

基于改进时间卷积网络的股票均线价格预测

作者姓名：	计金运
指导教师：	李志刚 副教授
专业领域：	电子信息
答辩日期：	2023 年 3 月 10 日

分类号	TP183	密 级	公开
UDC		单位代码	10146
		学 号	202085400456

硕士学位论文

基于改进时间卷积网络的股票均线价格 预测

研究生姓名：计金运

指 导 教 师:	李志刚 副教授	工 作 单 位 :	辽宁科技大学
指 导 教 师:		工 作 单 位 :	
论 文 提 交 日 期:	2023 年 03 月 11 日	答 辩 日 期 :	2023 年 03 月 10 日
学 位 授 予 日 期:		授 予 单 位 :	辽宁科技大学
论 文 评 阅 人:		工 作 单 位 :	
论 文 评 阅 人:		工 作 单 位 :	
答 辩 委 员 会 主 席:	王介生 教授	工 作 单 位 :	辽宁科技大学

**Prediction of Stock Average Price Based on Improved
Time Convolution Network**

**A thesis Submitted to
University of Science and Technology Liaoning**

**by
Ji Jinyun
(Majoring in Electronic Information)**

Supervisor: Prof. Li Zhigang

March 11, 2023

中文摘要

随着经济全球化的到来,股票这种新型的投资方式受到了许多投资者的关注,如何准确的预测出股票市场的走势情况成为投资者备受关注的课题。股票预测属于一种特殊的时间序列预测问题,由于其自身的非线性、不确定性,导致在进行股票预测时需要考虑诸多因素。时间卷积网络 TCN 在股票预测中表现出良好的预测效果。为了更好的解决股票预测方面的难题,使投资者在股票市场中获得更多的收益,进一步增强 TCN 网络的股票预测能力,本文对 TCN 网络中的时间卷积模块和连接方式提出了两种改进的方法。通过对沪深 A 股前 10 支股票所构建的数据集进行周线 ma_5、月线 ma_20、季线 ma_120 均线价格预测的消融实验后,实验结果表明,两种改进方式的时间卷积网络对 3 条均线的预测准确率都有所提升,具有很强的实用价值。本文的主要内容:

首先,提出了一种基于多尺度卷积和残差连接的股票均线价格预测模型 Res-TCN。在标准 TCN 网络的时间卷积模块 Temporal block 中加入多尺度卷积 Inception 结构,多尺度卷积中的多种卷积核结构使模型提取的特征更加完整,增强了模型对特征的提取。紧接着扩充网络的层数,使用残差连接将每一层的 TCN block 连接到一起,残差连接有利于防止网络训练时出现梯度下降的现象。本文设计的 Res-TCN 网络对周线 ma 5、月线 ma 20、季线 ma 120 的预测准确率分别为 70.58%、72.14%、74.53%,相对于 TCN 网络分别提高 5.89%、0.62%、1.98%,得到了不错的预测效果。

其次,为了使模型的预测效率更高,特征的提取效果更好,提出了一种基于全局注意力和密集连接的网络模型 Dens-TCN-GAM。将全局注意力模块 GAM 加入每一层 Temporal block 中组成 TCN-GAM block,将每一层的 TCN-GAM block 与 3×3 卷积作为 Dense block 中的 conv block,使网络利用特征的能力达到最大程度。采用 Adam 优化算法实验后,改进模型 Dens-TCN-GAM 在三种均线的预测准确率分别为 75.48%、77.33%、79.05%,相对于 Res-TCN 网络分别提高 4.90%、5.19%、4.52%,预测效率和准确度都有所提升。

最后,为了使投资者少走弯路,设计了基于多影响因子的股票选股系统。将多种影响股票的影响因子作为数据集的输入特征,同时扩充数据集中股票的种类,选取不同种类的股票加入股票池中。通过预测模型输出后,将周线、月线、季线的预测最佳的 10 只股票选出来组成股票池,投资者可以根据自身需求进行选择。

关键词: 股票预测; 时间卷积网络; 多尺度卷积; 注意力机制

ABSTRACT

With the advent of economic globalization, this new type of investment in stocks has received a lot of attention from investors, and how to accurately predict the stock market movements has become a matter of great concern to investors. Stock forecasting is a special kind of time series forecasting problem, and its own nonlinearity and uncertainty lead to many factors to be considered when making stock forecasts. Temporal convolutional network shows good prediction results in stock prediction. In order to better solve the challenges in stock prediction, enable investors to gain more profit in the stock market, and further enhance the stock prediction ability of TCN, this thesis proposes two improved methods for the temporal convolution module and the connection method in TCN. The data set constructed by the top 10 stocks of Shenzhen A-share was used to conduct the ablation experiment on the weekly line ma_5, monthly line ma_20 and seasonal line average price prediction of stocks. The experimental results show that the two improved time convolution networks have improved the prediction accuracy of the three moving averages, and have strong practical value. The main content of this article:

First, a stock mean price prediction model Res-TCN based on multiscale convolution and residual connectivity is proposed. The multi-scale convolutional Inception structure is added to the Temporal block of the temporal convolutional module of the standard TCN. The multiple convolutional kernel structures in the multi-scale convolutional make the features extracted by the model more complete and enhance the extraction of features by the model. Immediately after expanding the number of layers of the network, the TCN blocks of each layer are connected together using residual connections, which help prevent gradient degradation during network training. The prediction accuracy of the Res-TCN network designed in this thesis is 70.58%, 72.14%, and 74.53% for weekly ma_5, monthly ma_20, and quarterly ma_120, respectively, which improve 5.89%, 0.62%, and 1.98%, respectively, relative to the TCN, and obtain a good prediction result.

Secondly, in order to make the model more efficient in prediction and better in feature extraction, a network model Dens-TCN-GAM based on global attention and dense connectivity is proposed. The global attention module GAM is added to each layer of Temporal block to form a TCN-GAM block, and the TCN-GAM block of each layer is convolved with 3×3 as the conv block in the Dense block to maximize the

ability of the network to utilize the features. After the experiments using Adam optimization algorithm, the prediction accuracy of the improved model Dens-TCN-GAM for the three mean lines is 75.48%, 77.33%, and 79.05%, which is 4.90%, 5.19%, and 4.52% higher than that of the Res-TCN network, respectively, and the prediction efficiency and accuracy are improved.

Finally, a multi-influence factor-based stock selection system is designed to enable investors to take fewer steps. Multiple impact factors affecting stocks are used as input features of the dataset, while expanding the types of stocks in the dataset and selecting different types of stocks to be added to the stock pool. After the output of the prediction model, the 10 stocks with the best weekly, monthly and quarterly predictions are selected to form the stock pool, and investors can choose according to their needs.

Key Words: Stock forecast, Time convolution network, Multiscale convolution, Attention mechanism

目 录

1. 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 线性回归预测方法的研究现状.....	2
1.2.2 机器学习预测方法的研究现状.....	3
1.2.3 深度学习预测方法的研究现状.....	3
1.3 研究内容与论文安排.....	4
1.3.1 研究内容.....	4
1.3.2 论文安排.....	5
1.4 本章小结.....	6
2. 相关理论与技术.....	7
2.1 引言.....	7
2.2 股票分析方法研究.....	7
2.2.1 股票分析的数据集.....	7
2.2.2 股票预测模型.....	9
2.2.3 股票预测方向.....	10
2.3 卷积神经网络.....	10
2.3.1 卷积层.....	11
2.3.2 池化层.....	13
2.3.3 激活函数.....	14
2.4 时间卷积模型.....	16
2.4.1 因果卷积.....	16
2.4.2 膨胀卷积.....	17
2.4.3 残差连接.....	18
2.5 注意力机制.....	20
2.6 本章小结.....	22
3. 基于多尺度卷积和残差连接的 TCN 股票均线价格预测.....	23
3.1 引言.....	23
3.2 多尺度卷积 Inception 结构.....	23
3.3 残差 ResNet 结构.....	25
3.4 多尺度卷积融合和残差连接的 TCN 网络结构.....	26
3.5 实验结果与分析.....	28
3.5.1 实验流程框架.....	28
3.5.2 实验数据集.....	29

3.5.3 实验数据预处理.....	30
3.5.4 实验平台	31
3.5.5 实验评价指标.....	31
3.5.6 结果与分析.....	32
3.6 本章小结.....	37
4. 基于全局注意力机制和密集连接的 TCN 股票均线价格预测	38
4.1 引言.....	38
4.2 注意力模型 GAM	38
4.3 密集 DenseNet 结构.....	40
4.4 全局注意力机制和密集连接的 TCN 网络结构	41
4.5 实验结果与分析.....	43
4.5.1 实验优化算法.....	43
4.5.2 结果与分析.....	43
4.6 本章小结.....	48
5. 基于多影响因子下的量化选股策略	50
5.1 引言.....	50
5.2 量化指标数据.....	50
5.2.1 基础指标数据.....	50
5.2.2 技术指标数据.....	51
5.2.3 新闻指标数据.....	54
5.3 多影响因子选股系统的构建	55
5.4 实验结果与分析.....	56
5.4.1 实验流程.....	56
5.4.2 结果与分析.....	57
5.5 本章总结.....	60
6. 总结与展望	61
6.1 总结.....	61
6.2 展望.....	62
参考文献.....	63

1. 绪论

1.1 研究背景及意义

股票起初源自于西方发达国家，在股票的发展初期，没有正规的股票市场交易所，这样使得股票不便于交易流通。大约 17 世纪时，荷兰成立了世界上第一个最早的股票交易所，股民可以通过股票交易所进行股票的买入和卖出。到达 18 世纪后，随着有关股票交易制度的颁发和市场监管部门监管力度的增加，股票市场进入鼎盛时期，在这其中，世界上赫赫有名的美国纽约华尔街纽约证券交易会进入大众的视野之中。

我国的股票市场形成于 1989 年，在开设股市试验点后，就吸引了国内业内人士和大众投资者的关注^[1]。随着股票市场的规模扩大，关于股票的研究也越来越成熟。对于投资者来说，准确预测出股票的发展趋势从而获得收益，或者是可以把握股票的波动情况，从而在投资时规避风险，这些都显得尤为重要。由于许多大众投资者把目光都投入股票的市场中，股票的走势变得错综复杂，出现了各种影响股票的因素，许多目光投向了股票走势的影响因素上面。

随着全球经济的快速发展，股票可以看作宏观经济的写照，因为股票市场行情往往随着国家的宏观经济政策、上市公司的财务状况、新闻舆论和投资者情感的变化而变化^[2]。所以股票具有很大的波动性和不稳定性，现在股票预测问题也成为许多学者们研究的焦点。进入 21 世纪后，人工智能逐渐出现在大众的视野中，许多研究领域都使用了人工智能的技术。在人工智能中，深度学习最具有代表性，深度学习不仅使用在图像处理问题中，还广泛的应用到时间序列预测领域，而恰恰股票就属于一种典型的时间序列。深度学习里面的神经网络可以将股票的历史趋势特征反复学习后，进行训练输出网络对股票未来预测的趋势。现如今，许多研究者都在使用神经网络算法来对股票进行研究，股票预测的快速发展，加快对股票研究的脚步。国内股票市场的面向群体大多为中小投资者，这些中小投资者缺乏投资经验，缺少专业的知识容易盲目的跟风操作，不容易把握股票的趋势，最后导致在股票市场中亏损。

在进行股票的研究时，可以将股票的周期划分为短期、中期和长期，分时间段来看股票的发展趋势，从股票的均线可以看出股票每段时间的价格变化。本文所设计的股票均线的预测是基于股票的三种均线来分析股票的每个时期的价格趋势，使用改进的时间卷积模型对均线进行预测，为普通散户投资者提供投资参考，同时设计了量化选股的策略系统，投资者可以选择合适的股票池进行交易。总体来说，本文给投资者更加可靠的股票投资意向，具有很强的实用价值。

1.2 国内外研究现状

股票作为一种非平稳的时间序列数据,发展到迄今为止,它的变化受多种因素的影响,对于股票预测研究可以从多个角度来进行。股票的预测方法也在不断的更新变化,由传统数学模型里的线性回归预测方法到机器学习,再到现如今使用最多的深度学习。各种预测模型不断优化,它们都具有各自的特点,本章对这些股票预测方法的国内外研究现状进行逐一介绍。

1.2.1 线性回归预测方法的研究现状

传统线性回归模型可以作为股票预测的一种方法,其中包括自回归移动平均模型 ARIMA^[3]和广义自回归条件异方差模型 GARCH^[4]等,这两种模型是经典的线性回归预测模型^[5-6]。线性回归模型在进行时间序列预测的建模时,先判断出预测序列是否平稳,然后进行差分运算,紧接着设置相关性函数,最后通过构建白噪声对模型进行验证。线性回归模型进行股票预测的方法现如今依旧被国内外研究者所使用^[7]。

Junior 等人^[8]使用 ARIMA 模型对巴西股票指数 (Ibovespa) 的未来变化进行预测,通过实验后发现 ARIMA 模型比其他线性回归模型取得了更低的 MAPE 值,从而表明了 ARIMA 在股票预测方面有很大的实用性。Zheng 等人^[9]利用 ARIMA 和干预分析技术来识别影响滞后时间的过程,通过在酒店模块和标准的普尔 500 指数的对比实验后,验证了不同种类的股票之间的存在相互影响作用。Miah 等人^[10]使用不同滞后顺序的 GARCH 模型对四家孟加拉公司股票的每日收益情况进行预测,实验结果表明了 GARCH 模型对股票的每日收益预测较为优秀。Sun 等人^[11]将两种线性回归的模型和人工神经模型进行融合成 ARMA-GARCH-NN 模型,通过对美国股市的高频数据进行实证分析,证明了 ARMA-GARCH-NN 在大规模财务数据中的实用性。高天^[12]先使用最优小波包变换将股票数据进行分解后放入 ARIMA 与 SVR 相结合的预测模型中,通过在深证 A 股的一只股票实验中证明了此方法的有效性。吴玉霞等人^[13]使用 ARIMA 模型以华泰证券的收盘价作为数据集对创业板模块的股票趋势进行预测,实验在股票的短期预测取得了不错的效果。熊政等人^[14]通过融合线性回归模型 ARIMA-GARCH-M 模型后,在预测短期的收盘价上具有不错的效果。综上所述,这两种经典的线性回归模型用于解决非平稳的线性时间序列问题时,是可以完成股票进行预测任务。但是随着股票的数据增多时,线性模型不能有效的捕捉价格信息,这时模型的预测效果一般,所有需要更加深层的模型才可以提高提取特征的能力,从而提高模型预测的准确率。

1.2.2 机器学习预测方法的研究现状

随着股票的数据量剧增,传统的线性模型无法容纳大量的数据^[15],如何寻找可以处理大数据的模型成为研究者的研究方向。随着机器学习的发展,越来越多的机器学习模型呈现在大众面前,其中最为常用的是支持向量机 SVM^[16]模型,还有随机森林 RF^[17]和决策树 Decision Tree^[18]等模型。机器学习的优点是可以处理复杂的大量数据^[19-20],而且对于传统的线性回归模型,这些机器学习模型可以拟合非线性的数据。在模型预测的效果来说,机器学习预测股票的准确率更高^[21-23]。

Trafalis 等人^[23]将 Cortes 和 Vapnik 在 1995 年开发的 SVM 技术用于股票预测的实验中,他们使用 SVM 与径向基函数 RBF 在股票价格预测的实验中,发现 SVM 在预测股票的效果优于其他的预测网络。郭辉^[24]使用支持向量 SVM 机对股票的大盘和个体股走势进行训练,将模型与 DMDI 模型融合到 SVM 中,实验表明了带有 DMDI 的模型预测中短期股票更有效果。闫政旭等人^[25]使用 Pearson 特征选择对 RF 模型进行改进,实验在股票价格的短期预测上面表现的效果优于标准的 RF 模型。林娜娜等人^[26]使用相关性分析对预测指标进行二次筛选后,使用 RF 对 A 股的涨跌进行预测,实验证明了 RF 比二元 logistic 回归预测准确度更高。Khaidem 等人^[27]使用 RF 预测苹果、亚马逊和微软的股的日线价格,实验得出了不错效果。虽然机器学习可以在股票预测的实验中得到不错的效果,但是当加大股票的历史价格数据时,机器学习的预测效果和预测效率会大大降低,大数据时代的到来也间接的推进了股票预测的进程。

1.2.3 深度学习预测方法的研究现状

随着股票数据量日益增多,其中股票衍生出几十种衍生数据,机器学习无法对大量的数据进行学习,迫切需要一种需要容纳大量数据的学习技术,深度学习由此诞生。起初使用深度学习对股票预测的模型为人工神经网络 ANN,随着循环神经网络 RNN 的出现,许多由 RNN 发展出来的模型出现在大众面前,比如长短时间记忆网络 LSTM。

ANN 作为一种深度学习的预测方法,也可称为多层感知器模型 MLP^[28]。De Oliveira 等人^[29]使用改进的人工神经网络 ANN 对巴西国家石油公司股票收盘价进行预测,取得了不错的效果。Dhar 等人^[30]使用 MLP 模型来预测股票的未来走势,但 MLP 在预测时存在梯度消失等问题,而且模型往往会训练出现过拟合现象^[31]。随着网络的快速发展,RNN 在解决时间序列问题上表现的尤为突出^[32],在此基础之上所设计的改进模型 LSTM,它可以防止 RNN 存在的梯度消失问题

[33]。Back 等人^[34]使用具有预防过拟合模块和预测模块 LSTM 组成的新模型对两种不同数据集下都变现了比较好的效果，伴随着卷积神经网络 CNN 在处理图像分类任务的大放光彩^[35-36]，出现了 CNN 与 LSTM 相结合的 **CNN-LSTM 模型**。Kim 等人^[37]在使用 CNN-LSTM 模型对标准普尔 500 指数的预测效果优于单一模型的效果，CNN 有助于提升 LSTM 无法提取的单一特征，两者结合可以缩短时间序列，可以是 LSTM 模型更快、更好的学习到数据的特征。时间卷积网络 TCN 首次由 Lea 等人^[38]在 2016 年提出，结合 RNN 与 CNN 两种网络的优点，用于视频分割任务。LSTM 在长时间序列预测时占用大量内存，从而出现模型运行时间长、处理效果差等缺点。Bai 等人^[39]提出一种通用的时间卷积网络 TCN，它的预测效果优于 LSTM 和 GRU，网络的结构也更加简单清晰。Guo 等人^[40]使用改进的 **TCN 模型**去预测上证 A 股前 50 只股票的效果优于 MLP、LSTM 等模型，准确率和误差表现比较好的效果。

1.3 研究内容与论文安排

1.3.1 研究内容

在股票预测的研究中，研究者大多数都是探究如何改进股票预测的模型，很少关注预测的对象。而股票的预测对象多种多样，其中有对股票的价格直接进行预测，也有将股票的未來走势划分上涨、平稳和下跌从而预测股票的趨勢。由此可知，可以从两个方面进行股票预测的研究，一方面是从股票预测模型，另一方面是股票的预测对象。

目前在股票价格预测中最具有代表性的股票价格就是股票的收盘价，由于股票的收盘价可以看作股票在当天交易时间截止后最后一笔交易的价格，它可以反映股票当天的价格波动情况^[41]，是一个具有短期价格的趋势的参考值。在股票交易中，短期交易不能完全反映股票的市场，中期和长期也是股票预测需要研究的方向。在股票的市场数据中，可以反映出短期、中期、长期的价格趋势的变化量为股票的均线，它也被称为股票的移动平均线，期限对应的均线值为周线、月线、季线。本文所研究的就是以均线价格来预测股票的趨勢情况。

本文的具体工作如下：

1.以时间卷积网络为基础模型，对标准时间卷积模型进行了两种方法的改进，一种是将多尺度卷积与 TCN 融合后使用残差连接的方式设计了 Res-TCN。另一种是将全局注意力机制 GAM 加入每一层的 TCN 模块中，使用密集连接设计了 Dens-TCN-GAM 模型。将沪深 A 股代码前十只股票的 2011 年 1 月 4 日到 2022 年 1 月 4 日的历史日线 OHCL 和三种周线的价格作为实验数据集，将传统模型

和 TCN 网络与改进后的 TCN 进行消融实验,得出改进后的模型在预测股票的均线价格上表现的更加出色。

2.对实验股票种类进行扩充,加入多种类股票数据组成股票总池,在数据集中加入量化指标数据作为股票的影响因子,使用预测模型对股票总池的股票进行预测,输出预测准确的股票作为股票的选股池。股民可以通过对不同市场的交易需求进行选股,短期股民可以观察近周线的股票历史价格来进行交易。中期和长期的股民通过月线和季线两条均线的价格趋势判断出股票未来的走势情况,从而使收益达到最大化。

