

分类号: TP399

单位代码: 11057

密 级: 公开

学 号: 222109252006



**浙江科技大学**  
ZHEJIANG UNIVERSITY OF SCIENCE & TECHNOLOGY

## 硕士学位论文

论文题目: 基于多特征融合的图注意力股票  
走势预测研究

作 者 黄 琨

指导教师 李晓明 教授

学科专业 应用统计

二级学院 理学院

提交日期 2023 年 12 月 22 日

**A Thesis Submitted to  
Zhejiang University of Science and Technology for  
Master's Degree**

**Research on Stock Trend Prediction based  
on Multi-feature Fusion with Graph  
Attention Neural Network**

**Candidate: Huang Kun**

**Supervisor: Li Xiaoming**

**Zhejiang University of Science and Technology**

**Hangzhou China**

**December 2023(时间)**

基于多特征融合的图注意力

股票走势预测研究

论文作者签名:\_\_\_\_\_

指导教师签名:\_\_\_\_\_

论文评阅人 1: \_\_\_\_\_ 匿名评阅

评阅人 2: \_\_\_\_\_ 匿名评阅

评阅人 3: \_\_\_\_\_

评阅人 4: \_\_\_\_\_

评阅人 5: \_\_\_\_\_

答辩委员会主席: \_\_\_\_\_ 刘仁平 教授/硕导 浙江工商大学

委员 1: \_\_\_\_\_ 章迪平 教授/硕导 浙江科技大学

委员 2: \_\_\_\_\_ 盛宝怀 教授/硕导 浙江越秀外国语学院

委员 3: \_\_\_\_\_ 吴黎军 教授/硕导 浙江越秀外国语学院

委员 4: \_\_\_\_\_ 吴炎崑 教授/硕导 浙江越秀外国语学院

委员 5: \_\_\_\_\_

答辩日期: \_\_\_\_\_ 2023.12.02



**Research on stock trend prediction based on**  
**multi-feature fusion with graph attention networks**

**Candidate's signature:** Huang Kun

**Supervisor's signature:** Li Xiaoming

Thesis reviewer 1: Anonymous Reviewer

Thesis reviewer 2: Anonymous Reviewer

Thesis reviewer 3: \_\_\_\_\_

Thesis reviewer 4: \_\_\_\_\_

Thesis reviewer 5: \_\_\_\_\_

Chair: Liu Renping Professor ZJSU  
(Committee of oral defence)

Committeeman 1: Zhang Diping Professor ZUST

Committeeman 2: Sheng Baohuai Professor ZYU

Committeeman 3: Wu Lijun Professor ZYU

Committeeman 4: Wu Yankun Professor ZYU

Committeeman 5: \_\_\_\_\_

Date of oral defence: December 2,2023

## 摘 要

股票走势预测有助于投资者做出更合理的投资决策,这个经典且具有挑战性的问题,受到经济学家和计算机科学家的广泛关注。金融时间序列数据具有高度波动的特点,受到市场内外多种因素影响。而股价的变动不仅受自身状态的改变影响,还与相关公司状况紧密相连。然而,以往的研究通常只关注股票金融时序数据特征,未充分考虑金融文本特征,忽略股票间的关系特征。因此,引入和整合更全面的信息特征对于提高预测准确性至关重要。本文在传统深度学习模型基础上,融合股票市场多源特征,进一步通过图注意力机制学习股票关系特征,从而提升预测模型性能。

首先,构建了基于多级注意力机制的股票趋势预测模型。通过优化网络结构,采用多级注意力机制作为主体模型,在股票金融时序数据特征基础上,引入金融新闻特征和预定义股票关系特征。多级注意力机制可以对不同类型特征赋予不同的权重,从而有效聚合不同类型的特征,并将其融合到新的特征节点表示中。

然后,设计了一个新的多源特征融合的双图注意力模型。该模型在原来模型获取的金融时序特征和文本特征的基础上,引入低秩多模态融合模块提取股票深度特征。通过双图注意力机制聚合到由社区检测挖掘的股票集群结构构成的股票关系网络的节点表示,避免了将公司相关性限制在仅由先验信息定义的行业关系中。

通过长时间序列上的广泛实验证明本文所提出模型在股票趋势预测中具有显著的优势。融合的多源特征和股票关系特征提高了预测准确性,特别是通过社区检测挖掘得到的股票关系,更加准确地捕捉了市场中股票之间的集群结构,进一步提升了模型的学习能力。这项研究突破了以往仅关注时间序列数据本身的局限性,将多源特征融合到预测模型中,为投资者提供了更全面、准确的决策依据。

**关键词:** 股票预测, 图注意力网络, 社区检测, 股票关系, 多特征融合

## ABSTRACT

Stock trend prediction is critical to enable rational investment decisions, and this problem has attracted significant attention from economists and computer scientists. Financial time series data is highly volatile and affected by various market and external factors. The stock price varies not only due to the changes of its own state, but also closely associated with the conditions of related companies. However, previous studies have limited their focus only on the features of stock financial time series data and failed to fully consider financial text features and relationship features between stocks. Therefore, it is essential to integrate additional informative features to enhance prediction accuracy. This study employs traditional deep learning models and integrates multiple features from the stock market, further learning stock relationship features using Graph Attention Networks, thus improving the model performance.

Firstly, a multi-level attention mechanism-based model for stock trend prediction was constructed. By optimizing the network structure and employing the multi-level attention mechanism as the primary model, financial news and predefined stock relationship features were included based on the stock financial time series data. The multi-level attention mechanism assigns different weights to different feature types, effectively aggregating and merging diverse features into new representations.

Then, a novel dual-graph attention model with multi-source feature fusion was designed. This model incorporates low-rank multimodal fusion to extract deep stock features based on the financial time series features and text features obtained by the original model. Through the dual-graph attention mechanism, it aggregates the node representation of the stock relationship network formed by stock clusters mined through community detection and avoids limiting the relevance of companies to industry relationships that are defined by prior information. Extensive experiments with long time series data have confirmed that the proposed model has a significant advantage in predicting stock trends. The integration of multi-source features and stock relationship features has led to an improvement in prediction accuracy. In particular, the clustering structure among stocks in the market is more accurately captured by mining stock relationships through community detection, further enhancing the model's learning ability. This research overcomes the limitations of focusing solely on time series data in previous studies by integrating multi-source

features into the prediction model, providing investors with a more comprehensive and accurate decision-making reference.

**Key Words:** stock prediction, graph attention network, community detection, stock relationship, multi-feature fusion

# 目 录

第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 研究现状.....	3
1.2.1 股票预测.....	3
1.2.2 图神经网络在金融领域的应用.....	5
1.2.3 股票网络建模.....	6
1.3 研究内容和创新点.....	7
1.3.1 研究内容.....	7
1.3.2 创新点.....	8
1.4 论文章节安排.....	9
第 2 章 相关技术与理论 .....	10
2.1 新闻文本表征技术.....	10
2.1.1 Word2Vec .....	10
2.1.2 BERT .....	11
2.2 深度学习算法.....	12
2.2.1 多层感知机.....	12
2.2.2 卷积神经网络.....	13
2.2.3 长短期记忆网络.....	14
2.2.4 图神经网络.....	15
2.3 股票网络的构建.....	16
2.3.1 基于元路径的关系信息提取.....	16
2.3.2 基于社团检测的关系提取.....	17
2.4 本章小结.....	18
第 3 章 基于多级注意力机制的股票预测模型 .....	19
3.1 问题描述.....	19
3.2 特征提取.....	20
3.2.1 基于 LSTM 的历史价格编码器.....	21
3.2.2 基于 BERT 的金融新闻编码器.....	21
3.2.3 基于 meta-path 的关系编码器.....	22
3.3 多级图注意力机制.....	23
3.4 分类预测及模型训练.....	25
3.5 实验设计和结果分析.....	25



3.5.1 数据集.....	25
3.5.2 实验评估方法.....	27
3.5.3 实验环境和参数设置.....	28
3.5.4 实验结果与对比分析.....	28
3.6 本章小结.....	34
第4章 双图注意力和多源特征融合的股票预测模型 .....	35
4.1 问题描述.....	35
4.2 特征提取.....	36
4.2.1 股票特征.....	36
4.2.2 关系特征.....	36
4.3 双图注意力机制.....	38
4.3.1 多模态特征融合层.....	38
4.3.2 序列特征嵌入层.....	39
4.3.3 关系特征嵌入层.....	39
4.4 分类预测及模型训练.....	41
4.5 实验设计和结果分析.....	42
4.5.1 数据集.....	42
4.5.2 评价指标.....	42
4.5.3 实验环境和参数设置.....	43
4.5.4 实验结果与分析.....	44
4.6 本章小结.....	49
第五章 结论与展望 .....	50
5.1 全文总结.....	50
5.2 未来展望.....	51
参考文献.....	52

# 第1章 绪 论

## 1.1 研究背景及意义

### 1.1.1 研究背景

中国的经济规模已经快速扩大，成为全球经济的一个重要引擎。自1990年12月上海证券交易所开市以来，中国股市发展迅猛，成为全球第二大股票市场，上市公司数量超过4800家，总市值巨大。股市的快速发展有助于促进中国金融市场的稳定性、提高中国金融市场的透明度和效率。投资者通过股票市场不仅可以激发投资热情，还可以引导社会资金向实体经济投资，进一步促进经济的可持续发展。因此，中国股市在国家经济和金融市场中扮演着越来越重要的角色，股票趋势预测有利于辅助投资者做出更合理的投资决策，是一个经典而又具有挑战性的研究热点问题。

