



单位代码: 10484

编 号:

密 级: 公 开

# 河南财经政法大学

## 硕 士 学 位 论 文

论 文 题 目: 基于 Trans-T2V 的量化选股策略研究

学科、专业名称: 金融

研 究 方 向: 量化投资

申请学位类别: 金融硕士

申 请 人 姓 名: 李 瞿 成

导师姓名、职称: 喻军 副教授

定 稿 时 间: 2022 年 5 月

中国·郑州

## 学位论文独创性声明

本人郑重声明，所呈交的学位论文是我个人在导师指导下，独立进行研究工作所取得的成果。论文中除了已经注明引用的内容外，不包含其他个人或集体已经发表或撰写过的研究成果，也不包含为获得河南财经政法大学或其它教育机构的其他学位或证书而使用过的材料。对本学位论文的研究做出任何贡献的个人和集体，均已在文中作了明确的说明并表示了谢意，因本文引起的法律后果完全由本人承担。

特此声明

学位论文作者签名：  日期： 2022 年 5 月 8 日

## 学位论文版权使用授权书

作为学位论文的作者，本人完全了解河南财经政法大学有关保留、使用学位论文的规定，河南财经政法大学有权保留所学位论文（纸质文本和电子文本），可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采取影印、缩印、扫描和拷贝等复制手段保存、汇编学位论文（纸质文本和电子文本），允许论文被查阅和借阅。本人授权河南财经政法大学向国家图书馆、科研信息机构、数据收集机构等提供本学位论文，同意研究生处将本学位论文提交中国学术期刊（光盘版）电子杂志社和万方数据电子出版社全文出版，并编入 CNKI《中国知识资源总库》与万方《中国学位论文全文数据库》，传播学位论文的全部或部分内容。（涉及保密内容的学位论文在解密后适用本授权书）

学位论文作者签名：  日期： 2022 年 5 月 8 日

指导教师签名：  日期： 2022 年 5 月 8 日

## 摘 要

近年来,全球人工智能的研究热潮愈演愈烈,从 AlphaGo 到 Master 再到 AlphaGoZero,人工智能已在一部分场景下具备了战胜人类、不断进化的能力,并渗透在人们日常生活中的方方面面。深度学习作为当代人工智能的新潮流,已被各个基金公司尝试应用于量化投资领域。本质上,量化投资是应用金融领域中的各个指标进行建模,以预测金融资产的回报,并根据预测结果制定投资策略的一种投资方式。将深度神经网络的学术成果运用在量化投资场景之中,其一能够帮助广大投资者更准确地把握股票收益的变化趋势,有利于投资者在一定的风险水平下获取更为可观的溢价,其二有利于为广大的投资者提供一种新的量化研究视角。本文以在自然语言处理、计算机视觉等方面大放异彩的 Transformer 深度神经网络作为研究对象,根据 Time2Vector 中对于时间特征的描述构建了时间嵌入层 Time2VectorEmbedding,对 Transformer 深度神经网络进行改进并将其命名为 Trans-T2V,让模型更适用于处理金融时序信息并测试该模型在量化选股中的表现。

首先,本文通过添加时间维度将股票原始的 2 维数据处理为 3 维数据以便于模型识别,进行训练与预测;其次为了提升模型泛化性能,本文采取滚动的方式对模型进行训练与预测,将滚动训练集长度设置为 12 个月,利用每 12 个月的数据训练模型并预测下个月的超额收益情况,以月作为滚动窗口对数据集采取滚动训练,合计滚动 120 次。然后,本文根据实验环境及参数寻优的结果确定了注意力头数为 3、多头注意力层数为 1、注意力层神经元数量为 512、前馈层神经元数量为 256、dropout 率为 0.2、学习率为 0.01 的模型参数来构建模型。之后本文将 Trans-T2V 模型的均方误差与 Transformer、递归神经网络(recursive neural network, RNN)、长短期记忆神经网络(Long Short-Term Memory, LSTM)的均方误差进行对比,最后基于 Trans-T2V 模型构建相应的选股策略进行分层回测,观察回测结果,并与 Transformer、RNN 和 LSTM 神经网络为基准的选股策略的回测结果再次进行比较分析。

实证表明,本文基于 Trans-T2V 模型构建的量化选股策略在回测期内获得了超过 4 倍的累计收益,年化收益及年化超额收益分别可达 18.76%、15.30%,年

化波动率为 8.21%，相对基准月胜率高达 63%，超额收益最大回撤为 32.84%。本文设计了一种多空投资组合策略来进一步地评价该模型的股票优选能力，实验结果表明，多空策略的年化收益达到了 17.74%，年化超额收益为 14.29%，年化波动率仅为 2.32%，超额收益最大回撤为 14.37%。对比 Trans-T2V 模型与 Transformer、RNN 神经网络及 LSTM 神经网络在量化选股上的表现后发现 Trans-T2V 的风险指标及收益指标均高于其他 3 个模型。因此本文所提出的 Trans-T2V 是一个有效的量化选股模型，能够为投资者构建主动型投资组合提供一定的参考。

**关键词：**量化选股；深度学习；Trans-T2V 模型

## Abstract

In recent years, the global artificial intelligence research boom is getting more and more intense. From AlphaGo to Master to AlphaGoZero, artificial intelligence has been able to defeat human beings in part of the scene and constantly evolve, and infiltrated into all aspects of People's Daily life. As a new trend of contemporary artificial intelligence, deep learning has been applied to quantitative investment by various fund companies. In essence, quantitative investment is an investment method in which various indicators in the financial field are modeled to predict the return of financial assets and formulate investment strategies based on the predicted results. The depth of the academic achievements of neural networks used in quantitative investment scene, one is able to help the investors to more accurately grasp the trend of the stock returns to investors under a certain level of risk for a more substantial premium, secondly to for the general investors to provide a new quantitative research perspectives. This paper takes Transformer deep neural network which is outstanding in natural language processing and computer vision as the research object. According to the description of time features in Time2Vector, the Time2VectorEmbedding is constructed. The Transformer deep neural network is improved and named trans-T2V to make the model more suitable for processing financial time series information and test the performance of the model in quantitative stock selection.

First of all, this paper processed the original 2d data of stock into 3D data by adding time dimension to facilitate model recognition, training and prediction. Secondly, in order to improve the model generalization performance, this paper adopt the method of rolling for training and prediction model, sets the rolling length of the training set to 12 months, using the data of every 12 months training model and predict the excess earnings of the next month, month as rolling window to take the scroll training data sets, a combined rolling 120 times. Then, according to the results of experimental environment and parameter optimization, this paper determines the model parameters of 3 attention heads, 1 multi-attention layers, 512 neurons of

attention layer, 256 neurons of feedforward layer, dropout rate of 0.2, and learning rate of 0.01 to construct the model. Then, this paper compared the mean square error of trans-T2V model with that of Transformer, Recursive Neural Network and Long short-term Memory neural network. Finally, based on the Trans-T2V model, the corresponding stock selection strategy was constructed for hierarchical back test, and the back test results were observed, and compared with the results of Transformer, RNN and LSTM neural network as the benchmark.

The empirical results show that the quantitative stock selection strategy based on trans-T2V model in this paper obtained more than 4 times cumulative returns in the backtest period, with annual returns and annual excess returns reaching 18.76% and 15.30%, respectively. The annual volatility is 8.21%, the relative benchmark monthly win rate is as high as 63%, and the maximum retracement of excess returns is 32.84%. In this paper, a long-short portfolio strategy is designed to further evaluate the stock selection ability of the model. Experimental results show that the annual return of the long-short strategy is 17.74%, the annual excess return is 14.29%, the annual volatility is only 2.32%, and the maximum withdrawal of excess return is 14.37%. After comparing the performance of Trans-T2V model with Transformer, RNN neural network and LSTM neural network in quantitative stock selection, it is found that the risk index and return index of Trans-T2V are higher than the other three models. Therefore, trans-T2V proposed in this paper is an effective quantitative stock selection model, which can provide a certain reference for investors to construct an active portfolio.

**Key words:** Quantitative stock selection; Neural network; Trans-T2V Model

## 目 录

1	绪 论.....	9
1.1	研究背景.....	9
1.2	研究意义.....	10
1.2.1	理论意义.....	10
1.2.2	实际意义.....	11
1.3	文献综述.....	11
1.3.1	量化投资模型发展文献综述.....	11
1.3.2	深度学习在量化投资应用文献综述.....	14
1.3.2	文献综述总结.....	16
1.4	研究内容和方法.....	16
1.5	本文主要贡献和创新.....	18
2	相关理论介绍.....	19
2.1	Transformer 介绍.....	19
2.2	Time2Vector 介绍.....	27
2.3	Trans-T2V 模型结构介绍.....	28
2.4	本章小结.....	33
3	基于 Trans-T2V 的股票预测模型.....	34
3.1	模型设计流程.....	34
3.1.1	数据选取.....	34
3.1.2	特征和标签提取.....	34
3.1.3	特征预处理.....	34
3.1.4	训练集和交叉验证集的生成.....	36
3.1.5	交叉验证调参.....	36
3.1.6	样本外测试.....	36
3.1.7	模型评价.....	37
3.2	Trans-T2V 模型及对比模型设置.....	37
3.2.1	Trans-T2V 模型参数及结构设置.....	37
3.2.2	Transformer 模型参数及结构设置.....	40
3.2.3	循环神经网络模型参数及结构设置.....	41
3.2.4	长短期记忆神经网络模型参数及结构设置.....	41
3.3	基于 Trans-T2V 模型的量化选股策略.....	42
3.4	本章小结.....	45
4	实证分析.....	46
4.1	实证环境及数据选取.....	46
4.1.1	开发环境.....	46
4.1.2	数据选取.....	46
4.2	评价指标.....	47
4.2.1	模型评价指标.....	47
4.2.2	策略评价指标.....	47

4.3	各模型损失函数分析.....	49
4.4	基于各模型的策略回测结果对比分析.....	50
4.4.1	不同投资组合对比.....	50
4.4.2	不同选股策略对比.....	52
4.5	本章小结.....	54
5	研究结论.....	55
5.1	研究总结.....	55
5.2	不足与改进.....	56
	参考文献.....	57
	附录.....	62
1	所用因子展示.....	62
	致谢.....	66



# 1 绪 论

## 1.1 研究背景

近年来,我国股权分置改革在不断地深化,证券市场也取得了长足发展,在这发展的洪流之中,众多的机构投资者与散户投资者期望以较低的风险获取可观的收益。那么,如何在庞大的股票市场中选取一些优质股份构建适合自己的投资组合就是一个被大家所关心的问题。量化投资是使用数理模型、计算机技术和庞大的数据来进行建模从而区分自身与传统的基本面、技术面分析的投资差异,这种新兴投资方式旨在利用数理统计的规律来战胜市场。表面上看,与深度学习有关的研究似乎起始于 21 世纪,从谷歌团队设计的 AlphaGo 起,深度学习迅速地攻占了大众的视野高地。然而,“深度学习”早在 20 世纪 50 年代就已出现,当时的研究人员使用感知机算法将输入的多维数据进行二分类。但在 1969,有人提出,感知机模型本质上仅是一种线性模型,即使在遇到最简单的非线性分类问题 XOR 时,也无能为力。这使得之后的几十年间,深度学习进展缓慢,直到近年来才蓬勃发展。原因源于两方面,其一在于深度学习若要建立一个泛化能力强、具有现实意义的模型,必须要有海量的数据来支撑,而近年来,广大学者有着蓬勃发展的大数据技术作为支持,使他们能够运用更加庞大而丰富的数据为自身深度神经网络模型的训练提供帮助;其二在于计算机硬件设备不断地推陈出新,不同于 20 世纪,21 世纪的学者们之所以能够在研究深度神经网络时不断地改进模型的广度与深度,原因在于当前的计算机硬件设施有着性能更优秀的 CPU、GPU、更大的显卡内存和更好的分布式计算架构的技术支持,这无疑促使了众多学者创造出了许多经典的神经网络,为深入研究深度学习奠定了坚实的基础。在深度学习大热之前,人们在选择构建投资组合的建模方法时,大多都使用诸如马科维兹均值方差模型之类的经典线性回归模型,这些传统的数学建模方法因其有着简单明了、具有很强的可解释性而流传至今。然而,传统的线性回归模型仅在处理金融时间序列中解释变量与被解释变量是线性关系时有优势,对于更为复杂的非线性关系,线性回归模型往往会陷入欠拟合的尴尬境地。而近几年来有赖于深度学习研究的不断深入以及数据的愈加丰富,越来越多的量化研究员拥有了将深度学习的研究成果变现的能力。深度神经网络模型不同于经典线性回归

模型的地方在于,该型结构具有层次更深、参数更多、整体更复杂的特点,这在一定程度上有利于模型在复杂的金融市场找到股票收益率与各因素之间的非线性复杂函数关系,因此在处理解释变量与被解释变量的关系具有高度复杂性和非线性特点的任务上更具优势。

在有效市场理论中,一个强有效市场会将与股票价格相关的所有信息都反应在当前时点的股票价格中,但现实中的股票价格并不完全符合这一假说。从现实来看,股票市场中影响股票收益率变动的因素众多,它们作为股票收益率的被解释变量,既存在着线性的可能又存在着非线性的可能。同时,影响因素之间也可能存在着关联性,对股价有着正向或负向的影响。在这样的背景下,投资者通过使用深度神经网络来预测股票价格收益的变动情况相对于线性模型来说更具优势。在现有使用机器学习的非线性模型来指导量化投资的文献中,大多是有关支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest)、BP神经网络等机器学习算法和 Xgboost、Adaboost 等集成优化算法。目前众多学者对深度学习的耕耘主要还是集中在了计算机视觉、自然语言处理、语音识别与合成、目标检测等领域,将深度神经网络的成果放在量化选股上进行研究还远非主流。另外,由于循环神经网络 RNN 在处理时间序列数据方面具有一定的优点,因此其中大部分的量化研究都与 RNN 相关。当前,Transformer 因其有着更高效的计算速度,更强大的预测精度而在自然语言处理、图像识别等各个领域大放异彩,但如何利用该神经网络对金融时间序列进行建模与预测,如何将 Transformer 模型应用到量化选股领域,相关研究的文献却是一片空白,因此本文所做的研究对深度学习的金融应用方面来说是必要的。

## 1.2 研究意义

### 1.2.1 理论意义

近几年,人们不断将深度学习与股票涨跌趋势预测结合起来,但大多数研究者都将目光集中在以递归神经网络 RNN 和长短期记忆网络 LSTM 为代表的循环神经网络结构上,本论文通过将近两年大放异彩的 Transformer 神经网络与时间特征 Time2Vector 相结合,提出一种适用于处理金融时间序列数据的模型,将其命名为 Trans-T2V 并应用于股票预测当中,据此构建出相应的量化选股策略。因此,本文的研究既是对 Transformer 如何应用于金融中量化投资领域的探索,又

应用了时间特征 Time2Vector 来优化建模，为后续金融时间序列建模的优化方面的研究提供了一定的参考价值。

### 1.2.2 实际意义

本文为投资者提供了一种新的深度神经网络模型结构来处理金融时间序列数据，实验证明该结构能够有效提升投资者的股票预测能力。本论文对量化选股模型的构建过程进行了详细地阐述，具体过程包括数据源的寻找与下载、数据预处理，训练集、验证集与测试集的划分、目标值的标记方法、模型架构的设计及相关参数的选取过程、基于 Trans-T2V 股价预测模型构建的量化选股策略设计以及分层回测方法与评价指标等。对投资者而言，可参考本文中的模型构建思路自行搭建模型并进行回测，从而为自己的投资决策做出更为理性的判断和指导，使投资者能够在一个较低的风险水平中获取更高的收益。

## 1.3 文献综述

### 1.3.1 量化投资模型发展文献综述

在海外，量化投资已发展了40余年，据Preqin数据库中的数据显示，2020年国际上对冲基金top10中有4家以量化投资为主。在美国，量化对冲基金所占市场规模已超过三分之一，而量化投资的起源则可以追溯到20世纪50年代，当时Markowitz(1952)<sup>[18]</sup>建立了一种资产组合理论，它试图定量地确定风险与收益间的关系，用方差刻画风险，均值描绘收益，并给出了一种分散化的投资组合方法，这种方法是利用无风险资产和风险资产，构造出在单位收益水平上风险最低或在单位风险水平上获得最优收益的资产组合，这一理论成为了量化投资的基础理论，为后来诸多量化模型提供了参照。之后Sharpe (1964)<sup>[19]</sup>和Lintner(1965)<sup>[20]</sup>等人提出了资本资产定价模型（CAPM），该模型的提出有赖于资本市场理论和资产组合理论，它是建立在一系列假设基础上，通过建立股票的超额收益和市场的超额收益之间的关系函数来描述二者的关系，资本资产定价理论将风险成分划分为系统性风险beta和非系统性风险alpha，认为股票的收益情况与风险beta之间呈线性关系，应用该模型可对股票价格进行短期预测。接着，Ross (1976)<sup>[21]</sup>提出了套利定价理论，该理论在看待资产价格的定价因素时有着不同的角度，ross以一价定律和套利为视角来解释金融市场中资产定价的影响因素，为量化投资理