1.3.2 论文安排

本文的论文安排主要分为六章:

第一章:首先对股票预测的研究意义及背景进行介绍,从线性回归、机器学习、深度学习三个方面介绍股票预测当今的研究现状,最后介绍本文的研究内容和论文安排。

第二章:主要介绍股票预测的相关的理论和技術,先介绍了股票分析的数据集、股票预测模型、预测方向的股票分析方法研究。然后对卷积神经网络每层的作用,同时介绍时间卷积模型的组成进行详细的介绍。最后介绍注意力机制的原理和在神经网络中的作用。

第三章:介绍多尺度卷积 Inception 结构、残差 ResNet 结构,将两者融合到时间卷积网络中去,设计了多尺度卷积融合和残差连接的 Res-TCN 模型。改进的模型提高了特征的提取,加深了网络的层数,使模型在预测股票的均线价格时表现的更为出色。

第四章:首先介绍了全局注意力模型 GAM,将他与其他的注意力模型对比,紧接着介绍密集连接 DenseNet 结构。然后将全局注意力模型加入每一层的 TCN 模块中使用密集连接构成改进模型 Dens-TCN-GAM。最后通过消融实验证明了改进模型的预测效果更好。

第五章:介绍了基础、技术、新闻三种量化指标数据,将它们加入股票预测的数据集中组成多影响因子的数据集。构建了多影响因子的选股系统,通过加入多种类的股票组成股票总池,使用预测模型输出每只股票的预测误差,剔除预测效果不好的股票。最后将挑选预测效果好的股票组成选股股票池。

第六章:对本论文进行工作总结。总结本文所设计的改进的时间卷积模型在股票的均线价格上面的优点和不足,对未来的股票预测研究的展望。

1.4 本章小结

本章为论文的绪论。首先对论文的研究背景及意义进行介绍，紧接着介绍三种股票预测的方法国内外研究现状，详细的阐述了各种预测的方法的不足从而说明本文研究方法的优点，包括研究的模型和股票预测的对象。最后说明本文的研究内容和安排。

2. 相关理论与技术

2.1 引言

股票预测是一个庞大的深度学习任务，由于股票属于时间序列，其预测问题一直为深度学习中热门焦点问题。对股票预测需要提前了解股票相关的理论和技术，比如股票的数据类型、股票预测模型、股票预测方向等，选择合适的研究方法和方向可以使预测效果更好，以至于达到事半功倍的目的。深度学习的神经网络发展至今，用来处理时间序列的问题的模型层出不穷，本文使用的预测网络模型是以卷积神经网络为基础的时间卷积网络，将其与注意力模型相融合组成改进后的时间卷积网络。本章主要将以上所述的方面对所需的相关理论与技术做具体的介绍。

2.2 股票分析方法研究

2.2.1 股票分析的数据集

根据股票的大盘信息，股票可以看作为一段时间市场经济的发展情况，股票的各个板块的市场指数变化，也可以从中看出近期各种行业的上市公司的公司效益。股票发展至今，股票的分析方法有很多种方向，它们可以从多种角度对股票市场进行分析，通过股票带有的数据信息可以从中分析出股票的发展方向。其中股票的数据集可以分为市场的基础数据集、股票的文本数据集、宏观经济数据集、知识网络数据集、图像数据集、意向数据集等。

在股票的分析方法中，常见使用最多的就是股票的市场数据集，一般一只股票具有四种基础数据，它们分别为开盘价 `open`、最高价 `high`、收盘价 `close`、最低价 `low`，也可以将它们简称为 `OHCL`。由基础数据可以比较直观的观察出股票的具体情况，在股票预测的研究中，常见的预测对象是将目标预测数据的期望预测值设定为股票第二天的收盘价。将模型预测的第二天的收盘价值与真实的收盘价值进行对比，从而来衡量预测模型的预测效果。通过收盘价的预测可以让投资者在第二天收盘之前提前了解股票的最后交易价格，这样可以在收盘价之前进行合理的交易，从而中获得更大的收益，这也是所有预测模型的最终目的。由基础数据可以衍生出多种衍生数据，比较常见的衍生数据有均线 `MA`、异同移动平均线 `MACD`、相对强弱指标 `RSI` 等数据。从这些衍生数据可以看出股票的涨幅、跌涨等数据信息，可以在 `python` 库函数中调用相应的衍生数据。图 2.1 为 `choice` 金融终端软件中显示的股票市场数据信息图。图中为中国深圳证券交易所，简称

“深证 A 股”，其代码前 10 只股票在一天中的交易市场基本情况。

序号	代码 ↑	名称	●	最新	涨幅%	涨跌	成交量	现手	涨幅%	换手%	成交额	最高	最低	今开
1	000001	平安银行	↑	12.57	2.61	0.32	211万	14267	-0.08	1.09	26.53亿	12.75	12.23	12.27
2	000002	万科A	↑	17.01	1.31	0.22	71.12万	7612	0.00	0.73	12.13亿	17.30	16.63	16.75
3	000004	ST国华	*	9.75	0.62	0.06	2.44万	485	-0.10	2.09	2390万	9.97	9.64	9.64
4	000005	ST星源	↑	1.84	1.10	0.02	5.80万	320	0.00	0.55	1066万	1.85	1.81	1.82
5	000006	深振业A		4.13	0.73	0.03	8.73万	774	0.00	0.65	3623万	4.18	4.09	4.11
6	000007	全新好		8.16	4.08	0.32	33.15万	3506	0.12	10.73	2.64亿	8.33	7.59	7.78
7	000008	神州高铁	↑	2.48	0.40	0.01	16.34万	2912	0.00	0.61	4056万	2.50	2.46	2.46
8	000009	中国宝安	↑	16.39	-4.49	-0.77	120万	15453	0.00	4.69	20.01亿	17.33	16.37	16.91
9	000010	美丽生态		3.42	-0.58	-0.02	9.19万	721	0.00	1.76	3161万	3.50	3.40	3.44
10	000011	深物业A		10.41	1.76	0.18	7.67万	954	0.00	1.46	8000万	10.54	10.20	10.23

图 2.1 choice 金融终端软件中显示的股票市场数据信息图

Fig. 2.1 Stock market data information displayed in choice financial terminal software

股票的文本数据集是仅次于股票市场数据的第二种股票分析方向的数据信息，文本数据一般是由社交媒体、网络搜索、新闻等信息组成。这些信息可以组成一个数据量大的数据库中，通过文本挖掘的方法可以从数据库中将带有股票的文本数据挖掘出来。文本信息可能包括一段时间内股票的舆情变化，或者具体的政策信息导致的股票的市场变化，还有投资者自身的投资情绪，包括积极的和消极的情绪因素也会导致股票出现变化。在股票交易的软件中，投资者可以通过自己所选择交易的股票栏目中观察出股票的公司资讯信息，也可在贴吧、微博、论坛等网络社交网站中查询和讨论所选股票的其他股民的投资情况。图 2.2 为 choice 金融终端软件中平安银行的公司资讯信息，图中可以看出平安银行在近期的文本数据和银行行业的新闻信息。

公司资讯				更多>>	行业资讯				更多>>
08-20	[资讯]	“零售之王”业绩超预期！	银行中报密集披露 机构		08-20	[资讯]	苏州银行08月19日获深股通增持155.68万股		
08-19	[资讯]	平安银行信用卡2022年半年报：“五位一体”新模			08-20	[资讯]	齐鲁银行08月19日获沪股通增持143.56万股		
08-19	[资讯]	平安银行零售三大业务持续升级 打造零售智能银行3			08-20	[资讯]	杭州银行08月19日获沪股通增持173.27万股		
08-19	[资讯]	北向资金今日净买入17.13亿元 抢筹平安银行、五粮			08-20	[资讯]	江阴银行08月19日获深股通增持48.18万股		
08-19	[资讯]	天风证券给予平安银行买入评级 净利润维持高位 资			08-20	[资讯]	长沙银行08月19日被沪股通减持43.81万股		

图 2.2 choice 金融终端软件中平安银行的公司资讯信息

Fig. 2.2 Company information of PingAn Bank in choice financial terminal software

股票的宏观经济数据集一般包括一个国家的地区或者部门的经济的发展情况，在宏观经济数据中，通过定义带有经济的数学表达式，比如 GDP、CPI 等国际通用的数据。从一个方面上看，这些代表宏观经济的数据可以表示股票发展的将康

情况,也可在其中发现股市的趋势变化。不同公司之间可以构建知识图来将共同行业或不同行业的联系组织到一起,在研究中可以通过知识图构建知识图网络。在国外的最新研究中,将开源知识图的信息作为大众所研究的数据,也是一种股票分析数据的方向。

在深度学习任务中,图像处理任务一直是发展最为迅速的一个方向,在股票数据集中,K线图作为股票特有的图像数据集,它在卷积神经网络中被广泛使用。K线图也被称为烛台图,在K线图中,投资者可以清晰的看出一只股票在当日线、周线、月线、季线、年线中的发展趋势。由K线图的形状中,通过收盘价与开盘价之间的关系进行定义图像的颜色,若收盘价比开盘价高时图像为红色,反之则为黑色。图2.3为choice金融终端软件中平安银行的K线图。从图2.3中可以看出每日平安银行K线图的具体情况,由此可以得出平安银行的发展趋势。

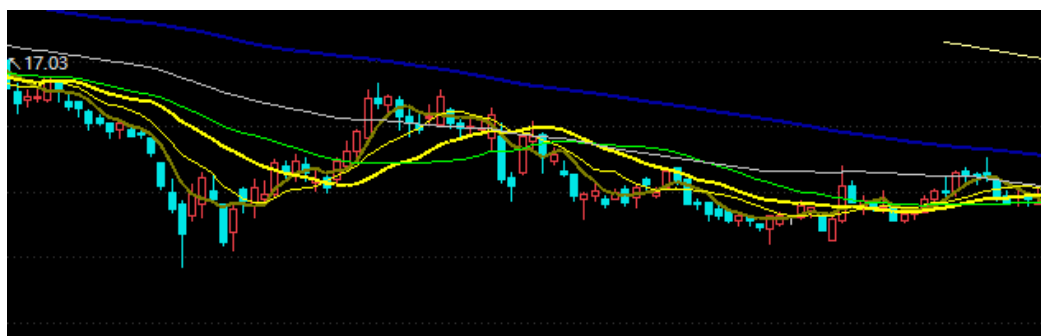


图 2.3 choice 金融终端软件中平安银行的 K 线图

Fig. 2.3 K-line diagram of PingAn Bank in choice financial terminal software

股票的意向数据集一般是指证券分析师或股票公司在报告中给予用户的买入和卖出的投资建议,数据集中包括对公司的商业交易的模式、销售活动等信息对股票的发展趋势分析数据。投资者可以通过购买分析数据来对股票做出明确的交易,虽然这些分析数据往往价格不菲,但也可以作为股票分析数据的一种。

本文使用的股票数据为股票市场数据中的 OHCL 和衍生数均线 MA,均线 MA 是通过公式变化后转换成所需要的均线数据,其中包括周线 ma_5、月线 ma_20、季线 ma_120。

2.2.2 股票预测模型

股票的数据属于时间序列,在时间序列预测模型中,大多数预测模型都为监督学习,只有在少数的预测模型使用半监督学习,模型基本可以分为以下三种^[41]。

第一种为标准模型，基本是以深度学习的为基础模型，包括 FFNN、CNN、RNN^[42]等模型，其中卷积神经网络和循环神经网络是使用最为广泛的两种基础模型。第二种为混合模型，预测模型可以为传统的机器学习模型与深度学习进行组合后形成新的模型，一般与深度学习模型混合的机器学习模型有 SVM、KNN 等，也可与一些常用的线性预测模型相结合，如 LR、ARIMA、GARCH。第三种为其他模型，研究者经过大量的实验发现一些与传统模型和深度学习不同的模型，这些其他模型包括 GAN、GNN、强化学习、迁移学习等。本文所使用的股票预测模型是通过卷积神经网络变形得出的时间卷积网络，改进后的 TCN 网络优于传统的时间卷积模型，在预测确定股票目标数据中表现良好。

2.2.3 股票预测方向

股票预测需要提前确定预测的方向，通过模型最终的评测指标来判断模型预测的好坏程度，比较常见的评测指标有分类指标、回归指标和利润指标等。分类指标是深度学习最为常见一个重要的指标，其中包括准确率、召回率、相关系数等，可通过这些分类指标来判断股票的趋势，用上涨、平稳和下跌来表示股票的趋势情况。回归指标常用于在股票的价格预测，通过对比预测模型的预测值与真实值之间的误差来确定误差值，也可以通过设置合适的容忍度，使用准确率来进行价格预测。利润指标是预测模型用于具体量化交易时评测的重要指标，通过股票的交易策略，使用最终的股票收益率来评测模型，这是评测模型最直接的方法。一个好的模型会通过股票交易策略给投资者带来最大的收益，将投资者的风险降为最低，获取更好的回报。

本文股票预测的方向为价格预测，在预测时设置相关的步长参数来调整预测的天数。股票采用的是 T+1 模型来进行交易，考虑到需要及时了解股票的市场变化，预测模型的默认步长设置为 1，这样做可以保证股票的实时性。

2.3 卷积神经网络

卷积神经网络是由 Alexander Waibel 等人在 1989 年提出，它不仅解决了前馈神经网络在训练时产生大量参数所导致的训练效率下降问题，还解决了训练出现的过拟合现象^[43]。卷积神经网络 CNN 作为深度学习使用最为广泛的模型之一，由 CNN 所展开的各种模型在其他深度学习任务中发展逐渐成熟，它被广泛地用于分类、目标检测、图像跟踪等问题中。在深度学习的任务中，CNN 一直为研究者所研究的重点模型。现如今比较经典的常见的 AlexNet、VGGNet、ResNet、GoogleNet 等网络模型，它们都是以 CNN 为基础进行改进的。而在预测时间序列的任务中，也出现以 CNN 为基础模型进行改进，由于 CNN 在特征提取时

可以提取循环神经网络不能提取的时间序列所带有的抽象的特征。本文所使用的时间卷积网络 TCN 就是在 CNN 基础之上所改进的。标准的 CNN 一般由卷积层、池化层、全连接层^[44]等组成,有时还可能会加入 Dropout 层和激励层、归一层,在训练时加大模型泛化的能力从而防止出现过拟合现象。图 2.4 为卷积神经网络的结构图。

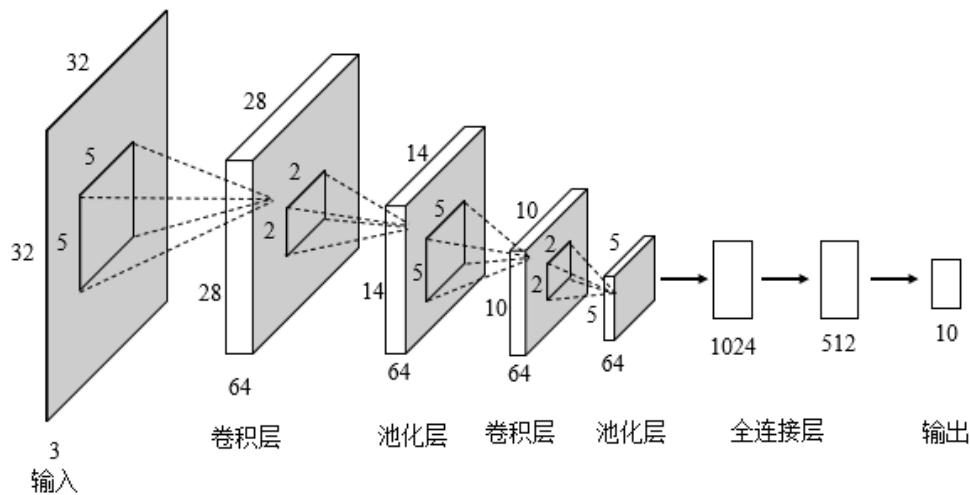


图 2.4 卷积神经网络的结构图

Fig. 2.4 Structure diagram of convolutional neural network

2.3.1 卷积层

卷积层作为卷积神经网络最为基础的一层,它主要由卷积计算组成,在深度学习模型的卷积层中,卷积计算有一维卷积计算和二维卷积计算。卷积层是由两个神经元通过卷积核来进行相互连接,这样做的目的是可以减少训练时出现参数暴增的现象,从而在训练时损失降到最小,同时拟合的速度也变快,以此防止出现过拟合的现象。在训练时使用共享的权重^[45],通过使用不同大小的卷积核来降低网络的复杂程度,达到简化网络模型的目的。卷积核的大小多为奇数^[46-47],常见的卷积核有 1×1 、 3×3 、 5×5 、 7×7 等。

卷积作为一种重要的计算方式,在时间序列中所使用的卷积为一维卷积,也可用于处理信号^[48],一维卷积的数学表达式为:

$$y_t = \sum_{k=1}^K \omega_k x_{t-k+1} \quad (2.1)$$

而二维卷积通常使用在处理图像任务中，二维卷积的数学表达式为：

$$y_{ij} = \sum_{m=1}^M \sum_{n=1}^N \omega_{mn} x_{i-m+1, j-n+1} \quad (2.2)$$

为了简化表达式，方便更加直接的表达出卷积计算的过程，将 ω 作为卷积计算的卷积核，通过卷积的输出为 y ， $*$ 表示卷积计算，卷积计算可以简化为：

$$y = \omega * x \quad (2.3)$$

可根据实例来表示卷积层中卷积的过程，图 2.5 为二维卷积运算示意图。可以看出输入特征的大小为 4×4 ，卷积核的大小为 3×3 ，设置卷积核移动的步长为 1。在本实例中，经过二维卷积计算后，输出的大小为 2×2 ，第 i 行 j 列输出值表示为 S_{ij} ，使用的是 Relu 激活函数，其计算过程为：

$$S_{11} = \text{Relu}(1 \times 1 + 0 \times 0 + 1 \times 1 + 0 \times 1 + 1 \times 1 + 2 \times 0 + 1 \times 1 + 0 \times 0 + 1 \times 1) = 5 \quad (2.4)$$

$$S_{12} = \text{Relu}(0 \times 1 + 1 \times 0 + 1 \times 1 + 1 \times 1 + 2 \times 1 + 1 \times 0 + 0 \times 1 + 1 \times 0 + 0 \times 1) = 4 \quad (2.5)$$

$$S_{21} = \text{Relu}(0 \times 1 + 1 \times 0 + 2 \times 1 + 1 \times 1 + 0 \times 1 + 1 \times 0 + 2 \times 1 + 2 \times 0 + 1 \times 1) = 6 \quad (2.6)$$

$$S_{22} = \text{Relu}(1 \times 1 + 2 \times 0 + 1 \times 1 + 0 \times 1 + 1 \times 1 + 0 \times 0 + 2 \times 1 + 1 \times 0 + 2 \times 1) = 7 \quad (2.7)$$

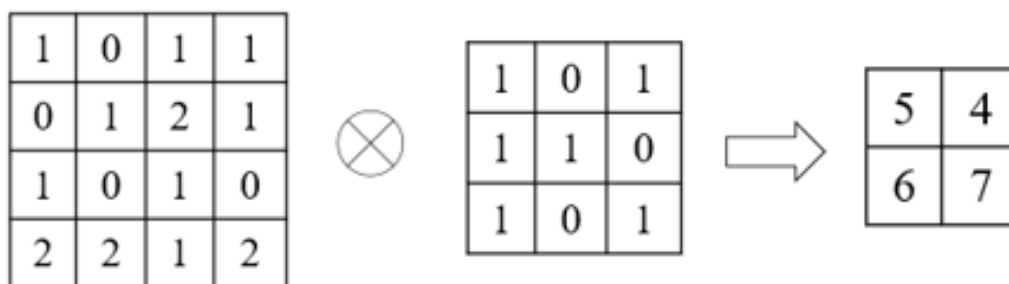


图 2.5 二维卷积运算示意图

Fig. 2.5 Schematic diagram of two-dimensional convolution operation

一维卷积多用于处理时间序列任务中，图 2.6 为一维卷积示意图。图中一维卷积的输入特征大小为 5，卷积核大小为 3，移动步长为 1，卷积核在一维卷积中只进行横向运算。经过卷积计算后的输出特征为 3，第 n 个输出的特征序列值

为 S_n ，其计算过程为：

$$S_1 = \text{Relu}(1 \times 1 + 0 \times 0 + 1 \times (-1)) = 0 \quad (2.8)$$

$$S_2 = \text{Relu}(0 \times 1 + (-1) \times 0 + 1 \times 1) = 1 \quad (2.9)$$

$$S_3 = \text{Relu}((-1) \times 1 + 1 \times 0 + (-2) \times 1) = 0 \quad (2.10)$$