在股票市场预测研究中，通常采用两种著名的分析方法，即基本面分析和技术分析<sup>[1-2]</sup>。基本面分析侧重于影响公司或行业的潜在因素。例如，公司的收入、费用、年增长率、市场地位以及其他包含在财务报表或报告中的信息。如果对代表众多公司股票的股票指数进行预测，就可以利用包括国家生产率、贸易、汇率或利率在内的同类信息和市场环境信息。这可能会对该指数所含公司的运营产生影响。相反，技术分析是侧重于对历史股票价格和成交量数据的研究，以预测股票价格的走势。技术分析是文献中最常见的方法，使用股票价格或由此衍生的指标作为输入。技术分析师认为，所有新信息，如新闻和宏观经济变量，都已经反映在股价中。因此，分析价格趋势的模式就足以预测股票市场。技术指标被广泛研究，并被用作股票信号，指示何时买进或卖出一只股票。

以往的研究大多采用基于历史数据的统计时间序列方法来预测股票价格和收益。其中，自回归条件异方差(ARCH)模型、自回归移动平均(ARMA)模型、自回归综合移动平均(ARIMA)模型、移动平均、卡尔曼滤波、指数平滑是最常用的技术。后来，随着人工智能的引入，深度学习在股票市场预测研究中越来越受到关注。与传统的时间序列方法不同，这些技术可以处理非线性、混沌、噪声和复杂的股票市场数据，从而实现更有效的预测。因此，这些方法代表了创新和有利的替代方法，这使它们具有吸引力，被研究人员采用的金融市场预测。大量研

究者用回归方法去预估股票未来价格；少部分学者通过分类方法去预测股票行情未来趋势。但是，越来越多的投资者更关心股票市场的未来走势。目前，已经有很多方法被应用到预测股票趋势，由传统的统计方法，到现在被引入股票市场处理复杂金融数据的深度学习，如卷积神经网络(CNN)，递归神经网络(RNN)和长短期记忆网络(LSTM)。虽然很多复杂的深度学习模型已被用于对时间序列建模且预测表现出色，但大多数以前的工作都只孤立地考虑了金融时间序列。因此，引入和融合更全面的信息至关重要，系统地聚合新闻事件、历史行情和股票间关系等信息是做出准确预测的关键。

随着全球化的趋势不断加强，各个行业间关系发展得越来越复杂，股票市场作为金融市场重要的组成部分，也变得日益复杂。股票市场不仅受到金融环境、经济政策和行业发展等因素的影响，同时也会受到国内外上市公司之间错综复杂的关系的影响。为了更好地研究股票市场的波动和变化，研究者们开始从多方面入手，包括股票行业关系、供应链关系、股票之间的相互关系等。在这些因素中，图神经网络模型被广泛应用于联合考虑各种影响因素。然而，当前的研究中大多数着眼于用预定义行业关系衡量股票之间的相互影响，而没有直接研究对股票收益率波动之间相关性的影响，这是需要我们进一步探究的问题。因此，我们需要在研究股票市场的过程中，多角度、全方位地考虑影响因素，在实际操作中，应该选择适当的模型和算法进行研究和分析。本文通过基于预定义行业关系和社团检测两种方式去挖掘股票关系，然后引入图注意力模型更新股票关系图节点特征，以此学习股票间的相关性。

### 1.1.2 研究意义

股票市场预测技术对于股票市场投资者、投资机构有巨大的经济价值，它可以帮助投资机构和投资者实现盈利，在一定程度可以规避投资风险，而股票市场预测技术的经济价值远不止于此。从社会层面来说，股票市场预测技术可以预防金融市场系统性风险，可以避免出现金融危机，股市崩盘等经济大萧条情况，帮助合理配置社会资金，为社会经济和谐稳定的发展作出贡献。另外股票数据具有自身特点和复杂性，一些预测技术方法不完全适用当下股票预测，怎么去充分融合结构化股票价格数据和非结构化市场数据的特征，成为了股票预测一大难点。因此此研究项目对技术提出了新的挑战，特别是其中的多特征融合技术不仅可以

用于股票市场预测,在医疗事件、能源需求预测、社交媒体行动以及网站流量预测等多个领域有广泛的借鉴意义,因为这些领域预测的本质都是对时间序列数据进行分析然后预测。

## 1.2 研究现状

本文通过不同方式挖掘股票关系构建股票关系网络图,然后引入多种注意力机制更新股票关系图节点特征,以此学习股票间相互影响的权重大小,从而对股票未来趋势进行预测。本节将从股票预测、图神经网络在金融领域的应用和股票网络建模三个方向整理和阐述国内外最新研究现状。

### 1.2.1 股票预测

股票市场预测是金融和科技交叉领域的经典研究问题,受到投资者和学者的广泛关注。股票市场上的投资行为通常受到某种形式的预测方法的指导,学者将这些预测方法主要分为两种,也就是基本面分析和技术分析<sup>[1-2]</sup>。基本面分析关注的是股票背后公司的运营状态,而不是股票本身<sup>[3]</sup>,而技术分析主要通过研究过去和现在的股价趋势来预测股票未来价格<sup>[4]</sup>。但是,著名的有效市场假说(EMH)<sup>[5]</sup>对此给出了一个悲观的观点,它认为技术分析和基本面分析不会给投资者带来超额利润,并暗示股票的价格反映了市场中所有可用信息。然后,许多研究者并不同意有效市场假说<sup>[6]</sup>。一些学者试图衡量成熟市场和新兴市场的不同效率水平,而另一些学者试图为股票市场建立有效的预测模型去预测股票趋势<sup>[7-10]</sup>。

股票的历史价格信息是一个明显的时序序列,所以早期的股票预测研究主要以时间序列分析为主。传统的时间序列模型依赖于计量经济学理论,如自回归模型、自回归移动平均模型、自回归综合移动平均模型及其扩展模型。如 Bollerslev<sup>[11]</sup>引入线性模型 GARCH 作为股票预测的解决方案;万健强<sup>[12]</sup>将 ARIMA 模型应用于香港股指的股票预测研究中。随着机器学习技术的发展,相关的预测模型层出不断,特别是深度学习模型,凭借神经网络强大的特征学习能力,可以更好的提取股票市场高度非线性和非平稳波动的特征,逐渐成为股票预测研究领域的主流。大量研究者用回归方法去预估股票未来价格<sup>[13-15]</sup>;少

部分学者通过分类方法去预测股票行情未来趋势<sup>[16]</sup>。传统机器学习方法侧重于技术分析，将历史股票价格等作为输入，并使用 SVM<sup>[17]</sup>和 RF<sup>[18]</sup>等机器学习算法对股票进行建模分析。Hoseinzade 等人<sup>[19]</sup>提出一个基于 CNN 的框架，成功的结合各种信息进行预测，实验表明深度学习方法优于浅层方法。Zhao 等人<sup>[20]</sup>提出基于 RNN 的预测模型，并引入注意力机制去聚焦关键信息。Ding 等人<sup>[21]</sup>提出一种基于 LSTM 的多输入多输出的关联深度递归神经网络模型，实验结果表明优于其他预测模型。随着股市公开信息的增加，研究者开始关注基本面分析，将影响公司或行业的潜在因素作为预测属性。Ding 等人<sup>[22]</sup>使用新型神经张量网络从新闻事件中提取特征，以预测股价走势。Wang 等人<sup>[23]</sup>提出新颖的框架可以有效地提高投资意见挖掘和个股预测的性能。

然而，之前的研究大多数侧重于单一或部分信息<sup>[16-24]</sup>，忽略了市场内其他信息。股票也不是独立存在的，股价的波动会受到相关公司突发事件的影响，比如龙头股可以带动整个板块行情上涨，相关股票间具有显著的联动性。然而目前研究大多数孤立地预测单只股票<sup>[20-25]</sup>，忽略股票间的相关性。所以系统地聚合新闻事件、历史行情和股票间关系等信息是做出准确预测的关键。实际上，有少量研究者整合了不同来源的信息去预测股票趋势。如 Zhang 等人<sup>[26]</sup>通过增强注意力的 LSTM 模型融合多种来源数据显著提高了股票预测性能。而金融市场的内在属性和股票间实际存在的相关属性，启发研究者引入图结构来更好的解决股票趋势的预测问题，旨在用图结构去处理市场中不同来源的信息，但这些工作仍处于起步阶段。Chen 等人<sup>[27]</sup>通过构建一个包含所有相关公司关系的图去更准确的进行预测，但只整合了企业相关信息；Li 等人<sup>[28]</sup>通过图卷积网络对股票关系建模去预测隔夜股票走势，只整合了新闻数据；Wang 等人<sup>[29]</sup>提出自适应分层时间关系图网络(HTAR)来表征和预测股票演变，但忽略了新闻事件的影响；Matsunaga 等人<sup>[30]</sup>和 Chen 等人<sup>[31]</sup>都运用图神经网络去对股票相关因素去进行建模分析，但忽略投资者情绪对预测的影响。虽然复杂的图模型已被引入处理金融领域信息，但目前方法都是简单地将不同来源的信息集成到图中，并不能高效利用股票市场信息。所以，如何高效挖掘并融合股票市场多源数据仍然是当下面临的重要挑战。

### 1.2.2 图神经网络在金融领域的应用

因为图数据保持了个体特征和复杂关系,因此在金融领域得到了深入的研究;而图神经网络(GNN)能够通过更新节点的表示来获取结构信息,因此得到了广泛的应用<sup>[32]</sup>。金融系统,特别是股票交易系统,是一个组成部分众多,关系复杂,更新频繁的复杂系统。为了表示金融领域中的各种信息和更好的利用其中的关系数据,通常构造相关图,比如用户评论图<sup>[33]</sup>,货币交易网络<sup>[34]</sup>和股票关系图<sup>[8,23]</sup>。通过图形表示,可以将各种任务表示为一个节点分类任务,然后在这些任务中可以使用 GNN 的方法。然而,金融系统的复杂性导致了数据源的多样性和图形结构的复杂性,这对数据序列或文本信息提出了挑战,金融数据需要谨慎处理以保持时间模式或语义。同时,金融关系的时变性和多面性使得很难构建一个图表来捕捉这些关系。