论带来了新的思路。但人们发现,如果简单地将上述的模型直接应用在现实的市场中,很难完整地解释资产价格的变化情况。于是,Fama和French(1992)<sup>[22]</sup>在此背景下提出了Fama-French三因子模型,三因子模型将市场因子、规模因子和价值因子作为解释变量,将股票收益率作为被解释变量来解释和预测股票的变化情况。通过实证检验,价值因子、规模因子相对于股票收益率的回归系数在统计学意义上显著,这意味着这些因子在当时的市场环境中能够捕获解释投资组合风险溢价的有关信息,因此业界和学术界迅速地接受并应用了该模型,使得量化投资这一概念变得易于实施,大大加快了量化投资的发展速度。之后,Fama和French的学生Carhart(1996)<sup>[23]</sup>在三因子模型的启发下,把动量因素当作新的被解释变量,组成动量因子、市场因子、规模因子和价值因子的四因子模型并应用于共同基金当中,观察其长期变化情况。自此之后,越来越多的学者陆续对公司的财务指标进行研究,不断发掘出新的因子,从而试图构造出新的多因子模型,以完善三因子模型的研究思路。但是, Titman、Wei和Xie(2004)<sup>[24]</sup>在相关文献中发现, Fama-French三因子模型和Carhart四因子模型所计算出来的数据都与市场中的真实数据存在出入,也就是说,这两种多因子模型在预测股票收益率方面的精度较差,为了优化模型,他们加入了投资机会及盈利能力这两个指标来进一步解释现实的股票收益率数据。到了2015年, Fama和French(2015)<sup>[25]</sup>在三因子模型的基础上进一步发展出了五因子模型,五因子模型的创新之处在于添加了盈利因子和投资因子作为新的被解释变量,与之前三个被解释变量共同解释个股的收益情况,通过实证, Fama和French证明了五因子模型能够在当时的市场上很好地解释股票收益率数据。

对于学术界而言,过往传统的时间序列数据建模方式多为线性回归方程,当前在处理时间序列数据时众多学者也在不断尝试着新的数理研究成果, black和 scholes(1973)<sup>[26]</sup>在7个重要的假设基础上,利用了正态分布变量的累积概率分布函数搭建了著名的B-S期权公式,使得期权定价得以量化,推动了量化研究的发展。Box和Jenkins(1976)<sup>[27]</sup>在20世纪80年代提出了差分整合移动平均自回归模型(ARIMA)用以进行时间序列数据的预测, Engle于(1982)<sup>[28]</sup>提出了自回归条件异方差模型(ARCH)用以解决方差恒定所引起的问题,之后, T.Bollerslev(1986)<sup>[29]</sup>又提出了广义自回归条件异方差模型(GARCH),该模型是对自回归条件异方差模型的一种改进,尝试进一步地解决时间序列方差恒定问题。随着21世纪计算机设备和数据网络的不断迭代,机器学习开始迅猛发展,这使得不少

学者将目光投向机器学习,试图利用更为复杂的模型来分析、预测金融时间序列数据。Cao等人(2005)<sup>[30]</sup>将人工神经网络(ANN)和Fama-French三因子模型应用于中国股票市场,分别利用两种模型对股票数据进行预测并做出对比,结果发现,人工神经网络在处理金融时间序列数据方面更优秀。Wang等人(2008)<sup>[31]</sup>对于金融时间序列数据的处理思路建立在支持向量机的模型基础上,通过两步核学习算法来获得多个候选核函数的稀疏线性组合从而对支持向量机的核函数进行了改进。Hadavandi等人(2010)<sup>[32]</sup>在人工神经网络(ANN)的基础上引入了遗传模糊系统(GFS),使遗传系统的核心优势得以运用,实证结果表明该模型结构优于其他对比结构。Ticknor(2013)<sup>[33]</sup>提出了贝叶斯正则化的人工神经网络用于预测估价,其通过减少过拟合来提高模型预测质量以及泛化能力,Sirohi等人(2015)<sup>[34]</sup>提出使用多核学习模型来预测股票价格的变动情况,他们的创新之处在于基核的构建,具体做法是将开盘价、收盘价、最高价、最低价等原始数据进行某种计算得到一系列新的技术指标,再以这些技术指标为特征建立不同的基核,然后将各个基核融入多核学习的架构中形成新的模型,从而达到提升模型预测精度的目的。Moghaddam(2016)<sup>[35]</sup>等人以纳斯达克2015年1月28日到2015年6月18日的股票每日换手率为研究对象,选择人工神经网络结构进行建模,实验结果证明了人工神经网络预测的有效性。Sezer等人(2017)<sup>[36]</sup>以技术分析指标作为股票的买卖信号,以人工神经网络和多层感知机作为模型基础,实验证明,在选择适当的技术指标作为买卖信号后,模型的结果与买入并持有的策略结果并无明显差异。YumoXu(2018)<sup>[37]</sup>等人针对股票市场的高随机性,噪声信息和时序依赖预测三个特点,提出了称为StockNet的网络架构,并利用tweet信息来增强股票预测。Sheng feng Mei(2020)<sup>[38]</sup>等人将股票价格与时间之间的定量关系定义为时间函数,根据股票的对数收益是白噪声序列的经验证据,建立了整数白噪声的数学模型来描述股票价格运动的现象。

在国内,乘着量化投资发展的迅猛之势,业界与学术界也将自身的目光投向了量化模型。然而,由于早期信息技术不够发达、金融市场不够完善,大部分研究都是依葫芦画瓢,将国外现有的理论研究换个场景,用中国市场的数据进行检验,理论缺乏创新,但随着时间的发展,研究也在持续深入,近年来基于三因子模型的创新性研究逐渐增多。如范振龙等人(2002)<sup>[1]</sup>首次把Fama-French三因子模型应用到中国金融时间序列数据的研究中,这项研究的选股范围为1995年7月到2000年6月期间全部A股股票的月收益率为研究对象,验证了三因子模型能够

较好的解释当时的中国股市。陈展辉等人（2003）<sup>[2]</sup>对于沪深A股市场股票收益率方面的研究也应用到了三因子模型，证明了当年的沪深A股股票价格受公司规模与股东权益账面市值比所影响。李传乐（2007）<sup>[3]</sup>采用均值-方差相交的方法构造了不同规模的模拟资产组合并作为因子模型的定价因子，结果表明大、中、小三个规模组合能够一定程度解释股票价格。兰峰（2010）<sup>[4]</sup>通过再生水行业关联整个市场，使用再生水行业特征对三因子模型进行改进，形成了四因子模型来解释收益率数据，结果表明该四因子模型可以解释再生水行业的收益率的变动情况。欧阳志刚等人（2016）<sup>[5]</sup>在三因子模型的基础上考虑了动量因子、反转因子对股票收益率的作用，以沪深A股月频数据为研究对象，构造出了符合中国股市的四因子模型。赵昕等人（2020）<sup>[6]</sup>针对股票市场的时变现象提出了时变三因子模型，以状态空间模型作为研究基础，以风险因子系数作为研究对象来研究股票市场超额收益的影响因素。此外，国内学者也对机器学习算法进行了探索，龙建成等人（2005）<sup>[7]</sup>根据非线性动力学系统理论构造了股票交易数据模型，使用并行神经网络进行学习来预测股票价格。刘海玥等人（2011）<sup>[8]</sup>分别构建了AR模型、RBF和GRNN神经网络模型对上证指数的价格进行预测，张秋明等人（2013）<sup>[9]</sup>在BP神经网络的基础上把灰色理论的研究思路引入其中，建立了一种新的股票价格预测模型，通过实证检验，这种预测模型可以有效地改善股价预测的准确性，也能够挖掘股票价格的变化规律。孙秋韵等人（2015）<sup>[10]</sup>基于改进遗传算法（GA）来优化支持向量回归机SVR，构建了股票预测模型，骆桦等人（2015）<sup>[11]</sup>基于朴素贝叶斯分类器对股票进行分类，构建了选股模型。孟叶（2018）<sup>[12]</sup>以Adaboost集成算法作为研究基础，以近邻传播聚类作为优化思路，将二者结合构建出了新的分类器，从而达到量化选股的目的。贾秀娟（2019）<sup>[13]</sup>以随机森林作为研究基础，模型结构上又融入了支持向量机，形成了改进的SVM模型结构，实验结果表明，该模型的分类性能优于其他对比模型。

### 1.3.2 深度学习在量化投资应用文献综述

随着深度学习的兴起，众多研究人员已将诸如人工神经网络（Artificial Neural Network, ANN）、循环神经网络（Recurrent Neural Networks, RNN）、长短期记忆神经网络（Long Short-Term Memory, LSTM）、门控循环神经网络（gated recurrent unit, GRU）和卷积神经网络（Convolutional Neural Networks,

CNN)等模型思路投入到股票市场的实证当中。在国外,Takeuchi、Lee (2013)<sup>[39]</sup>将深度神经网络的研究成果应用到了动量策略中,结果表明改进的动量策略其策略收益显著高于传统的动量策略。Rather 等人(2014)<sup>[40]</sup>将自回归移动平均模型、指数平滑模型和循环神经网络模型进行融合,相关最优权重通过遗传算法得到,由此构建出了一种新的股价预测模型,并通过实际数据验证了模型的有效性。Peng 和 Jiang (2015)<sup>[41]</sup>采用词嵌入方法和深度神经网络来预测股票的走势,发现通过将新闻文本中获得的特征添加到价格特征中可以提高预测性能。Türkmen 和 Cemgil (2015)<sup>[42]</sup>分析了各种机器学习算法对股票价格预测方面的有效性,并证明了使用 Stacked Denoising Auto-Encoders (SAE)的模型结构对预测股价的变动趋势更具精度。Xiong 等人 (2015)<sup>[43]</sup>以标普 500 指数的日收益率和日波动率以及谷歌趋势指标作为因子,以长短期记忆神经网络 LSTM 作为模型架构来预测标普 500 指数的波动率,这项研究表明了深度学习在股市时间序列预测具有巨大潜力。Qin 等人(2017)<sup>[44]</sup>采用具有注意力核心机制的双阶段循环神经网络架构进行时间序列数据分析,拥有了注意力核心机制的循环神经网络可以分阶段地提取与目标相关的输入数据,从而缓解长期依赖的问题。Fischer 和 CKrauss (2017)<sup>[45]</sup>以 1992 年至 2015 年标准普尔 500 指数成分股的变动趋势为研究对象,以长短期记忆神经网络架构为模型基础,实验证明 LSTM 神经网络的选股能力优于其他对比模型。Hakan Gunduz 等人 (2017)<sup>[46]</sup>在预测土耳其 BorsaIstanbul100 指数成分股时,使用了一个具有某种固定排序因子集的 CNN 架构,并通过因子筛选方法来减少训练时间以及模型复杂度,使模型的实证结果优于其他对比模型。Wilfredo Tovar (2020)<sup>[47]</sup>将生成对抗网络的思路融入模型,以双向长短期记忆(Bi-LSTM)和卷积神经网络(CNN)为基础模型构建出 Bi-LSTM-CNN 模型,生成与现有真实财务数据一致的合成数据,从而保留股票正面或负面趋势的特征,预测股票未来的趋势,结果表明所提出的解决方案优于以往的模型。Linyu Zheng 等人 (2021)<sup>[48]</sup>提出了一种把主成分分析(PCA)的方法融入循环神经网络 RNN 的混合预测模型,将主成分分析能够有效选取特征的优势与循环神经网络相结合,取得了不错的效果。Tuo Sun 等人 (2021)<sup>[49]</sup>构建了名为 ESAN 的模型来预测股票价格的变动,该模型包括了情感分析模块与时序预测模块,综合考虑了文本类情感数据与时间序列数据对股票价格的影响。

而在国内,目前将深度学习应用于量化投资的研究并不多,在有限的研究当中,绝大部分的学者都将目光投向了循环神经网络结构,以循环神经网络为基础

进行一定改进。如任君等人(2018)<sup>[14]</sup>将正则化方法应用于长短期记忆神经网络的模型构建中,通过优化模型的目标函数,构建出了更具鲁棒性的弹性网正则化 LSTM 模型来预测股票价格的变化情况。冯宇旭等人(2019)<sup>[15]</sup>将 LSTM 模型与 SVR、Adaboost 模型进行岭回归集成,实证检验表明了回归集成算法的有效性。赵红蕊等人(2020)<sup>[48]</sup>将注意力机制融入了长短期记忆网络(LSTM)和卷积神经网络(CNN)中,提出了混合模型 LSTM-CNN-CBAM 来预测股票价格变化情况,该融合模型的特征提取能力得到了显著加强。罗泽南(2021)<sup>[16]</sup>利用 Stacking 方法将随机森林、梯度提升树、XGBoost 和 BP 神经网络的设计思路进行融合,构建出了 RGXB-Stacking 模型来预测沪深 300 的股指变动,回测结果表明该模型优于其他模型。康瑞雪等人(2021)<sup>[17]</sup>将长短期记忆神经网络进行改进,通过融入注意力机制核心和多源数据构建出了 SA-LSTM 模型来处理股票价格的预测问题,该模型通过多源数据的融合以及注意力机制一定程度上解决了 LSTM 模型本身的长期依赖问题,提高了模型精度,实验证明,SA-LSTM 模型的股价预测能力优于其他对比模型。

### 1.3.3 文献综述总结

从国内外业界来看,基金公司中量化部门带来的收益逐年增加,从事量化研究的岗位需求也越来越多,这些都说明了量化投资行业具有广阔的发展前景,因而国内外有关量化投资的研究,也成为一个新的浪潮。在这期间,众多学者在进行金融时间序列建模时主要的视角包括了传统计量方法、经典机器学习方法、集成学习方法以及深度学习方法。但现有深度学习应用于金融时序建模的研究中主要使用基于循环神经网络的循环结构,该结构特有的顺序属性使得模型在训练过程中难以并行,同时对于处理跨度较长的时间序列时,由于内存的限制可能会使得模型丢失早期的信息。针对此问题,本文应用了目前在各领域广为流行的 Transformer 神经网络并结合 Time2Vector 时间特征向量提出了一种新的预测模型,将该模型命名为 Trans-T2V。一方面,Trans-T2V 包含了 Transformer 神经网络的注意力机制,能够一次性关注到所有序列信息并依靠输入输出序列的相似度来衡量信息的重要性而无需考虑序列中的距离,从而解决循环结构难以并行和难以应对长时间跨度数据的缺点;另一方面,Time2Vector 时间嵌入层的加入解决了 Transformer 神经网络本身无法捕获时间序列中位置信息的缺点,让模型更加适



用于处理金融时序数据。

## 1.4 研究内容和方法

本文主要使用时间特征 Time2Vector 对 Transformer 神经网络进行优化, 将其命名为 Trans-T2V, 使模型更适用于处理金融时间序列数据, 并基于 Trans-T2V 模型进行量化选股策略研究, 研究数据为全 A 股 2010 年 1 月至 2020 年 1 月的月频数据。研究的第一步是对数据进行预处理操作, 操作内容包括数据获取、提取特征、设置标签、特征预处理、划分训练集和测试集等, 最终得到 437026 个样本, 每个样本 209 个特征。然后将每个样本 209 个特征增添时间维度, 将数据从(batch,feats)的 2 维数据转化为(batch,steps,feats)的 3 维数据输入模型。本文参考《华泰人工智能系列》的数据划分方法, 在分出训练集和测试集后使用以月为单位窗口滚动的方式进行操作, 取每 12 个月的数据用于训练与验证, 取下 1 个月的数据作为样本外测试集, 对该月所有股票的超额收益率的变化情况进行预测, 因此滚动窗口为 1 个月, 按月依次滚动 120 次。之后将 Trans-T2V 神经网络与 Transformer 神经网络、RNN 神经网络以及 LSTM 神经网络进行分析比较, 最后将基于 Trans-T2V 神经网络构建的量化选股策略与基于 Transformer、RNN 和 LSTM 神经网络的量化选股策略分别进行分层回测, 将回测结果再次进行比较分析, 从而证明基于 Trans-T2V 模型构建的选股策略更具优势, 本文的改进模型具有一定的现实意义。

本文具体的论文写作思路如下: 第一章为绪论部分, 首先介绍本文的研究背景, 包括量化投资模型的发展状况和深度学习的研究进展, 随后介绍了本文的学术意义和现实意义, 表明了本文的研究内容具有必要性, 之后介绍了国内外深度学习、量化投资的研究进展, 引出了本文创作的必要性, 最后描述了本文行文的逻辑框架和论文的研究思路与方法。第二章为相关理论介绍部分。包括了 Transformer 的理论和结构, Time2Vector 的理论和结构, 以及根据理论构造出的 Trans-T2V 神经网络模型。第三章为基于 Trans-T2V 的股票预测模型设计。描述了如何构建模型, 包括股票数据的选取与处理、超参数的选取、数据集的滚动等。然后分别选用不同参数构建了 Trans-T2V 模型、Transformer 模型、RNN 模型和 LSTM 模型, 为后续的实证分析做准备。第四章为实证分析。首先介绍了本文实证环境的搭建和实证数据的选取, 之后介绍了模型及策略评价指标, 用来评判各

模型下量化选股策略的好坏,然后对策略进行分层回测,取出各模型的回测结果进行多指标对比,最后得出基于 Trans-T2V 模型的量化选股策略综合表现更优秀的结论。第五章是总结与不足部分,先是对论文的内容进行了总结并对主要结论进行阐述,之后分析本论文的不足之处并针对性地提出了未来的改进方向。