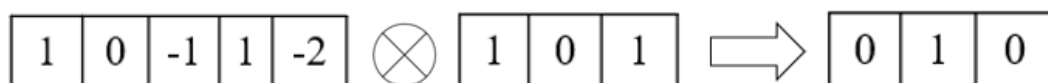


图 2.6 一维卷积运算示意图

Fig. 2.6 Schematic diagram of one-dimensional convolution operation

2.3.2 池化层

输入特征经过卷积层后，需要进行进一步的采样处理，故池化层也叫采样层。为了使模型具有更少的参数，不能通过减少神经元的方式。如果不采用池化层里的采样操作，可能会导致输出的维度过高，从而造成模型出现过拟合的不良后果，所以需要在连接 Softmax 之前加入池化层。池化分为最大池化和平均池化，最大池化是用一个区域内的所有神经元的最大活性值用来表示这个区域的总体^[49]，最大池化的表达式为：

$$y_{m,n}^d = \max\{x_i\}, \quad i \in R_{m,n}^d \quad (2.11)$$

式中， $R_{m,n}^d$ 代表一片区域， x_i 为区域 $R_{m,n}^d$ 里的神经元的活性值。

平均池化是用一个区域内的所有神经元的平均活性来表示这个区域的总体，平均池化的表达式为：

$$y_{m,n}^d = \frac{1}{|R_{m,n}^d|} \sum_{i \in R_{m,n}^d} x_i \quad (2.12)$$

为了更加清晰了解最大池化和平均池化的过程，图 2.7 为池化的过程示意图。图中输入特征的大小为 3×3 ，卷积核的大小为 2×2 ，步长为 1。

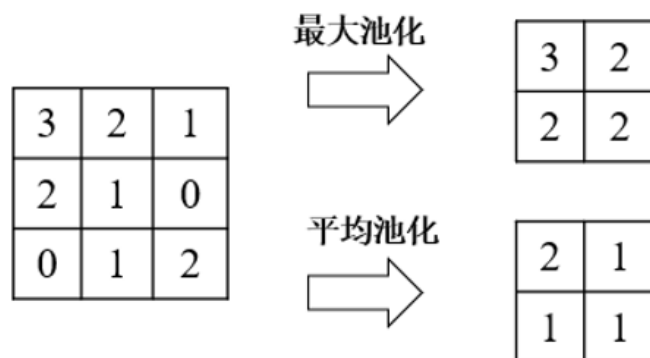


图 2.7 池化的过程示意图

Fig. 2.7 Process diagram of pooling

2.3.3 激活函数

在深度学习神经网络里面的神经元中，可以通过定义合适的激活函数来使模型输入的数据在偏置运算和特征权重后整合到同一函数上。通常在神经网络中的计算为线性计算，在网络的输出前加入激活函数使神经网络计算出的线性数据具有非线性的特征，这样可以使网络能够拟合更多的数据，从而提升网络提取特征的能力，同时也大大增加了网络的非线性表达。常见的激活函数如下：

(1) Sigmoid 函数

Sigmoid 函数可以使输入的数组映射到 0 到 1 之间，在神经网络中早期用于二分类中。其函数表达式为：

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2.13)$$

式中， x 为输入值， e 为自然对数。 x 趋近于无穷小时，输出值无限趋近于 0； x 趋近于无限大时，输出值无限趋近于 1。

图 2.8 为 Sigmoid 函数图像。输出值图像中的 Sigmoid 函数为平滑函数，平滑函数有益于 Sigmoid 函数进行求导计算，在训练时加快模型的梯度下降的速度。随着神经网络的层数增加，Sigmoid 函数无法用于隐含层，网络进行误差反向传播的同时，Sigmoid 函数容易使梯度变得越来越小，从而导致梯度消失的现象。

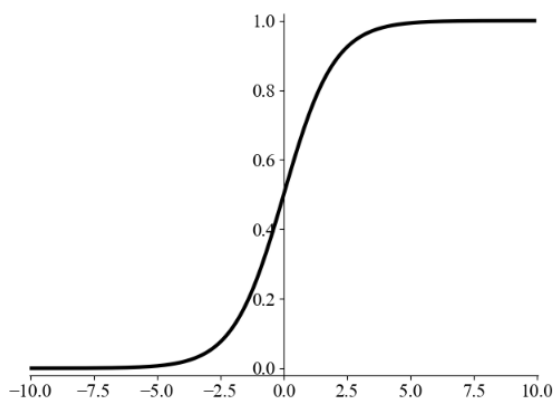


图 2.8 Sigmoid 函数图像

Fig. 2.8 Sigmoid function image

(2) tanh 函数

tanh 函数可以将输入的值映射到-1 到 1 之间，其表达式为：

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.14)$$

式中，x 趋近于无穷小时，输出值无限趋近于-1；x 趋近于无限大时，输出值无限趋近于 1。

图 2.9 为 tanh 函数图像。tanh 函可以看做一种特殊的 Sigmoid 函数，与 Sigmoid 函数相比，tanh 函数以 0 为中心的输出使用的更加频繁，其可以减慢梯度减小的速度，但是在特殊的情况下，tanh 函数也会出现梯度消失的问题。

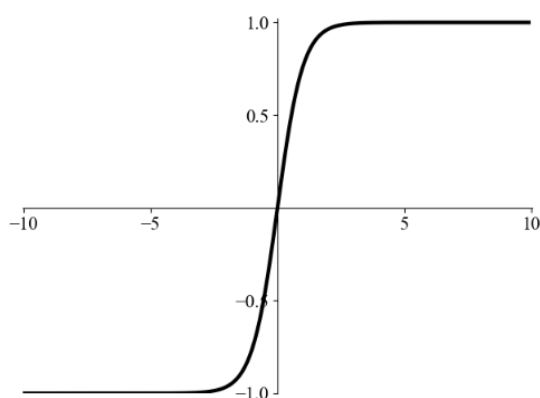


图 2.9 tanh 函数图像

Fig. 2.9 Tanh function image

(3) Relu 函数

相比于其他激活函数，Relu 函数现在广泛作为神经网络的激活函数，其表达式为：

$$\text{Relu}(x) = \max(0, x) \quad (2.15)$$

图 2.10 为 Relu 函数图像，它比 Sigmoid 和 tanh 激活函数计算更加简单，模型运算的速度更快，可以使模型具有更快的收敛速度，所以一直沿用至今。

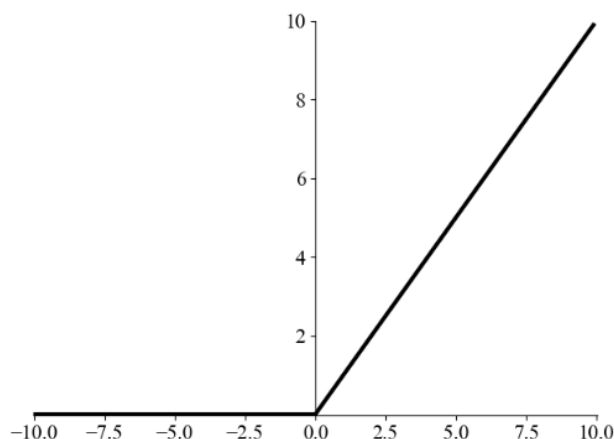


图 2.10 Relu 函数图像

Fig. 2.10 Relu function image

2.4 时间卷积模型

TCN 网络是为了解决 LSTM 在进行长时间序列预测时产生特征累积从而造成模型占据大量的运行内存而发展出来的。相比于其他的时间序列预测模型，TCN 在深度网络中更适合序列的预测，由于 TCN 使用的是共享卷积核，所以运行时占用内存少，训练的更快。TCN 网络主要由三个部分组成，它们分别为因果卷积、膨胀卷积、残差连接。

2.4.1 因果卷积

因果卷积是时间卷积网络的基础结构之一，它的结构与 WaveNet 网络中的因果卷积类似。在 WaveNet 网络中，网络的主要的组成为因果卷积，因果卷积以卷积神经网络为基础，传统的卷积神经网络不能直接处理时间预测问题，它可以看作卷积神经网络来模拟长短时间记忆网络。因果卷积的因果关系是在某一时刻的预测只与这一时刻之前有关，与未来的任何时刻无关，训练时前一个样本预测

值直接用于下一个样本预测，训练是串行进行的。图 2.11 为时间卷积中的因果卷积示意图。

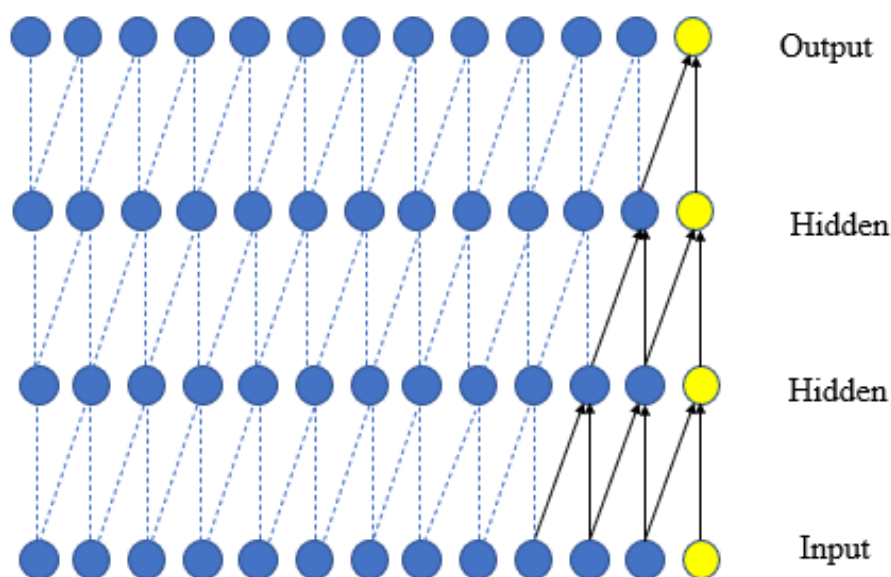


图 2.11 时间卷积中的因果卷积示意图

Fig. 2.11 Schematic diagram of causal convolution in time convolution

在图 2.11 中，输入序列为 $X = (x_0, x_1, x_2, \dots, x_{T-2}, x_{T-1}, x_T)$ ，输出序列为 $Y = (y_0, y_1, y_2, \dots, y_{T-2}, y_{T-1}, y_T)$ ，可以明显的看出是由输入序列 X 经过几层隐藏层去预测输出序列 Y 。隐藏层中的滤波器 $Z = (z_0, z_1, z_2, \dots, z_{N-2}, z_{N-1}, z_N)$ ，则输出序列 X 关于 x_T 处的因果卷积为：

$$(Z * X)(x_T) = \sum_{n=0}^N z_n x_T - N + n \quad (2.16)$$

2.4.2 膨胀卷积

由于因果卷积无法解决在处理长时间序列时每个时刻相互关联的问题，每一层经过线性的累加会加大计算量，不能根据传统的卷积网络在卷积层后加入池化层，这样会导致在运算中丢失部分原有信息。为了解决单一的因果卷积层从在传统的卷积相同的问题，膨胀卷积使用间隔采样的方法来解决此类问题，因此利用膨胀卷积可以使网络利用更少的层数来进行多特征的提取，减少因果卷积导致的

层数指数型的增长，而且可以增加长时间序列的感受野。第一层的采样间隔 Dilation 为 1 即膨胀系数为 1，依次第二层、第三层的膨胀系数依次为 2 和 4，由此得出第 n 层的膨胀系数为 2^n 。在标准的时间卷积网络中的层数为 3 即膨胀系数 Dilation 为 4。图 2.12 为时间卷积中的膨胀卷积示意图，时间卷积的每一层 T 时刻只与上一层有关，间隔式的采样随着膨胀系数的指数增加。

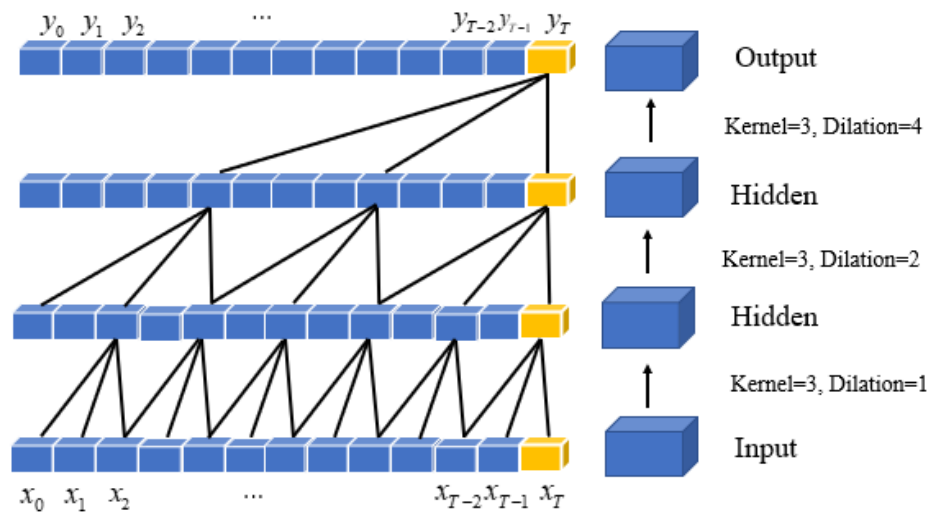


图 2.12 时间卷积中的膨胀卷积示意图

Fig. 2.12 Expansion convolution in time convolution

根据图 2.12 可知，膨胀系数 Dilation 为 d ，输出序列 X 关于 x_T 处的因果卷积为：

$$(Z * dX)(x_T) = \sum_{n=0}^N z_n x_T - (N - n)d \quad (2.17)$$

2.4.3 残差连接

残差连接是神经网络中比较常见的连接方式，对于深层次的网络，在网络增加的同时，网络训练会产生难以收敛和梯度消失的不良后果^[51]，残差连接就是解决此类问题。在 2015 年的 ImageNet 比赛中 ResNet 网络获得了比赛的冠军^[52]，至此残差连接逐渐出现在大众的视野之中。残差连接多用于解决神经网络梯度消失的问题之中，残差连接可以看作输入 x 与输入 x 的非线性变化相结合，在残差

结构中，神经网络是由输入 x 到 $F(x)+x$ 的映射的权重。在残差连接中，一般一个残差模块有两个分支，一个分支为直接映射分支 $h(x_l)$ ，而另一个分支为残差分支 $F(x_l, W_l)$ ，其数学表达式为：

$$x_{l+1} = h(x_l) + F(x_l, W_l) \quad (2.18)$$

在当残差为 0 时，网络的结构简化，但不利于网络学习更多的特征，所以一般不将残差设为 0。由此而知，残差连接相比于传统的级联连接，它更具有利于解决在网络训练时产生的网络退化现象。图 2.13 为残差连接结构示意图。

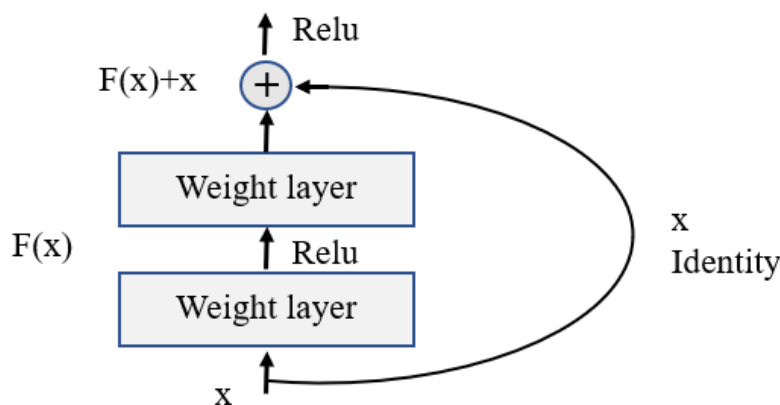


图 2.13 残差连接结构示意图

Fig. 2.13 schematic diagram of residual connection structure

在标准的时间卷积网络中，每层 TCN 网络由两个时间卷积模块 Temporal block 组成，这两个 Temporal block 模块外部采用级联方式连接。TCN 网络内部的 Temporal block 采用的是残差连接，这样做可以有效的解决模型在训练时出现梯度消失，残差连接也可使模型收敛的更快，从而增加训练时的效率。图 2.14 为每一个 Temporal block 残差连接方式示意图，残差的直接映射部分是经过 1×1 卷积后连接的，残差分支则是由两个时间卷积模块组成。时间卷积模块是由因果膨胀卷积 Dilated Causal Conv、规范化 WeightNorm、裁剪 Chomp、Dropout 组成。

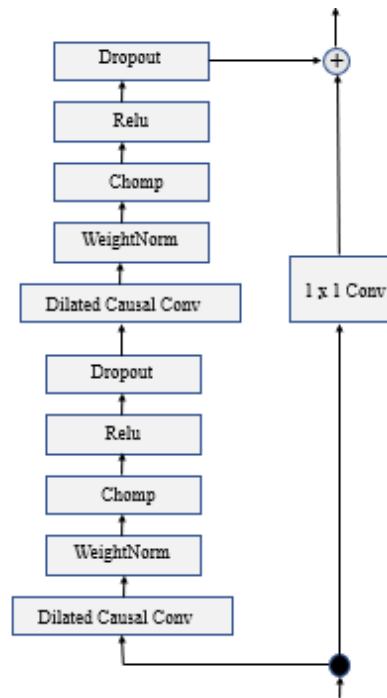


图 2.14 Temporal block 残差连接方式示意图

Fig. 2.14 Schematic diagram of residual connection of temporary block

根据图 2.14 中可知，在每一个时间卷积模块中，都使用残差连接的方式进行内部连接，残差连接在深度学习中被用于各种任务的模型中，是一种比较常用的连接方式。残差连接方式可以在训练过程中模型更新网络的参数时，防止由于网络变的深层而导致的模型梯度消失的现象，梯度消失可能导致训练参数无法更新。残差连接还有一个作用是防止网络由于层数深度增加所导致的严重退化现象，这种现象有时不一定是由于网络过拟合而产生的，可能是网络在训练时学习了不是恒等映射的参数，从而导致模型出现模型退化。

2.5 注意力机制

神经网络与生物神经系统类似，神经元储存信息和处理信息的容量是有限的，深度学习中的神经网络也是如此。正常情况下，神经网络需要存储更多的信息，这样需要增加神经元的数量来达到所需的效果，随着神经网络层数的增多，复杂程度也将变大。在网络进行训练的同时，信息量剧增带来的不良后果与人体大脑过载的情况相类似^[53]。注意力机制(Mechanism, AM)选择合适的处理信息和存储，可以有效的解决此类问题的出现，在神经网络中加入注意力机制可以提升网络自身对数据的分析能力。注意力机制一直是一个热点的研究方向，Google 公司的

DeepMind 团队在 2014 年首次将注意力机制与 RNN 相结合^[54]，取得了不错的效果。在处理深度学习图像任务中，注意力机制表现出良好的图像分类效果。图 2.15 为注意力机制的框架图，注意力机制通过权重大小来处理信息，集中处理高权重的重要信息，忽略不相关的低权重信息。对于模型来说，训练时调整更新合适的权重可以提高系统的鲁棒性^[55]。

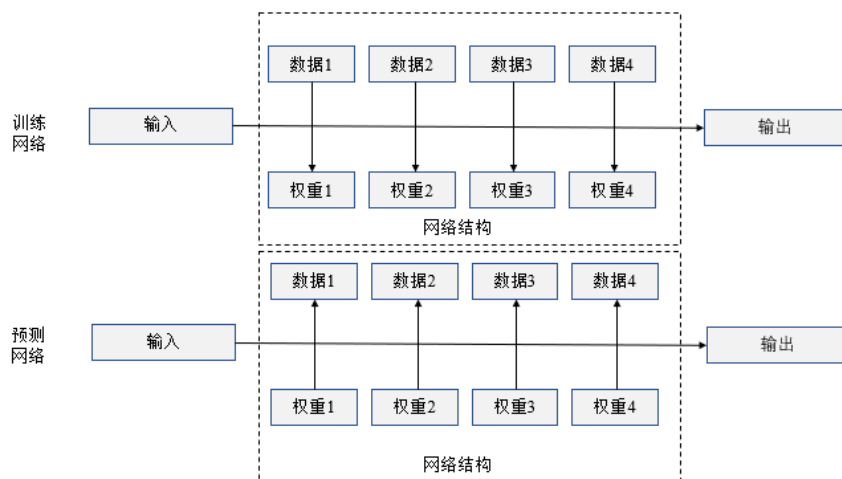


图 2.15 注意力机制的框架图

Fig. 2.15 Framework diagram of attention mechanism

卷积神经网络中的注意力机制可以分为选择注意力和自注意力机制两种类别，本文中所使用的注意力机制属于选择注意力机制，它多用于深度学习的语音、视觉任务中。在卷积神经网络里，比较有代表的选择注意力机制有 CBAM 注意力和 SE 注意力等。

