生变量考虑较少, 而投资实践证明以相关股收盘价为代表的外生变量对于股价有显著影响。此外, 目前针对模型在不同时期以及不同超参数下的预测效果差异的研究不多。

综上, 本文将融合注意力机制和 LSTM 神经网络的优点, 尝试构建基于注意力机制和 LSTM 神经网络的时序预测模型, 用于预测股票收盘价。同时将计算预测标的股票与其同行业股票的皮尔森相关系数, 将其绝对值从高到低排序, 取前十名作为此预测标的股票的相关股。通过加入相关股收盘价这一外生特征变量, 本文将探究模型的预测效果

是否得到改善。此外，本文还将研究模型在极端事件期间以及不同超参数组合下的预测效果。

3 研究方法概述

3.1 循环神经网络

循环神经网络（Recurrent Neural Network, RNN）是一系列以时间序列数据为输入，以时间序列顺序作为方向进行迭代，且所有神经元节点都是全连接的特殊递归神经网络。传统的神经网络是在每一个隐藏层间建立权值的连接关系，而与传统的神经网络相比，循环神经网络是一种建立在时间顺序基础上的递归神经网络结构，它充分考虑了过往的时间序列信息，它不仅在当前输入时刻考虑了前面时刻的所有输入，还运用了权值共享理念，能够捕捉时间序列的整体信息。循环神经网络模型的主要结构如图 3.1 所示：

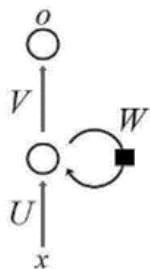


图 3.1 RNN 神经网络模型结构图

图 3.1 中一个圆圈代表隐藏层的神经单元，箭头的方向表示数据的循环方向和参数的更新方向，隐藏层神经单元的输出不仅会经过激活函数的处理形成当前时刻的隐藏状态输出，还会经过处理流向下一个时刻的隐藏层神经单元，进而和下一时刻的输入一起对于各个权重值产生影响，在迭代的过程之中，每一个时刻的权重是共享的，所以这种循环神经网络的构造可以实现对于时间序列的“记忆”。多个时刻下，循环神经网络的展开图如图 3.2 所示：

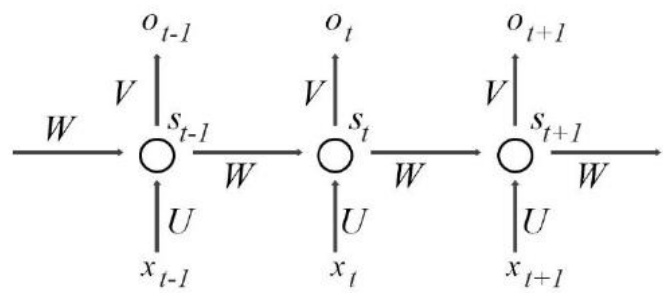


图 3.2 RNN 神经网络模型多个时间步展开图

图 3.2 是按照时间序列呈现的，时间序列的当前时刻记为 t ，那么 $t-1$ 代表的是当前时刻的前一时刻， $t+1$ 表示当前时刻的后一时刻， x 表示输入的特征变量数据， s_t 表示循

环神经网络在当前时刻的隐藏层神经元的状态值。在当前时刻，将当前时刻的特征变量输入到循环神经网络之中：

$$s_t = f(Ws_{t-1} + Ux_t) \quad (3.1)$$

其中 W 表示隐藏层神经单元的权重值， U 表示当前时刻的输入特征变量的权重， f 为预先设置的激活函数，激活函数的设置是为了让模型能够捕捉非线性变化，在实际操作过程中，激活函数的形式可以选用 \tanh ， ReLU ， sigmoid 等。

o_t 为当前时刻的输出：

$$o_t = g(Vs_t) \quad (3.2)$$

V 表示输出阶段的隐藏层神经单元的权重， g 为激活函数，在循环神经网络模型中，输出部分的激活函数通常是使用 \tanh 。

训练循环神经网络模型时，需要通过某种算法根据误差对于各神经单元的权重进行更新，在 BP 神经网络上，使用的算法是 BP 算法，它根据误差的梯度值进行反向传播更新各神经网络单元的权重。在循环神经网络模型上，依据误差更新权重的核心算法是时序反向传播算法，简称为 BPTT 算法。BPTT 算法在 BP 算法的基础上增加了时刻信息，因此被称为时序反向传播算法。时序反向传播算法的核心与 BP 算法是一样的，它们都是应用梯度下降法将误差进行反向传播来更新每一个神经网络单元的权重值，所以，BPTT 算法也是属于梯度下降法的一种，因此，通过合理的方法计算每一个神经网络单元的权重值的梯度非常重要。

因为参数 V 所在的神经单元实际上并不参与到时刻的计算之中，参数 V 的训练仅仅需要当前时刻的数据，与之前时刻的数据无关。因此，计算参数 V 的偏导数公式如式 3.3 到式 3.5 所示：

$$\frac{\partial L^{(t)}}{\partial V} = \frac{\partial L^{(t)}}{\partial o^{(t)}} \cdot \frac{\partial o^{(t)}}{\partial V} \quad (3.3)$$

$$L = \sum_{t=1}^n L^{(t)} \quad (3.4)$$

$$\frac{\partial L}{\partial V} = \sum_{t=1}^n \frac{\partial L^{(t)}}{\partial o^{(t)}} \cdot \frac{\partial o^{(t)}}{\partial V} \quad (3.5)$$

而参数 W 与参数 U 的训练则要利用到前面时刻的数据，这一点和 BP 算法不同，因为 BP 算法与时间无关，不考虑时刻的前后顺序对于权重值产生的影响。

W 与 U 的训练过程中，需要计算参数 W 与 U 的偏导数，不同于单纯的 BP 算法，需要按照时间来运用链式法则，所以计算的过程比较繁琐。

从简化角度来考虑，假设 $t=3$ ，即只有 3 个时刻参与模型的训练。则在最后一个时刻误差函数 L 对参数 W 的偏导数计算公式如式 3.6 所示：

$$\frac{\partial L^{(3)}}{\partial W} = \frac{\partial L^{(3)}}{\partial o^{(3)}} \cdot \frac{\partial o^{(3)}}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial W} + \frac{\partial L^{(3)}}{\partial o^{(3)}} \cdot \frac{\partial o^{(3)}}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial h^{(2)}} \cdot \frac{\partial h^{(2)}}{\partial W} + \frac{\partial L^{(3)}}{\partial o^{(3)}} \cdot \frac{\partial o^{(3)}}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial h^{(1)}} \cdot \frac{\partial h^{(1)}}{\partial W} \quad (3.6)$$

同样，误差函数 L 在最后一个时刻对参数 U 的偏导数计算公式如式 3.7 所示：

$$\frac{\partial L^{(3)}}{\partial U} = \frac{\partial L^{(3)}}{\partial o^{(3)}} \cdot \frac{\partial o^{(3)}}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial U} + \frac{\partial L^{(3)}}{\partial o^{(3)}} \cdot \frac{\partial o^{(3)}}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial h^{(2)}} \cdot \frac{\partial h^{(2)}}{\partial U} + \frac{\partial L^{(3)}}{\partial o^{(3)}} \cdot \frac{\partial o^{(3)}}{\partial h^{(3)}} \cdot \frac{\partial h^{(3)}}{\partial h^{(1)}} \cdot \frac{\partial h^{(1)}}{\partial U} \quad (3.7)$$

式 3.6 和式 3.7 表明，计算当前时刻的误差函数 Loss 对参数 W 和 U 的偏导数时，这一偏导数包含了之前所有时刻的误差函数 Loss 对参数 W 和 U 的偏导数。因为在误差的反向传播过程中，每一个时刻，每一个权重变量的偏导数都会参与到这一个过程中，所以当前时刻误差函数 Loss 对参数 W 与 U 的偏导数的值会非常复杂，这会导致整个误差函数 Loss 对参数 W 与 U 的偏导数计算量呈几何级数增加。经过总结规律，则误差函数 Loss 对参数 W 与 U 的偏导数的计算如式 3.8 和式 3.9 所示：

$$\frac{\partial L^{(t)}}{\partial W} = \sum_{k=0}^t \frac{\partial L^{(t)}}{\partial o^{(t)}} \cdot \frac{\partial o^{(t)}}{\partial h^{(t)}} \cdot \left(\prod_{j=k+1}^t \frac{\partial h^{(j)}}{\partial h^{(j-1)}} \right) \cdot \frac{\partial h^{(k)}}{\partial W} \quad (3.8)$$

$$\frac{\partial L^{(t)}}{\partial U} = \sum_{k=0}^t \frac{\partial L^{(t)}}{\partial o^{(t)}} \cdot \frac{\partial o^{(t)}}{\partial h^{(t)}} \cdot \left(\prod_{j=k+1}^t \frac{\partial h^{(j)}}{\partial h^{(j-1)}} \right) \cdot \frac{\partial h^{(k)}}{\partial U} \quad (3.9)$$

现在，将激活函数加入到误差反向传播的过程之中，计算公式如式 3.10 和式 3.11 所示：

$$\prod_{j=k+1}^t \frac{\partial h^{(j)}}{\partial h^{(j-1)}} = \prod_{j=k+1}^t \tanh^j \cdot W_s \quad (3.10)$$

$$\prod_{j=k+1}^t \frac{\partial h^{(j)}}{\partial h^{(j-1)}} = \prod_{j=k+1}^t \text{sigmoid}^j \cdot W_s \quad (3.11)$$

根据式 3.10 和式 3.11，激活函数的导数出现了按照时刻长度进行累乘的情况，这将导致模型出现“梯度爆炸”与“梯度消失”的不良后果。

下面以常见的 sigmoid 激活函数为例，详细介绍“梯度消失”与“梯度爆炸”产生的原因。sigmoid 激活函数及其导数的图像如图 3.3 和图 3.4 所示，sigmoid 激活函数将模型的输出控制在一定范围里，从图 3.4 可看出，sigmoid 函数的导数区间为(0, 0.25)之间。

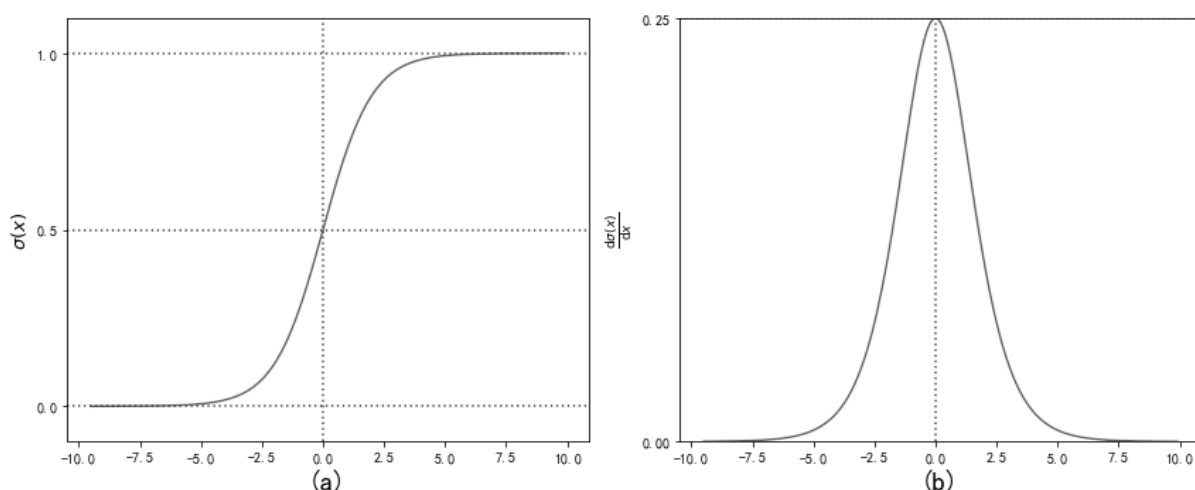


图 3.3 sigmoid 激活函数及其导数图

根据 BPTT 算法, 计算误差函数对各参数的偏导数是一个累乘的形式, 如果选定 sigmoid 函数为激活函数, 由于 sigmoid 函数的导数范围在(0, 0.25)之间, 会导致累乘的结果随着时刻的加长和网络深度的加深变得越来越接近于 0。这就是所谓的“梯度消失”现象。这一现象会造成时序反向传播算法中的核心要素梯度消失, 梯度消失后参数随着误差的负梯度方向更新将变为不可能, 这时, 参数所在的神经网络单元会变成没有作用的 dropout 状态, 实际上退出了迭代过程。所以实际应用中, 一旦时间步过长或者隐藏层的层数过多, 循环神经网络很容易陷入梯度消失现象, 导致效果迅速恶化。

BPTT 算法的应用使得循环神经网络针对于时间序列形式的数据有着独特的优势, 因为相较于对于忽略时间序列数据先后性的非循环神经网络来说, 循环神经网络能够捕捉时间序列数据各时间之间的非线性联系。然而, 梯度消失问题的产生使得循环神经网络的拟合和预测效果受到了很大影响。所以, 需要针对这个问题进行具体改进, 目前主流的改进办法有两种: 一是选择使用导数范围更加大的激活函数; 二是改造神经网络的单元构造。

就第一种改进方法而言, 线性整流 (ReLU) 函数成为了激活函数的首选, 线性整流函数的自变量小于零时, 它的导数始终为 0, 自变量大于零时, 它的导数始终为 1, “梯度消失”现象得到了缓解。但是由于线性整流函数的导数在自变量大于 0 时恒为 1 而不变, 又会导致另一种称作“梯度爆炸”现象的发生, 并且 ReLU 函数的负轴方向的导数恒为 0, 有可能会造成神经元权重不更新的情况。