针对金融市场以上特点,研究人员引入了图神经网络方法来学习公司之间的分布表示。Chen 等人<sup>[27]</sup>将 GNN 引入股票市场,根据真实市场投资事件构建相关公司图,通过图卷积神经网络捕获公司间关系,做出更准确的预测。Feng 等人<sup>[35]</sup>收集行业和公司之间的关系嵌入到 GNN 模型中,用时间敏感的方式捕获股票间关系,并通过时间图卷积网络为股票定制了一个排名系统。Cheng 等人<sup>[36]</sup>提出一种多模态图神经网络模型,学习多模态输入信息特征融合成异构图,以进行金融市场时间序列预测。Feng 等人<sup>[37]</sup>利用堆叠图神经网络捕获时序特征和全局信息,对股票关系进行动态感知,实现推荐高回报率股票的目的,该文也验证了图模型在股票市场的适用性和实用性。不幸的是,大多数在金融科技领域预测股票趋势的研究都孤立地考虑使用其自身历史价格的时间序列分析技术。他们往往假设股票未来的走势由历史价格信息决定,而忽略了来自其他股票波动的干扰,也就是金融学中的动量溢出效应<sup>[38-40]</sup>。

最近有研究试图通过图神经网络(GNN)对股票间动量溢出进行建模。Xu 等人<sup>[41]</sup>提出分层图神经网络合并了历史序列和股票关系进行股票预测;Feng 等人<sup>[35]</sup>通过提取 Wikidata 中公司实体关系来研究股票间动量溢出效应对股票预测的影响。这些研究把每只股票或者关联的公司作为图的一个节点,连接两个点的边由预定义的公司关系确定。然而,传统的图神经网络无法更新相关股票的节点状态和权衡不同关联程度的边<sup>[42]</sup>。其中,各种公司关系的重要性会随市场变化而变化,限制在一些由先验信息定义的关系不可避免地会产生干扰,从而影响预测

任务。

### 1.2.3 股票网络建模

随着研究者对股票市场研究的深入,也逐渐分析和证明了诸多性质,人们一步步认识到股票市场具有很多复杂系统的特性,比如动态性、敏感性、非线性等,所以股票市场是经济物理学领域中复杂系统的一个特例。近年来,复杂网络在金融市场的应用研究也逐渐发展起来,股票市场领域的研究者也开始引入复杂网络的概念和理论去对股票市场建模分析。在现有相关研究中,最常见的就是通过计算股票价格或收益率相关性来定义股票间关系矩阵。Wang 等人<sup>[45]</sup>通过 Pearson 系数构建了一个股票相关网络,以分析世界股票市场的相关性结构和演变。但是 Pearson 系数一般有前提条件,就是假设数据服从正态分布,而股票市场是具有非线性的特性。因此 Chen 等人<sup>[31]</sup>和 Zhong 等人<sup>[43]</sup>通过 Spearman 相关系数度量股票间非线性关系然后作为边去构建行业股票网络图,最后提出基于 GC-CNN 预测的新方法,综合考虑了个股信息和股票市场信息。

第一个股票网络图是由 Mantegna<sup>[46]</sup>构建,他们通过研究股票价格对数的日时间序列,发现了在金融市场上交易的股票的分层排列。而拓扑空间是一个亚占优的超参数空间,它与一个连接所分析投资组合股票的图相关联。从投资组合中所有股票对之间的相关系数矩阵出发,考虑每日股票价格对数之差的同步时间演化,得到了该图。而从 Mantegna 首次通过最小生成树分析 S&P 500 指数股票网络的层次结构后,大量学者开始研究股票网络的拓扑结构。Onnela 等人<sup>[47]</sup>通过引入最小生成树去描述股票间相关性,使用中心顶点的概念,选择作为树的最强连接节点,一个重要的特征是由平均占领层定义的。他们发现在崩盘期间,由于市场中强烈的全球相关性,树在拓扑上缩小,这可以通过平均占领层的低值来显示。黄玮强等人<sup>[48]</sup>以深证 100 指数中的股票为研究对象,分别通过平面最大过滤图算法和最小生成树算法构建相对应的股票关系网络,分析其拓扑结构和聚类性质。实验结果表明,总体上最大过滤图算法优于最小生成树算法。Pontes 等人<sup>[49]</sup>使用自回归向量模型以及方差分解和 Granger 关系去评估巴西股市的宏观经济指标的波动对资产网络拓扑结构的影响。他们在文章中主要研究巴西资本市场金融资产的复杂相关网络,同时考虑到巴西重要宏观经济指标的变化,如国内生产总值、利率等。Li 等人<sup>[50]</sup>运用复杂网络理论的前沿方法,将能源上市公司与股东

之间的股权关系视为一个异构的复杂网络,从全局的角度分析了基于这种关系的全球能源投资结构。他们构建了一个能源上市公司及其股东(两组行为主体)的原始投资网络。然后,基于两组行为主体和国家之间的双模式隶属关系,构建了一个 112 个国家(和地区)的衍生品投资网络。通过计算不同的拓扑特征,以全球能源上市公司的股权关系为基础,定量分析了各国对外和对内能源投资的差异性、能源投资偏好、国家(和地区)之间双边能源投资关系的强弱以及最强大的能源投资国群体。研究发现,绝大多数对外和对内投资关系仍然掌握在少数几个国家手中。

而对于研究股票市场这个复杂系统一种有前途的选择,就是采用复杂网络理论中的社区检测技术<sup>[44]</sup>。Chmielewski 等人<sup>[51]</sup>以复杂网络理论中的社团检测算法,分析股票间的相关性。然而,他们的研究只局限于更好的理解股票网络图之间的相互作用,没有将得到的相关性进一步纳入股票预测任务中。因此,在没有足够的先验知识情况下,如何考虑将股票市场这个复杂系统通过各种公司关系传递的动量溢出效应结合历史序列信息获得更好的预测性能仍然是具有挑战性的。

## 1.3 研究内容和创新点

### 1.3.1 研究内容

本文围绕“基于多特征融合的图注意力股票趋势预测”科学问题,从“如何有效挖掘历史序列信息特征并融合股票预定义关系进行预测”和“如何结合股票市场中动量溢出效应和多特征融合进行预测”两个关键问题展开,拟设定的研究技术路线框架如下图 1-1。两条研究主线并不是独立的,而是相互关联的,因此我们采用递进的方式进行研究。第一条研究主线用适合各种来源的特征提取器进行特征提取,然后通过元路径提取股票关系并嵌入股票图网络,解决之前股票预测研究仅仅依赖于历史价格或金融文本的不足,探索在历史序列信息特征基础上引入股票关系进行预测。第二条研究主线在此基础上,通过多特征融合提取股票深度特征,同时保留模态间的相关性。而针对以往研究预定义股票关系局限问题,探索基于社区检测算法构建股票关系图,弥补先验信息匮乏带来的不足。最后,基于两条主线研究内容,对两者的预测效果进行对比分析。



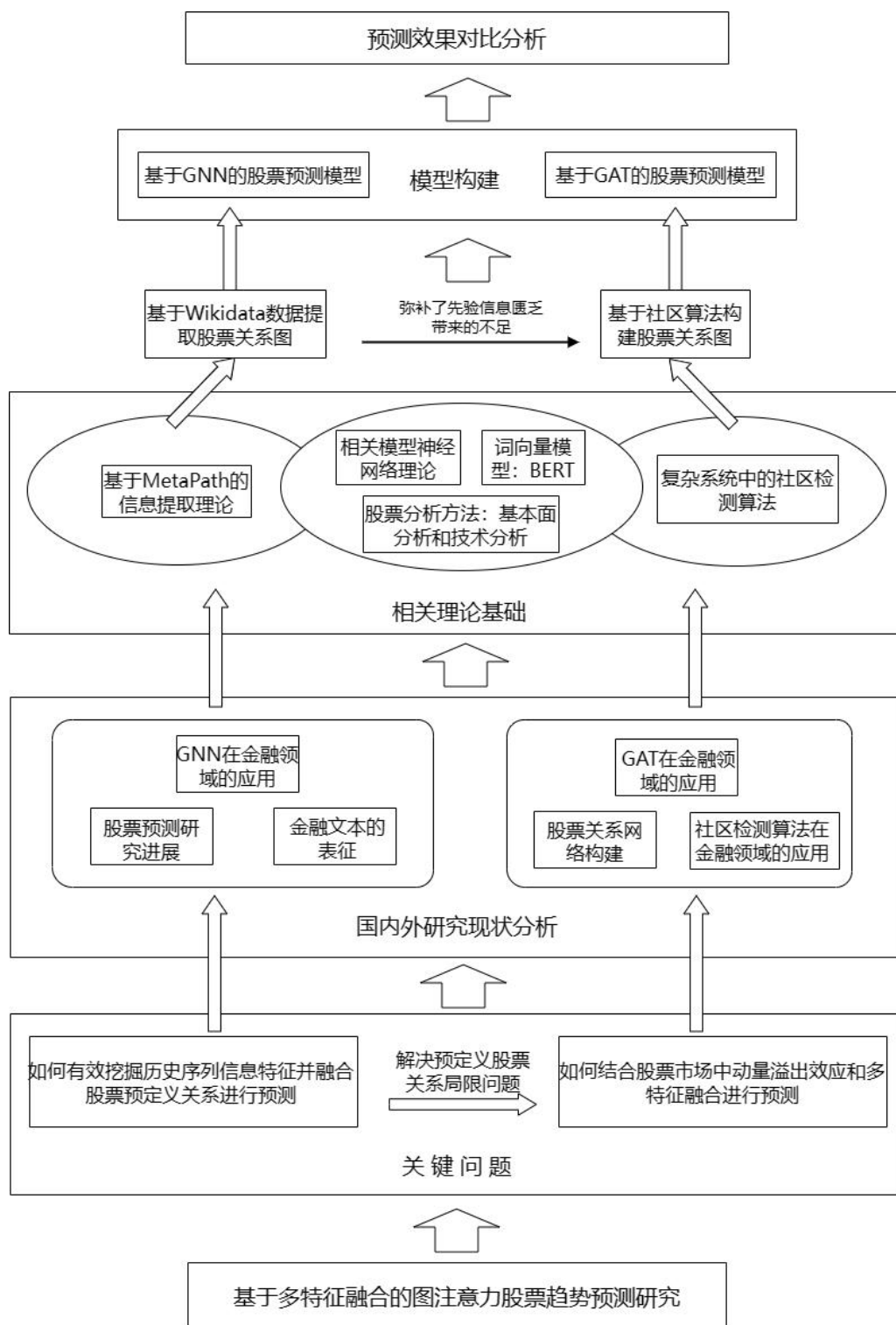


图 1-1 研究技术路线图

### 1.3.2 创新点

本文主要的创新点和贡献如下:

- (1) 提出一种新型股票多特征融合方法, 通过 LMF 模块捕获股票市场多种

模态信息间的交互特征,避免传统融合方式将特征向量直接拼接会忽略模态间相关性的问题,为股票预测任务提供更高效的股票市场信息特征。

(2) 设计了一种全新的股票关系建模模型 ML-GAT。与传统方法不同的是,ML-GAT 可以为不同的关系类型分配不同的权重,并有选择地将从特征提取模块中捕获到的节点表示聚合到股票关系图中。同时,使用股票网络图的拓扑信息和市场特征,该模型可以提高股票趋势预测的准确性和可靠性。

(3) 在(2)的基础上,提出了多特征融合的双图注意力模型(MF-DAT),与之前方法不同的是,我们不再限制股票之间的关系,通过挖掘社区结构的方法,我们能够避免过度依赖预定义的行业关系,为股票预测提供更加精确、全面的视角。通过社区检测算法,挖掘股票收益率波动相关性,并发现了股票集群结构。

## 1.4 论文章节安排

本文共有五个章节组成,此小节将对这些章节的内容进行一个简要说明:

第一章,绪论。本章首先从研究背景和研究意义出发,说明了股票趋势预测研究的大背景和深入研究这项任务的意义。然后基于领域内最新文献总结了国内外股票预测任务的研究现状。进一步地基于目前研究存在的问题提炼了本文的研究内容和创新点,最后就是简述论文各章节安排。

第二章,相关理论及算法概述。本章主要阐述了研究内容涉及的基础理论和相关算法,包括实验部分的基准模型:多层感知机、卷积神经网络、长短期记忆网络、图卷积神经网络和图注意力神经网络等。最后说明了分类任务的模型评价指标。本章的内容为本文的研究内容和模型设计及实现提供了充分的理论基础。

第三章和第四章,两种不同的股票预测模型设计及实验分析。首先就股票预测任务进行简单的问题描述,然后针对提出模型的各个模块的功能进行详细阐述,并且详细介绍了模型的搭建过程和总体框架。接下来介绍了本文实验用到的数据。然后列出了实验环境设计和模型参数设计,包括基准模型的实验设计。进一步地详细对比分析了各个模型实验的结果,展示了模型的准确率和回报率等。最后为了验证各个模块的有效性,对其进行了相对应的消融实验。

第五章,总结与展望。本章对本篇文章所开展的工作和研究成果进行了总结,然后分析了提出模型存在的不足和局限性,基于此提出未来可能的研究方向。

## 第 2 章 相关技术与理论

### 2.1 新闻文本表征技术

新闻文本向量表示实际上属于自然语言处理领域中常见的词嵌入任务范畴。目前，基于词嵌入的自然语言处理技术取得了很大的进展，这是一个积极的趋势，它可以应用于非常广泛的实际应用中，如计算词间的相似度，作为文本分类的特征，以及情感分析等不同的自然语言任务。词嵌入是一种特征学习技术，它将词汇表中的词转换为连续的低维实数向量。它是词的分布向量表示，又称语义向量空间，因为它从词的使用语境中获取了词的语义（意义）和句法（结构）信息。词嵌入大多是基于神经网络的，它首先对词向量进行随机初始化，然后在向量训练后对上下文进行最优预测，其中相应的单词往往出现并且语义相似单词具有类似的向量。词向量的生成有两种常见的方法，这两种方法都是相互关联的，第一种方法是基于词计数或词的上下文共现。第二种方法是根据上下文预测单词。

#### 2.1.1 Word2Vec

在文本/句子嵌入方面，有很多方法都是从 Word2vec<sup>[52]</sup>衍生出来的，它可以在数百万字典和数据集上进行有效的训练，可以度量词之间的相似度。Word2vec 的派生模型采用了生成词向量的思想，学习了篇章中关于句子流畅性的分布式句子级表征。Word2Vec 是 Mikolov 等人提出的一种基于原始文本的词嵌入学习方法，是 Word Embeddings 的方法之一。Word2Vec 的思想源于词的分布式表示概念，它利用浅层神经网络学习词的嵌入，并预测每个词与其上下文词之间的嵌入情况，从而使发生在相似上下文中的词发生关联。Word2Vec 内部有两种生成词向量的算法，Skip-gram（Continuous Skip-gram Model）和 CBOW（Continuous Bag-of-Words Model）。

CBOW 模型只有两个全连接层，并不是深层神经网络，如图 2-1a。它是基于周围词预测目标词（中心词）。上下文是由多个词表示的，要预测缺失的目标词，必须给出上下文词的顺序，然后根据窗口大小来预测上下文中的缺失词。CBOW 的目标函数如下：

$$\frac{1}{n} \sum_{i=1}^n \log (P(w_i|c_i)) \quad (2-1)$$

$$P(w_i|c_i) = \frac{\exp (\hat{e}(w_i)^n x)}{\sum_{j=1}^{|M|} \exp (\hat{e}(w_j)^n x)}, x = \sum_{j \in c} e(w_j) \quad (2-2)$$

式 (2-2) 中  $e(w_i)$  是词的输入向量,  $\hat{e}(w_i)$  是词的输出向量。而 Skip-gram 的思想和 CBOW 模型的原理类似, 如图 2-1b, 它在每一个估计步骤中, 都采用一个词作为中心词, 然后试着把它的上下文中的词预测到某个窗口大小, 模型要定义一个概率分布, 即单词出现在给定这个中心词上下文中的概率。最后要选择单词的向量表示, 这样就可以尝试最大化概率分布 (我们只有上下文单词的一个概率分布)。Skip-gram 的目标函数为:

$$\frac{1}{n} \sum_{i=1}^n \sum_{j \in c} \log (P(w_j|w_i)) \quad (2-3)$$

$$P(w_j|w_i) = \frac{\exp (\hat{e}(w_j)^n e(w_i))}{\sum_{m=1}^{|M|} \exp (\hat{e}(w_m)^n e(w_i))} \quad (2-4)$$

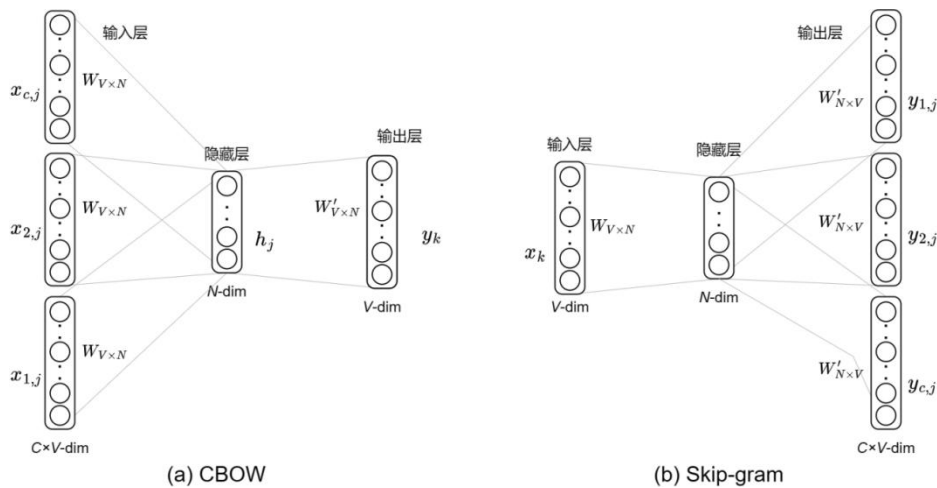


图 2-1 Word2Vec 训练方式<sup>[52]</sup>

### 2.1.2 BERT

作为 Word2Vec 的替代, Bert<sup>[53]</sup>使用 Transformer<sup>[57]</sup>作为算法的主要框架, 能够通过用于输入的编码器和用于输出的解码器建立单词之间的关系, 它可以彻底捕捉语句中的双向关系。传统的 NLP 模型将输入作为一个单词, 而基于 Transformers 的 Bert 模型将一次输入整个句子来学习隐藏在单词之间的语用意义。通过在海量语料库的基础上运行一个自监督学习模型, Bert 可以学习单词或句子的优势表征。

BERT 的基本原理是利用 Transformer 作为特征提取层，提取文本中的上下文特征，并采用了传统注意力模型的编码器-解码器结构。一个典型的 Bert 结构如图 2-2 所示。从图 2-2 中可以看出，上下文信息可以被 Transformer 充分利用。其中 E 表示嵌入，T 表示输出特征张量。

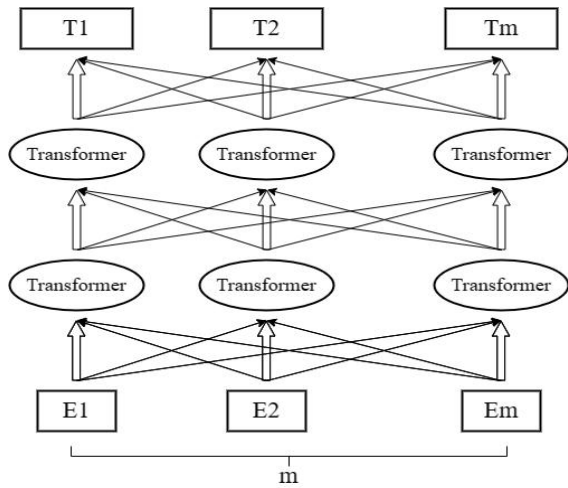


图 2-2 典型的 BERT 结构<sup>[53]</sup>

2.2 深度学习算法

2.2.1 多层感知机

多层感知器 (Multi-layer Perceptron)<sup>[54]</sup>是一种前向结构的人工神经网络 (ANN)，从一组输入向量映射到一组输出向量。MLP 可以被看做是一个具有多层节点的有向图，在输入和输出层之间有一个或多个层，每一层全连接到下一层。互连只允许在相邻的两个层之间。除了输入节点，每个节点都是一个带有非线性激活函数的神经元。使用 BP 反向传播算法的监督学习方法来训练多层感知机。MLP 是感知器的推广，克服了感知器不能对线性不可分数据进行识别的弱点，能够处理非线性可分离的问题。