在选择注意力机制中，注意力掩膜对输入数据的重要程度以达到对输入特征的是否加强或削弱，可以达到增强神经网络的非线性能力和鲁棒性的效果。注意力掩膜 $Att(x)$ 的表达式为：

$$Att(x) = F_{norm}(F_{\theta}(x)) \quad (2.19)$$

式中， x 为输入特征； $F_{norm}(x)$ 为对 x 进行归一化处理，可以由激活函数进行实现； $F_{\theta}(x)$ 为可学习函数，通过神经网络对输入 x 的学习程度来实现。

通过注意力模型输出后的信号可以看作输入信号 x 与注意力掩膜所对应元素相乘来表示，其表达式为：

$$\tilde{x} = Att(x) \odot x \quad (2.20)$$

式中， \tilde{x} 为输入 x 通过注意力模型后的输出； $Att(x)$ 为注意力掩膜。

2.6 本章小结

本章主要介绍股票预测的相关理论基础。首先介绍股票预测方法的研究，其中包括本文股票预测所使用的数据类型、预测模型、预测方向。紧接着介绍所使用是以卷积神经网络为基础的预测模型，具体介绍卷积神经网络的组成结构。最后介绍本文所使用的基础预测模型为时间卷积网络和注意力机制的原理，为本文的实验部分提供基础的消融实验神经网络模型，以便突出改进模型的优良之处。

3. 基于多尺度卷积和残差连接的 TCN 股票均线价格预测

3.1 引言

在标准的时间卷积网络中,时间卷积模块 Temporal block 采集的数据特征较为单一,只有 1×1 卷积核来进行特征提取。为了解决此类问题,本章在 Temporal block 中引入多尺度卷积的 Inception V2 结构,将 Inception V2 里面的多种卷积核来提升模型在特征提取的完整性。改进后的 Temporal block 结构相比标准 Temporal block 模型更容易避免网络出现的梯度下降现象。然后将改进后的 Temporal block 与 Inception V2 想融合产生新的 TCN block,其保留特征的完整性,提高模型在预测时准确率。为使模型更具有深层网络所具有的处理效率,对 TCN 网络框架进行拓展,将网络的层数变深,使用残差 ResNet 中的深层网络 ResNet18 结构与改进后的 TCN block 进行融合形成改进模型 Res-TCN。新模型 Res-TCN 更加适于深层网络训练强度,网络可以承受巨大的数据信息所带来的压力。在实验预设的数据集中,Res-TCN 模型表现出的股票均线的预测能力相比标准 TCN 误差小、准确率高。

3.2 多尺度卷积 Inception 结构

在神经网络中,多尺度卷积 Inception 结构源于 GoogLeNet 网络,该网络模型初次出现在大众视野中是在 2014 年的 ImageNet 比赛^[56],GoogLeNet 相对于传统模型 VGGNet 的训练参数变少。通过不同卷积核的多尺度卷积想融合的作用下,其与传统的卷积层相比更加具有对数据特征提取的效果,GoogLeNet 常用于深层的神经网络。GoogLeNet 里面的 Inception 结构现如今常用于深度学习领域中。随着研究者不断的探索研究,Inception 结构出现了多个版本,从最起初的 Inception V1 发展到 Inception V2、Inception V3、Inception V4 等版本,它们核心的思想都为 Inception 的多尺度卷积。通过不同大小的卷积核的卷积使网络提取不同大小的特征,最后使用拼接的方式连接,这样可以使网络具有更大的特征提取范围,更有利于模型的训练。其变形基本为增加网络深度所做出的改进方式,如调整卷积核的大小、添加特征池化后进行卷积、不同卷积融合等。不同版本的 Inception 结构完成不同情况下的任务。

图 3.1 为 Inception V1 结构。图中的 Inception V1 结构是由三种尺度的 1×1 、 3×3 、 5×5 卷积和 3×3 最大池化组成,输入数据通过 4 个分支后,通过 concatenation 计算融合作为输出处理层。

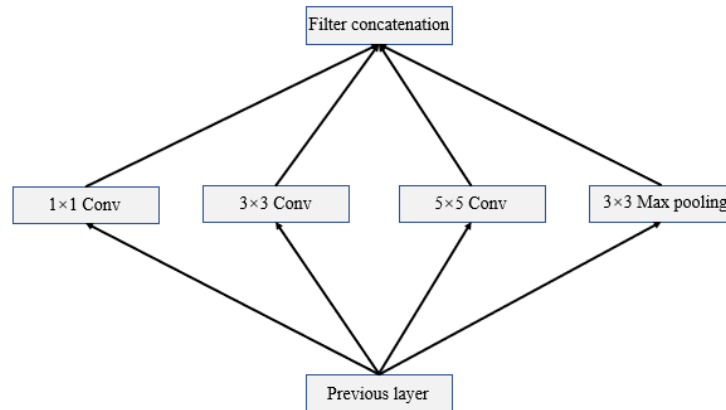


图 3.1 Inception V1 结构

Fig. 3.1 Inception V1 Structure

为解决 Inception V1 中 5×5 卷积分支中卷积核过大导致运行时造成的巨大计算量的负担, Inception V2 使用两个 3×3 卷积取代 5×5 卷积, 这种设计同时参考了 VGG 网络结构^[57]。与此同时在 3×3 卷积前面加入 1×1 卷积, 在 1×1 卷积分支前面加入平均池化, 从而起到降维的作用, 通过改变卷积核的大小来使特征采集更加完整, 加大特征的融合效果。图 3.2 为 Inception V2 结构, 相比于 Inception V1 结构, 其更具有实用价值, 本文所使用的多尺度卷积为 Inception V2 结构为基础下改进方法。

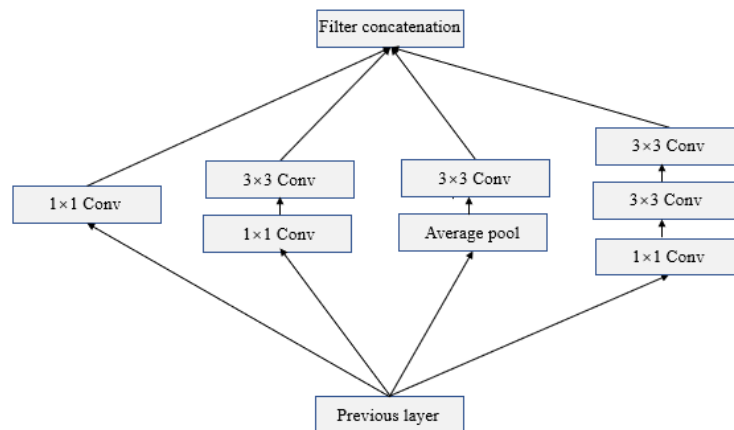


图 3.2 Inception V2 结构

Fig. 3.2 Inception V2 Structure

3.3 残差 ResNet 结构

随着机器视觉的飞速发展,卷积神经网络不断推出新的模型来完成更加复杂的任务,从最初用来完成简单的手写数据集的分类任务的 LeNet,再到在大赛 ImageNet 中大放光彩的 AlexNet,网络中首次加入 Dropout、Relu 激活函。随着网络层数的增加,神经网络深度和性能受到广大研究者所要考虑的问题,VGGNet 通过重复地累积 3×3 卷积来增加网络的层数有效地提升模型的准确率 [58]。

随着深度神经网络的迅速发展,VGGNet 网络的层数增加会出现梯度爆炸或者消失的现象 [59],从而导致网络的性能变差。ResNet 的有效的解决了以上问题,其中的恒等映射可以使网络在层数增加的同时保留每一层的信息。图 3.3 为 ResNet 中的残差模块 [52]。其中图 3.3(a)为标准的残差,图 3.3(b)为在标准残差基础之上改进后减少了训练的时间,图 3.3(c)为加入 1×1 卷积来增加网络的通道数从而可以增加模型的非线性表达和计算效率。

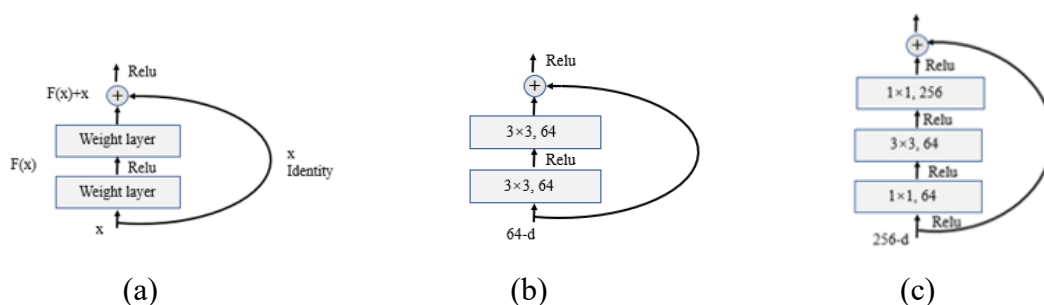


图 3.3 ResNet 中的残差模块

Fig. 3.3 Residual module in ResNet

在深层网络中,ResNet 表现出优良的性能,使用频率较高的 ResNet 深层模型有 ResNet18、ResNet34、ResNet50、ResNet101、ResNet152。表 3.1 为几种常见的 ResNet 结构,考虑到神经网络模型的承受能力和性能,所以本文使用的深层网络模型为 ResNet18 深层模型。

表 3.1 为几种常见的 ResNet 结构

Tab. 3.1 shows several common ResNet structures

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer

conv1	112×112	7×7,64, stride2				
conv2_x	56×56	3×3 max pool, stride2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				

3.4 多尺度卷积融合和残差连接的 TCN 网络结构

本章所设计的多尺度卷积融合和残差连接的 TCN 网络为 Res-TCN。与标准时间卷积网络里的时间卷积模块 Temporal block 相比，改进后的 Temporal block 模块在 1×1 卷积之后加入 BN 和 Relu，提升了整体网络的泛化能力，同时使用 Relu 作为激活函数避免网络发生梯度消失的现象^[60]。图 3.4 为改进时间卷积后的 Temporal block 模块。

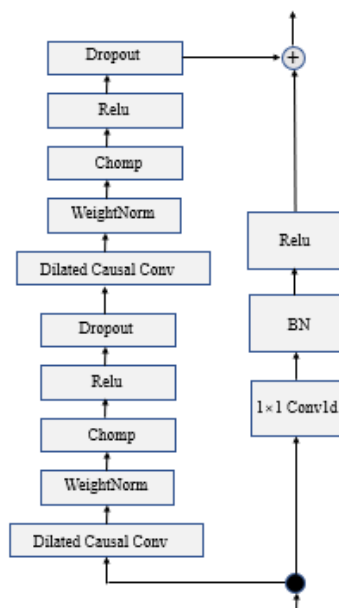


图 3.4 改进时间卷积后的 Temporal block 模块

Fig. 3.4 Temporary block module after improved time convolution

Res-TCN 的网络模块 TCN block 是使用改进的 Temporal block 模块与多尺度卷积 Inception V2 其他分支进行 **concentrate** 计算输出。之所以缺少 Inception V2 多尺度卷积的平均池化分支，考虑到时间序列模型使用的卷积类型为一维卷积，即 **conv1d** 不存在池化操作。输入特征经过多尺度的卷积后，相比标准模型对于输入特征提取更加完整，保留了特征的多样性，**concentrate** 计算与传统的 **add** 计算使提取不同输入特征进拼接在一起，增加输入特征的表达能力。时间卷积网络整体模型如图 3.5 TCN block 的结构示意图所示，图中 Improved Temporal block 为改进的 Temporal block 模块。

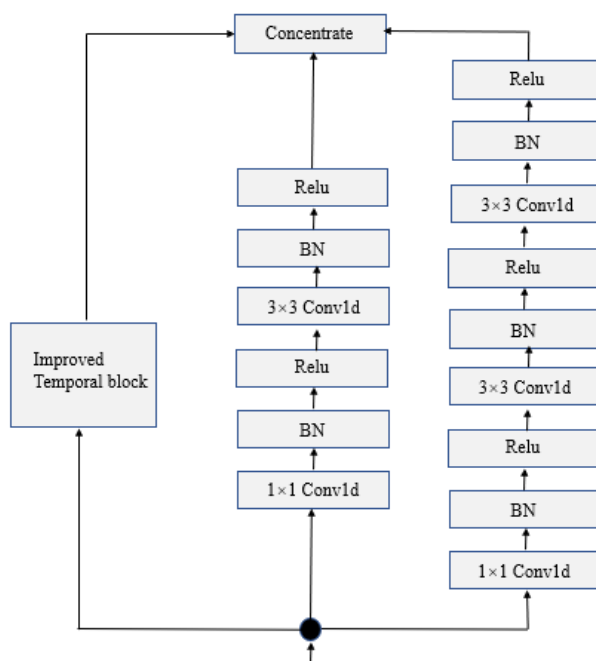


图 3.5 TCN block 的结构示意图

Fig. 3.5 Structure Diagram of TCN Block

改进的时间卷积网络 Res-TCN 的输入通道为 50，根据输入数据集的 10 只股票中每只 5 个特征一共 50 种特征。输入特征通过每一层 TCN block 与两层 3×3 卷积形成的残差模块后进行变化通道数，输入通道数由原始的通道数 50 逐步提升为 64、128、256、512。模型输出 10 之前是根据最后经过 **linear** 层将 10 只股票每一只所预测的均线值输出，即为本改进的模型 Res-TCN 的预测值，最后根据预测值进行实验分析。图 3.6 为 Res-TCN 网络整体框架。

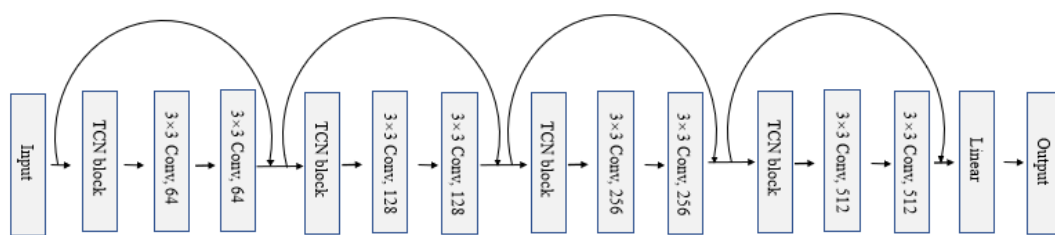


图 3.6 Res-TCN 网络整体框架

Fig. 3.6 Res-TCN network overall framework

在标准的时间卷积网络中，输入特征经过两个简单的时间卷积模块级联后，通过全连接层输出预测值。这种简单的级联连接可用于搭建层数较少的轻量化的模型，若模型在改进后导致网络层数增加到几十甚至几百时，这种简易的连接方式不利于应用此类深层的网络。若在训练模型参数增多后，采用一般的级联方式可能导致模型无法承受在训练时产生大量数据的压力，使用深层的网络可以减少模型的网络参数和训练运算量，通过变化通道的数量来增加模型的承受能力。该改进的时间卷积模型网络和标准模型相比，在网络层数增加时，通道数由一定的倍数增加，网络比原先的残差连接在训练时效率更高。

3.5 实验结果与分析

3.5.1 实验流程框架

本实验输入数据集为股票均线价格数据，股票数据的获取方式有很多种，一般可以从一些财经网站、股票软件、专业的 API 接口进行获取。实验是在 python 框架下进行的，采用的是专业 API 接口来下载实验所需要的股票数据。股票数据使用较多的免费网络接口为 Tushare，它为用户提供一个与 python 连接的数据接口。获取的股票数据一般都为基础的金融数据，其中比较通用的数据为 OHCL，OHCL 是股票基础的四种数据，分别为股票的开盘价、最高价、收盘价、最低价。实验所使用的均线价格包括当日的周线价格 ma_5、月线价格 ma_20、季线价格 ma_120 的价格。

实验开始时输入股票的均线数据，对输入数据进行数据预处理，然后定义输入特征数据集的标签，将数据集划分为训练集、测试集、验证集三种类型。紧接着设置模型的训练参数，其中包括设置合适的优化器来进行对训练的损失进行优化，还有采用合适的调整学习率的方法，设置模型迭代次数。以上流畅完成之后，开始正式进行模型的训练，若当前一次的训练结果比前一次效果好则保存当前训

练的权重，将此权重数据用于股票预测的实验中去。若效果差则重复上一步模型训练的步骤，直到实验所产生的数据符合要求，最后输出预测的股票均线数据。图 3.7 为本实验流程框架。

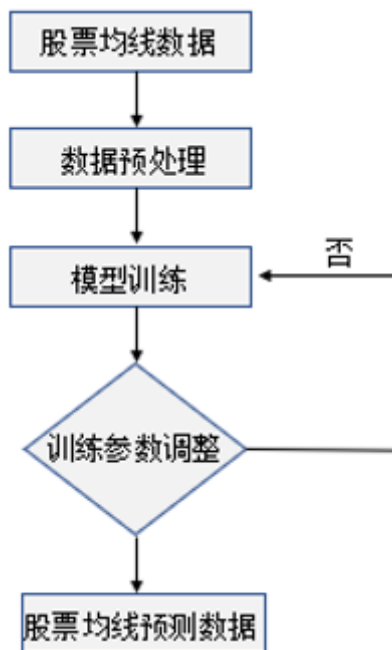


图 3.7 本实验流程框架

Fig. 3.7 Flow Framework of the Experiment

3.5.2 实验数据集

本实验所使用的是上海证券交易所和深圳证券交易所 A 股(简称“沪深 A 股”)股票代码前 10 的股票，通过 Tushare 平台获取的 2012 年 1 月 4 日至 2022 年 12 月 31 日共 10 年的日线数据。将数据集日期的前 90%作为训练集，后 10%作为测试集。实验数据集包含股票当日的 OHCL 基础数据，当模型预测周线价格时实验数据集的基础价格加上周线 ma_5 价格；当模型预测月线价格时实验数据集的基础价格加上月线 ma_{20} 价格；当模型预测季线价格时实验数据集的基础价格加上季线 ma_{120} 价格。基础数据称为 OHCL，其他的均线价格数据则是通过计算得出移动平均线值，它们为衍生数据。在股票的市场中，MA 均线价格是比较常见的衍生数据之一。本实验只考虑均线 MA 价格，它可以看作一段周期内的开盘价除以该周期，它们的表达式为：

$$ma_N = \frac{1}{N} \sum_{n=1}^N x_n \quad (3.1)$$

式中， N 为均线的天数， x_n 为第 n 日的开盘价。

本实验主要的预测对象为比较有代表意义的 3 种均线的值，它们分别为周线 ma_5 、月线 ma_20 、季线 ma_120 均值的价格。通过这 3 种均线价格的预测可以得知股票的发展趋势，是现在股票预测的一种重要指标。图 3.8 为本实验所使用的数据集。

date	code_1	open_1	close_1	high_1	low_1	ma5_1	...	code_10	open_10	close_10	high_10	low_10	ma120_10
2012/1/4	000001.SZ	15.59	15.16	15.59	15.15	15.37	...	000011.SZ	6.28	6.06	6.79	6.03	6.29
2012/1/5	000001.SZ	15.16	15.39	15.55	15.13	15.31	...	000011.SZ	5.75	5.46	5.98	5.45	5.66
2012/1/6	000001.SZ	15.35	15.35	15.51	15.15	15.34	...	000011.SZ	5.45	5.59	5.59	5.32	5.49
...
2021/12/29	000001.SZ	17.16	16.75	17.16	16.7	16.94	...	000011.SZ	11.59	11.48	11.63	11.48	11.55
2021/12/30	000001.SZ	16.76	16.82	16.95	16.72	16.81	...	000011.SZ	11.46	11.46	11.61	11.36	11.47
2021/12/31	000001.SZ	16.68	16.48	16.9	16.4	16.66	...	000011.SZ	11.5	11.66	11.75	11.47	11.6

图 3.8 实验所使用的数据集

Fig. 3.8 Data set used in the experiment

3.5.3 实验数据预处理

由于通过 Tushare 平台获得的数据集可能存在缺失值，需要对下载的数据进行缺失值补充处理。有时数据可能也会出现杂乱无序的情况，这种情况往往会导致在训练时的预测值不准确，在实验时必须避免此类情况的出现，实验前必须严格对日期信息与股票真实值进行一一对应。采用升序的排序方式对实验的输入数据集进行排序，使用的是 2012 年 1 月 4 日至 2022 年 1 月 4 日共 10 年的日线数据。由于股票的趋势价格往往与历史的价格有关，必须让数据进行升序排列，这样才可以根据前一日的价格历史来预测当日的价格变化。