循环神经网络的主要特点是能够将时间序列之间的固有联系通过递归的信息全部考虑在内。然而, 由于“梯度爆炸”和“梯度消失”现象的存在, 循环神经网络不能学习到时间序列之间的长时依赖关系, 针对这种问题的第二种改进方法被提出, 也就是 LSTM 神经网络。

3.2 LSTM 神经网络

长短期记忆神经网络（Long Short Term Memory Network, LSTM）由 Hochreiter 和 Schmidhuber（1997）提出，它的基础是循环神经网络，但是与传统循环神经网络不同的是，它的神经网络单元结构更加复杂，因此，它具有处理比较长的时间序列的能力，能够在一定程度上缓解“梯度爆炸”和“梯度消失”的问题，“长期依赖”问题得以部分解决。

由上一小节可知，传统循环神经网络的隐藏层神经单元只有一个隐藏状态 h ，由于循环神经网络的递归特性，所以在输入序列变长时，循环神经网络的效果会迅速恶化。所以，解决这个问题的思路就是从神经单元的内在结构入手，LSTM 神经网络在原有基础上新增加了一个细胞单元状态 c ，这样 h 储存短期状态， c 储存长期状态，从而实现对于时间序列短期和长期状态的保存，传统循环神经网络与 LSTM 神经网络的神经单元对比如图 3.5 所示：

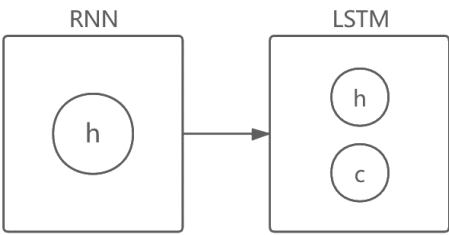


图 3.5 传统循环神经网络与 LSTM 神经单元对比图

为了更进一步的了解 LSTM 神经网络的具体结构，将图 3.5 按照时刻方向进行展开，如图 3.6 所示：

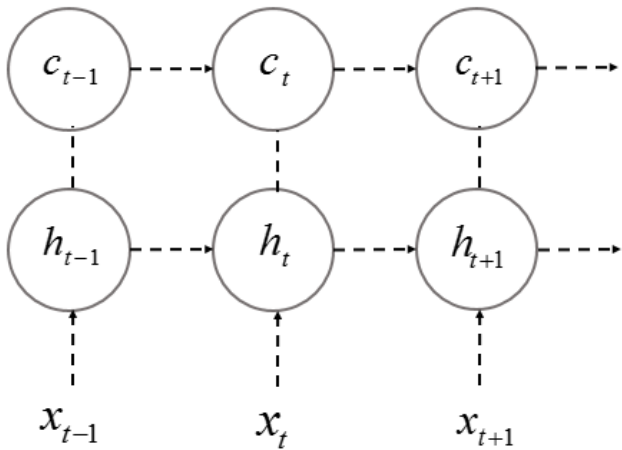


图 3.6 LSTM 神经网络按时间展开图

假设当前时刻为 t ，在当前时刻，有三种信息输入到 LSTM 神经网络的神经单元之中，它们分别是当前时刻的输入特征变量 x_t 、前一个时刻 LSTM 神经网络的隐藏状态输出值 h_{t-1} 和前一个时刻的细胞单元状态 c_{t-1} 。在当前时刻，LSTM 神经网络有两种信息输出，它们是当前时刻 LSTM 神经网络的隐藏状态输出值 h_t 和当前时刻的细胞单元状态 c_t 。

传统循环神经网络的隐藏层神经单元相当于一个简单的全连接层，因此无法储存长期状态，而 LSTM 神经网络的细胞单元 c 的特殊设计，使得这一点成为了可能。LSTM 神经网络通过三个门来控制流入和流出细胞单元 c 的信息。它们分别负责控制长期状态 c 的存留值，控制流入的长期状态 c 中的短期状态值以及长期状态 c 输出到隐藏层神经单元隐藏状态 h 的值。针对细胞单元 c 的控制如图 3.7 所示：

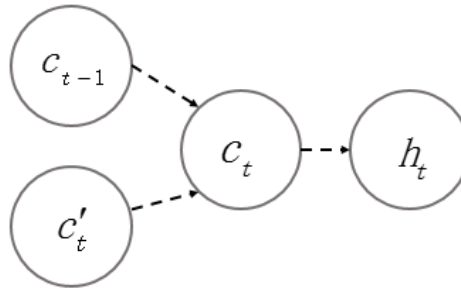


图 3.7 细胞单元 c 控制示意图

上述的门概念实际上就是一个全连接层，输入是一个向量，通过权重项和偏置项的处理，再经过 sigmoid 激活函数，其输出变为一个 0 到 1 之间的实数。门的输出公式如式 3.12 所示：

$$g(x) = \sigma(Wx + b) \quad (3.12)$$

其中， W 为门的权重项， b 为偏置项， x 为门的输入， σ 为 sigmoid 激活函数。

下面将详细介绍这三个门的具体构造，遗忘门 (forget gate) 的输入是前一个时刻的隐藏层神经单元状态 h_{t-1} 和当前时刻的输入特征变量 x_t ，输出是一个 0 到 1 之间的实数。输入门 (input gate) 和输出门 (output gate) 的原理类似，只不过是权重项和偏置项各不相同，在此不再赘述。三个门的计算如式 3.13 到式 3.15 所示：

$$\text{遗忘门: } f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.13)$$

$$\text{输入门: } i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.14)$$

$$\text{输出门: } o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.15)$$

其中 W_f, W_i, W_o 为各个门的权重向量, $[h_{t-1}, x_t]$ 表示向量之间的拼接, 便于同时学习两者的信息, b_f, b_i, b_o 为各个门的偏置项, σ 是 sigmoid 函数。三个门的输出实际上是决定信息的保存程度, 因此还需要计算将要保存到当前时刻的细胞单元状态的信息 c'_t , 公式如式 3.16 所示:

$$c'_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.16)$$

当前时刻的细胞单元状态 c_t 需要经过如下处理: 将前一个时刻的细胞单元状态 c_{t-1} 按元素和遗忘门的输出 f_t 相乘, 再用将要保存到当前时刻的细胞单元状态的信息 c'_t 按元素输入门 i_t 相乘, 最终将两部分求和, 公式如式 3.17 所示:

$$c_t = f_t * c_{t-1} + i_t * c'_t \quad (3.17)$$

这样, LSTM 神经网络将短期状态 c'_t 和长期状态 c_{t-1} 通过三个门的控制综合起来, 形成了保存在当前时刻细胞单元内的状态 c_t 。当前时刻 LSTM 神经网络最终的隐藏状态输出通过输出门决定, 公式如式 3.18 所示:

$$h_t = o_t * \tanh(c_t) \quad (3.18)$$

最终, LSTM 神经网络的全部结构如图 3.8 所示:

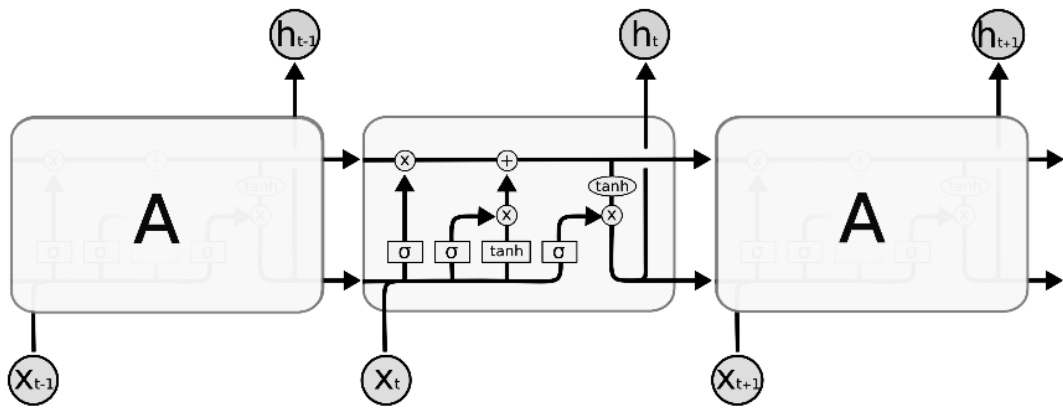


图 3.8 LSTM 全部结构图

LSTM 训练算法依然为前述的 BPTT 算法, 主要步骤如下所述:

首先沿着时刻方向计算每个神经网络单元的输出值, 在 LSTM 神经网络中, 它们是 f_t 、 i_t 、 o_t 、 c_t 、 h_t , 然后再计算每个神经网络单元的误差值, 将误差沿时刻的反方向进行传递, 同时, 如果存在多层隐藏层, 则将误差将上一层隐藏层进行传递, 最终根据相应的误差值计算对应权重变量的梯度, 并且将权重沿梯度的反方向进行更新。

3.3 注意力机制

注意力机制（Attention Mechanism）起源于科学家们对于人类大脑的研究，是神经科学之中研究大脑活动机制而产生的一个新兴的理论。注意力机制在被提出后，首先被应用于图像识别领域。Google 公司的 DeepMind 团队在循环神经网络模型上率先使用了该机制用于图像分类任务，并取得了很好的性能。之后，Bahdanau 等（2017）首次将注意力机制应用到自然语言处理领域中，他们在自然语言处理上利用注意力机制实现了原文和译文之间的注意力关联，改善了原有自然语言处理模型的效果。随后，注意力机制被广泛应用于自然语言处理领域，与卷积神经网络和循环神经网络都进行了融合。Google 公司在自家的机器翻译系统中大量使用注意力机制来学习原文和译文间的关系。目前广泛在深度学习模型中应用的注意力机制实际上是通过已有的信息学习一系列注意力权重系数。通过这一机制，模型有能力从众多信息中抽取出对预测变量影响最大，更为关键的信息。

深度学习中注意力机制的原理和人类的视觉注意力机制相似。当人类看到某个物体时，通常都会把注意力放在对于自己有价值或者潜意识认为关键的信息上，而对于其他的信息关注较少，这种注意力机制能够使得人们在短时间内充分利用自己宝贵的注意力资源来获取最大的信息收益，提高大脑处理海量信息的效率。在本文的金融时间序列预测领域同样如此，用来进行预测的时间序列输入特征变量可能非常长，最终的预测值可能不会都依赖于所有的输入特征变量，而是有所偏重，注意力机制的应用能够学习到时间序列的内在信息，分配不同的注意力权重，对于和预测变量的预测效果相关性更高的信息施加较高的权重，提取出重要信息，从而提高了模型的预测效果。

在不同的应用场景下，注意力机制施加于具体模型的方式不尽相同。最基本的注意力机制的思想如下：

循环神经网络模型处理时间序列数据的过程实际上是一个马尔科夫过程，当前时间步的隐藏状态只和前一时间步的隐藏状态有关。所以我们可以知道虽然很早以前的隐藏状态可以对当前时间步的隐藏状态产生影响，但实际上这种影响随着网络的迭代传递已经微乎其微，即神经网络很难提取到很早之前的信息的特征，而注意力机制可以有效地解决这个问题，注意力机制会把每一个隐藏状态保存下来，生成一个序列，当训练时会把最终输出和这个序列进行关联，根据和不同隐藏状态之间的关联性进行选择性的学习。其表达式如式 3.19 到式 3.21 所示：

$$e_t = a(h_t, h_t) \quad (3.19)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t=1}^T \exp(e_t)} \quad (3.20)$$

$$context = \sum_{t=1}^T \alpha_t h_t \quad (3.21)$$

其中, e_t 是时间步 t 上隐藏层输出向量 h_t 对应的注意力概率分布值, α_t 为 e_t 经过 softmax 处理后的值, context 是经过注意力权重加权之后的综合变量值, a 是注意力机制的载体函数。

所以注意力机制本质上是一种相似度的度量, 如果某一时刻的隐藏状态和所选定的比较值越相似, 那么该隐藏状态所占的权重就会越大, 对模型的贡献也越多, 即注意力机制有效地解决了时间长度对模型的限制, 每一时刻的隐藏状态按照一定的权重形成综合变量参与到模型之中, 长时依赖问题得以缓解, 由此可以获得更加准确的输出结果。

3.4 基于注意力机制和 LSTM 神经网络的时序预测模型

本文将 LSTM 神经网络与注意力机制相结合, 构建了一个融合两者优点的时序预测模型 (Att-LSTM)。

人脑处理时间序列的过程可以分为两个部分, 第一部分是获取用于预测的时间序列的历史数据, 这其中不仅包含了本身需要预测的时间序列变量的历史信息, 还包含了参与到预测过程中的其他特征变量的历史信息。人脑在接收到这些数据后, 对其进行审视, 将更多的注意力分配在对所需预测的时间序列变量影响较大的特征变量及其时间步上。第二部分是已经经过第一部分人脑思维处理的时间序列的历史数据, 进行综合判断, 在这一综合判断的过程中, 将所得到的整个综合信息再次与历史数据进行比较, 得出最终的预测结果。

在金融时间序列预测领域, 最常见的是对于股票时间序列的预测, 下面将以它为例具体说明人脑的思考过程, 影响股票价格的因素有很多, 投资者在预测股票时, 会关注很多方面的信息, 例如量价关系, 技术指标和板块轮动情况等等。投资者会根据每一支股票的具体情况具体分析, 决定将注意力更多地分配在某些重要的信息上, 而在某些不重要的信息上则分配较少的注意力。具体来说, 例如预测的标的股票是银行股, 由于它的股价波动比较小, 技术指标变化相对较为平稳, 投资者就可能更多关注银行板块的整体情况, 进而预测其未来走势; 当预测的标的股票是芯片股时, 由于它的股价波动比较剧烈, 成交量和其他技术指标都波动很大, 投资者就会更多关注其自身的量价关系以及技术指标等, 更多将它们用来作为预测后续走势的依据, 对于其他的信息关注则比较少。