## 1.5 本文主要贡献和创新

本文的创新点在于以下两点。第一,根据时间特征 Time2Vector 对 Transformer 神经网络进行改进,提出了一种更适用于处理时间序列数据的 Trans-T2V 神经网络模型,为研究者今后在自己的模型中融入时间嵌入层,提升模型鲁棒性提供了一种新的思路。第二,在当前的探索研究当中,研究者使用深度神经网络处理股价预测及量化选股方面的研究主要集中在循环神经网络结构,国内对于 Transformer 神经网络应用于量化选股方面的研究尚处空白,因而本文所进行的研究是对这一领域的一种探索。

## 2 相关理论介绍

### 2.1 Transformer 介绍

过去几年,循环(也称递归)神经网络(Recurrent Neural Network)、长短期记忆神经网络(Long Short-Term Memory)和门控循环神经网络(Gated Recurrent Unit)已在序列模型、语言模型、文本翻译等场景的应用中均得到了不错的成绩,在循环架构的基础上发展的语言模型架构和编码器-解码器(Encoder-Decoder)模型架构也在持续深入。但是,循环架构因其结构特点而产生的顺序属性使得样本间训练时难以并行,降低了计算效率,而对于长序列而言,内存的有限性大大制约了模型的样本处理效率,使循环结构处理需要长时间记忆的序列时,模型效果大打折扣。2017年,谷歌团队 Ashish Vaswani 等人(2017)<sup>[50]</sup>在机器翻译领域发表了一篇名为《Attention Is All You Need》的文章,该文章提出了一个结构与参数并不复杂的网络结构 Transformer,该结构将 encoder-decoder 的所有循环层换成了多头注意力层,在 BLEU 上面表现比当时最好的模型高出 2 个 BLEU Score,证明了 Transformer 神经网络的优秀之处。之后由于基于 Transformer 的 BERT、GPT 等模型在自然语言处理和 CV 图像识别领域应用后有着相当优秀的表现,使得 Transformer 名声大噪。而 Transformer 的核心,注意力机制(Attention)能够一次观察所有序列,在运行时并行化处理所有数据,同时以输入输出序列的相似度作为依据进行建模,无需考虑它们在序列中的距离,从而解决了上述循环结构中存在的问题。

在 Transformer 神经网络出世之前,学术界常常采用循环神经网络来处理时间序列数据,从而做出文本预测等工作。循环神经网络通过储存过去部分信息的隐藏状态  $h$  及代表当前信息的  $x$  相结合,在进行工作时遵循顺序属性,一步一步地进行预测,如图 2-1-1 所示。

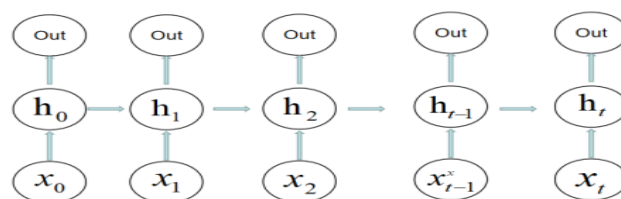


图 2-1-1: 循环神经网络的工作原理

在这种情况下，当相关信息和需预测任务距离较近时，循环神经网络可以较好地学习前面的信息并做出预测。但在更多的时候，预测目标任务需要使用更多的相关信息，这些信息却储存在距任务目标较远的位置，在此情况下，过去的信息要经过漫长的传递才能与预测目标相遇，而这传递链越长，过去信息丢失的可能性就越大，循环神经网络就越低效，这就是所谓的长期依赖问题。

针对循环神经网络的缺点，有学者提出了长短期记忆网络结构来处理长跨度信息缺失的问题。以日常生活为例，人们通常会对不同的事情确定不同的优先级，优先处理对他们来说更为重要的事情，选择性地取消一些不重要的事情，长短期记忆网络就利用了这样的思路，在长短期记忆网络中，过去的信息经过一种名为细胞状态的机制来选择性地记忆或遗忘一些重要或不重要的信息，通过细胞状态，那些重要的信息会被选择并一轮轮传递下去，如图 2-1-2 所示。

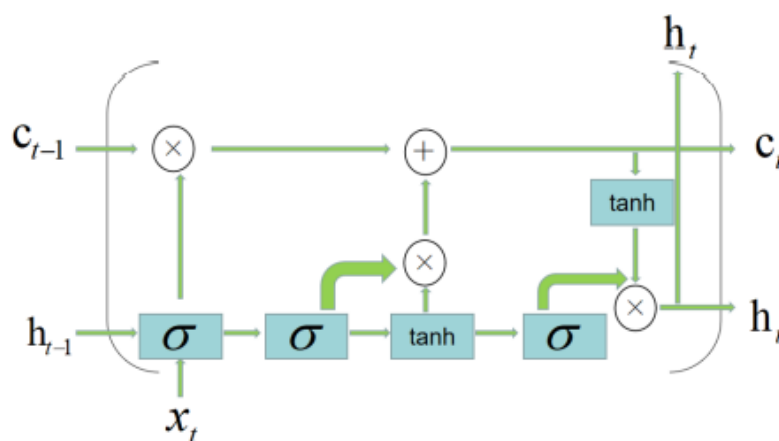


图 2-1-2: 长短期记忆神经网络的工作原理

但人们发现，长短期记忆网络在处理较长时序的数据时效果依然不是很好，主要的原因在于一方面，类似于循环神经网络，长短期记忆神经网络的模型架构也是顺序计算，难以并行；另一方面，长短期记忆神经网络虽然采用了细胞状态来处理信息，但依然没有彻底解决长期依赖问题。为了走出这种困境，Transformer 神经网络应运而生，该神经网络通过编码器-解码器的架构以及注意力机制实现信息并行，解决长期依赖问题。

Transformer 的结构可以分为编码器与解码器部分，编码器是由嵌入层(Input Embedding)、位置编码(Positional Encoding)、多头注意力层(Multi Head Attention)、前馈神经网络(Feed Forward)、残差连接和归一化层(Add&Norm)构成；解码器由嵌入层(Output Embedding)、位置编码(Positional Encoding)、加入掩码器的多头注意力层(Masked Multi-Head-Attention)、多头注意力层(Multi-Head-Attention)、前馈神经网络(Feed Forward)、残差连接和归一化层(Add&Norm)构成。如图 2-1-3 所示。

**编码器-解码器：**编码器的作用就是对输入词源进行编译，将句词输入编码器后经过各层作用进行非线性变换，从而转化成中间语义表示向量，使得计算机读懂原本无法理解的词源；解码器作为计算机语言的解释器，其作用在于根据编码器翻译好的中间语义表示和之前生成的历史信息共同生成  $t$  时刻的词源，实现计算机语言与任务需要语言之间的转换，实现预测目标。如图 2-1-4 所示。

**编码器(Encoder)：**Transformer 的编码器由嵌入层、位置编码以及 6 个 Layer 模块构成，其中 Layer 模块包含多头注意力层、残差连接及归一化层和前馈神经网络层。

**嵌入层(Input Embedding)：**该层主要是将各种类型的数据转换为模型可以理解的向量表示。对于文本序列来说，嵌入层会将文本数据转化成一个个向量，从而在高维空间捕获词汇之间的关系，每个向量代表着一个词，使得接下来的工作能够展开。

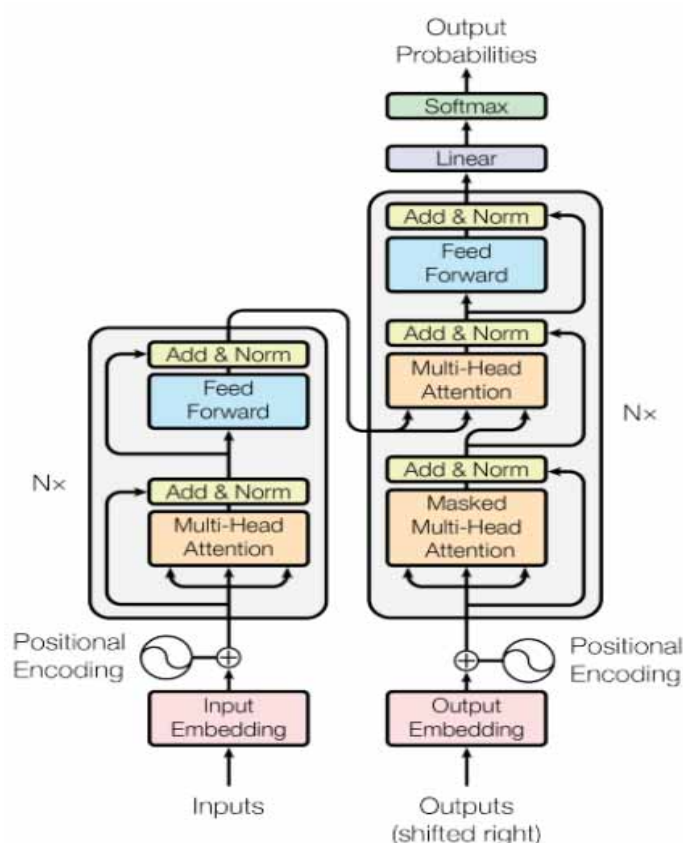


图 2-1-3: Transformer 模型结构

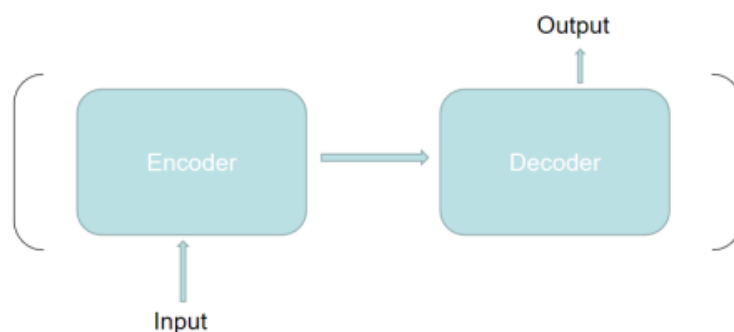


图 2-1-4: 编码器-解码器结构

**位置编码 (Positional Encoding):** 由于 Transformer 不采用循环神经网络的顺序结构, 而是使用全局信息, 因此难以捕获序列的顺序信息。但序列的顺序信息对语言处理、机器翻译等领域又十分重要, 所以作者加入位置编码层来捕获序列的相对或绝对位置信息。

位置编码层通过公式计算序列的相对位置信息, 并将结果直接与 Embedding 数据结合, 使得处理后的序列拥有了位置信息。定义如下:

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (2-1-1)$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (2-1-2)$$

其中,  $pos$  代表了单词在某句中的位置信息,  $d$  表示位置编码 (PE) 的维度 (即词向量的长度),  $2i$  表示偶数的维度,  $2i+1$  表示奇数维度。使用 这种公式计算位置编码的好处如下:

(1) 使 PE 能够灵活地处理样本外未曾训练过的句词长度, 提升模型鲁棒性。例如, 当训练集中句词长度最大值为 20 而样本外句词长度为 21 时, 使用公式计算的方法可以计算出第 21 位的 Embedding。

(2) 模型可以根据正、余弦函数高效地计算出句词的相对位置, 对于固定长度的间距  $k$ ,  $PE(pos+k)$  可由  $PE(pos)$  计算得来。

注意力层 (Self-Attention): 类似于人的思维, Attention 机制可以从庞杂的信息中找出与任务最相关的信息, 从而帮助编码器在对目标任务进行编码时查看所有输入信息中的其他信息。同样, 注意力层的核心思想是将输入值的关注点集中在与其最相似的全局信息上, 为每一个输入值学习一个权重, 该权重来自于单个输入值与其他输入值的相似度, 使得注意力层的输出值包含全局信息。如图 2-1-5 所示。

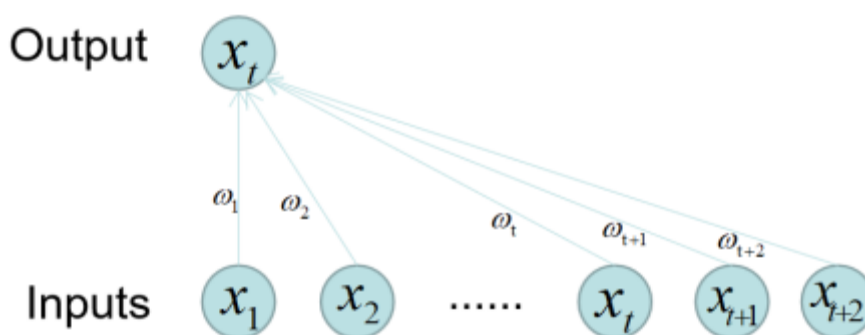


图 2-1-5: 注意力机制并行工作示意图

Attention 层的输入通常由 (Query, Key, Value) 来表示, 其中 Query、Key、Value 是由嵌入层的输出一分为三, 分别通过线性变换计算得到。权重由注意力分数来表示, 注意力分数的计算采用了 Scaled Dot-Product Attention 的方式: 令

Query 与 Key 内积，内积产生的值通过 softmax 函数变为权重，且每个权重处于 (0, 1) 之间，权重之和为 1。该内积值趋向于 0 则代表两者正交，相似度最差，越大则代表相似度越高。内积值除以  $\sqrt{d_k}$  的主要目的是使数值的分布均匀化，减少内积值向 0 或 1 靠拢的现象，使得 softmax 后的权重分布更加均匀，加速模型训练。

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2-1-3)$$

**多头注意力层 (Multi-Head-Attention):** 这一层的目的在于设置 h 个不同的线性变换对 Query、Key、Value 进行投影，进入到不同的注意力层中，再将多个注意力层的输出拼接起来得到多头注意力层的输出。多头注意力层由于线性层的可学习的神经元个数存在，使得每个“头”学习到的信息模式变得不同。

**残差连接和归一化层 (Add & Norm):** 残差连接 (Add) 指  $X + \text{Multi Head Attention}(X)$  常用于解决复杂网络训练容易出现的梯度消散的问题，如图 2-1-6 所示：

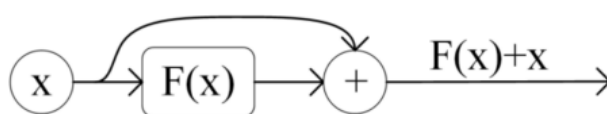


图 2-1-6: 残差连接

归一化 (Norm) 指 Layer Normalization，该结构的作用是根据隐藏层节点的数量实现各神经元输入的均值与方差的统一，以便加快收敛速度。

**前馈神经网络前馈神经网络以及残差连接和归一化层 (Fully Connected Feed-Forward Network):** 该层是由一个两层的全连接层构成，第一层的激活函数为 ReLu，第二层不使用激活函数。前馈层 (FFN) 将多头注意力层提炼好的向量再次投影到一个更大的空间，通过激活函数 (Relu) 更方便地提取需要的信息，最后再投影回向量原来的空间。前馈层提供了非线性变换的属性，FFN 的加入引入了非线性 (ReLu 激活函数)，变换了注意力层输出的空间，从而增加了模型的表现能力。前馈神经网络的工作原理及定义如图 2-1-7 和公式(2-1-4)所示。

$$\max(0, XW_1 + b_1) W_2 + b_2 \quad (2-1-4)$$



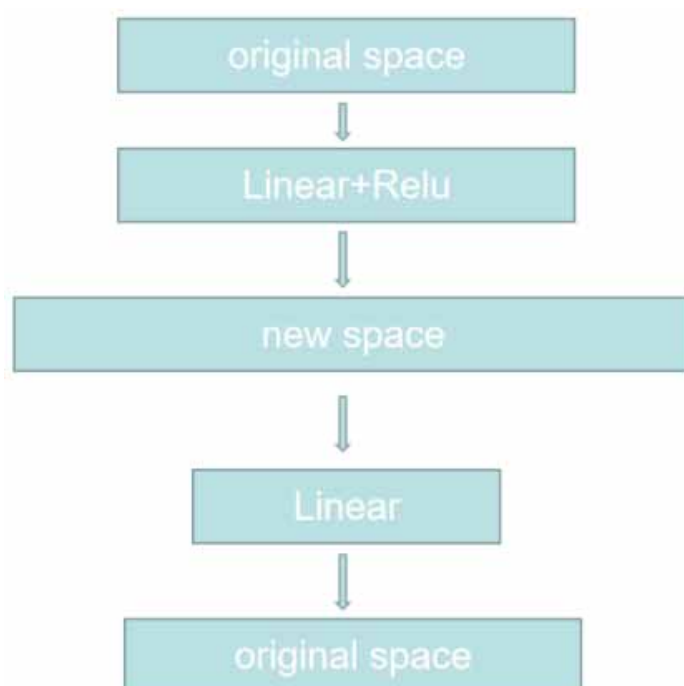


图 2-1-7: 前馈神经网络工作原理

**解码器 (Decoder):** 类似地, 解码器由嵌入层、位置编码层以及 6 个 Layer 模块构成。不同于编码器的 Layer 模块其注意力层的 Query、Key、Value 来自于相同的矩阵, 解码器的注意力层中 Key 与 Value 来自于编码器, Query 来自于解码器上一个位置的输出, 因此多了一个掩码的多头注意力层。Decoder 的输入输出和解码过程如下:

**输入:** 编码器的输出作为 Key、Value 加上对应  $i-1$  位置解码器的输出作为 Query。这体现出了编码器与解码器的核心差异, 编码器应用了自注意力机制, 而解码器仅为注意力机制, 它的 Key 和 Value 来自编码器, Query 来自上一位置解码器的输出。