实验所采用的 5 种股票价格有可能存在差异比较大的情况，比如当日的最高价格和最低价格两者是股票当日差异最大的真实值，两只不同的股票之间也存在大的差异。模型无法承受数据量大的数据进行训练，若这样会出现训练时训练速度慢、效率低的现象。因此需要对输入数据进行合适的标准化处理，本实验采用的是 **z-score 标准化**，它的表达式为：

$$z = \frac{x - \mu}{\sigma} \quad (3.2)$$

式中, x 为输入数据, u 为输入数据的平均值, σ 为输入数据的标准差。

3.5.4 实验平台

实验采用的硬件平台的 CPU 为 Intel @Xeon (R) Gold5218RCPU@2.10GHz $\times 40$, GPU 为 NVIDIA GTX3090 24G, 内存为 64G, 硬盘为 512G。软件平台的操作系统为 Ubuntu 20.04, python 编译器为 3.8.0, 在 pytorch1.9.1 深度学习框架下进行实验的。

3.5.5 实验评价指标

本实验所预测的股票值为第二天的均线价格, 其中需要几种实验指标来对改进的模型进行评估。在序列预测中比较常见的评估指标有均方误差 MSE、平均绝对误差 MAE、方根误差 RMSE、决定系数 R^2 , 对于其他深度学习的任务值中会使用准确率 Accuracy 来判断模型的效率。在本实验中也应用准确率这个指标来判断模型的优良性, 其中各种评测指标的表达式为:

$$MSE = \frac{1}{N} \sum_{n=1}^N (\hat{y}_n - y_n)^2 \quad (3.2)$$

$$MAE = \frac{1}{N} \sum_{n=1}^N |\hat{y}_n - y_n| \quad (3.3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{y}_n - y_n)^2} \quad (3.4)$$

$$R^2 = \frac{\sum_{n=1}^N (\hat{y}_n - \bar{y}_n)^2}{\sum_{n=1}^N (y_n - \bar{y}_n)^2} \quad (3.5)$$

式中, N 为实验样本个数; \hat{y}_n 为模型预测值; \bar{y}_n 为真实值的平均值; y_n 为真实值。

其中准确率 Accuracy 的表达式是根据如下方式定义的:

$$K = \frac{|T - P|}{T} \times 100 \quad (3.6)$$

$$KN = \text{sum}(K < 1) \quad (3.7)$$

$$Accuracy = \frac{KN}{TN} \times 100\% \quad (3.8)$$

式中, T 为真实值; P 为预测值; K 为真实值与预测值容忍度, KN 为容忍度为 1 时的模型预测正确的数量; TN 为真实值数量。

使用上述的实验指标来衡量模型预测股票的预测值与真实值之间的差距, 由误差表达式可知误差越小表示模型在预测股票的准确率越高, 模型在预测股票时效果越好。通过这些指标可以比较各模型的预测效果, 从而得出改进模型是否优于标准模型。

3.5.6 结果与分析

本实验所设计的改进时间卷积网络 Res-TCN 对股票均线的预测效果不仅与标准时间卷积网络相对比, 还与一些常见的序列预测模型相对比, 其中包括 RNN、LSTM。通过对比实验可以得出表现改进后的 TCN 模型更具有优良的效果。

实验模型在迭代 2000 次时, 模型的误差和准确率趋于稳定, 故将模型训练迭代的次数设置为 2000 次。在调整模型准确率的算法中, 共使用等间隔、指数衰减、余弦退火使用的方法来实验, 经过对比发现使用等间隔来调整学习率。其中等间隔的参数设置为每 200 次衰减 0.65, 设置初始学习率为 0.01, 这样模型预测效果最佳。表 3.2 为改进模型与其他模型在预测周线 ma_5、月线 ma_20、季线 ma_120 的效果。

表 3.2 改进模型 Res-TCN 与其他模型在预测周线 ma_5、月线 ma_20、季线 ma_120 的效果

Tab. 3.2 Comparison between the improved model Res-TCN and other models in the prediction weekly ma_5, monthly line ma_20, seasonal line ma_120 effect

Ma	Model	MSE	MAE	RMSE	R ²	Accuracy
ma_5	RNN	0.0240	0.0967	0.1549	0.9703	56.95%
	LSTM	0.0197	0.0823	0.1404	0.9756	58.97%
	TCN	0.0188	0.0698	0.1373	0.9855	64.69%
	Res-TCN	0.0078	0.0512	0.0882	0.9926	70.58%
ma_20	RNN	0.0198	0.0906	0.1406	0.9740	65.39%
	LSTM	0.0187	0.0811	0.1367	0.9755	65.45%
	TCN	0.0158	0.0510	0.1232	0.9884	71.52%
	Res-TCN	0.0026	0.0323	0.0513	0.9940	72.14%
ma_120	RNN	0.0199	0.0828	0.1411	0.9754	66.79%

ma_120	LSTM	0.0101	0.0718	0.1003	0.9787	67.37%
	TCN	0.0093	0.0408	0.0959	0.9923	72.55%
	Res-TCN	0.0011	0.0287	0.0298	0.9986	74.53%

由表 3.2 可知, Res-TCN 为改进的时间卷积模型, RNN 为表中模型预测效果最差的模型, LSTM 相比 RNN 表现的相对预测效果有些许改良但是还是不够良好。TCN 模型相比前两种模型来说, 提升效果大幅提升, 可以看出 TCN 模型对于预测股票的三种模型更加适合, 而 Res-TCN 模型预测效果表现出不同程度的提高。在预测周线 ma_5 时, Res-TCN 模型比标准模型提高大约 5% 的准确率, 可以看出在预测股票的短期价格趋势时, Res-TCN 模型更具有优势, 从此可以看出模型比较适合进行短期交易的投资者的需求。在短期的价格变动下, 及时预测出股票的趋势, 交易出手持的股票进行收益的最大化。而模型在预测月线 ma_20 时, Res-TCN 模型比标准模型提高大约 1% 的准确率。在预测季线 ma_120 时, Res-TCN 模型比标准模型提高大约 2% 的准确率。由此可知, 在预测中长期的股票价格时, Res-TCN 模型相对于其他序列预测模型提升准确率的效果没有短期预测的好, 但总体来说都有提升。Res-TCN 模型在预测三种均线的实验中, 表现出预测值误差小、准确率高的优点, 说明本章所设计的 Res-TCN 模型具有比较出色的股票预测能力。

在表 3.2 的各模型在预测在三种均线实验效果中, 为列出各模型的准确率对于损失下降的情况。为了探究模型的损失下降情况, 仅仅根据图表无法直观地发现各模型在训练时所进行的准确率上升和损失收敛的速度, 图标只能看出各模型训练最优的迭代准确率, 需要通过定义损失函数来观测模型训练的变化。图 3.9~3.11 改进模型与其他模型在预测周线 ma_5、月线 ma_20、季线 ma_120 的损失变化图。

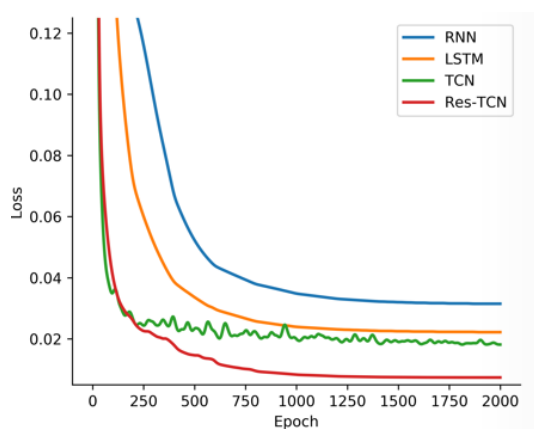


图 3.9 Res-TCN 模型与其他模型在预测周线 ma_5 的损失变化图

Fig. 3.9 Loss changes of the Res-TCN model and other models on the forecast weekly line ma_5

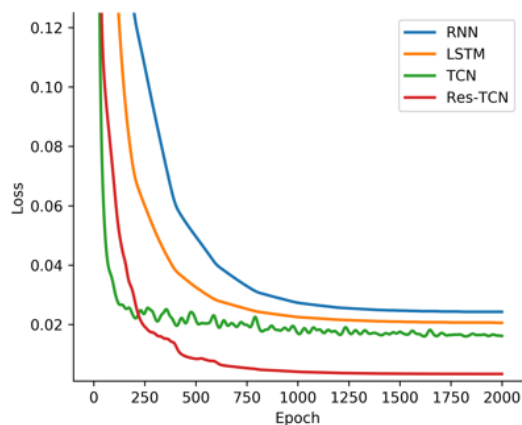


图 3.10 Res-TCN 模型与其他模型在预测月线 ma_20 的损失变化图

Fig. 3.10 Loss changes of the Res-TCN model and other models on the forecast monthly line ma_20

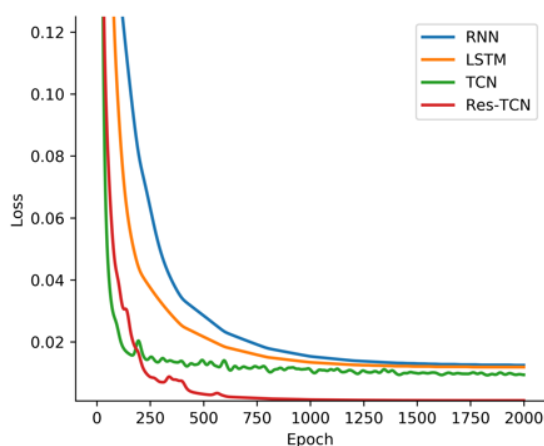


图 3.11 Res-TCN 模型与其他模型在预测季线 ma_120 的损失变化图

Fig. 3.11 Loss changes of the Res-TCN model and other models on the forecast seasonal line ma_120

根据实验训练的损失变化图可以看出, Res-TCN 模型与其他模型在预测三种均线的价格时表现的损失收敛速度更低、更快。标准 TCN 模型在加入多尺度卷积和残差后可以使模型更具有快速收敛的效果, 通过实验得出 Res-TCN 模型对标准模型更加有效的完成预测任务。

在深度学习的模型设计同时, 往往不仅要注重模型的最终表现出的效果, 还需注意模型在训练时的耗时情况。设计出既能更好的完成预测任务, 又能更快的处理出结果, 所以模型的效率问题也是实验所要考虑的问题。表 3.3 为实验各模

型的训练耗时情况。

表 3.3 实验各模型的训练耗时情况

Tab. 3.3 Training time consumption of each model in the experiment

Ma	Model	Training time(s)
ma_5	RNN	0.0037
	LSTM	0.0412
	TCN	0.0092
	Res-TCN	0.2033
ma_20	RNN	0.0044
	LSTM	0.0417
	TCN	0.0085
	Res-TCN	0.2044
ma_120	RNN	0.0040
	LSTM	0.0498
	TCN	0.0087
	Res-TCN	0.2021

由表 3.3 可知,实验所设计的 Res-TCN 模型在增加模型的网络层数后,它的训练时间提高,可以发现模型在训练一次迭代所消耗时间的只为 0.2s。结合它的提升效果,Res-TCN 模型的耗时是符合实验要求的,但需要后续实验进行改进。

为了更加直观表现 Res-TCN 模型的预测效果,选取训练数据集其中一只股票的数据作为实验验证集,用它来观察模型预测的预测值与真实值之间的差距大小。图 3.12~3.14 为 Res-TCN 模型与其他模型在预测周线 ma_5、月线 ma_20、季线 ma_120 的价格图。在时间预测模型的准确率实验中,若预测值越拟合真实值,则表示模型的预测能力越强。从图中可以看出 Res-TCN 模型相对于其他的预测模型的预测值最拟合真实值,所以可以说明本章所设计的 Res-TCN 模型比标准 TCN 模型在股票均线价格预测的准确性更好。

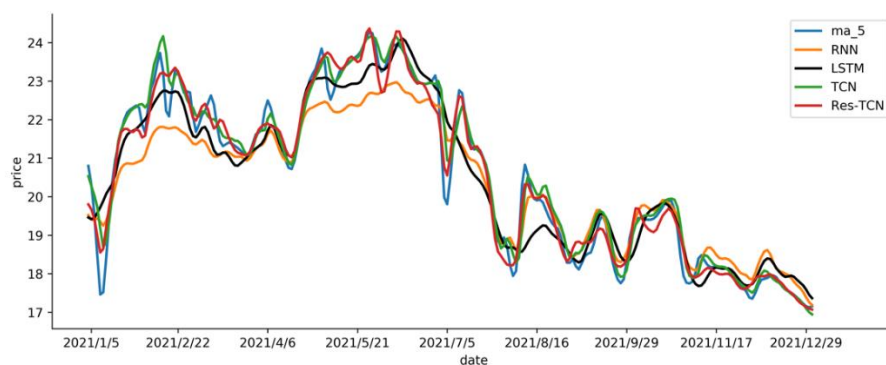


图 3.12 Res-TCN 模型与其他模型在预测周线 ma_5

Fig. 3.12 Comparison of the Res-TCN model and other models in the prediction weekly line ma_5

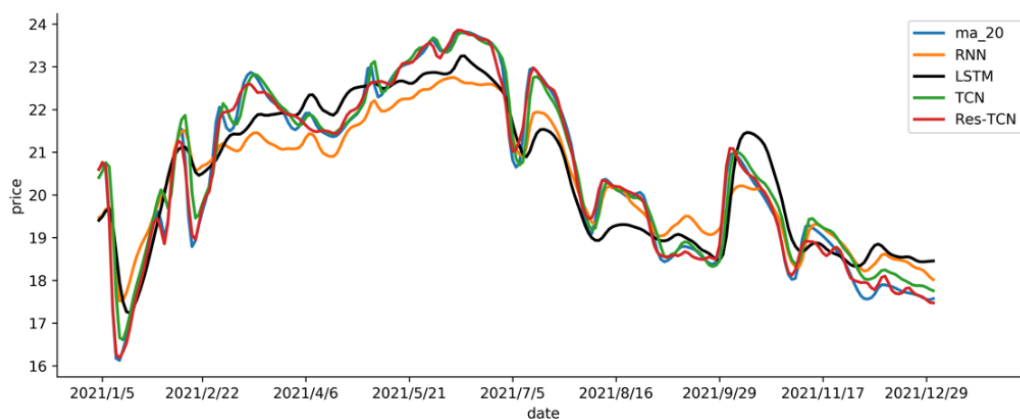


图 3.13 Res-TCN 模型与其他模型在预测月线 ma_20

Fig. 3.13 Comparison of the Res-TCN model and other models in the prediction monthly line ma_20

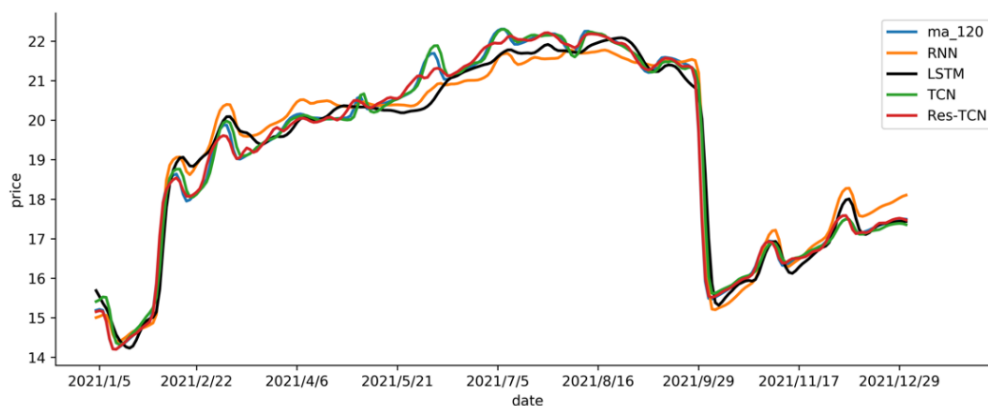


图 3.14 Res-TCN 模型与其他模型在预测季线 ma_120

Fig. 3.14 Comparison of the Res-TCN model and other models in the prediction seasonal line ma_120

验证集使用的是平安银行代码为 000001.Sz 的 2021 年 1 月 4 日至 2022 年 12 月 31 日的股票数据。由图 3.12~3.14 可知，横坐标为股票单位日时间，纵坐标为股票在当日的价格。实验表明 RNN 最偏离预测实际的股票值，其次为 LSTM、TCN 模型，Res-TCN 模型更加贴合实验的真实股票值。本章所设计的改进的时间卷积模型 Res-TCN 在预测股票三种均线价格时都表现的更好，具有对真实股票的预测价值。

3.6 本章小结

本章主要介绍所设计的多尺度卷积残差连接的 TCN 网络 Res-TCN，首先对多尺度原理进行介绍，其中包括多尺度 InceptionV1、InceptionV2 连接方式。紧接着介绍几种常见的残差结构，将多尺度和残差连接融合到 TCN 网络中。网络设计的方法是先将 TCN 中的 Temporal block 模块加入 1×1 卷积、BN、Relu 形成改进后的 Temporal block，从而加强网络的泛化能力，再将多尺度卷积里面的 Inception V2 网络结构与改进后 Temporal block 融合。其中多尺度卷积增加里面的不同大小卷积核的增强了基础 TCN 网络对数据特征的提取能力，融合了多尺度卷积的 TCN 网络 TCN block 与 3×3 卷积共同作为残差网络里面的残差模块，经过多层残差连接后输出。通过实验结果表明本章所设计的 Res-TCN 网络比 TCN 网络在股票均线价格预测实验表现的更为出色，但是实验所消耗的时间也变长，训练效率不高，总体来说实验的效果达到了本章所设计网络的目的。

4. 基于全局注意力机制和密集连接的 TCN 股票均线价格预测

4.1 引言

在上一章中的基于多尺度卷积的残差连接 TCN 的设计之中,改进模型 Res-TCN 相比 TCN 网络在预测股票三条均线的实验中表现出误差小、准确率高等优点,但模型存在预测耗时较大、预测效率变低的问题。为了克服这一问题,本章将引入全局注意力机制 GAM,在模型中加入注意力可以使模型对输入特征的提取能力进一步加强。同时本章将残差连接更换为密集连接结构,将最后总体框架中的残差模块更换为密集模块,将上述改进方法所设计出的模型称为 Dens-TCN-GAM 模型。Dens-TCN-GAM 模型相对于 Res-TCN 模型不仅在预测误差变小和准确率上有所提升,还在预测实验训练的耗时上面有所降低,模型的预测效率大幅度提升。本章实验分析部分在前一章的基础之上,加入实验优化算法的研究。通过实验对比后选用 Adam 作为实验的优化算法,使用与第三章一样的数据集,实验结果表明基于全局注意力机制的密集连接 TCN 模型 Dens-TCN-GAM 对于股票均线的预测具有实质性作用。

4.2 注意力模型 GAM

近些年的研究中,注意力模型一直是研究者比较感兴趣的一个焦点的研究方向,各种注意力机制逐渐展现在其在处理图像、视频、序列等常见的任务中的作用。在输入特征较为繁杂的数据集中,注意力模型可以在众多特征信息中提取对深度学习所需要的重要部分。选择注意力是比较常见的注意力模型,本章所设计的模型使用的是全局注意力模型(Global Attention Model, GAM)^[61],它属于选择注意力的一类。相对于其他的注意力模型 CBAM^[62]和 BAM^[63],GAM 是在它们基础之上进行改进的模型。在选择注意力模型中,虽然 CBAM 和 BAM 两种模型提高了 SENet^[64]在提取特征时抑制了不重要的特征,提高了模型的效率,但是由于它们没有利用通道注意力和空间注意力之间的相互联系,不能充分发挥两者的相互权重作用,从而不能有效的提高注意力模型的体征提取效率。GAM 将提取维度提高到三个维度,在这三个维度的作用下使得模型在捕捉特征更为有效。图 4.1 为全局注意力模型 GAM 的结构示意图,图中可以看出输入特征 F_1 与经过通道注意力机制 Channel Attention 输出的特征进行乘法计算后,使用相同的方式经过通道注意力 Spatial Attention 后输出特征 F_3 。

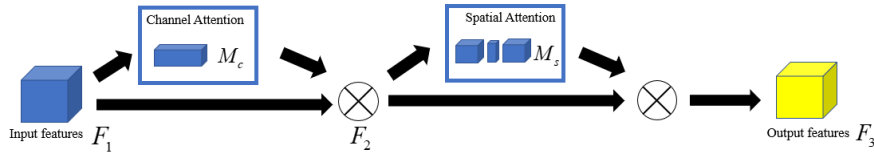


图 4.1 全局注意力模型 GAM 的结构示意图

Fig. 4.1 structural diagram of global attention model (GAM)