综上, 本文将 LSTM 神经网络和注意力机制相结合用于金融时间序列预测。从本质上来看, 不管是 LSTM 神经网络, 还是注意力机制, 它们都有共同的脑科学基础, 核心都是通过某种方式模拟人脑的思维过程。对于 LSTM 神经网络来说, 它参考了人脑处理信息的一种特殊过程: 人脑在接收到以序贯形式呈现的信息时, 会对某些不重要的信息进行有选择性的遗忘, 保留重要的信息, 与此同时, 人脑会对于序贯信息进行综合处理, 往往会更加关心与当前时间点相关的一些信息。对于注意力机制来说, 它是根据人在思考问题时, 总是根据事物之间的逻辑联系来分配所给予的注意力大小, 也就是把更多的注意力分配在问题的重要部分, 而问题的不重要部分则给予较少的注意力。

上面分别介绍了 LSTM 神经网络和注意力机制的脑科学基础, 可以得出结论: 它们在处理时间序列问题时, 都有各自的优势, LSTM 神经网络善于捕捉时间序列各个时间步之间的固有联系, 将人脑处理时间序列信息的过程抽象成一种对于序贯型信息的特殊处理机制, 注意力机制则可以模拟人脑的注意力分配过程, 使得注意力更多地分配在重要的时间步和其背后的隐含信息上, 而对于不重要的时间步及其背后的隐含信息则分配较少的注意力。如果能在时间序列预测过程中, 将两者的优势结合起来, 在理论上将会改善时间序列的预测效果。

下面将介绍本文所构建的 Att-LSTM 的详细结构。总体来说, 模型可以分为两大部分: 编码部分和解码部分。两个部分的主体结构均是一个双层 LSTM 神经网络。根据实践, 仅仅是训练单层 LSTM 神经网络就需要耗费大量的时间, 之所以使用双层 LSTM 神经网络的原因是, 在时间序列预测中, 往往需要输入非常多的特征变量进行预测, 单层 LSTM 神经网络无法充分识别和学习这些高维度输入特征变量所包含的信息, 使用双层 LSTM 神经网络虽然会耗费更多的时间和运算资源, 但是却能够有效提升模型识别输入特征变量所包含的信息的能力。编码部分和解码部分都融合入了以双线性模式作为载体的注意力机制。在编码部分融入的注意力机制主要模拟人脑处理时间序列的第一部分过程, 它能够更多的注意力分配在对所需预测的时间序列变量影响较大的特征变量及其时间步上。在解码部分融入的注意力机制主要模拟人脑处理时间序列的第二部分过程, 这一部分对于第一部分处理完毕的信息, 进行进一步综合判断, 得出最终的预测结果。

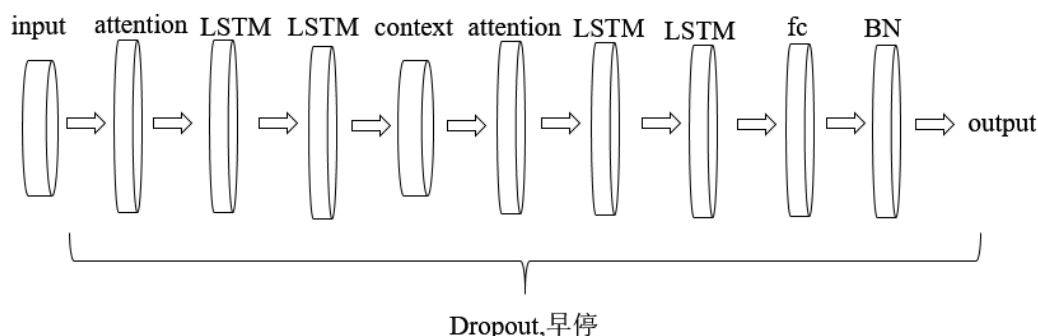


图 3.9 模型整体结构图

如图 3.9 所示, 以综合变量 `context` 为界, 综合变量前为模型的编码部分, 综合变量后为模型的解码部分, 两部分的主体结构均为一个双层 LSTM 神经网络, 在主体结构前均融入了以双线性模式作为载体的注意力机制。根据前面的分析, 这样的结构可以将 LSTM 神经网络和注意力机制的优势相结合, 有选择性地获取输入到模型的各时间维度特征变量中的相对重要的信息, 并且能够寻找时间序列间的联系, 进行综合处理, 形成最终的预测结果。

下面将详细介绍编码部分和解码部分的详细构造:

将输入模型之中的时间序列数据记为:

$$X = \begin{pmatrix} x_1^1 & \cdots & x_T^1 \\ \vdots & \ddots & \vdots \\ x_1^N & \cdots & x_T^N \end{pmatrix} \quad (3.22)$$

T 为用来预测时间序列数据的窗口期的长度, 即用过去 T 个时刻的历史数据来预测。 N 为用来预测时间序列数据的特征变量的个数。

编码部分的注意力机制是以双线性模式为载体, 其主要的构造如下:

$$e_t = h_{t-1} W_e x_t \quad (3.23)$$

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{k=1}^N \exp(e_t^k)} \quad (3.24)$$

其中 e_t 为注意力原始权重值, h_{t-1} 是前一个时刻的隐藏层, x_t 为当前时刻的特征变量序列, W_e 是需要通过对整个模型进行训练进而得到的双线性模式注意力机制的关键参数。式 3.24 实际上是一个在每一个输入时刻都会针对 e_t 施加的 softmax 变换, 它的存在保证了在每一个输入时刻, 实际运用在模型中的编码部分注意力权重的和为 1, 从而达到对于输入序列进行加权的目的:

$$\tilde{X} = \begin{pmatrix} \alpha_1^1 x_1^1 & \cdots & \alpha_T^1 x_T^1 \\ \vdots & \ddots & \vdots \\ \alpha_1^N x_1^N & \cdots & \alpha_T^N x_T^N \end{pmatrix} \quad (3.25)$$

其中, $\alpha_1^1 + \alpha_1^2 + \dots + \alpha_1^{N-1} + \alpha_1^N = 1$, 按照时刻以此类推。

将经过编码阶段的双线性模式注意力机制处理过的输入时间序列 \tilde{X} 作为新的双层 LSTM 神经网络输入:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_s \cdot [h_{t-1}, \tilde{x}_t] + b_s) \quad (3.26)$$

$$h_t = o_t * \tanh(c_t) \quad (3.27)$$

其中, c_t 代表当前时刻的细胞单元信息, f_t 代表遗忘门神经单元, i_t 代表输入门神经单元, h_{t-1} 代表前一个时刻的隐藏层信息, \tilde{x}_t 是当前时刻输入到 LSTM 神经网络中的特征变量, W_s 为上一时刻的隐藏层信息和当前时刻输入到 LSTM 神经网络中的特征变量横向拼接成的新向量的偏置项, b_s 为偏置常数项, o_t 为输出门神经单元。

需要说明的是, 上述两个偏置项均需要通过 BPTT 算法和 Adam 优化方法利用训练集的输入输出数据进行不断训练得出, 在训练的过程中, 将加入 Dropout 方法以及使用早停策略, 来防止过拟合以及提升模型的可靠性。

输入模型之中的时间序列数据 X 按照时间顺序 (t_1, t_2, \dots, t_T) , 依次通过双层的 LSTM 神经网络, 可以得到一组按照时间顺序排列的隐藏状态, $H = (h_1, h_2, \dots, h_T)$ 。每一个隐藏状态向量的长度均等于事先设定好的隐藏层的神经元数目。

解码部分主要是针对在编码部分所得到的时间序列的隐含信息, 进行综合处理, 形成综合变量, 在每一个输入时刻, 利用融合进的以双线性模式为载体的注意力机制, 对于原有综合变量进行加权, 得到新的综合变量, 再结合历史待预测变量的信息, 输入另一个双层 LSTM 神经网络, 将 LSTM 神经网络的输出结果通过一个全连接层, 以得出最终 $T+1$ 时刻的预测结果。

在 t 时刻, 解码部分的以双线性模式为载体的注意力机制可以表示为:

$$e'_t = h_{t-1}^T W'_e h_t \quad (3.28)$$

其中, e'_t 为注意力原始权重值, h_{t-1} 为解码部分双层 LSTM 神经网络在前一时刻的隐藏状态值, W'_e 是需要通过对整个模型进行训练进而得到的解码部分双线性模式注意力机制的关键参数。

与前面的做法一致, softmax 变换的存在保证了在每一个输入时刻的所有编码部分隐藏状态的注意力权重的和为 1:

$$\beta_t^i = \frac{\exp(e'_t)}{\sum_{i=1}^T \exp(e'_t)} \quad (3.29)$$

β_t^i 代表的是在 t 时刻下编码部分隐藏状态 h_i 的权重, 利用这个权重对于编码部分所有隐藏状态进行加权, 可以得到在 t 时刻的综合变量 $context_t$ 表示为:

$$context_t = \sum_{i=1}^T \beta_t^i h_i \quad (3.30)$$

在每一个时刻, 将综合变量与编码部分的双层 LSTM 神经网络输出的隐藏状态 h_t 进行拼接, 将其输入到解码部分的双层 LSTM 神经网络, 基本的结构与编码部分类似, 再

次不再赘述。最终我们可以得到一组解码部分双层 LSTM 神经网络的隐藏状态：

$$H' = (h'_1, h'_2, \dots, h'_r)。$$

最终将这组隐藏状态，通过一个全连接层，得到最终 $T+1$ 时刻的预测结果。

需要额外说明的是，在整个 Att-LSTM 模型的训练过程中，使用 Adam 对于学习率进行自适应调整，对于所有神经网络层均使用了 Dropout 方法来防止过拟合，并且还在每一个训练批次中加入了批次标准化层，以及使用了早停策略，来最大限度提升模型的可靠性。

3.5 LSTM 神经网络的优化策略

LSTM 神经网络的效果很大程度上依赖于如何优化训练的过程，它的重点是以什么方法来优化梯度下降法之中的学习率，如何处理训练批次间的关系以及处理过拟合问题。在本文的模型构造和训练过程中，主要使用以下三种优化策略：

3.5.1 自适应动量估计法

自适应动量估计法，英文叫做 Adam (Adaptive Momentum Estimation)。它是梯度下降法的一种，能够根据迭代得出的信息对于模型的学习率进行自适应地调整，使得每一维参数都有不同的学习率。它利用梯度的一阶矩估计和二阶矩估计对于每一维度参数的学习率进行动态调整。这种方法的优点主要在于经过动态调整之后，每一次迭代的学习率既不保持不变，又不大幅变动，从而使得参数的更新较为平稳。其主要公式如式 3.31 到式 3.35 所示：

首先，对于样本 i ，假设需要学习的参数为 w ，其误差函数为 $C_i(w)$ 。在 n 个样本构成的训练集上，其整体误差函数为：

$$C(w) = \sum_{i=1}^n C_i(w) \quad (3.31)$$

计算一阶动量 m_t ：

$$m_t = \alpha m_{t-1} + (1 - \alpha) \nabla C(w) \quad (3.32)$$

计算二阶动量 v_t ：

$$v_t = \beta v_{t-1} + (1 - \beta) \nabla C(w)^2 \quad (3.33)$$

为了在前几次迭代迅速降低误差，使用 t 次幂增大一阶动量 m_t 和二阶动量 v_t 的值：

$$\hat{m}_t = \frac{m_t}{1 - \alpha^t}, \hat{v}_t = \frac{v_t}{1 - \beta^t} \quad (3.34)$$

最终，参数按照负梯度方向进行更新：

$$w_{t+1} = w_t - \eta \frac{1}{\sqrt{\hat{v}_t}} \hat{m}_t \quad (3.35)$$

3.5.2 批次标准化层

在神经网络的训练过程中，需要有海量数据作为输入以确保模型的效果，但是随着输入的特征变量越来越多，训练速度就会越来越慢。在理想状态下，最好是每一次将一个样本输入模型进行训练，但是训练速度会非常慢，目前广泛采用的方法是预先设定一个训练批次数，将原始数据分为很多个批次分别输入模型进行训练，但是由于基于深度学习理念构建的神经网络包含很多隐藏层神经单元，如果根据 BPTT 算法对于每一层的参数进行更新就会导致后续隐藏层需要不断去重新适应其前面隐藏层的参数更新，这种情况既影响了模型的训练速度，又会引起过拟合的问题。

目前针对上述问题普遍采用的方法是添加批次标准化层，简称为 BN（Batch Normalization）层。

批次标准化层针对单个神经网络单元进行标准化。利用输入到神经网络中一个最小批次的的数据来计算该神经元 x_i 的均值与方差，计算公式如式 3.36 到式 3.37 所示：

$$\mu_i = \frac{1}{M} \sum x_i \quad (3.36)$$

$$\sigma_i = \sqrt{\frac{1}{M} (x_i - \mu_i)^2 + \epsilon} \quad (3.37)$$

其中 M 代表最小训练批次的数目。这样每个神经网络单元的输入经过平移缩放，最终满足均值为 0，方差为 1 的标准正态分布。为了保证模型的表达能力不因规范化而下降，对于输出进行全连接层处理，得到均值为 b ，方差为 g^2 的分布，公式如式 3.38 所示：

$$x' = g \cdot \frac{x - \mu}{\sigma} + b \quad (3.38)$$

批次标准化层独立地标准化每一个输入维度 x_i ，其标准化的参数是一个最小批次数据的一阶统计量和二阶统计量。这就需要每一个最小批次数据的分布和整体数据的分布类似。最小批次数据的分布差距对批次标准化层的使用有一定影响：若分布差距较小，批次标准化层的效果可以发挥得较好，若分布差距较大，会降低批次标准化层的效果。因此，批次标准化层比较适用最小训练批次数目较大且其数据分布差距不大的应用场景。