**输出:** 对应  $i$  位置的输出词的概率。

**解码:** 解码器的每一个预测都需要上一位置的输入当作 Attention 的 Query, 因此不同于编码器的并行计算, 一次性将结果全部输出, 解码器是将预测结果一个一个解出来。工作原理如图 2-1-8 所示。

**掩码的多头注意力层 (Masked Multi-Head-Attention):** 这一层结构用于确保解码器在预测第  $i$  个位置时不会接触到未来信息。掩码的加入会让 Attention 在

计算注意力分数时将未来信息的权重设置为 0，从而防止模型预测时模型提前查看 到全局信息，减少模型过拟合情况的发生。

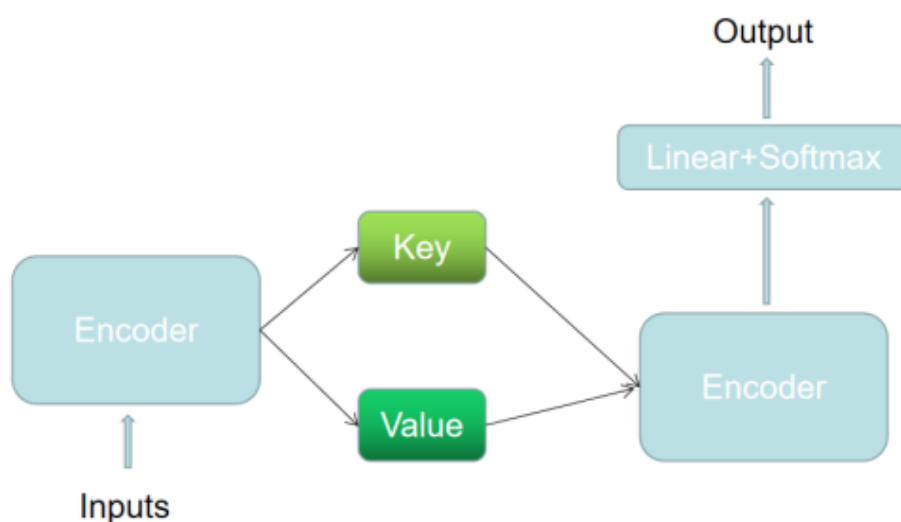


图 2-1-8：前馈神经网络工作原理

## 2.2 Time2Vector 介绍

Seyed Mehran Kazemi 等人（2019）<sup>[51]</sup>认为时间序列数据包含着时间特征，而时间特征应该具备周期性和非周期性、不受时间放缩的影响以及足够简单的重要属性。其中，周期性是指在许多情况下，一些事件会周期性的发生。例如，饺子的销量会在每年冬至这一天变得更高，商店的销售额往往会在周末或节假日变得更高，天气和温度在不同的季节遵循着周期性变化；非周期性是指一些事件的发生频率具备随机性，或随着时间的推移而变得更有可能发生。例如，某些疾病的传播具有非周期性，老龄人更容易得某些疾病等。这种周期性和非周期性的属性能够将时间的特点表达出来，因此应该使用一种能够捕获周期性和非周期性属性的特征来表示时间；时间放缩的不变性是指时间可以用不同的 度量单位来衡量（月、天、小时、秒等），即时间不会因为度量单位的改变而改变 自身的属性，因此时间特征另一重要属性为时间放缩的不变性；简单性是指时间 特征应该易于应用。例如，当时间特征表示为矩阵形状时可能难以与其他形状融 合和计算而难以应用，所以考虑将时间特征表示为向量形状。

根据以上时间特征三个重要的属性，文章的作者提出了向量形状的

Time2Vector 来表示时间特征，时间特征定义如下：

$$t2v(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i, & i = 0 \\ F(\omega_i \tau + \varphi_i), & 1 \leq i \leq k \end{cases} \quad (2-2-1)$$

对于一个给定的标量时间概念  $\tau$  来说, Time2Vector 的  $\tau$  定义为  $t2v(\tau)$ , 它是一个大小为  $k+1$  的向量。其中  $t2v(\tau)[i]$  为  $t2v(\tau)$  的第  $i$  个元素,  $F$  为激活函数,  $\omega$  和  $\varphi$  为可学习的参数。考虑到向量表示普遍应用于不同任务中, 时间特征使用向量表示使得它很容易与不同架构相结合。定义中的线性方程表示为时间的进程, 可用于捕获非周期性属性, 定义中在线性方程上加入  $F$  函数则用来捕获周期性属性。对于  $F$  函数的选取, 文章作者做了诸多实验, 最终选取正弦函数来捕获时间的周期性属性。当  $F=\sin$  时, 对于  $1 \leq i \leq k$ ,  $\omega_i$  和  $\varphi_i$  为正弦函数的频率和位移,  $\sin(\omega_i \tau + \varphi_i)$  的周期为  $2\pi/\omega_i$ , 即  $\tau$  和  $\tau + 2\pi/\omega_i$  具有相同的值, 因此, 正弦函数有助于捕获周期性。此外, 实验表明不同于其他激活函数可能会表现出奇怪的样本外行为, 正弦函数有望很好地外推到样本外的数据。

## 2.3 Trans-T2V 模型结构介绍

由于本人的研究环境无法复现原本十分庞大的 Transformer 结构, 并且原本的 Transformer 模型并不适用于处理金融时序数据, 因此本文将 Time2Vector 作为时间嵌入层融入 Transformer 模型并做出适当简化, 提出了 Trans-T2V 的模型架构, 用以提升 Transformer 在处理金融时间序列上的表现。

本文初始的数据形状为 (batch, steps, feats), 其中 batch=32 表示每次随机挑选 32 个批量大小的样本数; steps=12 表示时间步为 12, 以每 12 个月的特征值预测下 1 个月的目标值; feats=209 表示特征值个数为 209 个。

Trans-T2V 模型由时间嵌入层、模块层、全局池化层、线性输出层构成, 其中模块层包含多头注意力层、残差连接及归一化层和前馈神经网络层。如图 2-3-1 所示:

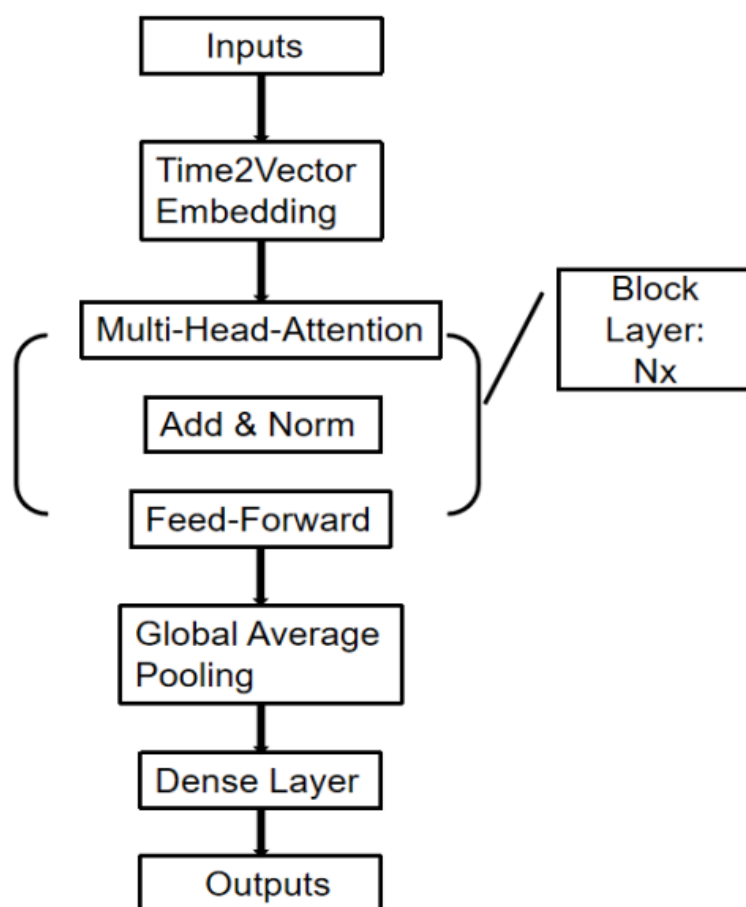


图 2-3-1: Trans-T2V 模型结构示意图

时间嵌入层 (Time2Vector Embedding): 时间嵌入层通过时间的周期性和非周期性来捕获数据的时间位置信息, 定义如下:

$$t2v(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i, & i = 0 \\ F(\omega_i \tau + \varphi_i), & 1 \leq i \leq k \end{cases} \quad (2-3-1)$$

该层的作用类似于 Transformer 的位置编码层, 用来获取金融时序数据的时间位置信息, 金融时序数据进入时间嵌入层后得到周期性和非周期性两个新的维度特征, 使数据形状从 (32,12,209) 变为 (32,12,211), 如图 2-3-2 所示:

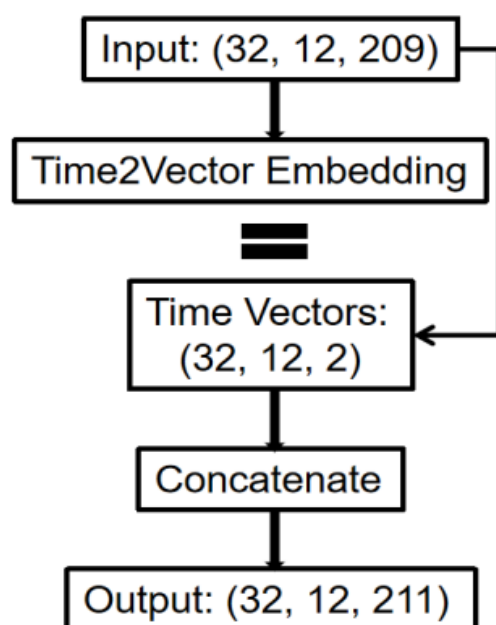


图 2-3-2: 时间嵌入层示意图

**模块层 (Layer):** 该层包含多头注意力层、残差连接及归一化层和前馈神经网络层。前馈神经网络由两个 1D 卷积层组成, 在这里卷积层起到了同全连接层一致的作用。第一层卷积层由 ReLu 函数激活, 每个子层后面都有一个子层, 每两个子层之间通过将初始 Query 输入添加到上一个子层的输出中来形成残差连接, 然后接入归一化层, 以加速训练过程。

**注意力层 (Self-Attention):** 该层的作用就是将整个输入序列中要注意的信息抓取出来并做汇聚, 得到的每一个输出向量都包含了整个序列的信息。

注意力层有三个输入 (Query, Key, Value), 三个输入均为上一步得到的输入矩阵, 然后分别输入三个分隔的全连接层 (Dense Layer), 每个 Dense Layer 包含 512 个神经元, 对因子时间向量进行 Embedding, 得到的输出矩阵分别用 Q, K, V 表示, 大小均为 (32, 12, 512)。

在数据经过初始的线性变换后, 需要计算注意力分数以求注意力层的输出。注意力分数决定了在预测未来股价时单一输入值将注意力集中在各个时间序列上的程度。本文的注意力分数计算手段与 Transformer 神经网络一致, 采用 Scaled Dot-Product Attention 的方式。首先通过 Q 与 K 内积得到内积值。之后, 将内积结果除以全连接神经元大小的开方, 再通过 softmax 函数生成一组范围在 (0, 1) 之间, 且和为 1 的权重。

最后, 把注意力分数矩阵与 Value 矩阵相乘, 得到自注意力层的输出。过

程 如图 2-3-3 所示:

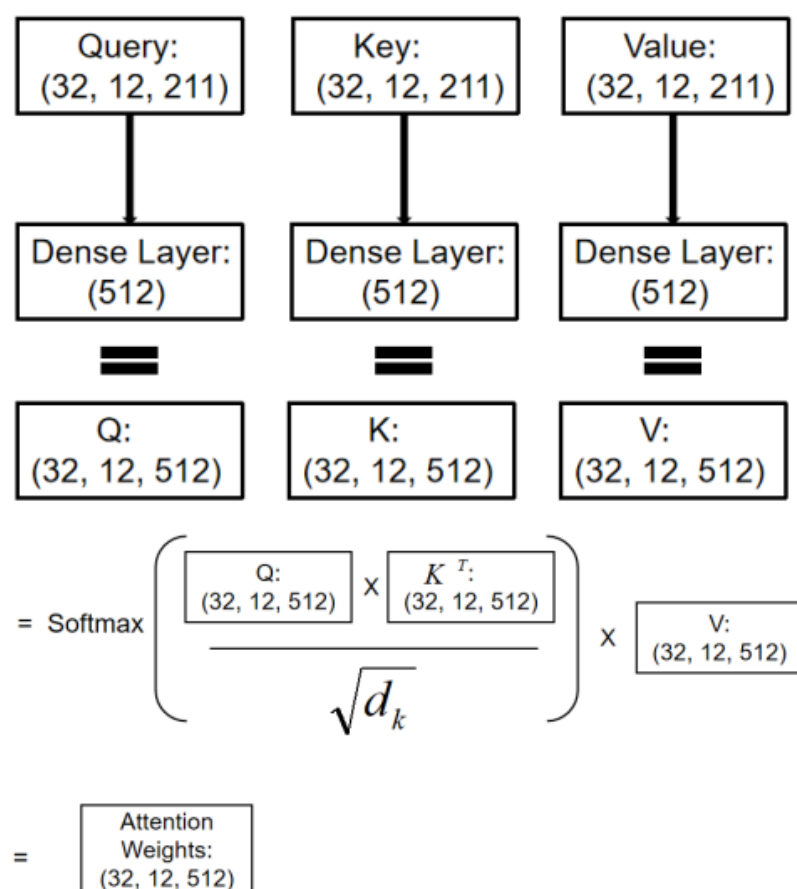


图 2-3-3: 注意力分数计算示意图

多头注意力层 (Multi-Head-Attention): 多头注意力层的功能是拼接  $n$  个单头注意力层的输出, 然后进入到全连接层进行非线性变换。多头注意力层将多个独立的单头注意力层输出作为输入, 各个注意力层有着不同的可学习的神经元个数, 因此能够学习到不同的时间序列特点。注意力头数量的增加会加强模型捕获远程依赖关系的能力。如图 2-3-4 所示:

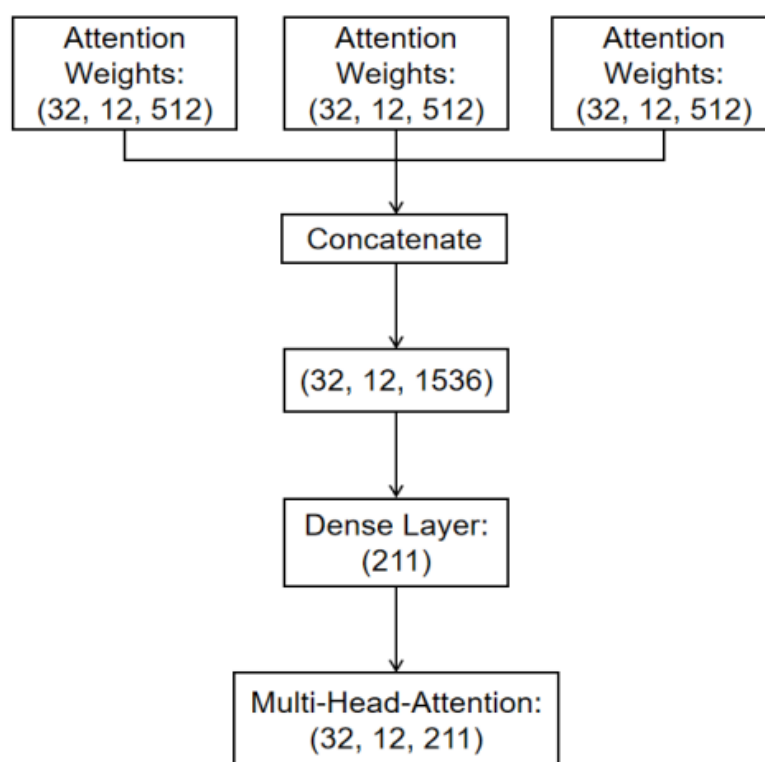


图 2-3-4: 多头注意力层示意图

前馈神经网络（Fully Connected Feed-Forward Network）：这一层结构由两个 1D 卷积层组成，在这里卷积层起到了同全连接层一致的作用。第一层的激活函数为 ReLu，第二层不使用激活函数。前馈子层提供了非线性变换的属性，该层的加入引入了非线性 (ReLu 激活函数)，变换了 attention output 的空间，从而增加了模型的表现能力。

残差连接和归一化层（Add & Norm）：残差连接 (Add) 指  $X + \text{Multi Head Attention}(X)$  常用于解决复杂网络训练容易出现的梯度消散的问题；归一化 (Norm) 指 Layer Normalization，该结构的作用是根据隐藏层节点的数量实现各神经元输入的均值与方差的统一，以便加快收敛速度。

全局平均池化层：该层允许输入值发生一定程度的偏移，使得输出后的结果不会随着输入值的微小变化而发生明显改变，同时也降低了参数个数，起到了全连接层的作用，从而有效地避免过拟合问题的出现，提升模型的泛化能力。