相对于单层感知器，MLP 多层感知器输出端从一个变到了多个，输入端和输出端之间也不只有一层节点。基于反向传播学习的模型是典型的前馈网络，其信息处理方向从输入层到各隐层再到输出层，逐层进行。隐藏层实现对输入空间的非线性映射,输出层实现线性分类，非线性映射方式和线性判别函数可以同时进行学习。如下图 2-3 为一个三层感知机的神经网络模型。

MLP 的训练过程大致如下：

- 1) 所有边的权重随机分配。
- 2) 前向传播：利用训练集中所有样本的输入特征，作为输入层，对于所有训练数据集中的输入，人工神经网络都被激活，然后经过前向传播，得到输出值。
- 3) 反向传播：利用输出值和样本值计算总误差，再利用反向传播来更新权重。
- 4) 重复 2~3 次，直到输出误差低于制定的标准。

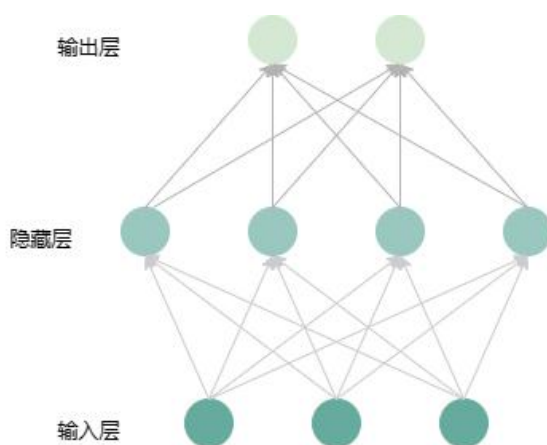


图 2-3 三层感知机的神经网络结构<sup>[54]</sup>

### 2.2.2 卷积神经网络

卷积神经网络（Convolutional Neural Networks, CNN）<sup>[55]</sup>是一种前馈神经网络，不仅在处理图像和视频，以及自然语言处理等领域都表现出优异的性能，还可以预测时间序列数据。CNN 的局部感知和权重共享能够大幅减少参数值，使模型能够提高学习效率。数据被提供给输入层进行预处理，该层的输出从该层传递到卷积层，数据通过池化层和扁平层，最后通过全连接层馈送至输出层。卷积过程应用于输入信号以提取其属性。此外，池化层可用于对输出特征图进行下采样，这允许卷积层对提取的特征图进行汇总。扁平层将所有给定的多维特征转换为一维向量，常被用在卷积层到全连接层的过渡。它的结构如图 2-4 所示。

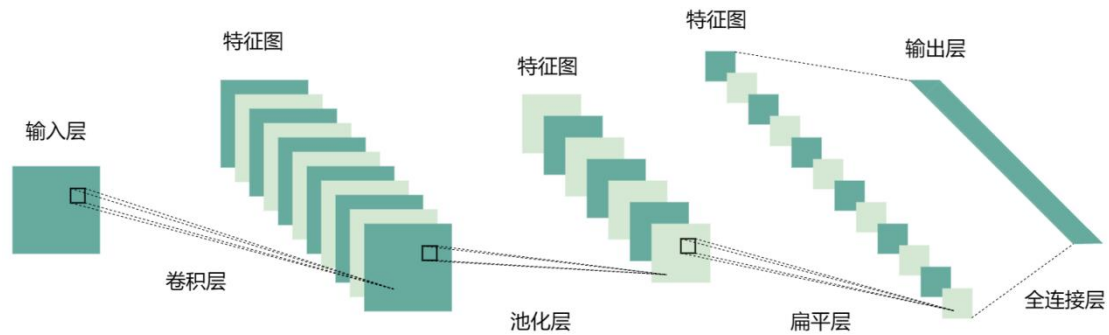


图 2-4 CNN 网络结构<sup>[55]</sup>

下面从主要的卷积层、池化层和全连接层介绍 CNN 网络结构。

(1) 卷积层。这一层对输入的原始信息进行特征提取。当输入的是股票的历史价格数据信息  $P$  时，通常采取通道设计，使用一维的卷积核对股票特征进行提取，最终卷积层提取到的特征  $F$  为：

$$F = f(P \otimes W + b) \quad (2-5)$$

式 (2-5) 中， $\otimes$  为卷积操作， $W$  为权重向量， $b$  为偏移量。

(2) 池化层。该层的目的是保留强特征数据信息，舍弃弱特征数据，从而实现降低特征维度的目的。每个特征图  $y_{m,n,z}^c$  的池化特征为：

$$y_{m,n,z}^c = \text{pool} (a_{x,y,z}^c), \forall (x, y) \in \varphi_{mn} \quad (2-6)$$

(3) 全连接层。该层是在经过卷积层和扁平层后的步骤，主要是为了进一步对前面几层提取到的特征信息进行最终整合，以获得更具有区分度的输出特征。

### 2.2.3 长短期记忆网络

长短期记忆网络 (Long Short Term Memory, LSTM) 是 Hochreiter 等<sup>[58]</sup>在 1997 年提出的，是 RNN 中一种流行的用于时间序列预测的深度学习技术。例如，LSTM 被用于分类和回归问题，不仅用于股票市场预测，还用于降雨径流建模、异常检测、移动交通预测。虽然标准 RNN 在保存信息方面优于传统网络，但由于消失梯度问题，它在学习长期依赖关系方面并不有效。而 LSTM 使用存储单元来克服渐变消失的问题，它由输入门、单元状态门、遗忘门和输出门组成。LSTM 体系结构的关键组件是在链中运行的单元状态，只有线性交互，保持信息流不变。LSTM 的门机制可以删除或修改单元状态信息。它是一种有选

择性地传递信息的方法，由 s 形层、双曲正切层和逐点乘法运算组成，具体网络结果如下图 2-5 所示。

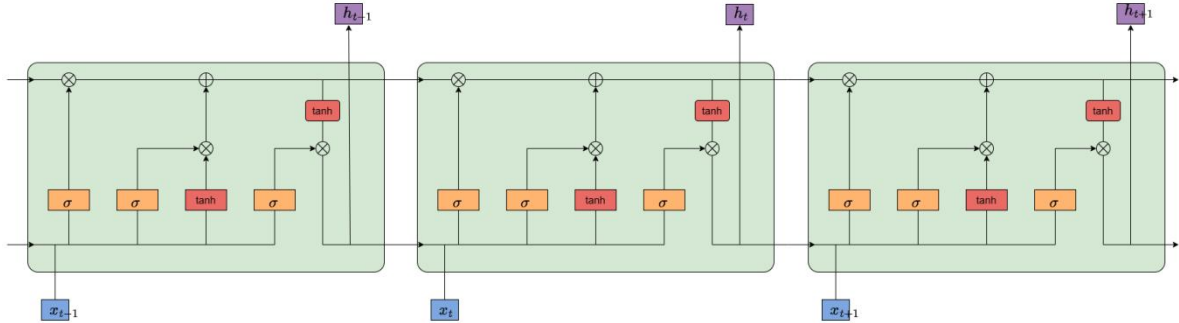


图 2-5 LSTM 网络结构<sup>[56]</sup>

对于给定的输入序列，存储单元 $c_t$ 使用三个门更新信息：输入门 $i_t$ 、遗忘门 $f_t$ 和改变门 $\tilde{c}_t$ 。LSTM 使用输出门 $o_t$ 和存储单元 $c_t$ 更新隐藏状态 $h_t$ 。在时间  $t$ ，各个门和层计算下列函数如下：

$$i_t = \sigma(W_i x_t + W_{hi} h_{t-1} + b_i) \quad (2-7)$$

$$f_t = \sigma(W_f x_t + W_{hf} h_{t-1} + b_f) \quad (2-8)$$

$$o_t = \sigma(W_o x_t + W_{ho} h_{t-1} + b_o) \quad (2-9)$$

$$\tilde{c}_t = \tanh(W_c x_t + W_{hc} h_{t-1} + b_c) \quad (2-10)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \quad (2-11)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (2-12)$$

其中， $\sigma$ 和  $\tanh$  分别表示 sigmoid 函数和双曲正切函数， $\otimes$ 是元素乘积， $W$ ， $W_h$ 是权矩阵， $b$ 是偏置向量。

#### 2.2.4 图神经网络

图神经网络(Graph Neural Network, GNN)是指能够处理图的神经网络模型的统称，根据采用的分类方法和技术的不同，又可以分为图卷积神经网络(GCN)和图注意力网络(GAT)等。图是一种强大的数据结构，通过各种方法学习图中有意义的节点表示，可以用来处理关系数据。与网络嵌入方法的主要目标是生成一个向量来表示每个节点不同，图神经网络模型被设计用于各种任务，包括节点分类、边缘预测和图分类。图神经网络模型由于其广泛的应用和优越的



性能而受到人们的广泛关注。随着人们对利用图结构化数据越来越感兴趣，人们对图中有意义的表示进行了大量的研究。大多数图神经网络可以分为谱型和非谱型。

基于谱图论的方法如 GCN 利用卷积神经网络来捕获图结构数据中的局部模式。GCN 应用频谱卷积滤波器来提取傅里叶域中的信息，而频谱卷积滤波器可以用数学公式表示如下：

$$f_{\theta}(M, x) = UMU^T x \quad (2-13)$$

式 (2-13) 中，图数据  $x \in \mathbb{R}^n$ ， $n$  表示股票公司个数； $M$  表示对角矩阵， $U$  是图拉普拉斯矩阵的特征向量矩阵。