3.5.3 早停策略

早停策略的本质是一种正则化方法，并且来防止过拟合方面表现得比很多其他正则化方法都要好。在 LSTM 神经网络的迭代过程中，输入特征变量数据以最小批次的形式被逐批次输入到模型之中，当所有批次的数据都被输入到模型之后，即对于整个训练数据集完成了一次遍历，所有需要通过训练学习的参数都会按照 BPTT 算法被更新一次，整个过程被称为一个迭代周期（epoch），epoch 的大小对于模型的整体效果十分重要，如果 epoch 过小，会导致模型无法充分学习到信息导致欠拟合，而 epoch 过大又会导致过拟合。

早停策略的示意图如图 3.10 所示，横轴是 epoch 的大小，纵轴是模型预测效果的衡量标准，纵轴的数值越大，代表模型的预测效果越好。可以看出，随着 epoch 的不断增大，训练集上的预测效果越来越好，而测试集上的预测效果则是一个开口向下的二次函数形式，随着 epoch 的增大，预测效果越来越好。因此，测试集衡量标准曲线的最高点所对应的 epoch 大小是理论上的最优值，当模型的迭代次是达到该值时，应该立即停止训练，因为这个时候，测试集的预测效果较好，而训练集也不差，如果 epoch 的大小超过了这个最优值，测试集的准确率会迅速下降，进入过拟合区域。

但是在实践中，训练集和测试集的预测效果不是像图 3.10 这样规则变化的，不一定只存在一个局部最高点，所以我们需要保存每一个 epoch 下的预测效果值，即使当前 epoch 的预测效果值低于上一个 epoch，依然进行迭代，仅当继续迭代的次数超过了预先设定的容忍值时，才停止迭代，本文将早停策略的容忍度设定为 15。

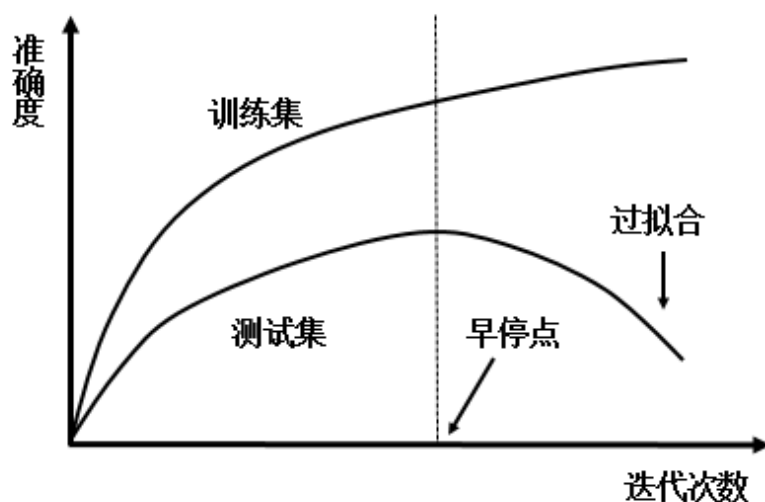


图 3.10 早停策略示意图

4 研究方案设计

利用神经网络模型进行时间序列预测，本质是利用过往的历史数据训练模型，使模型寻找出输入与输出之间的非线性关系，再根据两者之间的非线性映射关系，将待预测的输出所对应的特征变量代入模型，所得到的输出值即是预测结果。在整个预测研究过程中有以下几个要点：确定影响股票价格及涨跌走势的特征变量；合理选择研究所使用的数据集；数据集的预处理；超参数的选择以及网络优化策略确定；预测结果的评价方法。本章将对根据上述要点针对本文的研究方案设计进行详细介绍。

4.1 特征变量的构建

影响股票价格及涨跌走势的特征变量非常多。就本文研究而言，所用到的特征变量有三类：

第一类是基础指标，为开盘价（open）、收盘价（close）、最高价（high）、最低价（low）、成交量（vol）。开盘价，最高价，最低价，收盘价以及成交量是 K 线图的基本组成元素。本文所预测的变量就是股票收盘价，之所以将收盘价也纳入特征变量的范围的原因是：本文的研究领域属于时间序列预测，时间序列预测的显著特点就是当前时刻的值会受到前面时刻的影响，所以，模型的输入特征变量不仅是直接或者间接影响股票收盘价的相关因素，还包括收盘价本身。这一点不难理解，从预测时点看，过往的收盘价数据也是已知信息，也会对预测时点的收盘价产生影响。成交量与上述四个特征变量共同组成了 K 线图的基本形式，基于 K 线图的量价理论已经在投资界流行很久，非常多的投资者利用它来指导投资。以成交量为例，按照量价理论的观点，成交量增大，通常会推升股价，而成交量缩减则会对股价有负面影响。由此可见，各基础指标对于股票收盘价有间接的影响。

第二类是由基础指标衍生而来的 12 种技术指标，它们主要有指数平滑异同移动平均线（MACD）、乖离率（BIAS）、随机指标（KDJ）、布林带（BOLL）、相对强弱指标（RSI）等等。习惯使用技术分析理论指导自身投资行为的投资者往往会直接参考这一类数据，例如，按照技术分析理论，当 MACD 曲线从零轴以下上穿零轴到达零轴以上时，往往是中线的买入点。所以，习惯使用技术分析理论的投资者根据技术指标所发出的信号所进行的投资行为，会影响个股的供需关系，从而间接影响股票的收盘价，由此可以推出各种技术指标对于股票收盘价有着间接的影响。

第三类是相关股指标，由于股票市场是一个复杂的混沌系统，股价的涨跌不仅受到股票所代表的公司自身的经营好坏等内生因素的影响，同时还受到短期供求关系，市场

环境，同行业相关股票的走势表现等等诸多外生因素的影响。在诸多外生因素中，本文关注的是同行业相关股票的走势。在股票投资实践中，同行业相关股票的走势对于标的股票收盘价的影响主要有正向和负向两方面，正向影响指的是同行业股票上涨或者下跌时，标的股票通常会跟随，同步上涨或者下跌。负向影响指的是同行业股票通常与标的股票具有相互竞争的关系，所以当同行业股票上涨或者下跌时，标的股票会呈现相反的走势。基于上述分析，本文将相关股指标作为一个重要的特征变量，并且在下文的具体研究中验证其有效性。

4.1.1 基础指标及衍生技术指标的选取

关于基础指标的选取前文已经叙述，在此不再赘述，下面重点介绍衍生技术指标的选取。

股市投资中常用的技术指标主要有两大类：走势型指标和超买超卖型指标。走势型指标可以显示过去一个阶段股票价格变化的历史趋势，例如移动平均线、指数平滑异同移动平均线和布林带等等，超买超卖型指标的原理是利用股价有一定的正常波动范围的特征，找出股价的转折点，例如乖离率、随机指标和相对强弱指标等等。本文所选择的衍生技术指标如下：

（1）移动平均线 MA

移动平均线 MA 是一类最常用的技术指标，是指从当前时点前推 N 个交易日的收盘价的平均值，它可以帮助交易者判断和确认价格走势。计算公式如式 4.1 所示：

$$MA_N = \frac{\sum_{i=1}^N Close_i}{N} \quad (4.1)$$

当 N 值不同时，其所代表的移动平均线的含义也不同，本文选择短期，中期和长期三种移动平均线作为特征变量。N 值取 5 时为短期移动平均线，N 值取 20 时为中期移动平均线，当 N 值取 60 时为长期移动平均线。

（2）指数平滑异同移动平均线（MACD）

指数平滑异同移动平均线（MACD）是经过市场验证过的最有效的中线技术指标，同时也是目前使用最普遍的技术指标。指数平滑异同移动平均线是根据指数移动平均值演变而来，MACD 能够对市场行情及走势进行有效的掌控。本文选用计算 EMA 的参数值是目前市场最常用的 12 和 26，计算公式如式 4.2 到式 4.6 所示：

$$EMA_t(12) = \frac{11}{13} EMA_{t-1}(12) + \frac{2}{13} close_t \quad (4.2)$$

$$EMA_t(26) = \frac{25}{27} EMA_{t-1}(26) + \frac{2}{27} close_t \quad (4.3)$$

$$DIF_t = EMA_t(12) - EMA_t(26) \quad (4.4)$$

$$DEA_t = \frac{8}{10} DEA_{t-1} - \frac{2}{10} DIF_t \quad (4.5)$$

$$MACD_t = \frac{DIF_t - DEA_t}{2} \quad (4.6)$$

(3) 乖离率 BIAS

乖离率 BIAS 是一种用来衡量收盘价与移动平均线偏离程度的指标。移动平均线代表着当前市场的平均成本, 根据供求关系原理, 当股票价格与目前市场平均成本相差比较大的时候, 会出现向市场平均成本进行回归的走势。因为股价如果短期偏离程度过大, 会吸引短线投资者进场抄底或者卖出, 他们的买盘或者卖盘将会促使股价回归。所以乖离率是超买超卖指标的一种, 可以衡量人们在投资过程中的心理变化。本文中所用的参数 N 值是 12, 公式如式 4.7 所示:

$$BIAS_t = \frac{close_t - MA_N}{MA_N} * 100\% \quad (4.7)$$

(4) 随机指标 KDJ

随机指标 KDJ 的理论基础是数理统计学, 通过计算某一段周期内显示出来的最高价 (high)、最低价 (low) 和最后一个使用周期的收盘价 (close) 以及收盘价, 最高价和最低价三者的比例关系, 形成随机值 RSV, 最后再依据上一时刻的参数值求出所用的 K 值、D 值、J 值。本文中所用的参数 N 值是 12, 公式如式 4.8 到式 4.11 所示:

$$RSV_N = 100 * \frac{close_N - low_N}{high_N - low_N} \quad (4.8)$$

$$K_t = \frac{2}{3} K_{t-1} + \frac{1}{3} RSV_t \quad (4.9)$$

$$D_t = \frac{2}{3} D_{t-1} + \frac{1}{3} K_t \quad (4.10)$$

$$J_t = 3K_t - 2D_t \quad (4.11)$$

(5) 布林带 (BOLL)

布林带运用的是统计学之中的正态分布理论, 利用股价在过去一段时间的标准差和均值构建一条价格通道, 这条通道可以随着股价波动幅度的大小而自动调整, 对于考察未来一段时间的股价走势范围有着很大意义, 同时, 独特的上中下轨使得分析股价的支撑和压力易如反掌。本文中所用的参数 N 值是 12, K 值是 2, 它的公式如式 4.12 到式 4.16 所示:

$$MA_N = \frac{\sum_{i=1}^N Close_i}{N} \quad (4.12)$$

$$MD = \sqrt{\sum_{n=1}^{N-1} \frac{(close_n - MA_N)^2}{N}} \quad (4.13)$$

$$MP = MA_{N-1} \quad (4.14)$$

$$UP = MP + K * MD \quad (4.15)$$

$$DP = MP - K * MD \quad (4.16)$$

(6) 相对强弱指标 (RSI)

相对强弱指标 (RSI) 是利用某一段周期内股票价格涨跌幅比率来判断未来的股价走势, 它的计算方法是前 N 日收盘涨幅之和与 N 日收盘涨幅和跌幅之和的百分比值。这一指标表明了股价的变化强弱程度。本文取 N 值为 14, 公式如式 4.17 所示:

$$RSI_N = 100 * \frac{\sum_{i=1}^n up_i}{\sum_{i=1}^n up_i + \sum_{i=1}^n down_i} \quad (4.17)$$

4.1.2 相关股指标的选取

相关性通常是指多个数据之间关联程度的大小, 其中最简单直接的方法就是通过特定的方式来计算数据之间的“距离”, 然后通过“距离”值的大小来判断关联程度的大小。当多个数据在同一因素的作用下具有相同变化趋势时, 将之称为正相关, 反之若变化趋势相反时称为负相关, 如果对数据变化趋势并没有明显联系时则称为不相关。不同的股票走势的相关性直接通过股价来体现, 通常行业内股票的相关性大于行业间股票的相关性。

本文使用皮尔森相关系数这一传统指标来衡量股票之间的相关性, 具体的操作思路如下:

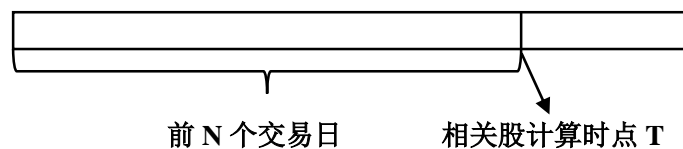


图 4.1 相关股指标选取示意图

如图 4.1 所示, 在相关股指标计算时点 T, 再向前推 N 个交易日, 利用这 N 个交易日的的数据计算市场中预测标的股票与同行业股票之间的皮尔森相关系数:

$$PCC(p, p_i) = \frac{\text{cov}(p, p_i)}{\sqrt{\text{Var}(p)\text{Var}(p_i)}} \tag{4.18}$$

其中 p, p_i 分别表示预测标的股票和某一只同行业股票的用来计算相关系数的收盘价序列。

在计算过程中如果遇到停牌的情况，导致 N 个交易日的数据不足，则再向前扩展数据，以满足计算相关系数的窗口期始终为 N 个交易日，本文研究中所选取的 N 的数值为 60。