## 2.4 本章小结

不同于其他神经网络，Transformer 神经网络的编码器与解码器的核心是由注意力机制组成，正是由于注意力机制的存在，Transformer 神经网络结构才能并行化处理全局信息，从而解决循环神经网络难以解决的问题。本章首先介绍了 Transformer 神经网络的工作原理及模型架构，Transformer 神经网络是由编码器-解码器组成，编码器结构包括了嵌入层、位置编码和模块层，模块层中包含多头注意力层、前馈神经网络层和残差连接及归一化层；解码器包括嵌入层、位置编码和模块层，解码器模块层包含了带掩码的多头注意力层、多头注意力层、前馈神经网络和残差连接及归一化层。之后介绍了 Time2Vector 中对于时间类型数据应该带有周期性和非周期性特征并据此进行设计的时间向量思路。最后阐述了在 Transformer 神经网络架构基础上，本文将时间向量作为时序位置编码替代 Transformer 中原本的位置编码，提出适用于处理金融时序数据的 Trans-T2V 模型结构并阐述了该模型结构的各层工作详情。



### 3 基于 Trans-T2V 的股票预测模型

#### 3.1 模型设计流程

##### 3.1.1 数据选取

本论文数据主要从 tushare 和 joinquant 以及 wind 上进行获取,所用到的数据为全 A 股的股票日频数据,并将数据频数处理为月频数据。时间范围为 2010 年 1 月至 2020 年 1 月共 120 个月,所涉及的股票因子类型包含估值、成长、财务质量、杠杆、动量反转、市值、波动率、股价、换手率、情绪、技术共 11 个因子类型。

##### 3.1.2 特征和标签提取

每个自然月的最后一个交易日计算上述提到的 11 类因子,将日频数据转为月频,并将 esg 评级和行业进行 one-hot 化,最终得到 209 个特征值,以个股下一个月的超额收益率(沪深 300 指数作为基准)作为样本的标签。

##### 3.1.3 特征预处理

在机器学习的任务中,数据质量的好坏对模型精度的提升具有重大影响,我们从数据平台中获取到的数据往往是未经加工的初始数据,常常存在着诸如数据类型不统一、量纲不统一、异常值、缺失值等问题,因此需要通过数据筛选、文本数据数字化并 one-hot 处理、去极值、缺失值处理、标准化等手段来提升数据质量。具体步骤如下:

(1) 数据筛选,为保证股票选取质量,本文剔除 ST 股、停牌股、当日涨跌停牌以及上市时间不满 3 个月的股票,由于原始数据为日频数据,因此为了保险起见,本文使用 100 天的时间跨度作为上市时间不满 3 个月的表示。

(2) 文本数据数字化与 one-hot 处理,由于行业类别、ESG 评级的数据类型与以 float 数据类型为主的特征不符,因此需要先将行业类别、ESG 评级转化为月频的 float 类型数据,之后进行 one-hot 处理,以便作为特征使用。

(3) 中位数去极值,本文使用绝对值差中位数法(Median Absolute Deviation, MAD)对所有特征设定上下限,避免极端值的干扰。假设某因子对

第  $T$  期所有个股上的暴露度序列为  $D_i$ ,  $D_m$  为该序列中位数,  $D_{m1}$  为序列  $|D_i - D_m|$  的中位数, 则将序列  $D_i$  中所有大于  $D_m + 5D_{m1}$  的数重设为  $D_m + 5D_{m1}$ , 将序列中所有小于  $D_m - 5D_{m1}$  的数重设为  $D_m - 5D_{m1}$ ;

在数理分析中, 中位数去极值法是对单变量数值型数据的样本偏差的一种鲁棒性测量, 是用来描述单变量样本在定量数据中可变的一种标准, 从本质而言, MAD 表示偏差绝对值的中位数:

$$MAD = \text{median}(|x_i - \text{median}(x)|) \quad (3-1-1)$$

对于标准差来说, 平方的计算特点使该方法在出现较大偏差时产生的权重较大, 难以忽视异常值对结果造成的影响。而 MAD 比标准差更能适应数据集中的离群值, 对 MAD 方法而言, 少量的离群值不会影响实验的结果, 因而更具鲁棒性。

(4) 缺失值处理, 对于某股票因子的缺失值, 使用该股票因子的行业内平均值进行填充, 若存在行业内某因子平均值缺失, 则使用该因子全行业平均值进行填充。

(5) 行业市值中性化处理, 将各个因子对行业虚拟变量及取对数后的市值做线性回归, 并将残差作为中性化后的新因子, 消除各个因子之间相关性对目标值的影响, 减少选股时对于行业、市值的偏好, 使得选股更加分散。

(6) 标准化处理, 使用 min-max 标准化将不同量纲的数据转化为同一数量级上, 消除数量级对选股的影响, 为下面的 Box-Cox 转换做准备。离差标准化 (Min-Max 标准化) 是将原始数据做线性变换处理, 使数据处于 0 到 1 之间的一种量纲统一方法, 其转换公式如下, 其中  $x_i$  为初始数据,  $y_i$  为转换后的无量纲数据:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3-1-2)$$

(7) Box-Cox 转换, 将归一化后的因子进行 Box-Cox 转换, 使因子的分布趋于正态化, 从而更加符合模型对数据分布的假设。

(8) 标准化处理, 使用 Z-Score 标准化将 Box-Cox 处理后的因子序列减去其当前均值并除以其标准差, 得到一个新的近似服从标准正态分布的序列, 保证各因子量纲上的统一。Z-score 标准化为最常用的标准化方法, 公式如下:

$$y_i = \frac{x_i - \mu}{\sigma} \quad (3-1-3)$$

(9) 数据维度转化, 由于原本的数据是只有样本维度和特征维度的 2 维数据 (batch, feats), 而神经网络模型的输入数据维数通常为 3 维, 因此本文新增时间步维度, 将 2 维数据变为 3 维数据 (batch, steps, feats), 以便模型识别。

### 3.1.4 训练集和交叉验证集的生成

本文将日频数据处理为月频后, 样本总量为 690826 条, 特征数量为 91 个, 经过特征预处理后的样本数总量为 437026 条, 特征数量为 209 个。由于深度神经网络需要优质的样本量, 若简单的将整个数据集切分为训练集和测试集, 不但无法充分利用数据, 而且因样本属于金融时序数据, 仅训练一个模型难以应对不同时间段的数据分布特点, 随着数据时间的发展, 模型的泛化性会越变越差。因此为了提高数据的使用效率和模型的鲁棒能力, 在划分训练集与测试集时使用窗口滚动的方式进行操作, 将当前时间点往前推 12 个月的样本合并作为训练集, 交叉验证集取全部的训练集。

### 3.1.5 交叉验证调参

模型训练过程中, 观察交叉验证集 loss 的变化, 在连续进行 5 轮 epoch 交叉验证集的损失都无明显变化时停止模型训练, 选择损失最低的一组参数作为模型的最优参数。由于本文开发环境的算力条件有限, 在算力允许的情况下设置最大迭代次数, 选择较高的训练次数可能会有更优秀的表现。

### 3.1.6 样本外测试

确定最优参数后, 将预处理后预测月份的因子值输入模型, 输出每个样本的预测目标  $f(x)$ , 然后把  $f(x)$  当作合成后的单个因子值进行单因子分层回测。

### 3.1.7 模型评价

本文以测试集的 loss、mse 值以及年化收益、年化波动率、最大回撤、夏普比率、信息比率、模型相对基准月胜率等分层回测结果作为模型评价指标。模型构建示意图如图 3-1-1 所示：

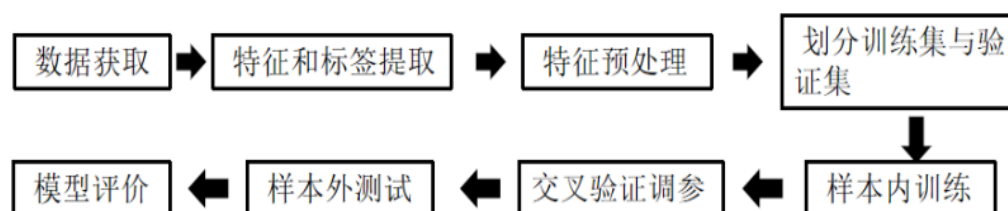


图 3-1-1：模型构建流程

## 3.2 Trans-T2V 模型及对比模型设置

### 3.2.1 Trans-T2V 模型参数及结构设置

近年来由于计算机算力的更新换代以及数据量的愈发丰富，多层感知机（MLP）、卷积神经网络（CNN）、循环神经网络（RNN）等深度神经网络及其变种飞速发展。由于基于循环结构的模型能够储存过去时间的信息，对于处理时间序列数据较为合适，因此过往对于金融时间序列数据的研究中，循环网络架构占据着主流，但循环架构却有着难以并行化和可能丢失早期信息的缺点，而 Translformer 模型将 encoder-decoder 的所有循环层换成了多头注意力层，使得该模型能够一次观察所有序列，在运行时并行化处理所有数据，同时以输入输出序列的相似度作为依据进行建模，无需考虑它们在序列中的距离，从而解决了上述循环结构中存在的问题。但 Transformer 模型本身的词嵌入层和位置编码并不适用于处理时间序列数据，因此本文通过构建时间嵌入层 Time2Vector Embedding 让模型捕获时间序列的位置信息，从而使新的模型在处理金融时间序列数据时更具优势。

由于本文的研究任务是回归问题，因此模型选取均方误差（Mean-Squared Loss, MSE）作为损失函数，该损失函数描述了样本真实值  $y$  与预测值  $\hat{y}$  的差异情况，在训练神经网络时，通过数次 epoch 不断地改变神经网络的有关参数，

从而降低损失值，得到预测精度更高的神经网络模型。选择 Adam 优化算法作为优化器训练模型，Adam 集合了动量梯度下降法（Momentum）、自适应梯度算法（Adagrad）和均方根传递（RMSProp）等算法的优点，具有收敛速度快、模型对学习率的变化不敏感、调参容易等特点。选取均方误差作为模型的评估指标，MSE 是反应估计量与被估计量之间差异的一个指标，MSE 的值越小，表示模型输出的数据与真实数据的偏差越小，模型精度越高。接下来是模型的一些参数设置，其一为参数 batch 的设置，batch 是指每个批量中样本数的多少，深度学习在训练时会选择一些优化算法对模型进行优化，优化算法在每次迭代中会选取 batch 个样本数量的数据计算模型损失并进行参数更新。参考 Dominic Masters 等人(2018)<sup>[52]</sup>对于 batch 的研究以及计算机算力的限制，本文将 batch 设置为 32，这样设置一方面能够在一个 epoch 中多次更新参数加速收敛，提高运行效率，另一方面是在每次参数更新时仅仅使用了 batch 大小的数据而没有遍历全局数据，从而人为地给训练带来了噪声，减少了模型陷入局部最优的可能；其二是可调超参数 epochs 的次数选择，每经过一轮 epoch，研究所选的全部样本数据便会训练一次模型，epoch 过小将难以充分训练模型使其达到收敛状态，epoch 过大又将使得模型陷入过拟合，反而使得模型泛化性能变差，本文使用交叉验证集进行调参，本文考虑到研究的数据集庞大，硬件配置有限和耗时昂贵方面的原因，将 epoch 设置为 40；其三是学习率 lr 的设置，如果说导数反映了梯度下降的方向，学习率则描绘了梯度向最优点前进的步伐，学习率太大容易出现超调现象，使得梯度跨过最优解或是剧烈震荡，学习率太小会导致梯度无法迅速找到正确的下降路径，收敛速度缓慢，经过参数调试后，Trans-T2V 模型的学习率 lr 设置为 0.01；其四是 Dropout 率的设置，在使用深度神经网络进行研究时，通常需要极为庞大的数据量对模型进行训练，从而得到一个泛化能力较好的模型，与此同时，epoch 次数往往也处于较大的水平。结果就是模型训练虽然花费了十分昂贵的时间成本，但模型可学习的参数量也成几何形式增长，使得时常会陷入过拟合，不但降低了模型的运算效率还难以获得较高的泛化能力，因此在机器学习中，Dropout 层是至关重要的。通常的做法是将丢弃发作用在隐藏全连接层的输出上，让一定比例的权重随机置 0 从而丢弃隐藏层的输出个数，控制模型的复杂程度。Trans-T2V 模型的 Dropout 率设置为 0.2。

Trans-T2V 模型的第一层为输入层(Inputs)，inputs 层作用主要在于输入参数及变量。初始输入数据形状为(batch, steps, feats)，其中 steps=12 表示从

当前时刻  $T$  开始, 使用过去 12 个月的因子数据。feats=209 表示每个样本有 209 个特征用来预测。由于本文采取按月滚动训练的方式, 因此每个滚动区间的训练集所含数据大小不同; 第二层为时间嵌入层 (Time2Vector Embedding), 初始数据由该层捕获周期性和非周期性两个时间特征向量, 输入由 (batch,12,209) 变为 (batch,12,211); 第三层为模块层 (Block Layer), 该层包含多头注意力层、残差连接及归一化层和前馈神经网络层。在模块层中, 注意力层 3 个输入 Dense Layer 的神经元大小均为 512, 注意力头数为 3, 前馈层神经元的大小为 256、激活函数为 relu, 归一化层的 epsilon 为  $1e-6$ ; 第四层为全局平均池化层 (Global Average Pooling), 这一层允许输入值发生一定程度的偏移, 使得输出后的结果不会随着输入值的微小变化而发生明显改变, 同时也降低了参数个数, 起到了全连接层的作用, 能够适当减少模型陷入过拟合的可能性, 提升模型的样本外预测能力; 第五层为 Dropout 层, 模型随机丢弃 20% 的神经元比例, 防止过拟合; 第六层为全连接层 (Dense Layer), 选择 tanh 作为激活函数, 权重和偏置向量使用 glorot\_uniform 初始化方法; 最后一层为输出层 (Outputs), 模型预测的目标是下个月的超额收益率, 因此是回归问题, 选择输出维度为 1。模型具体结构如图 3-2-1 所示。

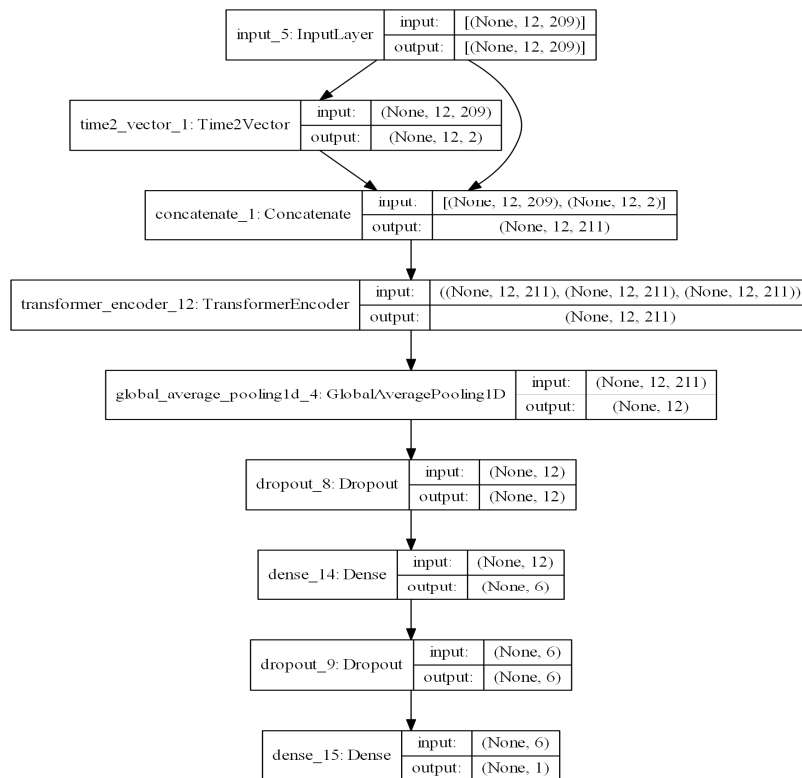


图 3-2-1: Trans-T2V 模型结构

### 3.2.2 Transformer 模型参数及结构设置

Transformer 模型相较于 Trans-T2V 模型仅仅是少了时间嵌入层 (Time2Vector Embedding)，损失函数、优化器、学习率以及训练次数等均与 Trans-T2V 模型保持一致，以便进行模型对比，体现时间嵌入层的优势，具体设置如下：

第一层为输入层 (Inputs)，inputs 层作用主要在于输入参数及变量；第二层为模块层 (Block Layer)，该层包含多头注意力层、残差连接及归一化层和前馈神经网络层，参数设置与 Trans-T2V 一致；第三层为全局平均池化层 (Global Average Pooling)，降低参数个数，避免过拟合；第四层为 Dropout 层，模型随机丢弃 20% 的神经元比例，防止过拟合；第五层为全连接层 (Dense Layer)，选择 tanh 作为激活函数，权重和偏置向量使用 glorot\_uniform 初始化方法；最后一层为输出层 (Output)，模型预测的目标是下个月的超额收益率，因此是回归问题，选择输出维度为 1。模型具体结构如图 3-2-2 所示。