然而，对于大量的图数据，计算图的拉普拉斯特征分解计算量太大。为了解决这个问题，Kipf 和 Welling 用基于切比雪夫系数  $\theta_k$  的切比雪夫多项式  $T_k(x)$  来近似谱滤波器，切比雪夫系数可以定义如下：

$$M \approx \sum_{k=0}^K \theta_k T_k(\tilde{\Lambda}) \quad (2-14)$$

式 (2-14) 中， $\tilde{\Lambda} = \frac{2}{\lambda_{\max}} \Lambda - I$  里的  $\lambda_{\max}$  表示拉普拉斯图的最大特征值。

另一方面，非谱方法直接定义卷积运算直接在图上，利用空间近邻。具体地，使用可学习的聚集函数迭代地聚集相邻节点的特征，其描述如下：

$$h_{\mathcal{N}(v)}^k \leftarrow \text{AGGREGATE}(h_u^{k-1}, \forall u \in \mathcal{N}(v)) \quad (2-15)$$

$$h_v^k \leftarrow \sigma(W^k \cdot \text{CONCAT}(h_u^{k-1}, h_{\mathcal{N}(v)}^k)) \quad (2-16)$$

式 (2-16) 中  $h_u^k$  表示节点  $u$  在第  $k$  次的迭代表示， $\text{AGGREGATE}$  是一个可学习的聚合函数。

## 2.3 股票网络的构建

### 2.3.1 基于元路径的关系信息提取

我们初步提取公开的 Wikidata 数据中企业关系。Wikidata 包含各种实体间（如国家，公司和个人）的关系，是一个具有不同类型的节点和边的异构图。而在本研究中，我们只对股票相关公司节点类型感兴趣，但是，这些公司之间往往存在几种类型的 edge，且它们之间的关系非常稀疏。针对这个问题，我们利用元路径去处理异构图，将复杂的异构图转换为只有股票（公司）节点的同构图。

我们将 Wikidata 表示为一个异构网络  $G = (V, E, T)$ , 其中每个节点  $v$  和每条链路  $e$  分布和它们的映射函数:  $\phi(v): V \rightarrow T_V$  和  $\phi(e): V \rightarrow T_E$  相关联。  $T_V$  和  $T_E$  表示节点对象和关系类型的集合。关系提取模块的任务是学习其中  $d$  维度潜在表征  $X \in \mathbb{R}^{|V| \times d}$  ( $d \ll |V|$ ), 捕捉目标节点之间的结构关系。然后, 我们展示了如何使用最大只有 2 跳的元路径引导来引导异构网络随机游走过程, 给定一个元路径方案  $\mathcal{P}: V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} V_3$ , 定义在第  $i$  步转移的概率如下:

$$\rho(v^{i+1}|v_t^i, \mathcal{P}) = \begin{cases} \frac{1}{|N_{t+1}(v_t^i)|} & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) = t + 1 \\ 0 & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) \neq t + 1 \\ 0 & (v^{i+1}, v_t^i) \notin E \end{cases} \quad (2-17)$$

式 (2-17) 中  $v_t^i \in V_t$  和  $N_{t+1}(v_t^i)$  表示节点  $v_t^i$  的领域的  $V_{t+1}$  类型。

### 2.3.2 基于社团检测的关系提取

股票  $s_q$  的收益率为  $R^{s_q} = (p_1^{s_q}, \dots, p_i^{s_q})$ , 对应的收益率区间可定义为  $[\min p_i^{s_q}, \max p_i^{s_q}]$ , 然后该区间被平均划分为  $m$  个子区间。计算股票  $s_q$  的收益率  $p_i^{s_q}$  落在第  $m$  个子区间的近似概率为:

$$p_m^{s_q} \approx \frac{f_m^{s_q}}{d} \quad (2-18)$$

式 (2-18) 中  $f_m^{s_q}$  为股票  $s_q$  落在第  $m$  个区间的频数,  $d$  为样本容量 (股票数量)。

因此股票  $s_q$  的收益熵为:

$$H(R^{s_q}) = -\sum_{m=1}^m p_m^{s_q} \log_2 p_m^{s_q} \quad (2-19)$$

定义股票  $s_s$  的收益率为  $R^{s_s}$ , 则其对应的收益率区间为  $[\min p_i^{s_s}, \max p_i^{s_s}]$ 。

然后将  $[\min p_i^{s_q}, \max p_i^{s_q}] \times [\min p_i^{s_s}, \max p_i^{s_s}]$  划分为  $M \times M$  个子区间。计算股票  $s_q$  和股票  $s_s$  的联合收益率  $(p_i^{s_q}, p_i^{s_s})$  落在子区间  $(k, l)$  的近似概率为:

$$p_{k,l}^{s_q, s_s} \approx \frac{f_{k,l}^{s_q, s_s}}{d} \quad (2-20)$$

于是联合收益率的联合熵为：

$$H(R^{s_q}, R^{s_s}) = - \sum_{k=1}^m \sum_{l=1}^m p_{k,l}^{s_q, s_s} \log_2 p_{k,l}^{s_q, s_s} \quad (2-21)$$

最后，得到股票  $s_q$  和股票  $s_s$  的相互信息表示：

$$I(R^{s_q}, R^{s_s}) = H(R^{s_q}) + H(R^{s_s}) - H(R^{s_q}, R^{s_s}) \quad (2-22)$$

为了方便比较互信息，一般会对互信息进行标准化处理，标准化互信息越大，表示股票间波动相关性越强。

$$NMI(R^{s_q}, R^{s_s}) = \frac{2I(R^{s_q}, R^{s_s})}{H(R^{s_q}) + H(R^{s_s})} \quad (2-23)$$

通过计算股票间互信息我们得到了一个全连接网络，此时网络中还存在冗余信息，因此本文选择阈值法[33]对网络进行初步过滤，筛选其中关键信息。通过调整阈值从而控制对网络信息的过滤。最后通过 Louvain 算法对股票网络进行社区检测，根据股票社区结构挖掘新的股票关系。

## 2.4 本章小结

本章概述了新闻文本表征技术、深度学习算法及股票网络构建方法。首先介绍了 Word2Vec 与 BERT 在文本表征中的应用。接着阐述了多层感知机、卷积神经网络、长短期记忆网络及图神经网络等深度学习算法。最后讲解了基于元路径的关系信息提取和基于社团检测的关系挖掘方法在股票网络构建中的应用。

本章为后续模型改进提供了理论基础。在下一章中，我们将根据本章相关背景知识，提出本文方法。

### 第 3 章 基于多级注意力机制的股票预测模型

在本章中，首先介绍提出方法的整体框架，然后详细阐述我们如何对金融市场中文本数据、数值数据和关系数据三个维度进行建模。针对新闻时讯等文本数据，应用 BERT 模型捕获金融市场文本特征。尽管目前已经有研究应用 BERT 去根据新闻预测股票趋势，但尚未探索 BERT 融合到基于图神经网络的方法。而传统基于图神经网络预测股票趋势，如 Feng 等人[1]忽略了 GNN 处理文本的局限性；本章通过自然语言处理先进技术对非结构化新闻提取特征进行优化，从而提高预测模型的预测能力。但这仍然不能捕捉股票内部之间关系，很难准确推断相关股票的走势。由此提出了基于多级注意力机制的预测模型（ML-GAT），这是一种新的基于图注意力网络的模块。原始数据经过处理后，通过不同的特征提取模块得到不同的特征向量，再聚合形成包含价格特征和文本特征的股票关系网络图。最后通过设计的多级别注意力机制，它可以对不同关系类型赋予不同的权重，并选择性地将特征提取模块中捕获到的节点表示聚合到股票图网络中。在对框架各部分描述之后，本章将详细说明提出的新型关系建模模块结构。

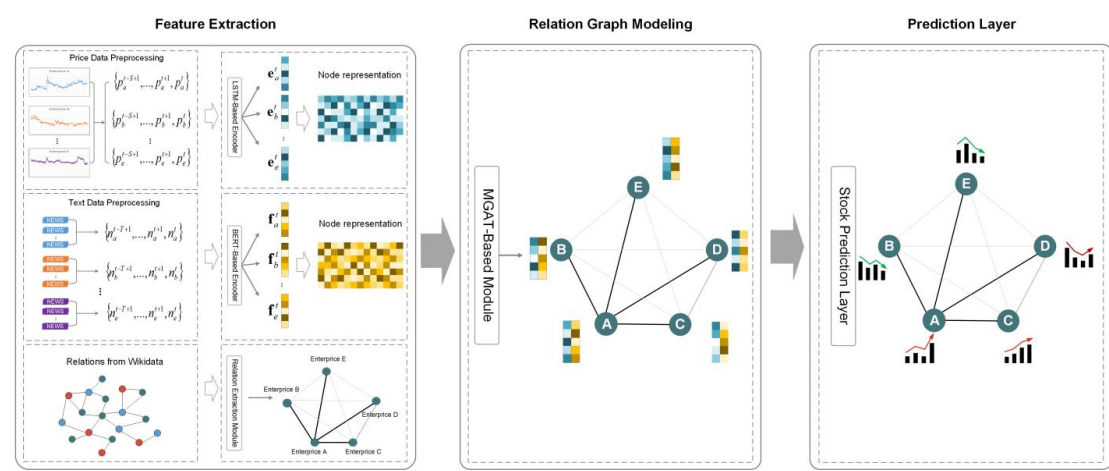


图 3-1 整体框架图

#### 3.1 问题描述

股票预测是金融科技交叉领域的热门研究问题，受到相关研究人员和投资者的广泛关注。如前面第一章国内外研究现状内容所介绍，之前的研究大多数侧重

于单一信息源作为股票特征，忽略了市场内其他信息。虽然目前已经有少量研究开始通过整合多源信息去预测股票未来走势，但是仍然没有考虑金融市场的内在联系对股价走势的影响。本文在研究如何高效挖掘并融合股票市场的多源数据基础上，通过学习股票间相关性以挖掘股票市场内外部的深度特征进行预测。本文的方法是基于以下两个方面的问题提出。