表 4.1 特征变量表

特征类型	特征名称	特征简称
基础指标	开盘价	open
	最高价	high
	最低价	low
	收盘价	close
	成交量	vol
衍生技术指标	短期移动平均线	ma5
	中期移动平均线	ma20
	长期移动平均线	ma60
	指数平滑异同移动平均线	macd
	乖离率	bias
	随机指标 K	kdj_k
	随机指标 D	kdj_d
	随机指标 J	kdj_j
	布林带 MP	boll_mp
	布林带 UP	boll_up
	布林带 DP	boll_dp
	相对强弱指标	rsi
相关股指标	相关股收盘价	cp1,cp2,...,cp10

对于数据集中的每一个时点，均进行上述操作，就可以得出最终的相关股指标。由于预测标的股票和其同行业股票之间的相关关系可能在不同时期存在不同的特点，例如，在某一时期预测标的股票选择跟随同行业股票的走势，那这一时期其与同行业股票之间的相关关系为正相关；在另一时期，预测标的股票和同行业股票之间的走势呈现此消彼长的态势，这一时期其与同行业股票之间的相关关系为负相关。因此，本文在实际选取

相关股时，将计算出的相关系数的绝对值从高到低排序，选择相关系数的绝对值最高的十只股票的收盘价作为最终纳入特征变量的相关股指标。

最终，本文选择使用的所有特征变量小结如表 4.1 所示。

需要额外说明的是：本研究使用的价格数据采用前复权处理。前复权就是使现有的价位保持不变，将以前的价格进行缩减，使除权前的 K 线向下移动，最终将图形吻合，保持股价走势的连续性。前复权是将当前的价格做为基准进行复权，使均线系统更加顺畅，这一点对于分析的准确非常重要。

4.2 样本选择与数据说明

本研究的预测标的是富时中国 A50 指数的成分股。由于富时中国 A50 指数的成分股定期会进行调整，所以为了最大限度地保证研究的时效性，本研究中所选择的是富时中国 A50 指数在 2020 年 8 月 30 日的成分股。选择富时中国 A50 指数的成分股作为预测标的原因是这一指数包含了沪深两市市值最大的 50 只股票，这些股票的走势能够反映出中国股市的基本运行趋势，其受到人为操纵的可能性较低，同时富时中国 A50 指数作为海外市场交投最活跃的股指期货标的，预测其成分股价格变动，也具有现实意义。

本研究采用滑动窗口的方法构建数据集，滑动窗口的长度设置为 10 个交易日。数据集的起始日期是 2011 年 4 月 7 日，数据集的终止日期是 2020 年 5 月 29 日。为了保证预测数据的完整性，本研究删除了在 2011 年 1 月 1 日之后上市的样本。最终的预测标的数目为 39 只。从上市板块分布上来看，33 只在上海主板或者深圳主板上市，4 只在深圳中小板上市，2 只在深圳创业板上市，预测标的不仅包含了主板上市公司，而且包含了中小创上市公司，反映了我国资本市场分层次协同发展的特征。从行业分布上来看，银行业标的有 12 只，占比最大，以证券业和保险业为代表的非银金融业标的有 5 只，包括上述行业在内，传统行业的标的有 34 只，这表明目前我国资本市场的主体依然是传统行业公司，但值得一提的是，有 5 只新兴行业的上市公司也进入到了预测标的之中，它们是高端制造业的立讯精密，生物医药业的智飞生物，免税服务业的中国中免，安防服务业的海康威视，快递物流业的顺丰控股。从市值占比来看，预测标的的总体市值覆盖率达到 30% 以上，这表明所选择的 39 只预测标的能够在一定程度上代表我国资本市场的总体情况，对于这些标的进行预测具有现实意义。详细的预测标的列表列示在附录 B 中，此处不再赘述。本研究采用各预测标的股票子数据集的前 70% 作为训练集，后 30% 作为测试集，以测试各模型的预测效果。

本研究的基础指标数据来源是 tushare 开源数据库，衍生技术指标是通过 TA-Lib 库计算获得，相关股指标是通过 python 程序计算获得，富时中国 A50 指数的成分股数据取自英国富时公司的官方网站。

4.3 特征变量的无量纲化

本文在预测股票价格时使用的数据集中的各个特征变量具有不同的量纲，例如股票价格和成交量在数量级上差异非常大。根据神经网络理论，如果其中一些特征变量的输入数值过大，则经过 BPTT 算法计算出的神经网络单元的权重更新值也会很大，这样会导致它对于神经网络输出的影响远远大于那些数量级较小的特征变量的输出，从而使数量级比较小的特征变量的数据信息实际退出了神经网络的训练过程，所以在把数据输入到神经网络中进行训练之前，先对数据集进行无量纲化处理十分必要。

本文所采用的无量纲化方法为归一化。计算公式如下：

$$\bar{X}_i = \frac{X_i - \min X_i}{\max X_i - \min X_i}$$

(4.19)

其中 X_i 指数据集中的第 i 个特征变量。通过归一化处理，可以将所有特征变量的取值压缩到 0 和 1 之间。

4.4 超参数寻优及模型训练

本文将上一章所构建的 Att-LSTM 分为三个子模型，其中同时加入基础指标，衍生技术指标以及相关股指标的 Att-LSTM 作为主模型，加入基础指标和衍生技术指标的 Att-LSTM 和仅加入基础指标的 Att-LSTM 作为对比模型，同时本文也加入了 ARIMA，LSTM 和 BPNN 作为对比模型。综上，本文所利用的主模型及对比模型如表 4.2 所示：

表 4.2 各预测模型简表

模型名称	特征变量	模型类型
Att-LSTM(B-T-P)	基础指标，衍生技术指标，相关股指标	主模型
Att-LSTM(B-T)	基础指标，衍生技术指标	对比模型
Att-LSTM(B)	基础指标	对比模型
LSTM	基础指标，衍生技术指标，相关股指标	对比模型
BPNN	基础指标，衍生技术指标，相关股指标	对比模型
ARIMA	收盘价	对比模型

在上述模型中，除了 ARIMA 模型的超参数确定是通过 AIC 准则以外，其余各模型的超参数的确定均将通过网格寻优法完成，具体操作如下：

首先根据各超参数的经验设置，将待寻优的超参数的取值范围限制如表 4.3 所示：

从表 4.3 可以看出，超参数的组合有 12 种。针对某只具体预测标的的股票，对于每一种参数组合，都进行 10 次完整的训练过程，记录下这一种参数组合下的 10 次训练过程所得到的模型参数，然后将测试集的数据代入到训练得到的模型，计算这 10 次训练过

程的测试集上的 RMSE 评价指标的平均值。最终选择的超参数组合为 RMSE 的平均值最小的超参数组合。需要说明的是,在寻优过程中本文利用上一章所介绍的早停策略来最大限度地提高所得超参数结果的可靠性。

本文训练各神经网络模型所使用的深度学习框架为 PyTorch。针对每一只预测标的股票的训练流程如下:首先,利用超参数寻优的结果设置好各模型的超参数,将训练集的数据输入模型进行训练,利用早停策略来控制模型训练的迭代次数,在训练完成后,将测试集的数据代入模型,得出预测结果。

表 4.3 超参数取值表

超参数	简称	取值
隐藏层神经元数	hidden_size	{64, 128}
每一训练批次的数据数量	batch_size	{128, 256, 512}
学习率	lr	{0.001, 0.01}

4.5 预测结果的评价与检验

4.5.1 预测结果的评价标准

本文的预测研究的本质是回归型预测,回归型预测有很多衡量预测效果的评价指标,如平均绝对误差(mean absolute error, MAE)、均方误差(mean square error, MSE)、均值绝对比例误差(mean absolute percentage error, MAPE)、均值绝对相对误差(mean absolute relative error, MARE)、均方根误差(root mean square error, RMSE)等。本文选取平均绝对误差、均值绝对比例误差和均方根误差作为对于模型预测效果的评价指标。

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{x}(t) - x(t)| \quad (4.20)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{\hat{x}(t) - x(t)}{x(t)} \right| \quad (4.21)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{x}(t) - x(t))^2} \quad (4.22)$$

其中 $\hat{x}(t)$ 为预测值, $x(t)$ 为真实值。

三者中均值绝对比例误差的稳定性最好, MAE、RMSE 最为常用。MAE、MAPE 和 RMSE 越小,模型的预测精度越高。

4.5.2 模型间预测效果差异的显著性检验

本研究采用配对样本 T 检验和 Wilcoxon 符号秩和检验来检验预测效果评价指标的均值和中位数是否存在显著差异。

配对样本 T 检验的原假设为： $H_0, \text{Mean}_{\text{model1}} = \text{Mean}_{\text{model2}}$ 。

其 T 统计量的计算公式为：

$$t = \frac{\bar{d}}{S/\sqrt{N}} \quad (4.23)$$

其中 \bar{d} 为配对样本间的差异的平均值， S 为配对样本间差异的标准差， N 为样本数。该统计量满足自由度为 $N-1$ 的 T 分布。

Wilcoxon 符号秩和检验的原假设为： $H_0, \text{Median}_{\text{model1}} = \text{Median}_{\text{model2}}$ 。

当样本容量大于 25 时，Wilcoxon 符号秩和检验依赖的二项分布趋向于正态分布，因此有：

$$\mu = \frac{n(n+1)}{4} \quad (4.24)$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} \quad (4.25)$$

其 Z 统计量的计算公式为：

$$Z = \frac{d - \mu}{\sigma} \quad (4.26)$$

其中 d 为配对样本间的差异值，该统计量满足标准正态分布。

5 基于 Att-LSTM 的股价预测研究

5.1 描述性统计分析

将本研究所使用的整体数据集进行描述性统计，结果如表 5.1 所示：

表 5.1 描述性统计表

特征变量	样本数	平均值	标准差	最小值	中位数	最大值
close	85371	25.53	81.51	0.39	10.04	1366.60
open	85371	25.49	81.36	0.39	10.03	1369.88
high	85371	25.87	82.43	0.40	10.20	1371.96
low	85371	25.15	80.45	0.38	9.88	1355.00
vol	85371	762.85	1565.45	0.00	345.75	51098.98
ma5	85371	25.48	81.24	0.40	10.05	1353.90
ma20	85371	25.30	80.25	0.43	10.04	1330.29
ma60	85371	24.85	77.99	0.47	9.91	1203.40
macd	85371	0.16	1.93	-30.89	0.01	49.93
boll_u	85371	26.30	83.72	0.41	10.36	1384.19
boll_m	85371	25.48	81.28	0.40	10.05	1354.56
boll_d	85371	24.66	78.86	0.37	9.69	1331.41
bias	85371	0.30	4.00	-34.17	0.10	60.15
kdj_k	85371	49.97	25.86	0.00	50.63	100.00
kdj_d	85371	49.97	24.69	2.09	50.53	100.00
kdj_j	85371	49.96	35.91	-49.21	50.13	155.85
rsi	85371	51.38	11.86	9.30	50.98	98.36
cp1	85371	19.59	45.73	1.62	10.34	1366.60
cp2	85371	17.74	37.05	1.59	10.23	1338.00
cp3	85371	16.94	36.41	1.59	10.17	1358.00
cp4	85371	16.59	34.69	1.57	10.19	1366.10
cp5	85371	15.93	29.73	1.62	10.14	1346.11
cp6	85371	15.98	34.29	1.53	10.16	1346.21
cp7	85371	15.36	24.58	1.61	10.10	1276.00
cp8	85371	15.17	24.87	1.55	10.02	1193.00
cp9	85371	14.99	25.28	1.58	9.83	1250.56
cp10	85371	14.81	22.62	1.63	9.97	1163.00

从整个数据集来看，各基础指标的标准差都非常大，反映出预测标的之间有非常大的差异性，以收盘价变量为例，最小值为 0.39，最大值为 1366.6，表明富时中国 A50 指数成分股的股价分布跨度非常大，平均值为 25.53，大于中位数 10.04，这表明目前 A50 指数成分股的股价总体来说在高价区域的分布较少。各衍生技术指标的值都在合理范围

内，没有出现明显的离群值。十个相关股指标的分布情况和基础指标类似，它们的平均值比预测标的收盘价变量的平均值要低近十元，这表明从平均股价上来看，预测标的股票的收盘价比其相关股的收盘价要高许多，表明所选择的预测标的股票比较优质，受到了投资者的追捧，使得股价得到推升。

5.2 相关性分析

为了了解所选特征变量和待预测变量——股票收盘价之间的相关性，本文计算了各特征变量与股票收盘价变量之间的皮尔森相关系数，由于股票收盘价也是特征变量之一，故在相关性的计算中剔除。结果如表 5.2 和图 5.1 所示，各相关系数均通过了 1% 水平上的显著性检验。

表 5.2 收盘价与其余特征变量的相关系数表

特征变量	相关系数	特征变量	相关系数
open	1.000***	cp4	0.147***
high	1.000***	cp5	0.145***
low	1.000***	cp6	0.137***
ma5	1.000***	cp7	0.125***
boll_u	1.000***	cp8	0.121***
boll_m	1.000***	cp9	0.116***
ma20	0.999***	cp10	0.093***
boll_d	0.999***	rsi	0.066***
ma60	0.996***	kdj_d	0.058***
macd	0.477***	kdj_k	0.055***
cp1	0.193***	kdj_j	0.040***
cp2	0.185***	bias	0.030***
cp3	0.152***	vol	-0.089***