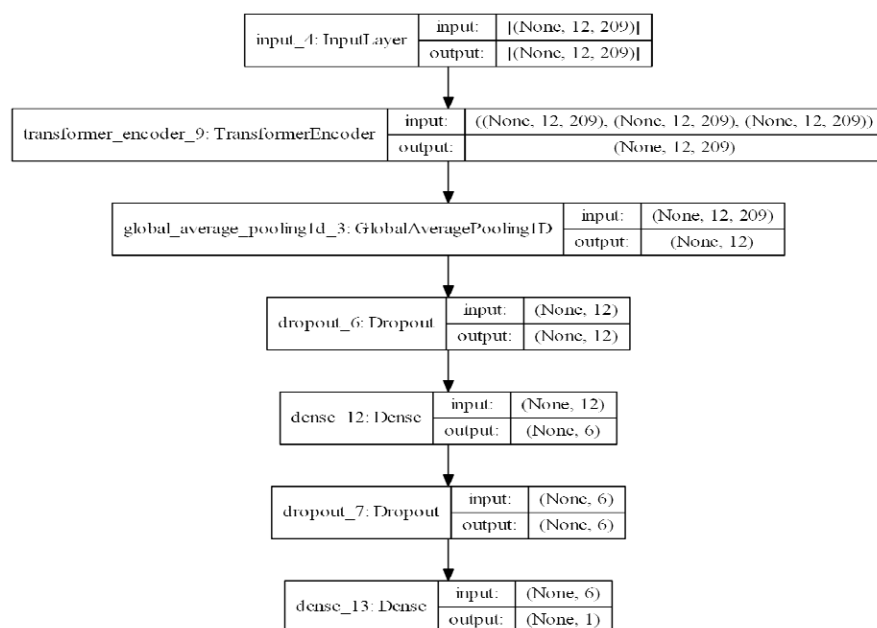


图 3-2-2: Transformer 模型结构

### 3.2.3 循环神经网络模型参数及结构设置

循环神经网络（RNN）是一种参考过去的信息对时间序列数据进行建模的深度学习算法，RNN 可以参考过去的信息对数据进行建模，本文设定的 RNN 模型结构的第一层为 input 层，第二层为一个两层的 SimpleRNN 层，第三层为 Dropout 层，Dropout 率为 0.1，最后一层为输出层（Output），模型预测的目标是下个月的超额收益率，因此是回归问题，选择输出维度为 1。通过参数寻优过程，以 Mean-Squared Loss 作为损失函数，采用 Adam 优化算法作为梯度求解的方法下，学习率 lr 设置为 0.01，训练次数 epoch 值设为 40 轮。模型具体结构如图 3-2-3 所示。

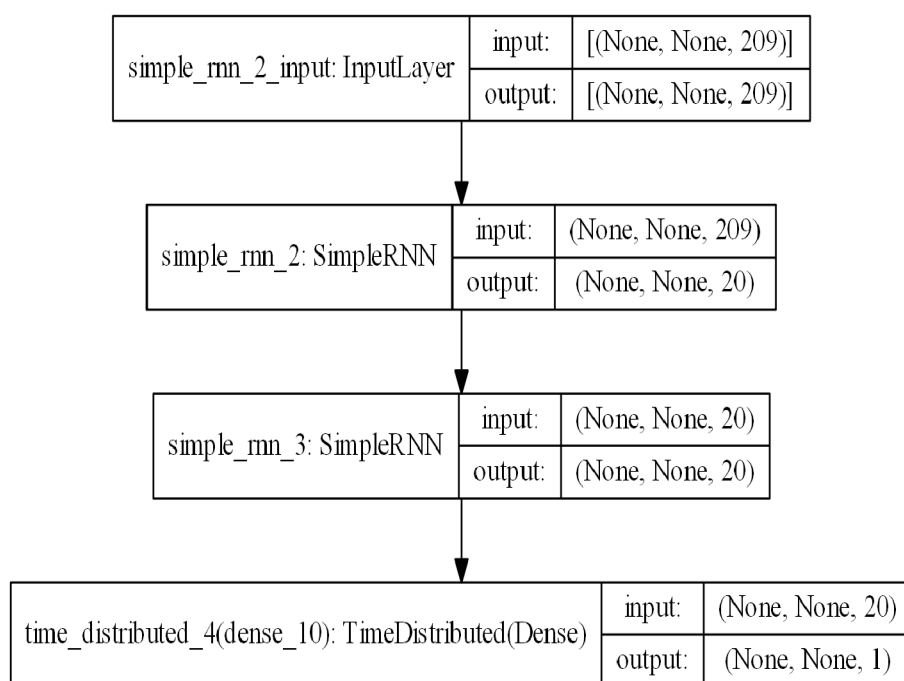


图 3-2-3: RNN 模型结构

### 3.2.4 长短期记忆神经网络模型参数及结构设置

长短期记忆神经网络（LSTM）是具有记忆长短期信息的能力的神经网络，由于 RNN 神经网络存在“长期依赖”的问题，常常会丢失过去较长时间跨度的数据信息，因此提出了 LSTM 神经网络，该神经网络引入了门（Gate）机制来控制特征的流通和损失。类似于 RNN 神经网络，本文设定的 LSTM 模型结构



的第一层为 input 层，第二层为一个两层的 LSTM 层，第三层为 Dropout 层，Dropout 率设为 0.1，最后一层为输出层（Output），模型预测的目标是下个月的超额收益率，因此是回归问题，选择输出维度为 1。通过参数寻优过程，以 Mean-Squared Loss 作为损失函数，采用 Adam 优化算法作为梯度求解的方法下，学习率 lr 设置为 0.01，训练次数 epoch 值设为 40 轮。模型具体结构如图 3-2-4 所示。

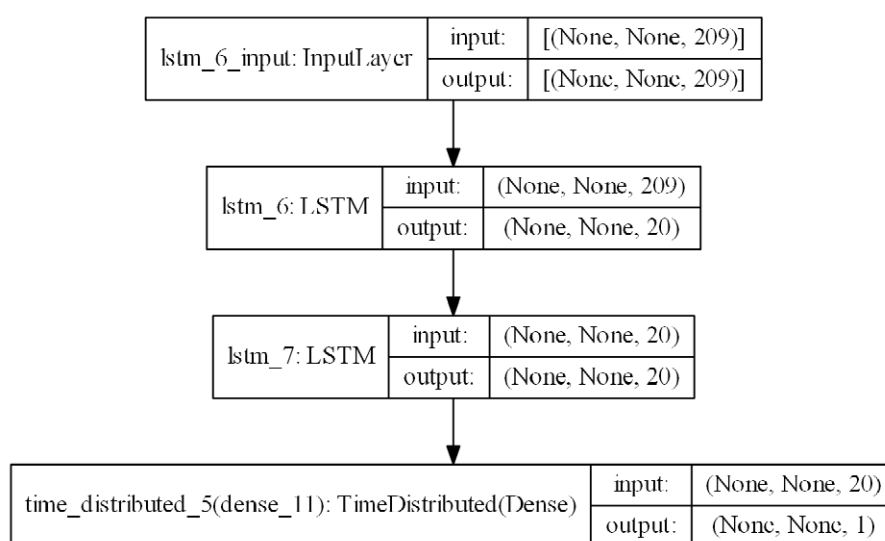


图 3-2-4: LSTM 模型结构

### 3.3 基于 Trans-T2V 模型的量化选股策略

一般来说，业界的量化投资应用思路有二，其一是量化选股，通过股票某一时点的截面数据对股票进行预测，从庞杂的股票池中筛选优质股票来辅助投资，其二是量化择时，从时间的角度设置买卖信号，使用模型对买卖信号进行预测并对该股票做出交易指令。本论文主要研究量化选股，使用模型以诸多因子值为自变量，以股票未来超额收益为因变量，对股票未来走势做出预测，挑选出未来收益更高的股票。本文的量化选股策略设计以分层划分组合为思路，首先从 T 时刻开始，取股票过去 12 个月的所有因子数据作为模型输入，通过模型预测输出每个样本的预测值  $f(x)$ ，该预测值是股票下一个月的超额收益率（以沪深 300 股指收益率为基准），随后将该值看作所有因子合成后的单因子，以该因子值按行业将股票划分为 5 个等级，进行单因子分层回测，共计形成 7 个投资组合。

其中,组合 1~组合 5 为按单因子值大小从高到低排序构造的分层组合。基准组合为沪深 300 股指组合。假设我国对股票的看空操作不设限,那么多空组合的设计为每月进行调仓时同时买入组合 1 和卖出组合 5。最后,在每个自然月的最后一个交易日计算单因子值,在下个自然月第一个交易日按当日收盘价进行调仓来完成回测。策略组合如表 3-1 所示:

投资组合	组合方式
组合 1	对每个行业,以预测值将股票排名从高到低划分出 5 个等级,选取所有行业股票等级为 1 的股票构建出组合 1。
组合 2	对每个行业,以预测值将股票排名从高到低划分出 5 个等级,选取所有行业股票等级为 1 的股票构建出组合 2。
组合 3	对每个行业,以预测值将股票排名从高到低划分出 5 个等级,选取所有行业股票等级为 1 的股票构建出组合 3。
组合 4	对每个行业,以预测值将股票排名从高到低划分出 5 个等级,选取所有行业股票等级为 1 的股票构建出组合 4。
组合 5	对每个行业,以预测值将股票排名从高到低划分出 5 个等级,选取所有行业股票等级为 1 的股票构建出组合 5。
基准组合	基准组合为沪深 300 股指组合。
多空组合	假设所有个股可以卖空的基础上,每月调仓时买入组合 1,卖空组合 5。

表 3-1: 策略组合划分

业界普遍使用单因子分层回测法来对因子进行测试,其原因主要在于该方法有着易于理解、操作简便、能够区分因子单调性的特点,单因子回测具体设置如下:

(1) 股票池与回测区间，股票池为剔除了 ST 股、停牌股、当日涨跌停股以及上市时间未满 3 个月的全 A 股，回测区间从 2010 年 1 月到 2020 年 1 月共 10 年，以月为窗口进行滚动回测，如图 3.3.1 所示。

(2) 调仓期，在每个自然月最后一个交易日计算该因子值，在下个自然月首个交易日按当日收盘价换仓，换仓频率为月频。

(3) 交易成本，本文以现实市场为参考，设置卖出印花税为千分之 1，买卖手续费为万分之 5，最小交易费用为 5 元。

(4) 分层方法，在每个一级行业内部对所有个股按因子大小进行排序，每个行业内均分成 N 个分层组合，如图所示，深蓝色方块代表各行业内个股初始权重，分层具体操作方法为 N 等分行业内个股权重累加值，例如图示行业 1 中，5 只个股初始权重相等（即每只股票的份额为 1/5），假设分 3 层对股票排序，则分层组合 1 在权重累加值 1/3 处截断，即分层组合 1 包含个股 1 和个股 2，它们的权重配比为  $1/5:(1/3-1/5)=3:2$ ，分层组合 2 包含个股 2、3、4，配比为  $(2/5-1/3):1/5:(2/3-3/5)=1:3:1$ ，同样地，分层组合 4 包含个股 4、5，配比为  $(1/3-1/5):1/5=2:3$ 。具体操作如图 3-3-2 所示。



图 3-3-1：滚动回测示意图

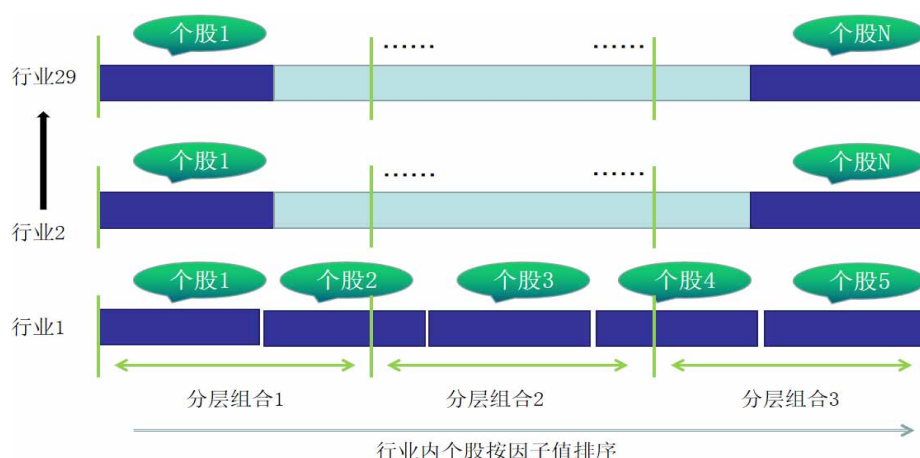


图 3-3-2: 单因子分层示意图

（5）评价方法：累积收益率、回测年化收益率、夏普比率、信息比率、最大回撤、相对基准月胜率等。

后续基于 Transformer、RNN 和 LSTM 模型的量化选股策略与基于 Trans-T2V 的选股策略设计思路保持一致。

### 3.4 本章小结

本章首先介绍了实证需要的基础设备，包括硬件、软件及相关三方库，接着介绍了模型总体的构建流程，包括股票数据获取、所用到的因子指标、特征和标签提取、数据预处理、训练集、验证集和测试集的设置、模型训练、测试以及评价指标，最后详细介绍了本文所设计的 Trans-T2V 模型具体各层结构和超参数设置以及 Transformer 模型、RNN 模型、LSTM 模型的详细设置，并以此构建了相应的模型选股策略，为之后的实证环节做了铺垫。

## 4 实证分析

### 4.1 实证环境及数据选取

#### 4.1.1 开发环境

本论文的开发与实验是在 Windows10 64 位操作系统下完成，其中 windows 配置为：CPU 为 6 核，Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz，独立显卡 为 NVIDIA GTX 1060 6G，集成显卡为 Intel(R)UHD Graphics 630 8G，使用 Jupyter notebook 作为 IDE 进行代码编写。数据预处理用到的三方库包含 numpy、pandas、os 库，因子的计算及处理主要使用了 TA-lib 库，绘制图像主要运用 matplotlib 和 seaborn 库，Trans-T2V 模型和对比模型 Transformer、RNN、LSTM 的构建主要使用 tensorflow2-gpu 库，模型进行参数寻优时主要应用 sklearn 库。

#### 4.1.2 数据选取

本论文数据主要从 tushare 和 joinquant 以及 wind 上进行获取，所用到的数据为全 A 股的股票日频数据，并将数据处理为月频数据。时间范围为 2010 年 1 月至 2020 年 1 月共 120 个月，所涉及的股票因子包含估值、成长、财务质量、杠杆、动量反转、市值、波动率、股价、换手率、情绪、技术共 11 个因子类型，具体涉及 91 个因子数据来对股票的超额收益进行预测，部分因子列表如表 4-1 所示，完整版见附录。

大类因子	具体因子	因子描述
估值	EP	净利润 (TTM)/总市值
估值	EPcut	扣除非经常性损益后净利润 (TTM) /总市值
估值	BP	净资产/总市值
估值	SP	营业收入 (TTM) /总市值
估值	NCFP	净现金流 (TTM) /总市值

估值	OCFP	经营性现金流（TTM）/总市值
估值	DP	近 12 个月现金红利（按除息日计）/总市值
估值	G/PE	净利润（TTM）同比增长率/PE_TTM
成长	Sales_G_q	营业收入（最新财报，YTD）同比增长率
成长	Profit_G_q	净利润（最新财报，YTD）同比增长率

表 4-1: 部分因子选取列表

## 4.2 评价指标

### 4.2.1 模型评价指标

机器学习中通常需要各种损失函数来评测模型性能，本文使用均方误差损失函数（MSE Loss）来对回归任务种模型的优劣做出评价，MSE 描述的是模型预测结果对目标值的偏离程度，其值越小说明拟合效果越好，MSE 的定义为计算每一个样本的预测值与真实值偏离的平方，然后求和再取平均值。MSE 的取值范围为  $[0, +\infty]$ ，一般来说，该值越小说明模型性能越佳。

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2 \quad (4-2-1)$$

### 4.2.2 策略评价指标

本文的研究通过查阅了过往有关投资策略的参考文献以及券商的金工研报，选用以下指标来评价各个选股策略好坏。

（1）年化收益率，年化收益率是指将不同投资期限的收益率按投资期限为 1 年统一起来，使不同期限的收益率之间更具可比性。

$$R = (1 + r)^{\frac{t}{12}} - 1 \quad (4-2-2)$$

(2) 年化波动率, 年化波动率是指金融资产收益的波动程度, 反映了金融资产的风险水平和不确定性。

$$\sigma_T = \frac{\sigma_R}{T} \quad (4-2-3)$$

(3) 夏普比率, 夏普比率表明了投资组合每承担一单位总风险时, 会获得多少风险溢价, 用以衡量金融资产的绩效表现, 夏普比率越大, 投资者为获得更高收益而付出的风险越值得。

$$sharp\_ratio = \frac{E(R_p) - R_f}{\sigma} \quad (4-2-4)$$

其中  $E(R_p)$  为投资组合的期望收益,  $\sigma_p$  为投资组合的标准差。

(4) 最大回撤, 最大回撤率是一个重要的风险指标, 该指标反映在回溯期的策略收益曲线上, 从任一收益率高点走到后续时间上的收益率最低点时的下跌幅度最大值, 体现出投资者极端情况下的收益损失情况。

$$DrawDown = \max\left(\frac{R_i - R_j}{R_i}\right) \quad (4-2-5)$$

其中  $R_i$  表示该时间段中收益的极大值,  $R_j$  表示该时间段中收益的极小值。

(5) 年化超额收益率, 投资组合超过市场基准收益部分的年化收益率记为年化超额收益率, 本文以沪深 300 股票指数收益率作为市场基准收益。

$$R = R_p - R_m \quad (4-2-6)$$

其中  $R_p$  为资产组合的年化收益率,  $R_m$  为沪深 300 股票指数的年化收益率。

(6) 年化跟踪误差, 年化跟踪误差是指组合年化收益率与基准年化收益率之间差异的收益率标准差, 反映了基金管理的风险。该值越大, 表明基金经理在进行基金管理时承担了更高的风险却难以获得相当的收益。

$$TE_i = \frac{\sqrt{\frac{1}{n-1} \sum_{t=1}^n (TD_{ti} - \overline{TD_i})^2}}{T} \quad (4-2-7)$$