（1）如何高效挖掘金融市场多种来源信息特征？准确预测股票趋势要求输入大量特征到模型训练，而金融市场的信息繁杂而多样，其来源涵盖了多个方面，如股票价格历史数据、金融新闻、社交媒体等，但并不是都是有效的，比如会有重复和虚假的新闻数据。因此，需要不断地更新和优化挖掘方法，以寻找更有效的信息特征并提高预测模型的准确性。

（2）如何构建更有效的股票关系网络？构建有效的股票关系网络是融合股票关系预测趋势的重要一环。股票关系网络本质上是一个连接各股票之间的关系网，不同股票之间的关系可以是同板块、同产业、同消息面等方面的联系。通过建立股票关系网络，有助于模型更好地了解不同股票之间的关联关系，进而有益于预测股票走势。在构建股票关系网络的过程中，需要注意数据的准确性和可靠性，避免错误或虚假的数据污染关系网络。同时，我们也需要考虑如何选取合适的网络连接节点和网络层级，以便更好地把握各个节点间的关系。通过持续优化和完善股票关系网络，我们可以更好地预测股票趋势，并做出更精准的股票交易决策。

（3）如何权衡各种特征对股价波动不同的影响？不同信息特征和不同类别关系对股票趋势具有不同的重要性，需要为不同的特征分配不同的权重。因此，我们需要不断优化和完善模型。

## 3.2 特征提取

在对股票市场图建模过程中，我们将每个公司的股票视为一个节点，每个节点的特征代表着每个公司在股票价格变动下的目前状态。随着时间的推移，节点特征也会因市场行情变化而更新。不同类型的数据会对股票波动产生不同的影响，本文使用历史价格和金融新闻等文本数据作为股价变动的指标，将 Wikidata 中公司关系数据作为股票间关系指标。

### 3.2.1 基于 LSTM 的历史价格编码器

在金融市场中,股票历史价格数据具有短期与长期存在依赖关系。时间跨度不同,股价表现形式也不同。因此我们引入了基于 LSTM 的编码模型,它能够充分捕捉长期依赖关系。股票历史价格有许多不同的原始特征,如开盘价和收盘价等,被输入到特征提取模块。在本文中,对股票历史每天收盘价进行预处理,计算得到的价格变化率作为 LSTM 的输入,价格变化率的计算公式为:

$$r_{i,j}^{s_q} = \frac{p_{i,j}^{s_q} - p_{i,j-1}^{s_q}}{p_{i,j-1}^{s_q}}。我们将股票 i 在时间步长 t 的历史价格变化率表示为  $R_i^t =$$$

$\{r_i^{t-S+1}, ..., r_i^{t+1}, r_i^t\} \in \mathbb{R}^{S \times D}$  (S 表示序列长度, D 表示每个时间步的特征维度), 将时间序列数据输入 LSTM 网络, 输出的最后一个隐藏状态( $h_i^t$ )作为后面网络的顺序嵌入( $e_i^t$ )( $h_i^t = e_i^t$ ),即我们有:

$$i_i^t, f_i^t, o_i^t = f_{\theta^i}, f_{\theta^f}, f_{\theta^o}(W[h_i^{t-1}; R_i^t] + b_{(i,f,o)}) \quad (3-1)$$

$$c_i^t = i_i^t \odot u_i^t + f_i^t \odot c_i^{t-1} \quad (3-2)$$

$$u_i^t = \tan(W^{(u)}R_i^t + U^{(u)}h_i^{t-1} + b^{(u)}) \quad (3-3)$$

$$h_i^t = o_i^t \odot \tanh(c_i^t) \quad (3-4)$$

简单来说,我们可以得到:

$$E^t = \text{LSTM}(\mathcal{R}^t)(\mathcal{R}^t = [R_1^t, ..., R_i^t, ..., R_M^t] \in \mathbb{R}^{M \times S \times D}) \quad (3-5)$$

式(3-5)中,  $E^t = [e_1^t, ..., e_i^t, ..., e_M^t] \in \mathbb{R}^{M \times U}$  表示 M 只股票的顺序嵌入, U 表示嵌入大小(LSTM 中隐藏单元的数量)。

### 3.2.2 基于 BERT 的金融新闻编码器

BERT 的输入嵌入由 token 嵌入,segment 嵌入和 position 嵌入组成。token 嵌入是将每个词转换为固定维度的向量。特别是,每个输入前面都有一个特殊标记“[CLS]”。segment 嵌入主要是区分两种不同的句子,输入中每个句子都有标记“[SEP]”。position 嵌入是通过训练获得的。

为了更详细地说明 BERT 在金融领域提取股票新闻文本数据特征,我们将股票 i 在时间段 t 中收集到的新闻,形成 T 份段落文本表示为  $N_i^t = \{n_i^{t-T+1}, ..., n_i^{t+1}, n_i^t\} \in \mathbb{R}^{T \times H}$  (T 表示文本段落数目, H 表示每个文本的特征维度), 将其输入到编码模型中预训练。BERT-Based encoder module 对输入句子作截断

或补齐预处理, 在输入的每段文本的句首增加[CLS]符号标识, 句尾增加[SEP]符号标识。股票  $i$  的相关新闻  $n_i^t$  会被转换为“[CLS]+ $n_i^t$ + [SEP]”, 其中[CLS]对应的输出向量可以代表该句子的语义特征信息。经过 BERT 模型多次深层次的特征学习, 输出分类结果前一层的[CLS]向量最能表示出文本的语义信息。对于新闻文本特征提取任务来说, 只需要输出分类结果的上一层提取出文本的特征, 即只提取句子前符号[CLS]对应 Transformer 的最后一层的特征向量  $\mathbf{f}_i^t$ , 即我们有:

$$\mathbf{F}^t = \text{BERT}(\mathcal{M}^t)(\mathcal{M}^t = [\mathbf{N}_1^t, \dots, \mathbf{N}_i^t, \dots, \mathbf{N}_M^t] \in \mathbb{R}^{M \times T \times H}) \quad (3-6)$$

式 (3-6) 中,  $\mathbf{F}^t = [\mathbf{f}_1^t, \dots, \mathbf{f}_i^t, \dots, \mathbf{f}_M^t] \in \mathbb{R}^{M \times V}$  表示  $M$  只股票的文本特征嵌入向量,  $V$  表示嵌入大小。

### 3.2.3 基于 meta-path 的关系编码器

在关系提取模块中, 我们初步提取公开的 Wikidata 数据中企业关系。Wikidata 包含各种实体间 (如国家, 公司和个人) 的关系, 是一个具有不同类型的节点和边的异构图。而在本研究中, 我们只对股票相关公司节点类型感兴趣, 但是, 这些公司之间往往存在几种类型的 edge, 且它们之间的关系非常稀疏。针对这个问题, 我们受<sup>[34]</sup>的启发, 利用元路径去处理异构图, 将复杂的异构图转换为只有股票 (公司) 节点的同构图。从 Wikidata 中提取的一阶和二阶公司关系的例子如图 3-2 所示。

我们将 Wikidata 表示为一个异构网络  $G = (V, E, T)$ , 其中每个节点  $v$  和每条链路  $e$  分布和它们的映射函数:  $\phi(v): V \rightarrow T_V$  和  $\phi(e): E \rightarrow T_E$  相关联。  $T_V$  和  $T_E$  表示节点对象和关系类型的集合。关系提取模块的任务是学习其中  $d$  维度潜在表征  $\mathbf{X} \in \mathbb{R}^{|V| \times d}$  ( $d \ll |V|$ ), 捕捉目标节点之间的结构关系。然后, 我们展示了如何使用最大只有 2 跳的元路径引导来引导异构网络随机游走过程, 给定一个元路径方案  $\mathcal{P}: V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} V_3$ , 定义在第  $i$  步转移的概率如下:

$$p(v^{i+1} | v_t^i, \mathcal{P}) = \begin{cases} \frac{1}{|N_{t+1}(v_t^i)|} & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) = t+1 \\ 0 & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) \neq t+1 \\ 0 & (v^{i+1}, v_t^i) \notin E \end{cases} \quad (3-7)$$

式 (3-7) 中  $v_t^i \in V_t$  和  $N_{t+1}(v_t^i)$  表示节点  $v_t^i$  的领域的  $V_{t+1}$  类型。



图 3-2 从 Wikidata 中提取的一阶和二阶公司关系的例子

3.3 多级图注意力机制

关系图建模模块的功能是更新上节关系提取模块转换的同构图中股票(公司)节点。图神经网络的主要功能是交换相邻节点间的信息，然后聚合相邻节点的信息，最后添加到每个节点表示中。而节点特征是图预测任务成功的关键，因此我们需要从收集到的不同节点的不同类型关系信息进行有效聚合。为此，本文提出一种新的基于 GNN 的多级图注意力网络的关系图建模方法（ML-GAT），如图 3-3 描述了 ML-GAT 框架的详细结构。通过增加多层不同级别注意力机制，为信息筛选分配不同的权重值，使我们的模型可以从不同的节点收集到不同类型关系的信息，过滤掉对趋势预测无效的信息。我们的方法对准确预测股票趋势很重要，因为股票间有很多不同类型的关系，其中有些信息与股票预测无关。