根据图 4.1 可以得出，其中模型输入特征为 F_1 ，中间状态为 F_2 ，输出特征为 F_3 ，通道注意力为 M_c ，空间注意力为 M_s ，乘法操作符号为 \otimes ，则计算过程为：

$$F_2 = M_c(F_1) \otimes F_1 \quad (4.1)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (4.2)$$

通道注意力 M_c 的表达式为：

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg}^c))) + W_1(W_0(F_{max}^c)) \end{aligned} \quad (4.3)$$

式中， $W_0 \in R^{c/r \times c}$ ， $W_1 \in R^{c/r \times c}$ 。

通道注意力通过采用改变三维数组排序组合的方法来保留模块上的信息，在中间加入一个双层的感知器 MLP，这个双层感知器是一种类似于编码器和解码器相结合的结果，这样可拓展通道之前的空间相互关联。相对于 BAM，GAM 具有相同的缩减比 r ，输入特征在进行三维通道变换后，通过双层感知器里的编码解码操作后，进行与前一步同样的逆向操作将三维通道变换为之前的类型。输出特征前通过采用 sigmoid 激活函数，从而使模型更具有泛化能力。图 4.2 为通道注意力模块结构图。

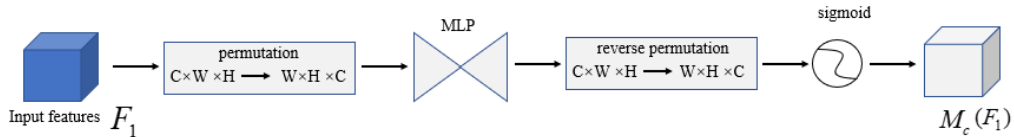


图 4.2 通道注意力模块结构图

Fig. 4.2 Structure of channel attention module

空间注意力图 M_s 的表达式为：

$$\begin{aligned}
 M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\
 &= \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s]))
 \end{aligned} \quad (4.4)$$

在空间注意力中,输出特征先经过一个 7×7 的卷积操作后将输入的通道数变化成与 BAM 相同的缩减比 r , 继续经过 7×7 的卷积后将通道数变换成之前相同的数量,最后在输出前使用 **sigmoid** 激活函数。由于模型在进行最大池化时会使输入特征丢失一些特征信息,对模型进行训练时造成预测误差变大的后果。在空间注意力中没有采用最大池化的操作,若模型增加空间注意力模块后,会使模型的参数增加,从而导致模型在训练时效率变低,图 4.3 为空间注意力模块结构图。

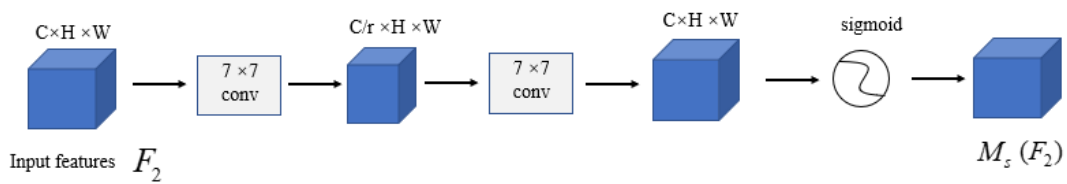


图 4.3 空间注意力模块结构图

Fig. 4.3 Structure of spatial attention module

4.3 密集 DenseNet 结构

随着残差连接 ResNet 在深度学习神经网络里面的应用,研究者发现可以在 ResNet 的基础知识之上研究出更实用于模型的网络连接方式。Huang G 等人^[65]在 2017 年为了解决残差连接存在的梯度消失和网络退化问题,提出一种的新的密集连接 DenseNet 网络结构。其设计的基础思想是使模型的每一层都连接,将不相邻的两层网络连接起来,每一层输都与前面每一层的输出有联系,加强了每一层之间的关系。这种密集连接的方式最大程度的保留了模型的特征,更加有利于每一层信息之间的传输,减少了网络传输的特征损耗。

DenseNet 的密集连接与传统的卷积神经网络不同,其网络之间的连接器不与网络的层数相等,具体的表达式为

$$x_i = H(x_{i-1}) + x_{i-1} \quad (4.5)$$

$$x_i = H([x_0, x_1, \dots, x_{i-1}]) \quad (4.6)$$

式中, x_i 为第 i 层的输入特征, $[x_0, x_1, \dots, x_{i-1}]$ 为第 0 层到第 $i-1$ 层输出特征的融

合， $H(\bullet)$ 表示 BN 层、Relu 激活函数和卷积层进行非线性变换后的函数。

密集连接由每一个 Dense block 密集块和过渡层连接组成，图 4.4 为经典 Dense block 的结构图，图中 Dense block 包含 BN、Relu 和 conv，其中 conv 里面常见的卷积核大小为 1×1 、 3×3 等。在 Dense block 中，后面中的一层都包含前面所有层的特征信息，过渡层中的作用是减少每一层中连接的特征数量，降低网络累积的冗余从而提高模型之间的连接关系。

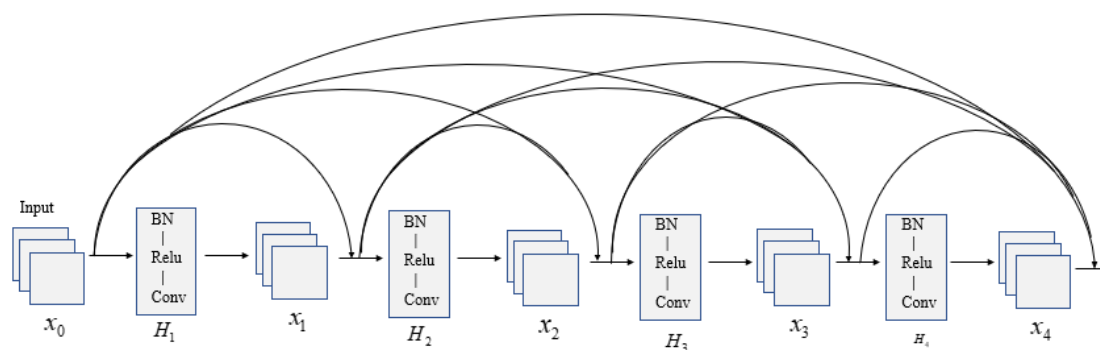


图 4.4 经典 Dense block 的结构图

Fig. 4.4 Structural diagram of classic Dense block

4.4 全局注意力机制和密集连接的 TCN 网络结构

本章所设计的改进的 TCN 模型 Dens-TCN-GAM。在每一层的 TCN block 层中融合 GAM 注意力模块为 TCN-GAM block，使每一层的卷积模块都具有 GAM 注意力模块，这样可以提高输入特征经过每一层时的运行效率其中每一层的 TCN block 模块为原先改进后的时间卷积模块。图 4.5 为 TCN-GAM block 的结构示意图。

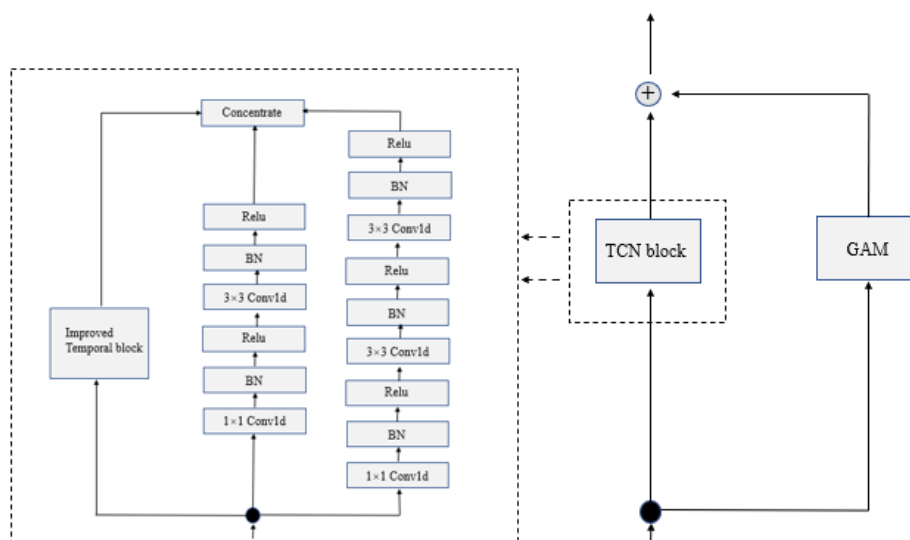


图 4.5 TCN-GAM block 的结构示意图

Fig. 4.5 Structure diagram of TCN-GAM block

全局注意力下的密集连接的时间卷积网络 **Dens-TCN-GAM** 总体框架是以原先的残差连接的基础之上，将残差连接改进为密集连接。在密集连接下，带有全局注意力 **GAM** 的改进模块 **TCN-GAM block** 与两层 3×3 卷积作为密集连接里的 **conv block** 模块。图 4.6 为具有全局注意力的密集连接的时间卷积网络 **Dens-TCN-GAM** 框架图。

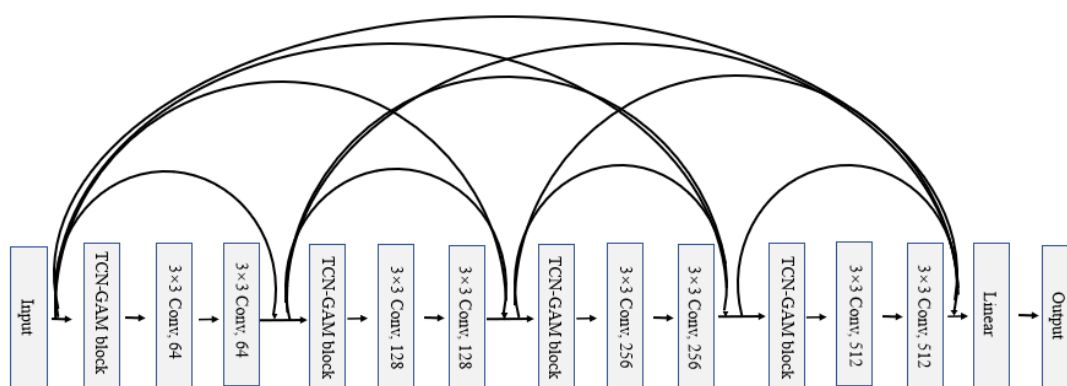


图 4.6 具有全局注意力的密集连接的时间卷积网络 Dens-TCN-GAM 框架图

Fig. 4.6 Time convolution framework of Dense connections with global attention

Dens-TCN-GAM

4.5 实验结果与分析

4.5.1 实验优化算法

在深度学习模型实验中,网络模型的选择不仅是实验的一部分,而且选取一个适合模型的优化器也十分重要,优化算法直接影响网络模型的收敛速度,在本实验之前,探讨最适合改进网络的优化算法。常见神经网络的优化算法有 SGD、RMSprop、Adam^[66]等。本章通过实验验证的方法对比这三种优化算法,从中选取最合适的加入后续实验中。图 4.7 为各种优化算法实验对比。

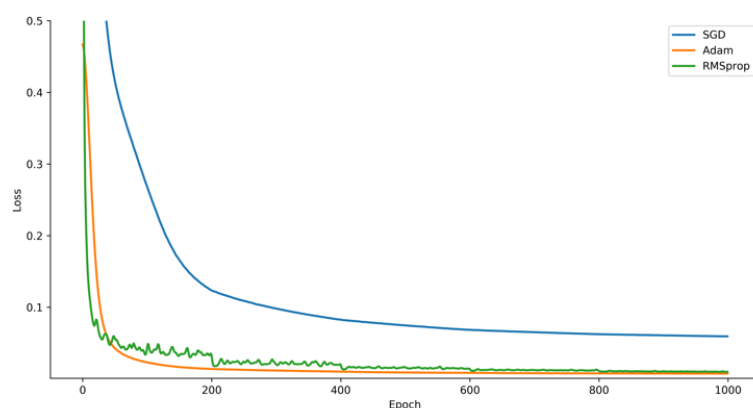


图 4.7 各种优化算法实验对比

Fig. 4.7 Experimental comparison of various optimization algorithms

各种优化算法的对比实验是在同种数据集、网路模型下的,所使用的模型训练是按照每一批次进行反向传播,梯度来更新权重信息,随着训练的批次的增加,学习率不断下降到达最佳。根据图 4.7 可以看出,在相同的批次下,使用 Adam 优化器时,模型的学习率下降最低,从而可以看出对比几种优化算法。Adam 优化器表现出相对于其他优化器更加良好,所以所有模型均使用 Adam 优化器进行对比实验。

4.5.2 结果与分析

本实验结果不仅包括本章中所设计的具有全局注意力机制下的密集连接方式的时间卷积网络与本文第三章中的改进模型相对比,还包括与其他的预测模型相对比,这样说明本章所设计的模型不仅优于传统的模型,还看出在第三章中的模型基础下进行进一步的改进可以使预测效果更好。实验中训练所设置的参数与

第三章相同，表 4.1 为各模型在预测 ma_5、月线 ma_20、季线 ma_120 的效果。

表 4.1 各模型在预测 ma_5、月线 ma_20、季线 ma_120 的效果

Tab. 4.1 Effect of each model on forecast weekly line ma_5、monthly line ma_20、seasonal line ma_120

Ma	Model	MSE	MAE	RMSE	R ²	Accuracy
ma_5	RNN	0.0240	0.0967	0.1549	0.9703	56.95%
	LSTM	0.0197	0.0823	0.1404	0.9756	58.97%
	TCN	0.0188	0.0698	0.1373	0.9855	64.69%
	Res-TCN	0.0078	0.0512	0.0882	0.9926	70.58%
	Dens-TCN-GAM	0.0012	0.0261	0.0348	0.9990	75.48%
ma_20	RNN	0.0198	0.0906	0.1406	0.9740	65.39%
	LSTM	0.0187	0.0811	0.1367	0.9755	65.45%
	TCN	0.0158	0.0510	0.1232	0.9884	71.52%
	Res-TCN	0.0026	0.0323	0.0513	0.9940	72.14%
	Dens-TCN-GAM	0.0007	0.0214	0.0281	0.9993	77.33%
ma_120	RNN	0.0199	0.0828	0.1411	0.9754	66.79%
	LSTM	0.0101	0.0718	0.1003	0.9787	67.37%
	TCN	0.0093	0.0408	0.0959	0.9923	72.55%
	Res-TCN	0.0011	0.0287	0.0298	0.9986	74.53%
	Dens-TCN-GAM	0.0002	0.0121	0.0156	0.9998	79.05%

由表 4.1 可以看出，本章所设计的带有 GAM 注意力机制下密集连接方式的时间卷积网络 Dens-TCN-GAM 相对于第三章中的改进模型 Res-TCN 在预测周线 ma_5 的实验中，准确率从 70.58% 提高到 75.48%。在预测月线 ma_20 的实验中，准确率从 72.14% 提高到 77.33%。在预测月线 ma_120 实验中，准确率从 74.53% 提高到 79.05%。总体看来，第四章所设计的带有注意力机制的密集连接网络模型 Dens-TCN-GAM 比第三章中的 Res-TCN 模型在预测股票均线的价格时，表现的更好，都有大约 5% 的提升效果。而且，在模型预测误差的实验来看，Dens-TCN-GAM 模型相比 Res-TCN 误差更小。

根据表 4.1 中的各模型在预测均线的效果实验结果为了更好的观察出 Dens-TCN-GAM 的损失下降情况，图 4.8~4.10 为各模型在预测周线 ma_5、月线 ma_20、季线 ma_120 的损失变化图。在预测三种均线的实验中，根据实验的损失变化图

可以看出 Dens-TCN-GAM 模型相对于 Res-TCN 模型的损失下降的更早, 大约在训练的第 250 轮后趋于稳定值, 这个稳定值比 Res-TCN 低得多, 后面损失也保持平稳, 没用出现过拟合和欠拟合的状态。总体来看, 本章所设计的 Dens-TCN-GAM 模型预测 3 种股票的均线表现的最为出色。

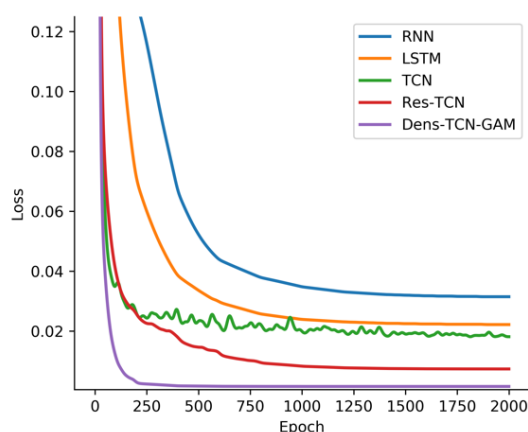


图 4.8 各模型在预测周线 ma_5 的损失变化图

Fig. 4.8 Loss changes of each model on forecast weekly line ma_5

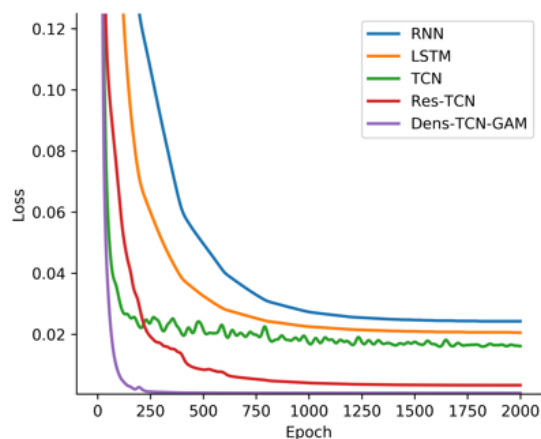


图 4.9 各模型在预测月线 ma_20 的损失变化图

Fig. 4.9 Loss changes of each model on forecast monthly line ma_20

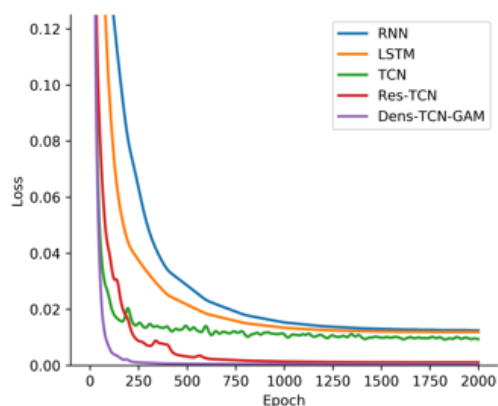


图 4.10 各模型在预测季线 ma_120 的损失变化图

Fig. 4.10 Loss changes of each model on forecast seasonal line ma_120

根据图 4.7~4.9 各模型在预测三种均线的损失变化图可以看出,本章所设计的 Dens-TCN-GAM 模型相比第三章中的 Res-TCN 模型在进行训练时,模型的收敛速度变快,而且变低。从而通过实验可以说明 Dens-TCN-GAM 模型比 Res-TCN 模型等预测股票均线价格效果更好。

为探究本章设计 Dens-TCN-GAM 模型的预测效率,将各模型每一轮的实验训练所消耗的时间进行一个对比实验。表 4.2 为各模型的训练耗时情况。

表 4.2 各模型的训练耗时情况

Tab. 4.2 Training time consumption of each model

Ma	Model	Training time(s)
ma_5	RNN	0.0037
	LSTM	0.0412
	TCN	0.0092
	Res-TCN	0.2033
	Dens-TCN-GAM	0.2388
ma_20	RNN	0.0044
	LSTM	0.0417
	TCN	0.0085
	Res-TCN	0.2044
	Dens-TCN-GAM	0.2510
	RNN	0.0040

ma_120	LSTM	0.0498
	TCN	0.0087
	Res-TCN	0.2021
	Dens-TCN-GAM	0.2353

相对于 Res-TCN 模型，Dens-TCN-GAM 模型实验训练耗时增加越 0.03s。但 Dens-TCN-GAM 模型相比 Res-TCN 来说，预测准确率增加 5%，而 Res-TCN 模型相对于 TCN 模型，准确率只增加 2%，Res-TCN 模型耗时比 TCN 大大增加，Dens-TCN-GAM 模型所造成的耗时情况完全在实验承受的范围之内。综上所述，本章所设计的 Dens-TCN-GAM 模型在预测效率方面也优于 Res-TCN 模型。