其中  $TD_{ti}$  表示组合  $i$  在时间  $t$  内的跟踪偏离度, 反映了组合的净值增长率 and 基准收益之间的差异收益率;  $TE_i$  表示组合的年化跟踪误差,  $\overline{TD_i}$  表示组合  $i$  跟踪偏离度的样本均值,  $T$  表示年份数。

(7) 信息比率, 信息比率表示了每单位主动风险下所获得的风险溢价, 衡量了基金经理主动投资的能力, 信息比率越高, 证明基金经理的主动管理能力越强。

$$IR = \frac{\alpha}{\omega} \quad (4-2-8)$$

其中  $\alpha$  为组合的超额收益, 由年化超额收益率表示;  $\omega$  为主动风险, 由年化超额收益波动率表示。

(8) 相对基准月胜率, 相对基准月胜率是指以月为单位, 投资者持有资产组合后收益超过基准收益 (超额收益大于 0) 的概率, 用以表明主动型投资由于被动型投资的程度。

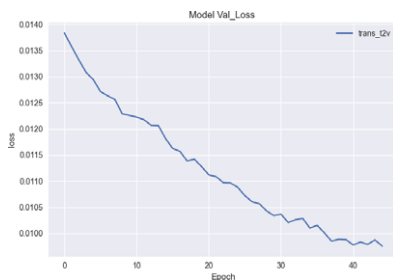
$$\begin{aligned} R_{tp} - R_{tm} > 0, y_t &= 1 \\ Y &= \sum y_t \\ win &= \frac{Y}{n} \end{aligned} \quad (4-2-9)$$

其中  $R_{tp}$  为第  $t$  月组合的收益率,  $R_{tm}$  为第  $t$  月沪深 300 股指的收益率;  $Y$  为组合收益超过基准收益的月份数量;  $n$  为投资期总月份数。

(9) 超额收益最大回撤, 超额收益最大回撤是指计算周期内, 投资组合收益高于基准收益部分的最大回撤率。

### 4.3 各模型损失函数分析

本文依照滚动训练的原则对前面提到的 Trans-T2V、Transformer、RNN 和 LSTM 模型进行测试, 观察各模型不同 epoch 参数下的损失函数变化情况, 为高效训练模型做铺垫。各模型验证集损失值变化如图 4-3-1 所示。通过模型的损失函数最小的变化情况来看, 在 epoch 小于 40 之前, 各模型验证集上的损失函数在迅速下降, 当模型的 epoch 进入 40 次后变化趋缓, 因此综合考虑算力与时间成本的影响, epoch 只需要进行到 40 次即可。同时可以看出, 当 epoch 进行到 40 次后, Trans-T2V 模型 MSE Loss 为 0.0099, 处于 4 种模型最低的位置, 模型性能最佳, Transformer 模型次之, MSE Loss 值为 0.0101, LSTM 模型和 RNN 模型 MSE Loss 分别为 0.0118 和 0.0108。

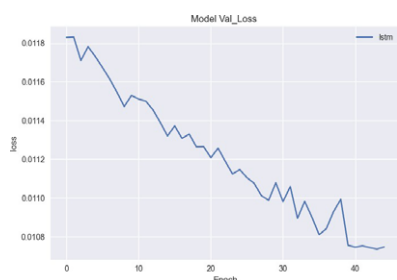


(a) Trans-T2V

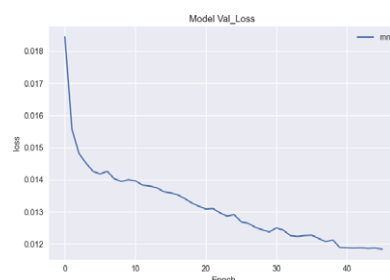


(b) Transformer





(c) LSTM



(d) RNN

图4-3-1: 各模型 MSE Loss 的变化情况

## 4.4 基于各模型的策略回测结果对比分析

### 4.4.1 不同投资组合对比

对于不同的投资组合，本文使用年化收益率、年化波动率、夏普比率、最大回撤、相对基准月胜率等诸多评价指标进行对比分析，以 Trans-T2V 模型所作策略为例，分层组合指标如表 4-2 所示。

投资组合	年化收益率	年化波动率	夏普比率	最大回撤	年化超额收益率	年化跟踪误差	信息比率	相对基准月胜率	超额收益最大回撤
组合 1	18.76%	8.21%	1.57	40.86%	15.30%	0.50%	1.86	63.33%	32.84%
组合 2	9.92%	7.93%	0.78	44.01%	6.46%	0.47%	0.81	58.33%	32.53%
组合 3	7.75%	7.95%	0.59	48.03%	4.29%	0.47%	0.54	53.33%	30.84%
组合 4	5.91%	8.17%	0.41	53.61%	2.45%	0.51%	0.30	49.17%	33.71%
组合 5	3.83%	8.66%	0.23	58.53%	0.38%	0.61%	0.04	51.67%	36.91%
基准组合	3.46%	2.06%	0.00	38.79%	0.00%	0.00%	0.00	0.00%	0.00%
多空组合	17.75%	2.32%	2.61	29.22%	14.29%	0.23%	6.16	61.25%	14.37%

表 4-2: Trans-T2V 模型分层组合指标分析

从表中可以看到，多空组合策略和分层组合策略 1 在年化收益率、夏普比

率、年化超额收益率、相对基准月胜率等方面均优于组合 2-组合 5，其中多空组合更是以风险指标更小、收益指标更大的特点显著优于其他组合。接下来将分层组合收益可视化，更为直观地观察各组合收益情况，如图 4-4-1 所示。

图 (a) 为分五层组合回测净值图。组合 1 到组合 5 的净值与基准组合（沪深 300）对比图。

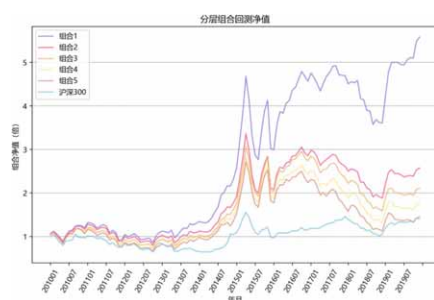
图 (b) 为分五层组合回测，为了便于更加清楚地展示各个组合在不同时间段的对比情况，本文对组合 1 到组合 5 采用了净值除以基准组合净值的处理方式。

图 (c) 为组合 1 相对沪深 300 月超额收益分布直方图。

图 (d) 为分五层时的多空组合收益图。

其中蓝线是基准沪深 300 指数的收益曲线，从图 1 和 2 可以看出，以 Trans-T2V 为基准的分层策略累计收益收益均跑赢了基准，组合 1 的累计收益更是超过 500%，同时，与模型预测的因子排序方向相同，组合 1 至组合 5 的策略整体表现也是由大到小依次下降，观察图 3，可以看到组合 1 相对沪深 300 月超额收益分布右偏，表示组合 1 的收益表现优于沪深 300 基准组合，展现出了 Trans-T2V 模型具有一定的选股能力，说明了以该模型为基准构建的投资策略的有效性。

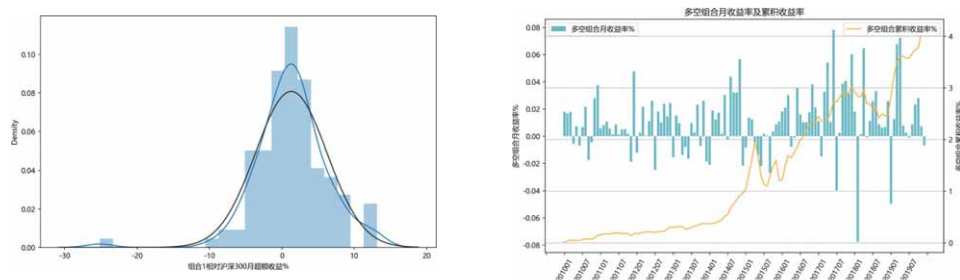
本文假设股票市场没有卖空限制，从而构建多空组合来进一步地验证基于模型的策略性能，策略的回测结果如图 (d) 所示，图中曲线为多空策略的累计收益率曲线，柱状图表示多空组合的月收益率情况，从图中可以看出，在进行多空对冲后，策略的 10 年的累计收益超过 400%，月收益情况多数为正，表现出了良好的选股能力。



(a) 分层组合回测净值示意图



(b) 各层组合净值除以基准组合净值示意图



(c) 组合 1 相对于沪深 300 月超额收益分布直方图 (d) 多空组合月收益率及累积收益率示意图

#### 4.4.2 不同选股策略对比

本文构建了全 A 选股策略并分别运用 Trans-T2V、Transformer、RNN 和 LSTM 模型对数据进行回测，从上文可知，投资组合 1 与多空组合的整体表现最好，考虑到中国股市的现实情况，选择组合 1 作为投资策略对比各模型下的策略优劣，调仓频率为月频，以模型预测的当月预测值顺序作为股票池中股票选择的顺序依据。各项指标见表 4-3。

	年化收 益率	年 化 波 动 率	夏 普 比 率	最大回 撤	年化超 额收益 率	年 化 跟 踪 误差	信 息 比 率	相对基 准月胜 率	超额收 益最大 回撤
Trans-T2V	18.76%	8.21%	1.57	40.86%	15.30%	0.50%	1.86	63.33%	32.84%
Transformer	11.87%	8.28%	1.03	39.13%	9.40%	0.49%	1.14	59.17%	32.93%
RNN	11.48%	8.42%	1.01	43.89%	9.08%	0.52%	1.08	59.17%	32.37%
LSTM	10.26%	8.42%	0.88	41.37%	7.75%	0.52%	0.92	57.50%	33.97%

表 4-3：不同选股模型指标对比

不同选股模型下策略组合 1 的收益情况如图 4-4-2 所示：

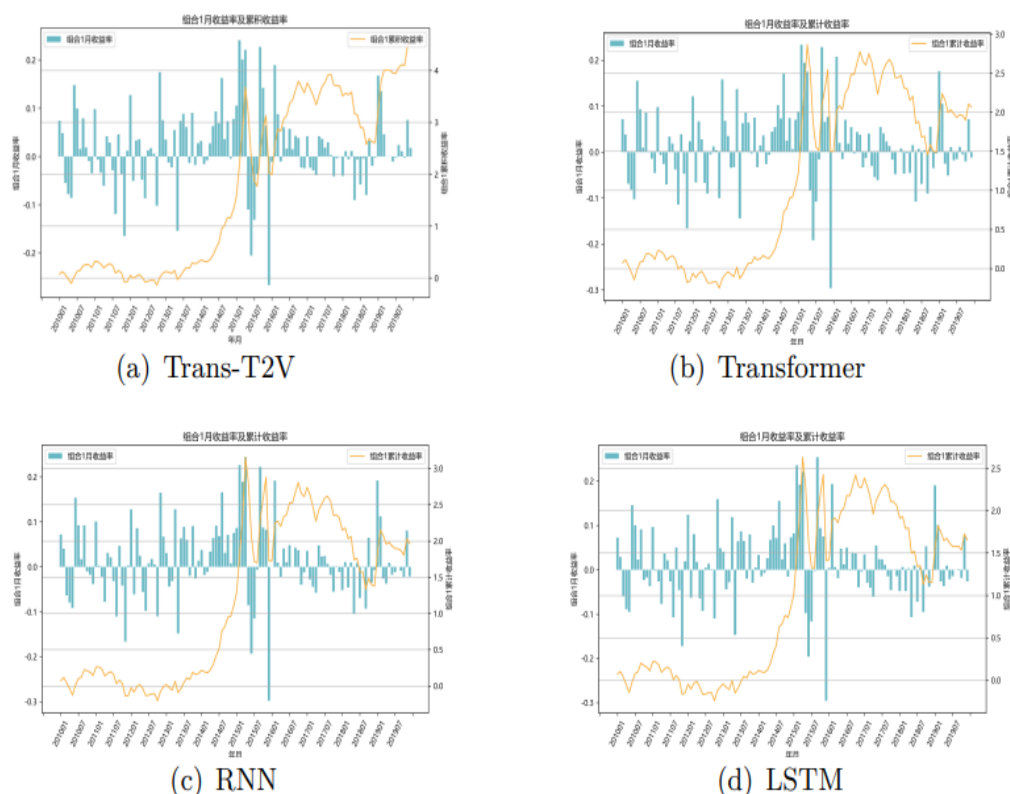


图 4-4-2: 各选股模型策略收益情况

可以看到,对于行业中性的全 A 选股,Trans-T2V 回归下的多头注意力算法在年化收益率、年化超额收益率、夏普比率、信息比率和相对基准月胜率方面显著优于其它三种模型模型,表明了 Trans-T2V 结构下的神经网络性能优于循环神经网络。然而,Transformer 模型的选股表现却与以 RNN、LSTM 模型为代表的循环神经网络结构相差不大,且与 Trans-T2V 模型的性能存在较大差距。可能的原因在于,一方面,Transformer 模型的复杂度较之 RNN 与 LSTM 模型更高,但本文使用 2010 年至 2020 年全 A 股的月频数据,数据量仅为 437026 条,对于庞大的 Transformer 架构来说,需要更多更丰富的数据量才能更好的发挥其性能;另一方面,Transformer 模型本身并不具备识别时序数据相对位置的能力,没有时间嵌入层的加入,模型对时间序列数据的利用率变得低下。因此,从上述指标对比可以看出,本文所改进的 Trans-T2V 模型在量化选股方面具备一定的实际意义。

另外,通过 Trans-T2V 与 Transformer 的指标对比可以发现,Transformer 神经网络在加入了时间嵌入层 Time2Vector 后,在处理金融时间序列数据方面确有提升,证明了时间嵌入层的有效性,为今后优化自身模型提供了一个思路。

## 4.5 本章小结

本章首先介绍了一些基于模型的策略评价指标，分为模型评价指标和策略评价指标，模型评价指标主要观察 4 种模型的均方误差，策略评价指标囊括了年化收益率、年化波动率、夏普比率、最大回撤、年化超额收益率、年化跟踪误差、信息比率、相对基准月胜率和超额收益最大回撤。接着本文按照前述方法构建分层组合并对组合 1 到组合 5 进行比较，结果表明组合 1 在策略评价指标方面显著优于其他 4 种分层组合。随后，本文构建多空策略来进一步观察 Trans-T2V 网络架构的选股有效性，通过做多组合 1 的股票，做空组合 5 的股票，策略累计收益超过 400%，年化超额收益达到了 14.29%，而最大回撤仅为 14.3%，最后本论文同样为 Transformer 回归、RNN 回归和 LSTM 回归构建相应的选股策略并以组合 1 为比较对象与 Trans-T2V 神经网络进行比较，结果发现基于 Trans-T2V 模型的量化策略不论从是模型评价指标来看或是从策略评价指标来看均优于其他 3 种模型，表明了本文改进模型具备一定的实际意义。

## 5 研究结论

### 5.1 研究总结

目前我国的量化投资正处于高速发展阶段，量化投资因其信息捕捉能力强、交易具备纪律性、数据分析广泛、数据处理快速响应等特点而逐渐受到广大投资者的欢迎，但现实来看，不同的量化投资策略其收益状况却大相径庭，这说明了基于不同算法框架下的量化投资策略其效果千差万别，因此一个好的算法结构起着至关重要的作用。本文利用近年大热的 Transformer 神经网络与时间特征向量 Time2Vector 相结合构建出适用于处理金融时间序列数据的 Trans-T2V 神经网络，然后介绍了深度学习在量化投资领域的发展情况，引出了本文所构建模型的理论意义和现实意义。随后本文介绍了 Transformer 相关理论与结构、Time2Vector 相关理论以及 Trans-T2V 的具体结构。接着，本文以从 tushare、joinquant 和 wind 下载的全 A 股股票的 70 个因子数据作为原始数据，通过 python 的 numpy、pandas 及 talib 等第三方库构建出 209 个因子数据并利用深度学习库 Tensorflow2-gpu 构建了 Trans-T2V、Transformer、RNN 神经网络模型，为了更充分利用样本数据，提高模型泛化性能，本文采用按月滚动的方式对模型进行训练与回测，然后本文分别介绍了各个神经网络模型的具体结构以及相应的参数设置，紧接着本文对改进模型 Trans-T2V 在全 A 选股且跨度为 10 年的环境下做出了系统的测试，初步得到以下几个结论：一、Trans-T2V 股票价格预测模型具备不错的预测能力。为了提高模型在不同时间阶段的泛化能力，本文采取滚动回测方法，按月对模型进行滚动训练，共计滚动 120 次，并选取 Transformer、RNN、LSTM 模型作为对照模型进行实证。结果表明，Trans-T2V 模型的均方误差在 epoch 进入 40 次后趋缓，测试集的均方误差仅为 0.0099，性能高于其他 3 种模型，二、本文以全 A 股为股票池，分别利用 Trans-T2V 回归模型、Transformer 回归模型、RNN 回归模型和 LSTM 回归模型构建选股策略。实证结果表明，Trans-T2V 回归模型能在较低的风险下获取更高的超额收益，策略指标显著高于其他 3 种模型，在收益方面，Trans-T2V 模型在多数选股策略下能获取更高的超额收益；而在回撤方面，Trans-T2V 模型相比于其他 3 种模型模型不具备明显优势。同时，本文还发现 Transformer 神经网络结构在没有