注：***，**，*分别代表在 1%，5%，10%的显著性水平上显著。

图 5.1 表明，除了成交量以外的基础指标和衍生技术指标中的布林带及各期均线与收盘价的相关系数接近于 1，呈现高度线性相关关系，macd 指标与收盘价的相关系数为 0.477，接近 0.5，呈现中度线性相关关系，各相关股指标与收盘价的相关系数位于 0.093 到 0.193 之间，显示在整个数据集上，相关股指标与收盘价的相关系数并不高。其余技术指标和收盘价的相关系数低于 0.066，线性相关程度非常低，成交量与收盘价间的相关系数为-0.089，呈现弱负相关关系，在整体数据集上呈现这一特点主要是因为中国股市的散户投资者非常多，股价过高会显著降低投资者的交易可能性，导致成交量与收盘价间呈现弱负相关关系。

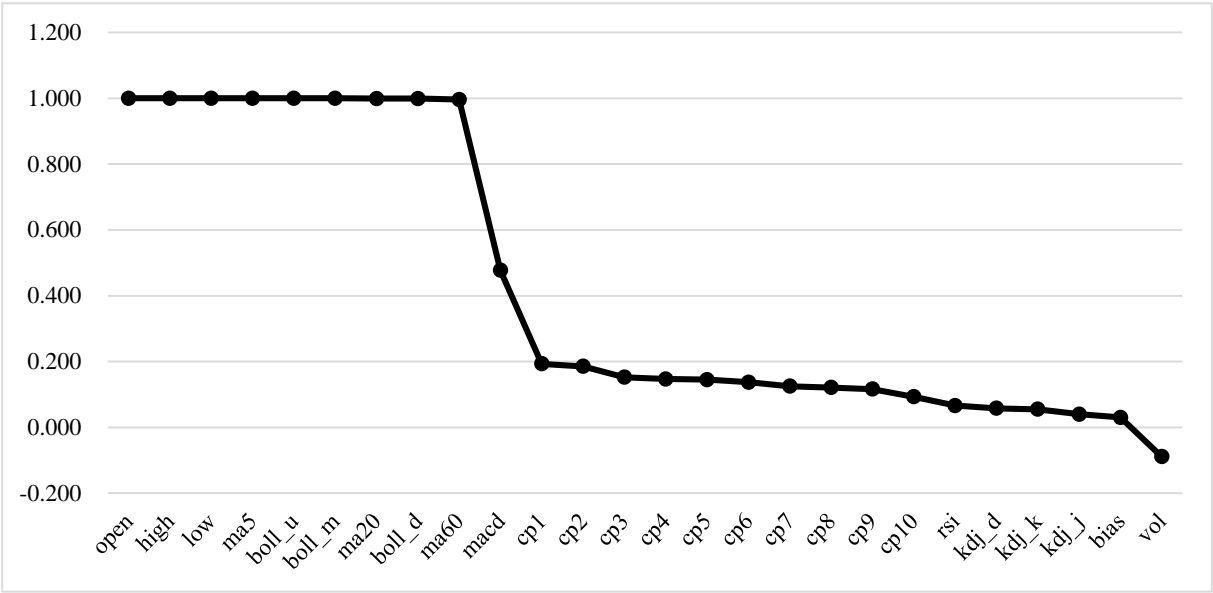


图 5.1 收盘价与其余特征变量的相关系数折线图

上文的相关性分析依赖的是整体数据集，但是由于本文实际上是根据每一个预测标的的子数据集进行预测，所以针对性研究各子数据集中收盘价和其他特征变量的相关系数情况也十分关键。限于篇幅考虑，本文选择银行业中的工商银行，白酒业中的贵州茅台，制造业的立讯精密这三个有代表性的个股分别进行相关性分析，探究其相关系数结果是否存在差异，结果如表 5.3 所示。

由表 5.3 可以看出，有部分特征变量的相关系数差异比较大，下面本文将把差异较大的特征变量的相关系数画图，如图 5.2 所示：

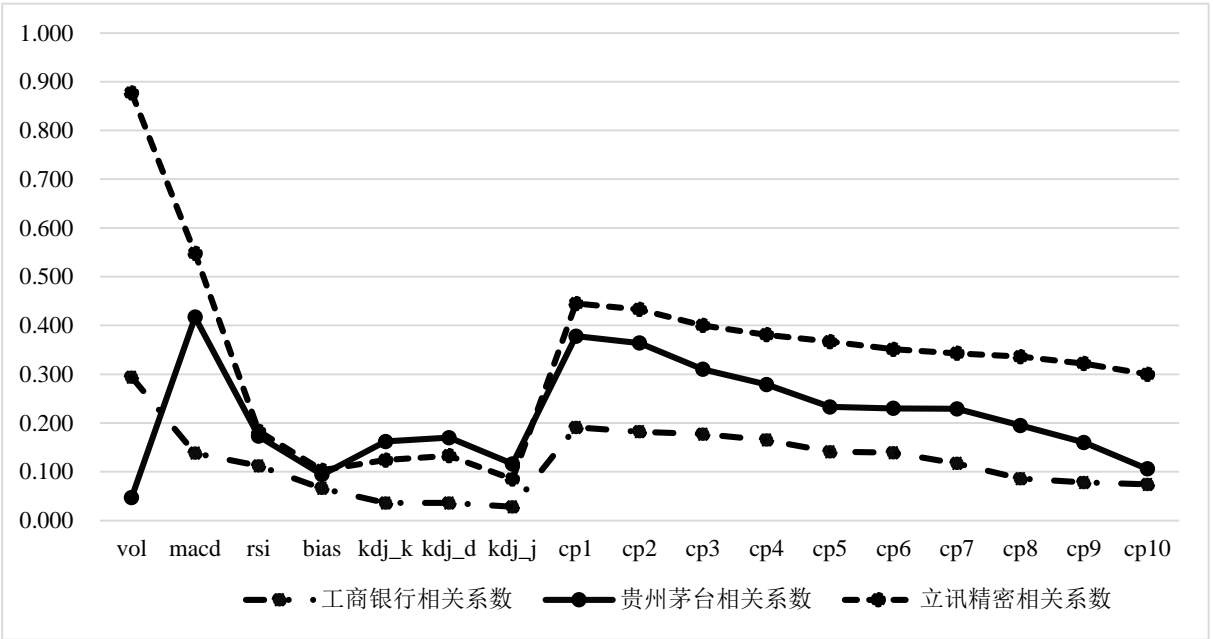


图 5.2 三只个股收盘价与其余特征变量的相关系数折线图

表 5.3 三只个股收盘价与其余特征变量的相关系数表

特征变量	工商银行相关系数	贵州茅台相关系数	立讯精密相关系数
open	0.999***	1.000***	0.999***
high	1.000***	1.000***	1.000***
low	0.999***	1.000***	1.000***
vol	0.294***	0.047**	0.877***
ma5	0.999***	0.999***	0.999***
ma20	0.993***	0.998***	0.995***
ma60	0.981***	0.993***	0.988***
macd	0.138***	0.417***	0.548***
rsi	0.112***	0.173***	0.183***
boll_u	0.997***	0.999***	0.998***
boll_m	0.998***	0.999***	0.999***
boll_d	0.996***	0.999***	0.998***
bias	0.066***	0.094***	0.103***
kdj_k	0.036*	0.162***	0.124***
kdj_d	0.036*	0.170***	0.133***
kdj_j	0.028	0.116***	0.085***
cp1	0.191***	0.378***	0.445***
cp2	0.182***	0.364***	0.433***
cp3	0.177***	0.310***	0.400***
cp4	0.165***	0.279***	0.381***
cp5	0.141***	0.233***	0.367***
cp6	0.139***	0.230***	0.351***
cp7	0.117***	0.229***	0.343***
cp8	0.086***	0.195***	0.336***
cp9	0.078***	0.160***	0.322***
cp10	0.074***	0.106***	0.300***

注：***，**，*分别代表在 1%，5%，10%的显著性水平上显著。

从成交量和收盘价的相关系数来看，贵州茅台仅为 0.047，而工商银行达到了 0.294，立讯精密则为 0.877，呈现高度相关关系。从传统的量价分析理论来看，这一结果表明，立讯精密的股价呈现明显的成交量驱动型，成交量越大，股价也越高；贵州茅台的成交量则和股价之间的相关性极弱，表明贵州茅台的股价攀升过程中，成交量保持相对平稳，没有明显放大；工商银行则介于两者之间，成交量和股价间是中度相关关系。从各衍生技术指标和收盘价的相关系数来看，工商银行最小，在 KDJ 各指标上，是贵州茅台最大，而在 macd 等指标上，是立讯精密最大。从相关股指标和收盘价的相关系数来看，立讯精密最大，第一相关股指标上达到了 0.445，属于中度相关，贵州茅台其次，工商银

行最低，表明立讯精密更容易受到同行业其他股票走势的影响，而工商银行的走势与同行业其他股票的同步性较低。

综上，仅从皮尔森相关系数角度来看，各预测标的的特征变量对于收盘价的影响各有不同。本研究在下文的模型对比分析中予以考虑，将分别研究加入不同类型的特征变量的模型的预测效果的差异。

5.3 超参数寻优

本文利用上一章所叙述的网格寻优法来进行超参数寻优工作。由于本文的待优化模型有 5 个，待预测标的有 39 个，所以需要重复运行的各模型的总次数为 1950 次。在本文的实际研究中，总的运行时间达到了 89 个小时。在每一次运行模型的过程中，本研究均使用了早停策略，虽然费时费力，但是对于超参数结果的可靠性提升非常大。

限于篇幅原因，本文将以上述三只股票为例在 Att-LSTM(B-T-P) 模型上的超参数寻优的过程报告如表 5.4 所示，针对每一只股票而言，其中加粗的评价标准值所对应的超参数组合为本文研究中所确定的最优超参数组合：

表 5.4 三只个股超参数寻优结果表								
工商银行				贵州茅台				
batch_size				batch_size				
lr	hidden_size	128	256	512	128	256	512	
0.001	64	0.0201	0.0496	0.0274	0.0734	0.0935	0.0750	
	128	0.0235	0.0307	0.0270	0.0538	0.0374	0.0419	
0.01	64	0.0372	0.0433	0.0477	0.2905	0.2068	0.1379	
	128	0.0991	0.0389	0.0248	0.2777	0.1311	0.0628	
立讯精密								
lr	hidden_size							
0.001	64	0.0335	0.0449	0.0557				
	128	0.0219	0.0214	0.0279				
0.01	64	0.1322	0.1026	0.0372				
	128	0.1101	0.1327	0.0585				

5.4 模型预测对比分析

本研究首先将已经计算好的整体数据集按照股票代码分成了 39 个子数据集，针对每个子数据集分别使用滑动窗口方法构建用于训练和预测的新的子数据集。在新子数据集构建完毕后，将其前 70%划分为训练集，后 30%划分为测试集。因为本文一共有 6 个待训练模型，所以还需要根据每一个模型的具体要求，选择特征变量的组合。除了 ARIMA 模型无需按照神经网络的一般模式训练以外，其余各模型均需要将训练集数据

输入，进行训练，各神经网络模型的超参数采用上一节确定好的超参数组合，其训练过程中依然使用早停策略，以降低过拟合的影响，来保证预测结果的可靠性。在训练过程结束后，再将测试集的数据代入已经训练好的各模型，分别得出预测结果并记录。然后，利用上一章所提到的三个评价指标的公式，计算每一个预测标的股票的每个评价指标的具体数值。最后，计算每个模型下各预测标的股票的预测评价指标的均值以及中位数。结果如表 5.5 所示，中括号内为中位数值，较优模型的评价指标加粗表示：

表 5.5 模型预测结果对比表			
模型	MAE	MAPE	RMSE
ARIMA	0.0282	0.2343	0.1679
	[0.0274]	[0.2214]	[0.1581]
BPNN	0.0021	0.0536	0.0417
	[0.0015]	[0.0476]	[0.0382]
LSTM	0.0020	0.0459	0.0398
	[0.0011]	[0.0376]	[0.0332]
Att-LSTM(B)	0.0011	0.0378	0.0307
	[0.0007]	[0.0350]	[0.0273]
Att-LSTM(B-T)	0.0010	0.0324	0.0285
	[0.0006]	[0.0314]	[0.0252]
Att-LSTM(B-T-P)	0.0005	0.0251	0.0209
	[0.0004]	[0.0249]	[0.0209]

表 5.5 的结果表明：本文所构建的 Att-LSTM 的三个子模型在三种评价指标上都优于其他基准模型。比较三个子模型，Att-LSTM(B-T-P)的预测效果在三种评价指标上都优于其他两种子模型，这表明对于 Att-LSTM 模型，同时加入基础指标，衍生技术指标和相关股指标有助于提升预测效果。同时我们发现加入基础指标和衍生技术指标的情况下，预测效果虽然也优于只加入基础指标，但是两者的差异并不大，例如 MAE 指标，同时加入技术指标以及衍生技术指标后，MAE 指标仅降低了 0.0001。

下面将对主要模型之间的预测评价指标差异进行配对 T 检验和配对 Wilcoxon 检验。结果如表 5.6 所示，其中对比序号 1 表示 Att-LSTM(B-T-P) - LSTM，对比序号 2 表示 Att-LSTM(B-T) - LSTM，对比序号 3 表示 Att-LSTM(B) - LSTM，对比序号 4 表示 Att-LSTM(B-T-P) - Att-LSTM(B-T)，对比序号 5 表示 Att-LSTM(B-T) - Att-LSTM(B)，对比序号 6 表示 Att-LSTM(B-T-P) - Att-LSTM(B)。

表 5.6 的结果表明，大部分对比模型组合的预测效果差异的配对 T 检验和配对 Wilcoxon 检验都在 5%的显著性水平上显著。对比序号 5，即 Att-LSTM(B-T)和 Att-