为探究本章所设计的模型在验证集中的预测效果，通过对比实验后，将各模型的实验结果绘制在同一张图中进行比较。图 4.11~4.13 为各模型在预测预测周线 ma_5、月线 ma_20、季线 ma_120 的价格图。由三种均线的预测图，可以看出本章设计的 Dens-TCN-GAM 模型在验证集上的预测效果优于 Res-TCN 模型和 RNN、LSTM、TCN 等其他模型。

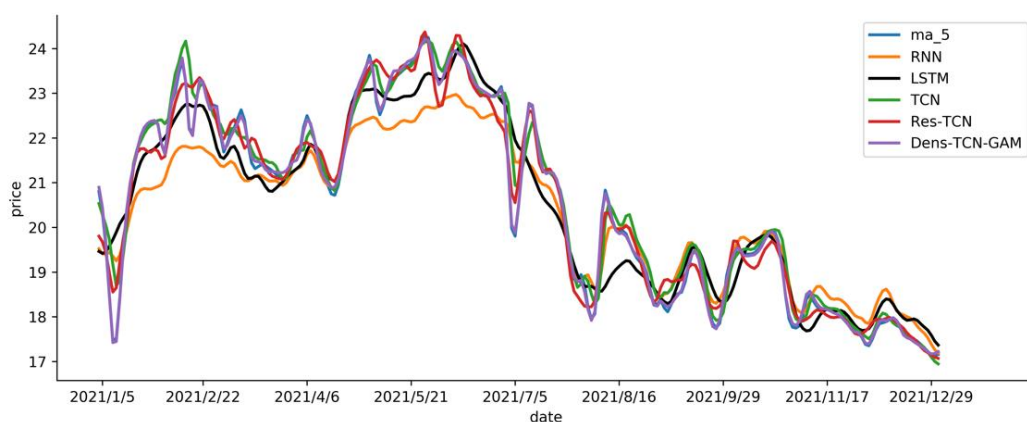


图 4.11 改进模型与其他模型在预测周线 ma_5

Fig. 4.11 Comparison of the improved model and other models in the prediction weekly line ma_5

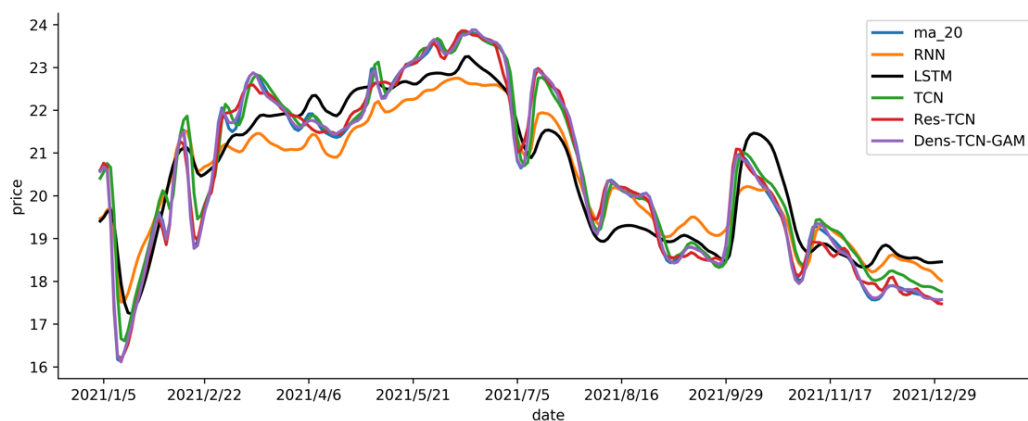


图 4.12 改进模型与其他模型在预测月线 ma_20

Fig. 4.12 Comparison of the improved model and other models in the prediction monthly line ma_20

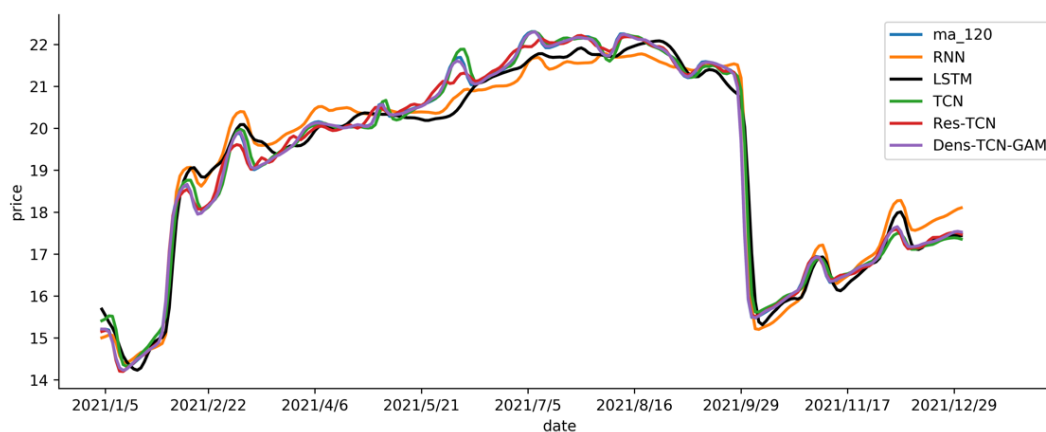


图 4.13 改进模型与其他模型在预测季线 ma_120

Fig. 4.13 Comparison of the improved model and other models in the prediction seasonal line ma_120

4.6 本章小结

本章主要介绍所设计的基于全局注意力机制的密集连接 TCN 的股票预测模型 Dens-TCN-GAM。首先对全局注意力机制 GAM 进行介绍，将 GAM 模块引入每一层的 TCN block 中形成 TCN-GAM block。相对于 TCN block，新的 TCN-GAM block 是由改进后的 Temporal block 模块相融合，这样提高了模型的特征提

取效率，在 GAM 里面的通道注意力和空间注意力作用于下，模型的泛化能力增强。紧接着将每一层的 TCN-GAM block 与两层 3×3 卷积作为 Dens-TCN-GAM 里面的 conv block，四层 conv block 组成一个基本的 Dense block。实验使用 Adam 优化算法的下，Dens-TCN-GAM 相比 Res-TCN 模型在预测 3 种股票均线的准确率上提高 5%，训练损失能变化到更低，预测误差也比 Res-TCN 小，可以有效的预测出股票均线的价格走势情况，极大的避免了股民所要承担的风险。

5. 基于多影响因子下的量化选股策略

5.1 引言

在股票的交易市场中，投资机构主导着股票的发展方向，普通的投资者在进行股票投资时需要规避一定的投资风险，这样就需要投资者在选择股票时擦亮双眼，选股的策略就显现的十分重要。为了使投资者在股票市场中少走弯路，将投资风险降到最低，从而获得更大的收益的目的，本章所设计的基于多种影响因子下的量化选股策略针对普通散户的需求。首先将 3 种股票的量化指标数据加入原有数据集里形成股票选股的数据集。这 3 种股票的量化指标数据属于三种不同的影响因子，组成的选股的数据集相比原先数据集的特征更加全面，模型更能采取输入特征的变化情况，更加准确的预测出股票的均线走势。其次构建出针对多影响因子的选股系统，将选股的数据集构建完毕。最后在股票池中加入新的种类股票组成选股股票池，将股票池中的预测效果最好的前 10 只股票选出。通过实验列出短期、中期、长期所对应的股票选择方案，提供给了不同投资方式的用户投资方向。

5.2 量化指标数据

股票的量化需要多项指标支撑，其基本思想就是在股票的多项先关指标中加入适合分析的股票指标，通过实验说明指标的对股票的量化有所帮助，最好可以选取合适的股票类型组成股票池，这样可以使股民在股票交易时获取最大的利益。本章所使用的股票量化指标数据不仅包括本文前两章中所使用的每日的行情数据 OHCL 和股票日线所衍生的三种均线 ma_5 、 ma_{20} 、 ma_{120} ，还加入三种股票量化指标数据。这三种股票数据可以分为基础指标数据、技术指标数据、新闻指标数据，通过实验对比后选取合适的股票池来进行市场交易。

5.2.1 基础指标数据

基础指标数据是描述一只股票的基础数据的重要指标，本章中设计的股票基础数据不仅包括第三章和第四章中的股票代码、股票名称，还加入股票所在地区、所属行业、市场类型、所述交易所、上市情况、上市日期、退市日期信息。通过加入这些基本指标，使模型在训练时捕捉更多特征信息，使预测更加准确，同时用户可以在股票池中选取同类型的股票进行买入和卖出。表 5.1 为股票的基础指标数据。

表 5.1 股票的基础指标数据

Tab. 5.1 Basic index data of stocks

名称	基础指标描述
code	股票代码
name	股票名称
area	所在地区
industry	所属行业
market	市场类型
exchange	所属交易所
list status	上市情况
list date	上市日期
delist date	退市日期

股票的基础指标中股票代码的数字一般为六位, A 股大部分股票属于上海证券交易所和深圳股票交易所, 在数据集中代码中的六位数字后缀的“SZ”代表深圳市场, 而“SH”代表上海市场。表中的市场类型一般包括主板、中小板、创业板、科技版等类型, 股票上市信息对于股票选择时也十分重要, 用户可以通过股票最新的上市状态分析出公司的近期发展情况, 做出交易判断。

5.2.2 技术指标数据

股票的技术指标数据主要为股票的行情数据 OHCL 通过计算衍生出的数据, 也被称为股票的衍生数据, 最常见的衍生数据有 MACD、KDJ、RSI 指标等数据。这些技术指标是根据实际股票的价格发展趋势所总结出的公式, 这些指标给投资者在选取股票时, 提供一种参考, 同时这些指标也可以看出某只股票最近的走势情况。表 5.2 为股票的技术指标数据。

表 5.2 股票的技术指标数据

Tab. 5.2 Technical index data of stocks

名称	技术指标描述
trade time	交易日期
timestamp	时间戳
macd	MACD

kdj	KDJ
rsi	RSI

在表 5.2 中，主要列出了三种比较有代表意义的股票技术指标数据，下面将对这几种指标进行具体的介绍：

(1) MACD 指标

MACD 是异同移动平均线 (Moving Average Convergence Divergence) 的首字母缩写，其首次出现在大众视野是在 1979 年，由著名的股票投资者 Gerald Appel 提出。MACD 指标一经问世就收到了许多股票投资者的认可，该指标使用起来具有灵活简单的优点，被称为股票指标中的“指标之王”，是一种形容股票趋势发展的重要指标。MACD 是以双指数移动平均线为基础，主要计算方法分为 3 个步骤，下面将对这 3 个步骤进行逐步介绍：

步骤 1：先计算出平滑移动平均线 EMA (Exponential Moving Average)，给出 EMA 的数学定义。对于一个序列 $\{x_n\}$ 来说，它截止第 n 项的周期 N 的平滑移动线 $EMA_N(x_n)$ 为：

$$EMA_N(x_n) = \frac{2}{N+1} \sum_{k=0}^{\infty} \left(\frac{N-1}{N+1} \right)^k x_{n-k} \quad (5.1)$$

然后，对于股票数据来说，第 t 天的 EAM 值表示为 $EMA_{(x,t)}$ ，其计算方法为：

$$EMA_{(x,t)} = \alpha \times (x - EMA_{(t-1)}) + EMA_{(t-1)} \quad (5.2)$$

式中， x 为起始价格， α 为平滑系数， $EMA_{(t-1)}$ 为前一天的 EMA 值。在计算 EMA 时，采用的是 T+1 交易模式进行的，使用最常见的 12 和 26 日 EMA 来进行计算，第 n 天的 $EMA(12)$ 值为：

$$EMA_n(12) = \frac{2}{12} (c_n) + \frac{11}{12+1} (EMA_{n-1}) \quad (5.3)$$

式中， c_n 为第 n 天当天的收盘价 close price， EMA_{n-1} 为第 $n-1$ 天的 EMA 值。

步骤 2：紧接着计算出离差值 DIF (Difference)，其计算方法为：

$$DIFF = EMA_{(p,12)} - EMA_{(p,26)} \quad (5.4)$$

式中， p 示为第 n 天当天的股票价格 price。

步骤 3：最后计算出 $EMA(9)$ ，即为 9 日的 EMA 值，其计算方法为：

$$EMA_{(x,t)} = \alpha \times (x - EMA_{(t-1)}) + EMA_{(t-1)} \quad (5.5)$$

第九日的 EMA 值就为 MACD 值，将此值命名为 DEA 或者 DEM。通过定义可以得知 MACD 值与离差值 DIFF 相关，将公式 5.3 中的两个离差值看做快慢 DIFF，由快慢 DIFF 可以看出股票的发展趋势。MACD 指标是现在使用比较广泛的一种指标。

(2) KDJ 指标

股票的 KDJ 指标是在 1957 年出现的，由美国 George Lane 博士提出，又被称为随机震荡指数，是一个重要的股票的技术指标。KDJ 是由 RSV (Row Stochastic Value) 发展出的，RSV 为未成熟随机值，其数学表达式为：

$$RSV_n = \frac{c_n - l_n}{h_n - l_n} * 100\% \quad (5.6)$$

式中，n 为天数，一般 n=9， c_n 为最后一天的收盘价， l_n 为 n 时间段内的最低价， h_n 为 n 时间段内的最高价。

KDJ 又被称为 KD 值，它是由 K 值和 D 值组成，K 值被称为快速平均值，D 值被称为慢速平均值，它们的取值范围为 0~100。第 n 天的 K 值和 D 值的定义：

$$K_n = \alpha * RSV_n + (1 - \alpha) * K_{n-1} \quad (5.7)$$

$$D_n = \alpha * K_n + (1 - \alpha) * D_{n-1} \quad (5.8)$$

式中， α 为平滑系数，一般取值为 0.5。

在股票市场中，一般由快慢速平均值的大小可以看出股票的交易信号。表 5.3 为 K 值与 D 值的变化情况对应的股票市场。

表 5.3 K 值与 D 值的变化情况对应的股票市场

Tab. 5.3 Stock market corresponding to the change of K value and D value

KD 值变化情况	股票市场
K 值>D 值	上涨趋势
K 值向上突破 D 值	买进信号
K 值<D 值	下跌趋势
K 值向下跌破 D 值	卖出信号
D 值>70	超买现象

D 值<30	超卖现象
K 线与 D 线交叉点>70	买卖可靠
K 线与 D 线交叉点>80	最佳卖点
K 线与 D 线交叉点<30	买卖可靠
K 线与 D 线交叉点<20	最佳买点

(3) RSI 指标

RSI (Relative Strength Index) 指标可以描述一个时间内的股票趋势变化的大小, 也被称为相对强弱指标, 它是由 Welles Wilder 在 1978 年提出的, 可以有效地反映出是股票市场的发展情况。RSI 一般取 14 天的变化情况, 经过研究发现 14 天的时间跨度最佳, RSI 的数学表达式为:

$$RSI_n = \frac{\sum_{i=0}^{n-1} (p_{t-i} - p_{t-i-1}) * M}{\sum_{i=0}^{n-1} |(p_{t-i} - p_{t-i-1})|} \times 100\% \quad (5.9)$$

式中, p_t 表示为第 t 时刻时股票的价格, 若前一时刻股票的值大于当前时刻, 即

$p_{t-i} - p_{t-i-1} > 0$, 则 M 为 1, 反之 M 则为 0。

RSI 指标的取值范围反映出了股票在一段时间内的变化情况。表 5.4 为 RSI 指标的取值范围对应的股票市场, 股票交易者可通过 RSI 的值分析出股票趋势情况, 使交易时做出合理的判断。

表 5.4 RSI 指标的取值范围对应的股票市场

Tab. 5.4 Stock market corresponding to the value range of RSI index

RSI 值变化情况	股票市场
RSI 值>80%	超卖现象
RSI 值≈50%	观望信号
RSI 值<20%	超买现象

5.2.3 新闻指标数据

新闻指标数据对股票的市场波动有着不可或缺的影响, 往往新闻影响股票的形式是多样性的, 比如国内的政策变化、公司的经营状况、国外的政治事件和自

然灾害等都可以成为股票市场变化的影响因素。新闻事件通过网络多媒体的传播后,股票投资者会对股票投资的决策产生影响,所以新闻指标数据也是分析股票趋势的重要部分。新闻数据可由 python 库函数进行爬虫获得,通过库函数将获取的新闻文本数据进行分词切分后,把处理后的分词保存到数据库中,对文本的情感极性进行区分,将情感极性分为积极、消极,其中积极为 1,消极为 0。表 5.5 为万科 A 的部分股票新闻指标数据。

表 5.5 万科 A 的部分股票新闻指标数据

Tab. 5.5 Some stock news index data of Vanke A

日期	新闻标题	情感极性
2022-10-27	1.79 亿元主力资金今日撤离房地产板块	0
2022-10-28	万科前三季度净利润 170.5 亿元,同比增长 2.17%	1
2022-10-29	中邮证券给予万科 A 推荐评级,业绩微增销售好转	1
2022-10-30	央行通报三季度拒收现金处罚情况万科物业被罚	0
2022-10-31	港股万科企业尾盘跌超 10%	0
2022-11-1	西南证券给予万科 A 买入评级,坚持稳健经营多元业务	1
2022-11-2	龙湖、万科美元债大涨	1

5.3 多影响因子选股系统的构建

股票的价格市场受多种影响因子决定,在交易时,投资者往往陷入进退两难的境地,为了解决投资者无法选择合适的股票进行投资,本章设计了多影响因子选股系统。首先将第三、四章中所使用的原始均线的数据集加入本章所提到的量化指标数据,这样组成新的股票数据集,对新的数据集进行数据清洗后补充缺失值。然后将处理完成的新数据集放到第四章中设计的股票预测模型 Dens-TCN-GAM 中进行训练,输出每一只股票的预测的准确率和误差值,选取预测准确率最好的股票进行重组。最后得出最佳的股票组合类型,这样就可以根据实验结果来推荐投资者所选择的股票种类。图 5.1 为多影响因子选股策略系统的流程框架。

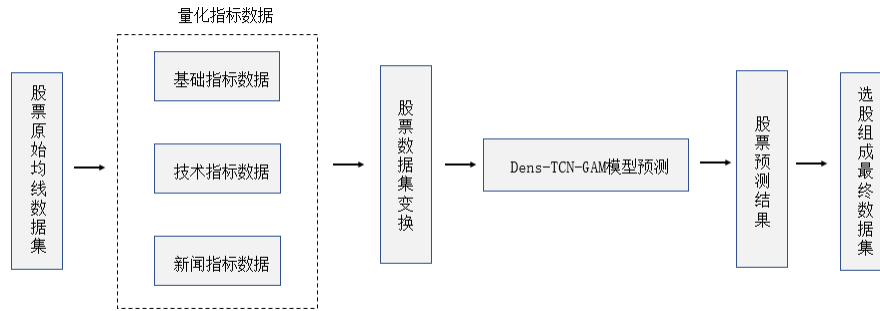


图 5.1 多影响因子选股策略系统的流程框架

Fig. 5.1 Process framework of stock selection strategy system with multiple influencing factors

5.4 实验结果与分析

5.4.1 实验流程

本章实验的目的是在股票池中选取合适的股票组合。首先将股票所有组合的数据集组成股票总池，然后将股票总池放入预测模型中，将每一只股票预测的结果列出，经过比较排序后选出选股股票池。图 5.2 为选股实验流畅框架。实验的股票总池不仅包括第三章、四章中的沪深 A 股前十的股票，还包括新加入的 30 只股票。这样做的目的是若实验股票的波动较为小，不能提现实验模型的实用环境。为了加大实验模型的泛化能力，在合理的情况下进行选股，实验考虑到从两个方面加入合适的数据集。一方面为股票行业的类型，在股票交易市场中经常会出现同行业股票在一段时间内同时出现上涨和下跌的情况，本实验新加入的新股票有 3 种行业的股票类型，分别为白酒、生物医药、5G 概念行业下的股票。另一方面，将在同一行业中的股票按照市值的占比划分成 3 种盘口类型，分为大盘股、中盘股和小盘股，这 10 只股票的盘口分布的比例为 3:4:3，这样选择主要考虑不同盘口的股票会出现不同的波动，在选择股票使更有说服力。