时间嵌入层的加入下,对金融时序数据的处理能力相较于循环神经网络来说并不具备优势,可能的原因在于:一方面本文现有的月频数据量不够丰富,难以充分发挥模型性能;另一方面是 Transformer 本身不具备识别时序数据相对位置的能力,对数据的利用率大大降低。总的来说,Trans-T2V 不仅适用于股价预测,而且比 Transformer 神经网络和循环神经网络的表现更好,证明了本文所提出的改进模型应用于选股方面的有效性。

## 5.2 不足与改进

从模型实证结果来看,在 2010 年 1 月至 2020 年 1 月这一历史回测区间上,基于 Trans-T2V 模型的量化选股策略显著跑赢了基准策略(沪深 300 股指),其年化超额收益和夏普比率等风险与收益指标都显著高于本文其他对照策略。但该模型与其他机器学习模型相同的一个缺点是模型的前期盈利能力并不突出,基本与基准收益持平,这也表明了深度学习神经网络需要充分而优质的数据量来支撑。另外,该模型的最大回撤并不优于其他 3 种模型,说明该模型在股市风格发生异变时并没有显著的抗风险能力,未来仍然有较大的提升空间。本文深度学习算法的优化方向主要从四个方向去考虑,第一个方向是考虑使用更高频率的数据使模型性能得到充分发挥,如使用周频、日频甚至分钟数据去训练模型,提高模型精度,第二个方向是优化量化投资策略,本文仅简单地依靠模型回归将股票分 5 层构建投资组合,可以尝试以模型分类的方式预测股票涨跌方向并更为细致地对股票进行分层。第三个方向是改进神经网络的架构以及相关的超参数,这需要研究者有扎实的数学功底、高效的硬件设施以及付出昂贵的时间去不断试错,第四个方向是从因子改良的思路入手,这方面的研究是量化策略研究的一个热点方向,可以着重研读券商的金工研报实现改进。因此,未来策略的目标是使得本文的模型不论在经济上行时期或经济下行时期都能获得较高的收益,提升模型的抗风险能力。

## 参考文献

- [1] 范龙振, 余世典. 中国股票市场的三因子模型[J]. 系统工程学报, 2002, 17(6): 537-546 (引用页: 5).
- [2] 杨炘, 陈展辉. 中国股市三因子资产定价模型实证研究[J]. 数量经济技术经济研究, 2003(12): 137-141 (引用页: 5).
- [3] 李传乐. 基于均值-方差相交的定价因子研究——兼论 GMM 中估计 BlockBootstrap 模拟的区组选择[J]. 南方经济, 2007(8): 74-84 (引用页: 5).
- [4] 兰峰. 基于 Fama-French 三因素改进模型的再生水行业期望收益率研究[J]. 统计与决策, 2010, 2010(3): 53-55 (引用页: 5).
- [5] 欧阳志刚, 李飞. 四因子资产定价模型在中国股市的适用性研究[J]. 金融经济研究, 2016, 2: 84-96 (引用页: 5).
- [6] 赵昕, 崔峰, 丁黎黎. 时变三因子模型风险系数的动态研究[J]. 统计与决策, 2020(6): 5-10 (引用页: 5).
- [7] 龙建成, 李小平. 基于神经网络的股票市场趋势预测[J]. 西安电子科技大学学报, 2005, 32(3): 460-463 (引用页: 5).
- [8] 刘海玥, 白艳萍. 时间序列模型和神经网络模型在股票预测中的分析[J]. 数学的实践与认识, 2011, 41(4): 14-19 (引用页: 5).
- [9] 张秋明, 朱红莉. 灰色神经网络在股价预测中的应用研究[J]. 计算机工程与应用, 2015, 49(12): 29-34 (引用页: 5).
- [10] 孙秋韵, 刘金清, 刘引, 等. 基于改进 GA 参数优化的 SVR 股价预测模型[J]. 计算机系统应用, 2015, 24(9): 29-34 (引用页: 5).
- [11] 骆桦, 张喜梅. 基于贝叶斯分类法的股票选择模型的研究[J]. 浙江理工大学学报, 2015, 33(3): 418-422 (引用页: 5).
- [12] 孟叶, 于忠清, 周强. 一种安全的量化选股策略——来自 A 股市场的实证[J]. 金融理论与实践, 2018, 24(8): 102-107 (引用页: 5).
- [13] 贾秀娟. 基于随机森林的支持向量机量化选股[J]. 区域金融研究, 2019(1): 27-30 (引用页: 5).
- [14] 任君, 王建华, 王传美, 等. 基于正则化 LSTM 模型的股票指数预测[J]. 计算机应用与软件, 2018, 35(4): 44-48 (引用页: 7).



- [15] 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(7): 308-315 (引用页: 7). [48] 赵红蕊, 薛雷. 基于 LSTM-CNN-CBAM 模型的股票预测研究[J]. 计算机工程与应用, 2021, 57(3): 203-207 (引用页: 7).
- [16] 罗泽南. 基于集成树模型的 Stacking 量化选股策略研究[J]. 中国物价, 2021 (2): 81-84 (引用页: 7).
- [17] 康瑞雪, 牛宝宁, 李显, 等. 融合多源数据输入具有自注意力机制的 LSTM 股票价格预测[J]. 小型微型计算机系统, 2021: 1-9 (引用页: 7).
- [18] MARKOWITZ H. Portfolio Selection[J]. The Journal of Finance, 1952, 7(1): 77-91 (引用页: 3).
- [19] SHARPE W F. Capital asset prices: A theory of market equilibrium under conditions of risk[J]. The journal of finance, 1964, 19(3): 425-442 (引用页: 3).
- [20] LINTNER J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets[J]. The Review of Economics and Statistics, 1965: 13-37 (引用页: 3).
- [21] ROSS S A. Mutual fund separation in financial theory—The separating distributions[J]. Journal of Economic Theory, 1978, 17(2): 254-286 (引用页: 3).
- [22] FAMA E F, FRENCH K R. The cross-section of expected stock returns[J]. The Journal of Finance, 1992, 47(2): 427-465 (引用页: 3).
- [23] CARHART M M. On persistence in mutual fund performance[J]. The Journal of finance, 1997, 52(1): 57-82 (引用页: 3).
- [24] TITMAN S, WEI K J, XIE F. Capital investments and stock returns[J]. Journal of financial and Quantitative Analysis, 2004, 39(4): 677-700 (引用页: 4).
- [25] FAMA E F, FRENCH K R. A five-factor asset pricing model[J]. Journal of financial economics, 2015, 116(1): 1-22 (引用页: 4).
- [26] BLACK F, SCHOLES M. The Pricing of Options and Corporate Liabilities [J]. Journal of Political Economy, 1973, 81(3): 637-654 (引用页: 4).
- [27] BOX G E, JENKINS G M. Time series analysis. Forecasting and control[J]. Holden-Day Series in Time Series Analysis, 1976 (引用页: 4).
- [28] ENGLE R F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation[J]. Econometrica: Journal of the ec

- onometric society, 1982: 987-1007 (引用页: 4).
- [29] T B. Generalized autoregressive conditional heteroskedasticity[J]. Journal of Econometrics, 1986, 31(3): 307-327 (引用页: 4).
- [30] 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 (引用页: 4).
- [31] WANG L, ZHU J. Financial market forecasting using a two-step kernel learning method for the support vector regression[J]. Annals of Operations Research, 2010, 174(1): 103-120 (引用页: 4).
- [32] HADAVANDI E, SHAVANDI H, GHANBARI A. Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting[J]. Knowledge-Based Systems, 2010, 23(8): 800-808 (引用页: 4).
- [33] TICKNOR J L. A Bayesian regularized artificial neural network for stock market forecasting[J]. Expert systems with applications, 2013, 40(14): 5501- 5506 (引用页: 4).
- [34] SIROHI A K, MAHATO P K, ATTAR V. Multiple kernel learning for stock price direction prediction[C]//2014 International Conference on Advances in Engineering & Technology Research (ICAETR-2014). [S.l. : s.n.], 2014: 1-4 (引用页: 4).
- [35] MOGHADDAM A H, MOGHADDAM M H, ESFANDYARI M. Stock market index prediction using artificial neural network[J]. Journal of Economics, Finance and Administrative Science, 2016, 21(41): 89-93 (引用页: 4).
- [36] SEZER O B, OZBAYOGLU A M, DOGDU E. An artificial neural network based stock trading system using technical analysis and big data framework [C]//Proceedings of the southeast conference. [S.l. : s.n.], 2017: 223-226 (引用页: 4).
- [37] XU Y, COHEN S B. Stock movement prediction from tweets and historical prices[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). [S.l. : s.n.], 2018: 1970- 1979 (引用页: 5).

- [38] MEI S, GAO H. The time function of stock price[J]. ArXiv preprint arXiv:2008.11806, 2020 (引用页: 5).
- [39] TAKEUCHI L, LEE Y Y A. Applying deep learning to enhance momentum trading strategies in stocks[G]//Technical Report. [S.l.]: Stanford University Stanford, CA, USA, 2013 (引用页: 6).
- [40] RATHER A M. A hybrid intelligent method of predicting stock returns[J]. Advances in Artificial Neural Systems, 2014, 2014 (引用页: 6).
- [41] PENG Y, JIANG H. Leverage financial news to predict stock price movements using word embeddings and deep neural networks[J]. ArXiv preprint arXiv:1506.07220, 2015 (引用页: 6).
- [42] TÜRKMEN A C, CEMGİL A T. An application of deep learning for trade signal prediction in financial markets[C]//2015 23rd Signal Processing and Communications Applications Conference (SIU). [S.l. : s.n.], 2015: 2521-2524 (引用页:6).
- [43] XIONG R, NICHOLS E P, SHEN Y. Deep learning stock volatility with google domestic trends[J]. ArXiv preprint arXiv:1512.04916, 2015 (引用 页: 6).
- [44] QIN Y, SONG D, CHEN H, et al. A dual-stage attention-based recurrent neural network for time series prediction[J]. ArXiv preprint arXiv:1704.02971, 2017 (引用页: 6).
- [45] FISCHER T, KRAUSS C. Deep learning with long short-term memory networks for financial market predictions[J]. European Journal of Operational Research, 2018, 270(2): 654-669 (引用页: 6).
- [46] GUNDUZ H, YASLAN Y, CATALTEPE Z. Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations[J]. Knowledge-Based Systems, 2017, 137: 138-148 (引用页: 6).
- [47] TOVAR W. Deep Learning Based on Generative Adversarial and Convolutional Neural Networks for Financial Time Series Predictions[J]. ArXiv preprint arXiv:2008.08041, 2020 (引用页: 6).
- [48] ZHENG L, HE H. Share price prediction of aerospace relevant companies with recurrent neural networks based on pca[J]. Expert Systems with Applications, 2021, 183: 115384 (引用页: 6).

- [49] SUN T, ZHENG W, YU S, et al. ESAN: Efficient Sentiment Analysis Netl  
work of A-Shares Research Reports for Stock Price Prediction[J]. ArXiv pre  
print arXiv:2112.11444, 2021 (引用页: 7).
- [50] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J].  
Advances in neural information processing systems, 2017, 30 (引用页: 10).
- [51] KAZEMI S M, GOEL R, EGHBALI S, et al. Time2vec: Learning a vector r  
epresentation of time[J]. ArXiv preprint arXiv:1907.05321, 2019 (引用 页: 1  
6).
- [52] MASTERS D, LUSCHI C. Revisiting small batch training for deep neural n  
etworks[J]. ArXiv preprint arXiv:1804.07612, 2018 (引用页: 25).

## 附录

### 1 所用因子展示

大类因子	具体因子	因子描述
估值	EP	净利润 (TTM)/总市值
估值	EPcut	扣除非经常性损益后净利润 (TTM)/总市值
估值	BP	净资产/总市值
估值	SP	营业收入 (TTM)/总市值
估值	NCFP	净现金流 (TTM)/总市值
估值	OCFP	经营性现金流 (TTM)/总市值
估值	DP	近 12 个月现金红利 (按除息日计)/总市值
估值	G/PE	净利润 (TTM) 同比增长率 /PE_TTM
成长	Sales_G_q	营业收入 (最新财报, YTD) 同比增长率
成长	Profit_G_q	净利润 (最新财报, YTD) 同比增长率
成长	OCF_G_q	经营性现金流 (最新财报, YTD) 同比增长率
成长	ROE_G_q	ROE (最新财报, YTD) 同比增长率
财务质量	ROE_q	ROE (最新财报, YTD)
财务质量	ROE_ttm	ROE (最新财报, TTM)
财务质量	ROA_q	ROA (最新财报, YTD)
财务质量	ROA_ttm	ROA (最新财报, TTM)
财务质量	grossprofitmargin_q	毛利率 (最新财报, YTD)

财务质量	grossprofitmargin_ttm	毛利率（最新财报，TTM）
财务质量	profitmargin_q	扣除非经常性损益后净利润率（最新财报，YTD）
财务质量	profitmargin_ttm	扣除非经常性损益后净利润率（最新财报，TTM）
财务质量	assetturnover_q	资产周转率（最新财报，YTD）
财务质量	assetturnover_ttm	资产周转率（最新财报，TTM）
财务质量	operationcashflowratio_q	经营性现金流/净利润（最新财报，YTD）
财务质量	operationcashflowratio_ttm	经营性现金流/净利润（最新财报，TTM）
杠杆	financial_leverage	总资产/净资产
杠杆	Debt equity ratio	非流动负债/净资产
杠杆	Cash ratio	现金比率
杠杆	Current ratio	流动比率
市值对数	total_mv_log	总市值取对数
总市值	total_mv	总市值
动量反转	HAlpha	个股 60 个月收益与上证综指回归的截距项
动量反转	return_Nm	个股最近 N 个月收益率，N=1, 3, 6, 12
动量反转	wgt_return_Nm	个股最近 N 个月内用每日换手率乘以每日收益率求算术平均值，N=1, 3, 6, 12
动量反转	exp_wgt_return_Nm	个股最近 N 个月内用每日换手率乘以函数 $\exp(-x_i/N/4)$ 再乘以每日收益率求算术平均值， $x_i$ 为该日距截面日的交易日个数，N=1, 3, 6, 12

波动率	std_FF3factor_Nm	特质波动率——一个股最近 N 个月内用日频收益率对 F-F 三因子回归的残差的标准差, N=1, 3, 6, 12
波动率	std_Nm	个股最近 N 个月的日收益率序列标准差, N=1, 3, 6, 12
股价	ln_price	股价取对数
beta	beta	个股 60 个月收益与上证综指回归的 beta
换手率	turn_Nm	个股最近 N 个月内日均换手率 (剔除停牌、涨跌停的交易日), N=1, 3, 6, 12
换手率	bias_turn_Nm	个股最近 N 个月内日均换手率除以最近 2 年内日均换手率 (剔除停牌、涨跌停的交易日) 再减去 1, N=1, 3, 6, 12
情绪	rating_average	wind 评级的平均值
情绪	rating_change	wind 评级 (上调家数-下调家数)/总数
情绪	rating_targetprice	wind 一致目标价/现价-1
股东	holder_avgpctchange	户均持股比例的同比增长率
技术	MACD	DIF=快的指数移动平均线-
	DEA	慢的指数移动平均线;
	DIF	DEA=DIF 的 9 日加权移动均线; MACD= 2× (DIF-DEA)
技术	RSI	该指标是根据一段时间内股票价格涨幅与股票价格波动

		的平均水平来衡量多空势力的强弱，从而给出相应的建议。周期取 20 日
技术	PSY	该指标将一段时间投资者的心理预期进行量化，计算出股价的未来涨跌情况，周期取 20 日。
技术	BIAS	该指标由收盘价与移动平均线的差异计算而来，以反映一定时期的价格偏离程度的周期取 20 日。

---



## 致 谢

光阴似箭日月如梭，两年前，我踏入河南财经政法大学的大门成为了一名研究生，转瞬间，我却要再次离开校园，心中难免生出不舍，不舍这优美的校园，不舍这清脆的铃声，不舍这诲人不倦的老师，不舍这相亲相爱的同学，但生活总是向前，我们终究要离开母校，走向人生的下一站。疫情似乎让时间停留在 2020 年，这两年中，关于郑州的记忆并不是很多，但学校的教室、图书馆和自习室却留下了我的身影，让我收获许多。在这研究生生涯结束之际，我想感谢一些人，第一个就是引我入门的喻军老师，若不是老师孜孜不倦地教导，我也不会迅速成长起来。每周的研讨会，喻军老师都会让我们学习某本书、某篇论文从而分享各自的收获，这让我接触到了各个金融领域，让我思路开阔，也让我逐渐步入量化投资与机器学习的新世界。此外，老师勤奋、专注、爱好阅读以及严谨的学术研究习惯也在潜移默化地影响着我。其次，要感谢的是给我带课的各位老师，让我从一个土木行业的门外汉进一步了解到金融是什么，也让我形成了金融思维。最后，我要感谢我的家人，我的女朋友以及我的同窗们，感谢家人对我学习与生活的默默付出，感谢女朋友在我失落及彷徨时给予的陪伴，感谢同窗们在讨论中给予我的灵感，让我得以克服路上的一切困难，顺利完成论文。