LSTM(B) 的两种差异检验均在 5% 的显著性水平上不显著, 这表明, 在本文所构建的 Att-LSTM 中, 加入衍生技术指标对于模型的预测效果提升不明显。

表 5.6 模型差异对比及显著性检验表

序号	MAE	T/Z	MAPE	T/Z	RMSE	T/Z
1	-0.0015	-4.893***	-0.0208	-6.780***	-0.0189	-6.753***
	[-0.0007]	-5.275***	[-0.0127]	-5.177***	[-0.0122]	-5.233***
2	-0.0010	-3.612***	-0.0134	-5.310***	-0.0114	-4.964***
	[-0.0005]	-4.535***	[-0.0062]	-4.912***	[-0.0080]	-4.689***
3	-0.0009	-2.744***	-0.0081	-2.149**	-0.0092	-2.800***
	[-0.0004]	-2.400**	[-0.0026]	-2.023**	[-0.0058]	-2.484**
4	-0.0005	-3.410***	-0.0074	-4.326***	-0.0076	-4.138***
	[-0.0002]	-4.186***	[-0.0065]	-4.103***	[-0.0043]	-4.312***
5	-0.0001	-0.485	-0.0053	-2.025*	-0.0022	-0.920
	[-0.0001]	-1.088	[-0.0036]	-2.01**	[-0.0021]	-1.242
6	-0.0006	-4.212***	-0.0127	-6.309***	-0.0098	-5.471***
	[-0.0003]	-4.842***	[-0.0101]	-5.08***	[-0.0064]	-4.940***

注: ***, **, *分别代表在 1%, 5%, 10% 的显著性水平上显著。

以工商银行 (601398.SH) 为例, 图 5.3 和图 5.4 分别是本文所构建的 Att-LSTM(B-T-P) 模型和单纯的 LSTM 神经网络的预测效果对比图。图 5.3 和图 5.4 的时间区间为测试集的最后 60 个交易日。

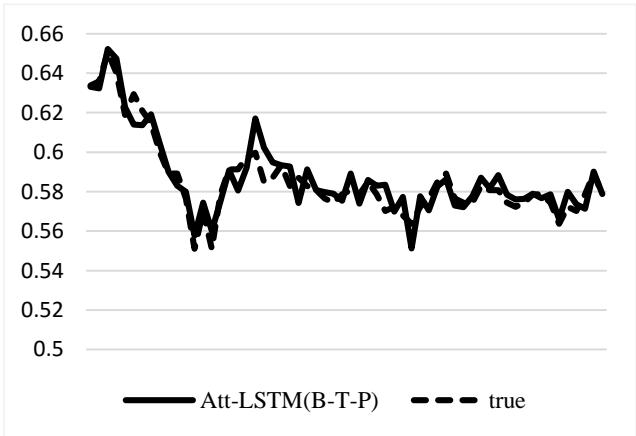


图 5.3 Att-LSTM(B-T-P) 模型实际预测图

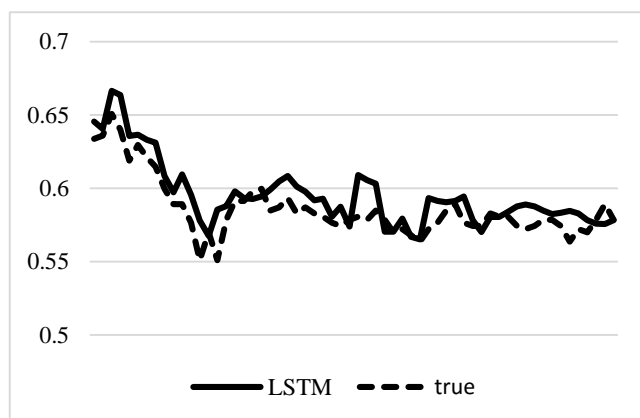


图 5.4 LSTM 模型实际预测效果图

从图 5.3 和图 5.4 可以直观地看出，本文所构建的 Att-LSTM 神经网络模型所预测的波动趋势和真实值基本一致，相较于单纯的 LSTM 神经网络模型的拟合效果要更好，但是依然在某些时间点存在预测值滞后的现象。针对存在预测值滞后的原因，本文认为是有部分影响股价的关键因子没有被考虑进本文所构建的模型之中，这一点有待以后进行进一步改进。

总的说来，本研究构建的改进模型的确能够提高预测的总体精度，达到了预期效果。

5.5 进一步研究

5.5.1 极端事件期间预测效果的扩展研究

极端事件的发生往往会对股价形成比较大的冲击，2008 年汶川地震使得我国 A 股市场短期暴跌，2018 年初中美之间发生的贸易战使得上证指数形成的阶段高点至今没有突破，年初突发的新冠疫情更是在年后引发了千股跌停。股票预测的复杂性表现在个股的走势非常容易受到极端事件的影响。

从上面可以得出，极端事件对于股价的冲击可能影响模型的预测效果。为了研究极端事件可能对模型的预测效果产生的影响，本文研究了各模型在疫情期间的预测效果，以钟南山院士宣布确定人传人的时点（2020 年 1 月 20 号）为起点，2 月底为终点，在前面各模型进行训练并且代入测试集数据得到的预测结果序列中按照上述日期范围提取出疫情期间的预测结果序列，计算各评价标准的值，并进一步计算它们的平均值和中位数，结果如表 5.7 所示，其中较优模型的评价标准值加粗表示，中括号内为中位数值。

从表 5.7 可以看出，疫情期间各模型的预测效果都存在一定的下降，这可能是由于疫情事件，各预测标的股票的波动比以往要剧烈许多，模型捕捉暴涨暴跌的能力相对较低，因此导致在疫情期间的预测效果下降了。同时我们发现在三个子模型之中，仅加入基础指标的模型体现出了更好的效果，其主要可能的原因从衍生技术指标的角度看是由于股价短期波动剧烈导致各衍生技术指标的滞后性变大，此时加入衍生技术指标给模

型带来了负效用；从相关股指标的角度看，相关股在疫情期间和预测标的股票的走势类似，加入相关股指标，带来了更多的噪音，从而导致模型的效果下降。

表 5.7 疫情期间模型预测结果对比表

模型	MAE	MAPE	RMSE
ARIMA	0.0343	0.2723	0.1979
	[0.0329]	[0.2587]	[0.1746]
BPNN	0.0037	0.0619	0.0508
	[0.0035]	[0.0532]	[0.0493]
LSTM	0.0031	0.0588	0.0487
	[0.0028]	[0.0457]	[0.0481]
Att-LSTM(B)	0.0009	0.0347	0.0288
	[0.0007]	[0.0293]	[0.0271]
Att-LSTM(B-T)	0.0021	0.0476	0.0390
	[0.0010]	[0.0378]	[0.0323]
Att-LSTM(B-T-P)	0.0025	0.0509	0.0401
	[0.0018]	[0.0523]	[0.0425]

表 5.8 疫情期间模型差异对比及显著性检验表

序号	MAE	T/Z	MAPE	T/Z	RMSE	T/Z
1	-0.0006	-2.066**	-0.0079	-2.924***	-0.0086	-1.995**
	[-0.0010]	-2.177**	[0.0066]	2.149**	[-0.0056]	-2.386***
2	-0.0010	-2.547**	-0.0112	-3.249***	-0.0097	-2.419**
	[-0.0018]	-2.223**	[-0.0079]	-2.221**	[-0.0158]	-3.181***
3	-0.0022	-2.771***	-0.0241	-4.921***	-0.0199	-3.268***
	[-0.0021]	-2.972***	[-0.0164]	-4.073***	[-0.0210]	-3.917***
4	0.0004	1.205	0.0033	1.819*	0.0011	0.919
	[0.0008]	1.707*	[0.0145]	2.703***	[0.0102]	2.791***
5	0.0012	2.535**	0.0129	3.126**	0.0102	3.043***
	[0.0003]	1.088	[0.0085]	3.196***	[0.0052]	2.112**
6	0.0016	3.040***	0.0162	5.268***	0.0113	4.985***
	[0.0011]	4.177***	[0.0230]	5.191***	[0.0154]	5.247***

注：***，**，*分别代表在 1%，5%，10%的显著性水平上显著。

下面将对疫情期间主要模型之间的预测评价指标差异进行配对 T 检验和配对 Wilcoxon 检验。结果如表 5.8 所示, 其中对比序号 1 表示 Att-LSTM(B-T-P) - LSTM, 对比序号 2 表示 Att-LSTM(B-T) - LSTM, 对比序号 3 表示 Att-LSTM(B) - LSTM, 对比序号 4 表示 Att-LSTM(B-T-P) - Att-LSTM(B-T), 对比序号 5 表示 Att-LSTM(B-T) - Att-LSTM(B), 对比序号 6 表示 Att-LSTM(B-T-P) - Att-LSTM(B)。

表 5.8 的结果表明, 大部分对比模型组合的预测效果差异的配对 T 检验和配对 Wilcoxon 检验都在 5% 的显著性水平上显著。对比序号 4, 即 Att-LSTM(B-T-P) 和 Att-LSTM(B-T) 的 MAE 指标的两种差异检验均在 5% 的显著性水平上不显著, 这一对比模型组合的另外两种评价指标的配对 T 检验也在 5% 的显著性水平上不显著。这表明, 在疫情期间, 在本文所构建的 Att-LSTM 中加入相关股指标对于模型的预测效果提升不明显, 甚至可能降低模型的预测效果。这一结果再次证明, 处于以疫情为代表的极端事件情况下, 相关股指标不适合作为本文所构建的模型的输入特征变量

5.5.2 超参数对于预测效果的影响的扩展研究

由于上文中神经网络的隐藏层神经元数目, 训练批次数以及学习率大小都会对于最终的预测效果有影响, 所以有必要进行进一步研究, 探究在不同的超参数设置下, 网络的平均误差将如何变化。

在实践中证实, 隐藏层的神经元数目 (hidden_size) 作为首先确定的超参数, 对于模型的预测精度影响非常大。如果隐藏层的神经元数目过少, 网络不能具有必要的学习能力和信息处理能力。反之, 若过多, 不仅会大大增加网络结构的复杂性, 网络在学习过程中更易陷入局部极小点, 而且会使网络的学习速度变得很慢。隐藏层的神经元数目的选择问题一直受到神经网络研究工作的高度重视。

训练批次数 (batch_size) 是神经网络训练过程中一个重要的参数, 它虽然与神经网络模型本身无关, 却与预测精度紧密相关, 因为神经网络训练过程中采用的梯度下降法的梯度下降方向是根据每一个训练批次的具体情况来选择的。所以理论上存在一个较优的训练批次数值, 使得预测精度较高。

本节在前文构建的 Att-LSTM(B-T-P) 模型的基础上, 为了简化起见, 首先将 lr 固定为 0.001, 评价标准选择为 RMSE, 将隐藏层的神经元数目的选择范围进一步扩展为 {32, 64, 128, 256}, 将训练批次数的选择范围进一步扩展为 {32, 64, 128, 256, 512}, 以探究隐藏层的神经元数目和训练批次数对于本文中的 Att-LSTM(B-T-P) 模型的预测精度的影响。

分析结果如表 5.9 所示: 对于 Att-LSTM(B-T-P) 模型, 随着隐藏层神经元数目的增加, 总体上 RMSE 值呈现先下降再上升趋势, 保持训练批次数不变, 当隐藏层神经元数目设定在 128 时, 整体模型的效果较优。

从理论上来说，如果神经元的数量远远大于特征维度，那么训练过程中出现过拟合的概率也会大大增加，从而影响预测效果。上文的结果也印证了这一理论，存在某一个最优的隐藏层神经元数目。

对于 Att-LSTM(B-T-P) 模型，随着训练批次数的增加，总体上 RMSE 值呈现先下降再上升趋势，保持隐藏层神经元数目不变，当训练批次数设定在 128 时，整体模型的效果较优。

从结果来看，同样存在一个最优的训练批次数，使得模型的预测效果最好。

表 5.9 超参数扩展分析表（RMSE 为准则）

hidden_size	batch_size				
	32	64	128	256	512
32	0.0554	0.0654	0.0379	0.0412	0.0388
64	0.0625	0.0434	0.0230	0.0476	0.0285
128	0.0588	0.0287	0.0245	0.0327	0.0244
256	0.0744	0.0524	0.0525	0.0657	0.0743

6 结论与展望

6.1 结论

本文基于 LSTM 神经网络，将注意力机制引入了基础的 LSTM 神经网络之中，构建了基于注意力机制和 LSTM 神经网络的时序预测模型（Att-LSTM），详细分析了 Att-LSTM 模型的基本结构，论证了在 LSTM 神经网络中加入注意力机制以进一步改进金融时间序列分析长时依赖问题的可行性，综合理论分析和实证检验，得到以下结论：

（1）本文构建的 Att-LSTM 模型，在一定程度上缓解了时序神经网络所固有的长时依赖问题，尝试针对多预测因子赋予不同的权重，结果表明：改进后的模型的预测效果相较于基准模型更佳，在三个评价标准下，改进后的模型的三个子模型都打败了基准模型，说明这一改进是有效的。

（2）在本文所构建的 Att-LSTM 模型情况下，选择适当的隐藏层神经元数目和适当的训练批次数对于预测精度的提升作用较大，具体来说，预测精度随着隐藏层神经元数目的升高呈现先降低再上升的趋势，随着训练批次数的升高，呈现先降低再上升的趋势。但是它们都有一定的限度，盲目降低隐藏层神经元数目和提高训练批次数，不仅不能提高预测精度，反而会损害预测的精度，使得模型的预测能力变得非常差。