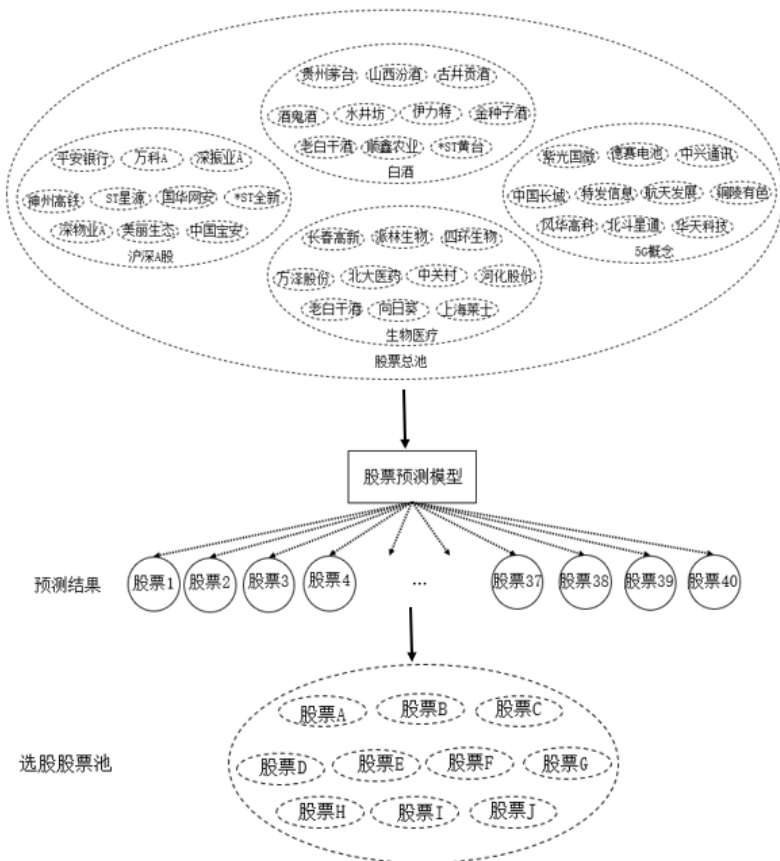


图 5.2 选股实验流程框架

Fig. 5.2 Smooth framework of stock selection experiment

5. 4. 2 结果与分析

本实验是在第四章中所设计的股票均线预测模型 Dens-TCM-GAM 中进行的，实验的各项参数设置也与第三、四章中的训练相同，将各种类的股票数据放入模型中进行实验。为了选择查看预测效果好的股票，本实验使用前两章中的绝对误差 MAE 来评估预测的效果。表 5.6~5.9 为各种类股票的预测效果。

表 5.6 沪深 A 股每只股票的预测效果

Tab. 5.6 Prediction effect of each stock in Shanghai and Shenzhen A stock

股票名称	代码	MAE_{ma_5}	MAE_{ma_20}	MAE_{ma_120}
平安银行	000001.SZ	0.0372	0.0237	0.0131
万科 A	000002.SZ	0.0407	0.0293	0.0140

国华网安	000004.SZ	0.0419	0.0290	0.0153
ST 星源	000005.SZ	0.0335	0.0213	0.0124
深振业 A	000006.SZ	0.0287	0.0227	0.0122
*ST 全新	000007.SZ	0.0315	0.0262	0.0147
神州高铁	000008.SZ	0.0250	0.0187	0.0126
中国宝安	000009.SZ	0.0236	0.0216	0.0144
美丽生态	000010.SZ	0.0273	0.0232	0.0159
深物业 A	000011.SZ	0.0401	0.0317	0.0166

表 5.7 白酒每只股票的预测效果

Tab. 5.7 Prediction effect of each stock in liquor stock

股票名称	代码	MAE_{ma_5}	MAE_{ma_20}	MAE_{ma_120}
贵州茅台	600519.SH	0.0406	0.0293	0.0190
山西汾酒	600809.SH	0.0405	0.0277	0.0256
古井贡酒	000596.SZ	0.0424	0.0268	0.0204
酒鬼酒	000799.SZ	0.0471	0.0332	0.0184
水井坊	600779.SH	0.0389	0.0247	0.0185
伊力特	600197.SH	0.0332	0.0317	0.0242
金种子酒	600199.SH	0.0407	0.0250	0.0172
老白干酒	600559.SH	0.0339	0.0265	0.0224
顺鑫农业	000860.SZ	0.0301	0.0226	0.0185
*ST 皇台	000995.SZ	0.0557	0.0292	0.0177

表 5.8 生物医疗每只股票的预测效果

Tab. 5.8 Prediction effect of each stock in biomedical stock

股票名称	代码	MAE_{ma_5}	MAE_{ma_20}	MAE_{ma_120}
长春高新	000661.SZ	0.0360	0.0544	0.0242
派林生物	000403.SZ	0.0283	0.0382	0.0150
四环生物	000518.SZ	0.0425	0.0294	0.0151
万泽股份	000534.SZ	0.0263	0.0227	0.0242
北大医药	000788.SZ	0.0379	0.0226	0.0204
中关村	000931.SZ	0.0288	0.0254	0.0243

河化股份	000953.SZ	0.0362	0.0145	0.0135
华兰生物	002007.SZ	0.0307	0.0574	0.0211
向日葵	300111.SZ	0.0212	0.0190	0.0135
上海莱士	002252.SZ	0.0557	0.0390	0.0138

表 5.9 5G 概念每只股票的预测效果

Tab. 5.9 Prediction effect of each stock in 5G concept stock

股票名称	代码	MAE_{ma_5}	MAE_{ma_20}	MAE_{ma_120}
紫光国微	002049.SZ	0.0277	0.0229	0.0161
德赛电池	000049.SZ	0.0313	0.0259	0.0229
中兴通讯	000063.SZ	0.0411	0.0296	0.0206
中国长城	000066.SZ	0.0454	0.0317	0.0151
特发信息	000070.SZ	0.0335	0.0262	0.0152
航天发展	000547.SZ	0.0412	0.0386	0.0254
铜陵有色	000630.SZ	0.0309	0.0246	0.0154
风华高科	000636.SZ	0.0452	0.0380	0.0185
北斗星通	002151.SZ	0.0335	0.0340	0.0203
华天科技	002185.SZ	0.0424	0.0390	0.0179

在上述的表格中， MAE_{ma_5} 、 MAE_{ma_20} 、 MAE_{ma_120} 分别表示模型预测周线 ma_5、月线 ma_20、季线 ma_120 的绝对误差。根据实验的预测结果将每一种均线的预测效果前 10 组成 3 个选股股票池，将预测股票的周线 ma_5、季线 ma_20、月线 ma_120 分别看作成股票的短期、中期、长期的发展情况。表 5.10 为预测模型在短期、中期、长期股票上推荐的选股股票池，表中由上到下依次为模型预测效果的排序，投资者可以根据自己的需求自行选择交易的股票池。

表 5.10 预测模型在短期、中期、长期股票上推荐的选股股票池

Tab. 5.10 Stock pool recommended by the prediction model for short-term, medium-term and long-term stocks

短期		中期		长期	
股票名称	代码	股票名称	代码	股票名称	代码
向日葵	300111.SZ	河化股份	000953.SZ	深振业A	000006.SZ

中国宝安	000009.SZ	神州高铁	000008.SZ	ST星源	000005.SZ
神州高铁	000008.SZ	向日葵	300111.SZ	神州高铁	000008.SZ
万泽股份	000534.SZ	ST星源	000005.SZ	平安银行	000001.SZ
美丽生态	000010.SZ	中国宝安	000009.SZ	河化股份	000953.SZ
紫光国微	002049.SZ	顺鑫农业	000860.SZ	向日葵	300111.SZ
派林生物	000403.SZ	北大医药	000788.SZ	上海莱士	002252.SZ
深振业A	000006.SZ	深振业A	000006.SZ	万科A	000002.SZ
中关村	000931.SZ	万泽股份	000534.SZ	中国宝安	000009.SZ
顺鑫农业	000860.SZ	紫光国微	002049.SZ	*ST全新	000007.SZ

5.5 本章总结

本章所设计的是基于多影响因子下的量化选股策略,主要针对的是普通用户在选择股票时的难题,股票受多项影响因子决定,在量化时考虑股票具有的量化指标数据,将其分为基础指标、技术指标和新闻指标。实验将这些影响因子加入数据集中的输入特征中组成新的数据集,考虑到股票数据中的股票种类单一,对数据集进行扩充,加入三种类型的股票组成股票总池。将股票总池放入预测模型 Dens-TCN-GAM 中,经过模型预测后输出每一只股票的预测误差,选择预测效果最好的 10 只股票作为选股的股票池。对于预测周线来说,每种均线对应的股票发展的周期不同,将周期划分为短期、中期和长期,经过选股后的股票都具有预测效果好的特点,投资者这样就可以规避投资风险,获得更大的收益。

6. 总结与展望

6.1 总结

本文先对股票预测的研究背景和意义进行了详细的阐述,股票的预测对于股票投资者的实用价值。然后说明了股票预测的国内外研究的现状,各种传统预测方法的优势和不足紧接着随着深度学习模型的提出,使股票预测的研究更进一步发展,其中 TCN 模型在股票预测的模型中的表现传统的模型和其他深度学习模型好。为了进一步的使 TCN 模型在股票预测的准确率更高、误差更小,以 TCN 为实验基础模型,进行了两种不同方式的改进方法。在实验预测的对象中选取周线 ma_5、月线 ma_20 和季线 ma_120,分别对于股票市场的期、中期、长期价格变化趋势,对于股票价格预测提供了新的预测方法,使投资者在短中获取良好的收益。

本文的主要工作可以分为:

1.本文所进行的股票预测研究是以均线价格来预测股票的趋势情况,股民可以通过预测的数据对不同市场进行交易。短期需求的股民可以观察周线的股票历史价格来进行交易,使收益最大化,中期和长期需求的股民通过月线和季线均线的价格趋势图也可判断出股票即将上涨或下跌。实验只研究三种股票第二天价格的变化,即为股票预测值与当日股票的真实值对比的情况。由于模型在预测较为长久的时间序列时会导致模型在误差进行累积后,最终的误差值爆炸性增长,所有没有考虑改进模型在更加长久的步长中预测的效果。

2.在标准的 TCN 时间卷积模块 Temporal block 中,模型在进行特征提取时会丢失带时间的股票价格特征,导致模型预测股票的未来的价格时,股票的预测值与真实值有所偏差,基于此问题,本文提出了一种提高模型在提取特征完整的方法。将多尺度卷积与时间卷积模块相融合形成新的模块,然后将带有多尺度卷积的时间卷积的模块作为残差模块,使用残差连接的方式将每一层的残差模块连接组成改进后的时间卷积模型 Res-TCN。通过使用沪深 A 股前 10 只股票的训练数据集实验后,改进的模型取得了更好的预测效果。

3.本文介绍了全局注意力模型 GAM,将 GAM 加入每一层的时间卷积模块中。为了防止网络出现梯度消失和网络退化的现象,在网络的连接部分采用密集连接的方式取代残差连接,这样就设计出了第二种改进后的模型 Dens-TCN-GAM。实验前优化模型的优化算法,选取预测效果好的 Adam 优化算法,通过实验表明具有全局注意力密集连接的改进模型对于股票的均线价格预测更有优势,提取股票均线价格特征的效率更高。

4.根据不同投资者的需求,让股票预测更具准确,设计了基于多影响因子的选股策略系统。对实验的数据集的输入特征进行扩充,将多种影响股票的影响因子加入股票数据中,同时加入不同种类的股票数据进入股票总池。使用预测模型对股票进行预测,剔除预测效果不好的股票,选出预测合适的预测股票组成股票的选股池。提供给短期、中期和长期交易方式的投资者的投资意向,实现股票收益的最大化。

6.2 展望

本文提出两种改进的时间卷积网络来对股票的均线价格进行预测,在通过消融实验后,说明改进模型在预测效果上面有所提升,但依然从在一下的问题,需要在进一步的研究进行改进:

1.实验的预测方式是基于股票交易的“T+1”模式下进行的,采用的数据集是股票的日线数据,未考虑其交易模式下的预测问题,可以设置滑动窗口探究股票预测模型在那一段时间段内预测效果更好。

2.本文只对股票的历史数据大于 10 年的股票类型进行预测,若上市公司的时间较短,实验的数据集还需要进行扩充才可以进行预测。

3.设计的两种实验模型都存在实验的消耗时间过长的问题,后续还需要对模型的预测的效率进行改进。

4.在进行量化时为考虑股票的收益率问题,可以加入具体的收益率大小,这样对选股的效果更有说服力。

参考文献

- [1] 傅捷,华生,汲铮. 关键历史节点与资本市场股票发行制度演进[J]. 东南学术, 2022(5):206-214.
- [2] Fama E F. Efficient capital markets: II[J]. The journal of finance, 1991, 46(5):1575-1617.
- [3] Hyndman R J, Athanasopoulos G. Forecasting: principles and practice[M]. OTexts, 2018.
- [4] Bollerslev T. Generalized autoregressive conditional heteroskedasticity[J]. Journal of econometrics, 1986, 31(3):307-327.
- [5] 杨佳鹏. 基于注意力机制融合多特征的股票预测研究[D]. 北京:北京交通大学, 2021.
- [6] 欧仁鹏. 人民币实际有效汇率对中国商品出口的影响[D]. 东北财经大学, 2017.
- [7] Bauwens L, Laurent S, Rombouts J V K. Multivariate GARCH models: a survey[J]. Journal of applied econometrics, 2006, 21(1): 79-109.
- [8] Junior P R, Salomon F L R, de Oliveira Pamplona E. ARIMA: An applied time series forecasting model for the Bovespa stock index[J]. Applied Mathematics, 2014, 5(21):3383.
- [9] Zheng T, Farrish J, Kitterlin M. Performance trends of hotels and casino hotels through the recession: An ARIMA with intervention analysis of stock indices[J]. Journal of Hospitality Marketing & Management, 2016, 25(1):49-68.
- [10] Miah M, Rahman A. Modelling volatility of daily stock returns: Is GARCH (1, 1) enough[J]. American Academic Scientific Research Journal for Engineering, Technology, and Sciences, 2016, 18(1):29-39.
- [11] Sun J, Xiao K, Liu C, et al. Exploiting intra-day patterns for market shock prediction: A machine learning approach[J]. Expert Systems with Applications, 2019, 127:272-281.
- [12] 高天. 基于最优小波包变换、ARIMA 与 SVR 的股票价格预测研究[J]. 贵州财经大学学报, 2015(6):57-69.
- [13] 吴玉霞, 温欣. 基于 ARIMA 模型的短期股票价格预测[J]. 统计与决策, 2016(23):83-86.
- [14] 熊政, 车文刚. ARIMA-GARCH-M 模型在短期股票预测中的应用[J]. 陕西理工大学学报(自然科学版), 2022,38(4):69-74.
- [15] 徐浩然, 许波, 徐可文. 机器学习在股票预测中的应用综述[J]. 计算机工程与应用, 2020, 56(12):19-24.
- [16] Alpaydin E. Introduction to machine learning[M]. MIT press, 2020.
- [17] Breiman L. Random forests[J]. Machine learning, 2001, 45(1):5-32.
- [18] Quinlan J R. Induction of decision trees[J]. Machine learning, 1986, 1(1):81-106.
- [19] 李红娇, 陈红艳. 基于 WGAN 的移动恶意对抗样本生成研究[J]. 信息网络安全,

2020(11):51-58.

[20] 王义锋. 基于 XGBoost+RF 的个人信贷风险预测研究[D]. 重庆:重庆大学,2020.

[21] 方毅, 陈煜之, 卫剑. 人工智能与中国股票市场——基于机器学习预测的投资组合量化研究[J]. 工业技术经济, 2022, 41(8):83-91.

[22] Cao Q, Leggio K B, Schniederjans M J. A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market[J]. Computers & Operations Research, 2005, 32(10):2499-2512.

[23] Trafalis T B, Ince H. Support vector machine for regression and applications to financial forecasting[C]//Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium. IEEE, 2000, 6:348-353.

[24] 郭辉. 支持向量机选择及其在股票走势预测中的应用[J]. 重庆师范大学学报(自然科学版), 2007, 24(4):45-49.

[25] 闫政旭, 秦超, 宋刚. 基于 Pearson 特征选择的随机森林模型股票价格预测[J]. 计算机工程与应用, 2021, 57(15):286-296.

[26] 林娜娜, 秦江涛. 基于随机森林的 A 股股票涨跌预测研究[J]. 上海理工大学学报, 2018, 40(3):267-273, 301.

[27] Khaidem L, Saha S, Dey S R. Predicting the direction of stock market prices using random forest[J]. arXiv preprint arXiv:1605.00003, 2016.

[28] Tolstikhin I O, Houlsby N, Kolesnikov A, et al. Mlp-mixer: An all-mlp architecture for vision[J]. Advances in neural information processing systems, 2021, 34: 24261-24272.

[29] De Oliveira F A, Nobre C N, Zárate L E. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil[J]. Expert systems with applications, 2013, 40(18):7596-7606.

[30] 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 Communication and Computational Intelligence (INCOCCI). IEEE, 2010:597-602.

[31] 尹光花, 刘小明, 张露等. 基于 LSTM 特征模板的短文本情感要素分析与研究[J]. 电子科技, 2018, 31(11):38-41, 46.

[32] 张晨荣. 基于联合学习的知识库问答研究[D]. 内蒙古:内蒙古科技大学, 2020.

[33] 黄齐禹. 基于 LSTM 神经网络认知无线电频谱预测方法的研究[D]. 黑龙江:哈尔滨工业大学, 2020.

[34] Baek Y, Kim H Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module[J]. Expert

Systems with Applications, 2018, 113:457-480.

[35] 张珂, 冯晓晗, 郭玉荣等. 图像分类的深度卷积神经网络模型综述[J]. 中国图象图形学报, 2021, 26(10):2305-2325.

[36] 颜荔. 基于卷积神经网络的遥感图像飞机目标识别研究[D]. 安徽:中国科学技术大学, 2018.

[37] Kim T, Kim H Y. Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data[J]. PloS one, 2019, 14(2):e0212320.

[38] Lea C, Flynn M D, Vidal R, et al. Temporal convolutional networks for action segmentation and detection[C]//proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:156-165.

[39] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:1803.01271, 2018.

[40] Guo W, Li Z, Gao C, et al. Stock price forecasting based on improved time convolution network[J]. Computational Intelligence, 2022, 38(4): 1474-1491.

[41] 田秀杰. 基于流动性与价格波动的香港内地股市整合的实证研究[D]. 辽宁:东北财经大学, 2006.

[42] 路季超. 基于深度学习的股票预测模型与交易策略研究[D]. 陕西:西北大学, 2021.

[43] Thakkar A, Chaudhari K. Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions[J]. Information Fusion, 2021, 65:95-107.

[44] 赵世林. 网络大数据驱动的应用程序智能分类与识别[D]. 中国科学院大学(中国科学院深圳先进技术研究院), 2020.

[45] 陈玄, 朱荣, 王中元. 基于融合卷积神经网络模型的手写数字识别[J]. 计算机工程, 2017, 43(11):187-192.

[46] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition[M]//Competition and cooperation in neural nets. Springer, Berlin, Heidelberg, 1982:267-285.

[47] 嵇文萍. 基于集成学习的恶意软件多特征检测技术研究[D]. 北京交通大学, 2021.

[48] 刘栋. 深度学习框架下时空关联的基坑变形预测方法研究[D]. 山东农业大学, 2022.

[49] 张西宁, 刘书语, 余迪等. 改进深度卷积神经网络及其在变工况滚动轴承故障诊断中的应用[J]. 西安交通大学学报, 2021, 55(6):1-8.

[50] 王阳, 牛长流, 马国昊等. 基于深度学习的多人语音识别研究[J]. 数字技术与应用, 2021, 39(7):71-74.50

[51] 汪亚航. 基于深度学习的人脸活体检测算法研究[D]. 江南大学, 2021.

- [52] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770-778.
- [53] Nabipour M, Nayyeri P, Jabani H, et al. Deep learning for stock market prediction[J]. Entropy, 2020,22(8):840.
- [54] Mnih V, Heess N, Graves A. Recurrent models of visual attention[J]. Advances in neural information processing systems, 2014, 27.
- [55] 任欢, 王旭光. 注意力机制综述[J]. 计算机应用, 2021, 41: 1-6.
- [56] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:1-9.
- [57] 贺然. 基于视频内容感知的视频质量评价方法研究[D]. 陕西:西安电子科技大学, 2019.
- [58] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409. 1556, 2014.
- [59] 冯庆东. 基于神经网络的情感分析研究[D]. 重庆:重庆邮电大学, 2021.
- [60] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. PMLR, 2015:448-456.
- [61] Liu Y, Shao Z, Hoffmann N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions[J]. arXiv preprint arXiv:2112. 05561, 2021.
- [62] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018:3-19.
- [63] Park J, Woo S, Lee J Y, et al. Bam: Bottleneck attention module[J]. arXiv preprint arXiv:1807. 06514, 2018.
- [64] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:7132-7141.
- [65] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:4700-4708.
- [66] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412. 6980, 2014.