（3）本文发现在整体数据集上，加入基础指标，衍生技术指标以及相关股指标的 Att-LSTM 模型效果最好，但是在极端事件期间，仅加入基础指标的效果最好，主要原因可能是由于股价短期波动剧烈导致各衍生技术指标的滞后性变大，此时加入衍生技术指标给模型带来了负效用，相关股在极端事件期间和预测标的股票的走势类似，加入相关股指标，带来了更多的噪音，从而导致模型的效果下降。

6.2 展望

本文虽然针对传统的 LSTM 神经网络模型进行了改进，取得了一定效果，但是只是对下一个交易日的收盘价进行预测，属于单步预测，并没有进行多步预测，且最终的预测结果具有滞后性，实际能否指导投资，产生收益，还需要进一步实际验证。针对存在预测值滞后的原因，本文认为是有部分影响股价的关键因子没有被考虑进本文所构建的模型之中，这一点有待以后进行进一步改进。

同时，本文中针对具体超参数的寻优部分的研究，采用的是简单的网格寻优法，这一方法虽然效果比较好，但是随机性依然比较强，稳定性不够高，在后续的研究中，可以尝试采用遗传算法或者粒子群算法等方法进行寻优。

参考文献

- [1] Abu-Mostafa Y S, Atiya A F. Introduction to Financial Forecasting[J]. Applied Intelligence, 1996, 6(3): 205-213.
- [2] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014(5): 1-15.
- [3] Bengio Y, Lamblin P, Popovici D, et al. Greedy Layer-wise Training of Deep Networks[J]. Advances in Neural Information Processing Systems, 2007, 19: 153-160.
- [4] Bengio Y, Simard P, Frasconi P. Learning Long-Term Dependencies With Gradient Descent is Difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.
- [5] Chen K, Zhou Y, Dai F. A LSTM-based Method for Stockreturns Prediction: A Case Study of China Stock Market[C]. Proceedings of the 2015 IEEE International Conference on Big Data, Santa Clara, CA, 2823-2824, October, 2015.
- [6] Ching M M, Siok K S. Comparing the Performances of GARCH-type Models in Capturing the Stock Market Volatility in Malaysia[J]. Procedia Economics and Finance, 2013, 5(1): 478-487.
- [7] Cinar Y G, Mirisae H, Goswami P, et al. Time Series Forecasting using RNNs: an Extended Attention Mechanism to Model Periods and Handle Missing Values[J]. Neural and Evolutionary Computing, 2017, 2(1): 12-22.
- [8] Desimone R, Duncan J. Neural Mechanisms of Selective Visual Attention[J]. Annual Review of Neuroscience, 1995, 18: 193-222.
- [9] Dhar S, Mukherjee T, Ghoshal A K. Performance Evaluation of Neural Network Approach in Financial Prediction: Evidence from Indian Market[C]. 2010 International Conference on IEEE, 2010: 597-602.
- [10] Di Persio L, Honchar O. Artificial Neural Networks Architectures for Stock Price Prediction: Comparisons and Applications[J]. International Journal of Circuits, Systems and Signal Processing, 2016(10): 403-413.
- [11] Di Persio L, Honchar O. Recurrent Neural Networks Approach to the Financial Forecast of Google Assets[J]. International Journal of Mathematics and Computers in Simulation, 2017(11): 7-13.
- [12] Green J, Hand J R M, Zhang X F. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns[J]. The Review of Financial Studies, 2017, 30(12): 4389-4436.

- [13] Hammad A A A, Ali S M A, Hall E L. Forecasting the Jordanian Stock Prices using Artificial Neural Networks[M]. Intelligent Engineering Systems through Artificial Neural Networks, U.S. : American Society of Mechanical Engineers, 2007: 502-505.
- [14] Hansen J V, Nelso R D. Neural Networks and Traditional Time Teries Methods: A Synergistic Combination in State Economic Forecasts[J]. IEEE Transactions on Neural Networks, 1997, 8(4): 863-873.
- [15] Harvey C R, Y Liu, H Zhu. The Cross-Section of Expected Returns[J]. The Review of Financial Studies, 2016, 29(1): 5-68.
- [16] Heaton J B, Polson N G, Witte J H. Deep Learning for Finance: Deep Portfolios[J]. Applied Stochastic Models in Business and Industry, 2017, 33(1): 3-10.
- [17] Heaton J B, Polson N G, Witte J H. Deep Learning in Finance[J]. arXiv preprint arXiv: 1602. 06561, 2016.
- [18] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786): 504-507.
- [19] Hochreiter S, Schmidhuber J. Long Short Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [20] Hou K C, L Zhang. Replicating Anomalies[EB/OL]. The Review of Financial Studies, 2019, <https://doi.org/10.1093/rfs/hhy131>.
- [21] LeCun Y, Bengio Y, Hinton G. Deep Learning[J]. Nature, 2015, 521(3): 436-444.
- [22] Linnainmaa J T, Roberts M R. The History of the Cross-Section of Stock Returns[J]. Review of Financial Studies, 2018, 31: 2606-2649.
- [23] Liu S, Zhang C, Ma J. CNN-LSTM Neural Network Model for Quantitative Strategy Analysis in Stock Markets[C]. International Conference on Neural Information Processing. Springer, Cham, 2017: 1-5.
- [24] Mabrouk A H. Stock Price Forecasting in an Interest Free Economy[J]. International Journal of Economics, Management and Accounting, 2011, 19(3): 137-141.
- [25] McLean R D, Pontiff J. Does Academic Research Destroy Stock Return Predictability?[J]. Journal of Finance, 2016, 71(1): 5-32.
- [26] Mnih V, Heess N, Graves A, et al. Recurrent Models of Visual Attention[C]. Advances in Neural Information Processing Systems (NIPS 2014), 2014: 2204-2212.
- [27] Rumelhart D E, Hinton G E, Williams R J. Learning Representations by Backpropagating Errors[J]. Nature, 1986, 323(9): 533-536.
- [28] Samarawickrama A J, Fernando T G. A Recurrent Neural Network Approach in Predicting

Daily Stock Prices an Application to the Sri Lankan Stock Market[C]. Proceedings of the 2017 IEEE International Conference on Industrial and Information Systems, Peradeniya, Sri Lanka, 2017: 1-6.

[29] S Chen, L Ge. Exploring the Attention Mechanism in LSTM-based Hong Kong Stock Price Movement Prediction[J]. Quantitative Finance, 2019(9): 1507-1515.

[30] Ticknor J L. A Bayesian Regularized Artificial Neural Network for Stock Market Forecasting[J]. Expert Systems with Applications, 2013(14): 5501-5506.

[31] Xiong R, Nichols E P, Shen Y. Deep Learning Stock Volatility with Google Domestic Trends[J]. arXiv preprint arXiv: 1512. 04916, 2015.

[32] Zakaria J, Mueen A, Keogh E. Clustering Time Series using Unsupervised Shapelets[C]. IEEE, International Conference on Data Mining. Brussels: IEEE Computer Society, 2012: 785-794.

[33] 巴曙松, 蒋峰.“违约潮”背景下的信用风险测度研究[J].湖北经济学院学报, 2019, 17(06):5-13+127.

[34] 包振山, 郭俊南, 谢源, 张文博.基于 LSTM-GA 的股票价格涨跌预测模型[J].计算机科学, 2020, 47(S1):467-473.

[35] 冯宇旭, 李裕梅.基于 LSTM 神经网络的沪深 300 指数预测模型研究[J].数学的实践与认识, 2019, 49(07):308-315.

[36] 谷丽琼, 吴运杰, 逢金辉.基于 Attention 机制的 GRU 股票预测模型[J/OL].系统工程: 1-11[2020-09-26].

[37] 林杰, 康慧琳.基于注意力机制的 LSTM 股价趋势预测研究[J].上海管理科学, 2020, 42(01):109-115.

[38] 刘国旗.非线性 GARCH 模型在中国股市活动预测中的应用研究[J].统计研究, 2000(1):49-52.

[39] 刘宁宁, 张量.基于 PCA-GRU 的股票指数预测模型的研究[J].计算机应用研究, 2020, 37(S1):113-115.

[40] 乔若羽.基于神经网络的股票预测模型[J].运筹与管理, 2019, 28(10):132-140.

[41] 乔忠学.基于双向 LSTM 的深圳成分指数的对比研究[J].产业创新研究, 2020(14):83-84.

[42] 宋睿.基于 CNN-LSTM 的股票中长期趋势预测[J].电子技术与软件工程, 2020(4):160-161.

[43] 王理同, 薛腾腾, 王惠敏, 刘震.基于循环神经网络的股指价格预测研究[J].浙江工业大学学报:2020, 47(2):186-191.

- [44] 王卫红, 卓鹏宇. 基于 PCA-FOA-SVR 的股票价格预测研究[J]. 浙江工业大学学报, 2016, 44(4): 399-404.
- [45] 王霄鹏. 基于 LSTM 改进模型的股票预测研究[D]. 重庆理工大学, 2020.
- [46] 韦艳华, 张世英. 金融市场的相关性分析——Copula-GARCH 模型及其应用[J]. 系统工程, 2004(04): 7-12.
- [47] 文宝石, 颜七笙. 数据多维处理 LSTM 股票价格预测模型[J]. 江西科学, 2020, 38(04): 443-449+472.
- [48] 武大硕, 张传雷, 陈佳, 向启怀. 基于遗传算法改进 LSTM 神经网络股指预测分析[J]. 计算机应用研究, 2020, 37(S1): 86-87+107.
- [49] 许兴军, 颜钢锋. 基于 BP 神经网络的股价趋势分析[J]. 浙江金融, 2011(11): 57-59+64.
- [50] 杨青, 王晨蔚. 基于深度学习 LSTM 神经网络的全球股票指数预测研究[J]. 统计研究, 2019, 36(03): 65-77.
- [51] 朱永明, 邵庚云. 基于 BP 神经网络的股价趋势预测——以房地产开发上市公司为例[J]. 财会月刊, 2013(14): 76-79.

致 谢

韶华易逝，青春不再。研究生三年转瞬即逝，虽然有困难，有挫折，但终归坦途，在这里我要感谢我的导师，同学和朋友，他们无私地在最困难的时候帮助我，成为了照亮我前进的光芒。此文付梓之际，我感到格外感激和感动。

首先，我要感谢我的导师安智宇副教授。在学习和生活方面，安老师都对于我悉心指导和关心。安老师严谨的学术精神和追求卓越的论文写作态度对我影响颇深。在毕业论文撰写过程中，安老师给我提供了很多思路，无私传授了很多撰写技巧，此文离不开安老师的帮助，在这里我想对安老师衷心说一声感谢。

其次，我要向各位授课老师和同学表示谢意，三年时光虽然白驹过隙，但是众多美好的回忆我将一直永留心中，在校的一切我都会铭记于心。

最后，我要向家人说一声谢谢，谢谢你们在日常生活中给我的关心，给予我鼓励，让我全身心地投入到毕业论文创作中。

感谢在百忙中抽空评审我论文的各位老师，希望你们不吝赐教。

附 录 A

攻读硕士学位期间发表的论文和出版著作情况：

- [1] Tianxiang Leng, Zhiyu An. Motivation Research of High-split of Listed Companies in China——Based on Logistic Model[C]. In: Pin Wang, eds. 2020 International Conference on Financial Economics and Investment Management (FEIM 2020). Taiyuan: FEIM, 2020. 135-139.

附录 B

表 B1 预测标的表

股票代码	简称	地区	行业	上市地
000002.SZ	万科 A	深圳	全国地产	主板
000001.SZ	平安银行	深圳	银行	主板
600887.SH	伊利股份	内蒙	乳制品	主板
000651.SZ	格力电器	广东	家用电器	主板
600104.SH	上汽集团	上海	汽车整车	主板
000858.SZ	五粮液	四川	白酒	主板
600000.SH	浦发银行	上海	银行	主板
600276.SH	恒瑞医药	江苏	化学制药	主板
600016.SH	民生银行	北京	银行	主板
600309.SH	万华化学	山东	化工原料	主板
600028.SH	中国石化	北京	石油加工	主板
600519.SH	贵州茅台	贵州	白酒	主板
600585.SH	海螺水泥	安徽	水泥	主板
600036.SH	招商银行	深圳	银行	主板
600050.SH	中国联通	北京	电信运营	主板
600030.SH	中信证券	深圳	证券	主板
600900.SH	长江电力	北京	水力发电	主板
601988.SH	中国银行	北京	银行	主板
600048.SH	保利地产	广东	全国地产	主板
601398.SH	工商银行	北京	银行	主板
601628.SH	中国人寿	北京	保险	主板
601166.SH	兴业银行	福建	银行	主板
601318.SH	中国平安	深圳	保险	主板
601998.SH	中信银行	北京	银行	主板
601328.SH	交通银行	上海	银行	主板
601939.SH	建设银行	北京	银行	主板
601088.SH	中国神华	北京	煤炭开采	主板
601857.SH	中国石油	北京	石油开采	主板
601601.SH	中国太保	上海	保险	主板

续表 B1

601668.SH	中国建筑	北京	建筑工程	主板
601888.SH	中国中免	北京	旅游服务	主板
002304.SZ	洋河股份	江苏	白酒	中小板
002352.SZ	顺丰控股	深圳	仓储物流	中小板
300059.SZ	东方财富	上海	证券	创业板
002415.SZ	海康威视	浙江	电器仪表	中小板
601288.SH	农业银行	北京	银行	主板
601818.SH	光大银行	北京	银行	主板
002475.SZ	立讯精密	深圳	元器件	中小板
300122.SZ	智飞生物	重庆	生物制药	创业板