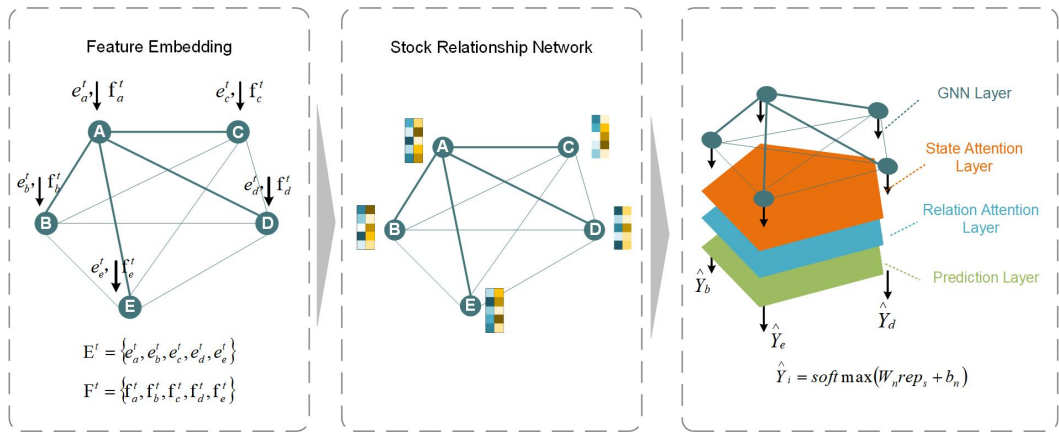


图 3-3 ML-GAT 框架的详细结构

通过股票数据特征提取模块，可得到股票  $i$  在时间  $t$  的价格嵌入向量  $e_i^t \in \mathbb{R}^h$  和新闻文本嵌入向量  $f_i^t \in \mathbb{R}^h$ 。为了简单化，在后文中统一省略上标  $t$ ，假设所有向量表示都是在同一时间  $t$  计算的。此外，因为模型是基于图神经网络实现的，



我们需知道在每种关系类型中目标节点  $i$  的相邻节点集。我们将节点  $i$  的关系类型  $m$  表示为  $r_m \in N_i^{r_m}$ , 关系类型  $r_m$  的嵌入向量表示为  $e_{r_m} \in \mathbb{R}^d$ 。我们的目标是选择性地从相邻节点收集关于不同关系的信息。

在第一级状态注意力层, ML-GAT 从一组相邻节点中选择同一类型关系的相关信息。注意力机制主要用于根据选择的关系类型  $r_m$  计算不同的权重。在计算状态注意力系数前, 需要将关系类型嵌入向量  $e_{r_m}$  和节点  $i, j$  的节点表示串联到一个向量中, 其中  $j \in N_i^{r_m}$ 。我们将这个连接向量表示为  $x_{ij}^{r_m} \in \mathbb{R}^{2h+d}$ 。计算目标节点  $i$  和目标节点  $j$  之间关系类型  $r_m$  注意力系数公式为:

$$\alpha_{ij}^{r_m} = \frac{\exp\left(\text{LeakyReLU}\left(\begin{smallmatrix} \rightarrow \\ a_{r_m} \end{smallmatrix}^T [W_s x_{ij}^{r_m} + b_s]\right)\right)}{\sum_{k \in N_i^{r_m}} \exp\left(\text{LeakyReLU}\left(\begin{smallmatrix} \rightarrow \\ a_{r_m} \end{smallmatrix}^T [W_s x_{ik}^{r_m} + b_s]\right)\right)} \quad (3-8)$$

式 (3-8) 中  $W_s \in \mathbb{R}^{2h+d}$  和  $b_s \in \mathbb{R}$  是可学习参数, 用于计算状态注意力系数的。

基于计算得到的状态注意力系数, 我们将公司关系  $r_m$  的输出特征计算为具有所有节点加权平均值, 公式为:

$$Z_i^{r_m} = \sigma\left(\sum_{j \in N_i^{r_m}} \alpha_{ij}^{r_m} W_s e_j\right) \quad (3-9)$$

通过以上公式计算, 就得到了股票间每种关系的向量表示, 其可以看作是关系的汇总信息。向量  $Z_i^{r_m}$  包括来自股票  $i$  关系类型  $r_m$  的汇总信息。例如, 子公司关系向量表示可以概括目标股票公司所有子公司情况。和人的投资决策行为一样, 我们的模型应该做到从所有汇总信息中优先考虑那些利于交易决策的重要信息。在 ML-GAT 的第二级注意力层中, 就是引用注意力机制去为众多信息分配重要性权重。

我们将关系的汇总信息向量  $Z_i^{r_m}$ , 股票公司的当前节点表示  $e_i$  和  $f_i$  以及关系类型的嵌入向量  $e_{r_m}$  连接起来, 并将连接后的向量表示为  $v_i^{r_m} \in \mathbb{R}^{2h+d}$ , 作为下一注意力层的特征输入。与前文公式类似, 在第二级注意力层中, 注意力系数由公式计算:

$$\alpha_i^{r_m} = \frac{\exp\left(\text{LeakyReLU}\left(\begin{smallmatrix} \rightarrow \\ a_r \end{smallmatrix}^T [W_r v_i^{r_m} + b_r]\right)\right)}{\sum_{k \in \Phi} \exp\left(\text{LeakyReLU}\left(\begin{smallmatrix} \rightarrow \\ a_r \end{smallmatrix}^T [W_r v_i^{r_k} + b_r]\right)\right)} \quad (3-10)$$

式(3-10)中 $W_r \in \mathbb{R}^{2h+d}$ 和 $b_r \in \mathbb{R}$ 是可学习参数,每种类型的加权向量相加形成一个聚合关系表示,将聚合关系计算为具有 sigmoid 函数的源隐藏特征的加权平均值,公式为:

$$e_i^r = \sigma \left( \sum_{k \in \Phi} \alpha_i^{(r_k)} W_r Z_i^{r_k} \right) \quad (3-11)$$

最后,添加节点其节点表示,得到目标节点  $i$  的表征:

$$rep_s = e_i^r + e_i + f_i \quad (3-12)$$

### 3.4 分类预测及模型训练

鉴于从 3.2.2 节中学习到的目标股票节点表示,对于目标股票价格预测任务,我们使用浅层神经网络实现。我们将预测任务表现为一个分类问题,也就是把价格未来走势分为三类{up,neutral,down},在后面对此任务设置进行了详细说明。预测网络是一个简单的线性变换层,被定义为:

$$\hat{Y}_i = \text{softmax} (W_n rep_s + b_n) \quad (3-13)$$

式(3-13)中 $W_n \in \mathbb{R}^{d \times l}$ 和 $b_n \in \mathbb{R}^l$ 分别是权重矩阵和方差。 $l$ 是目标预测类别的数目,在本文中, $l=3$ 。

最后,我们使用输出层中似然的交叉熵损失在全部股票关系数据上训练模型:

$$Loss_{\text{node}} = - \sum_{i \in Z_n} \sum_{c=1}^l Y_{ic} \ln \hat{Y}_{ic} \quad (3-14)$$

式(3-14)中 $Y_{ic}$ 为股票  $i$  在 $c_{th}$  时间的真实值标签, $Z_u$ 表示数据集中所有股票集合。

### 3.5 实验设计和结果分析

在本节中,我们进行了广泛的实验来研究提出方法的有效性和局限性,并且将我们的方法和近几年金融领域流行的几种方法进行了比较。

#### 3.5.1 数据集

##### (1) 股票历史价格数据集

本文的研究集中于世界上发达的股票市场,分别选取了 2 个真实世界的数据集: NASDAQ[3]和 S&P 500[4],下表描述了这些数据集的详细信息。两个数据集来自于美国股票市场,包含了 2013/02/08 到 2018/03/27 期间的每日股票交易

数据。股票历史价格数据源于雅虎财经网站(<https://finance.yahoo.com/>)。股票关系数据来自 Wikidata，而 Wikidata 并不包含指数中所有股票，因此剔除一些在 Wikidata 中与其他公司（股票）不存在关系的股票。例如，对于 S&P 500 指数，去掉无关股票后，我们将剩下的 423 家公司的股票作为目标股票。本文的历史价格数据集详细描述如表 1。很多研究工作都是直接将历史价格的原始特征，如开盘价和收盘价等直接馈送到特定的特征提取模块，而我们使用历史价格变化率作为 LSTM 的输入。股票在时间  $t$  的价格变化率可以通过： $R_i^t = \frac{P_i^t - P_i^{t-1}}{P_i^{t-1}}$  计算，其中  $P_i^t$  和  $P_i^{t-1}$  分别为股票  $i$  在时间  $t$  和  $t-1$  时的收盘价。

表 3-1 历史价格数据统计

Markets	S&P 500	NASDAQ
Stocks	423	1026
Training Period	08/02/2013-23/05/2017 1080days	08/02/2013-23/05/2017 1080days
Validation Period	24/05/2017-27/03/2018 213days	24/05/2017-27/03/2018 213days
Testing Period	27/03/2018-29/08/2019 316days	27/03/2018-29/08/2019 316days

(2) 股票金融新闻数据集

对于金融新闻，我们选择和历史价格数据集的时间范围保持一致，在相同时间内从 Benzinga.com 上获取目标股票相关的新闻（Benzinga 是一家金融数据公司，每天会发布 50-60 篇文章）。最终我们的金融新闻数据集由 130 万余条新闻标题组成。



#	title	date	stock
Index key	Article headline	Release timestamp in UTC-4 timezone	Stock ticker (NYSE/NASDAQ/AMEX only)
			
0	843062 unique values	15Feb09 12Jun20	6193 unique values
0	Stocks That Hit 52-Week Highs On Friday	2020-06-05 10:30:00-04:00	A
1	Stocks That Hit 52-Week Highs On Wednesday	2020-06-03 10:45:00-04:00	A
2	71 Biggest Movers From Friday	2020-05-26 04:30:00-04:00	A
3	46 Stocks Moving In Friday's Mid-Day Session	2020-05-22 12:45:00-04:00	A
4	B of A Securities Maintains Neutral on Agilent Technologies, Raises Price Target to \$88	2020-05-22 11:30:00-04:00	A

图 3-4 金融新闻数据示例

### (3) 股票关系数据集

Wikidata 作为实体关系的丰富来源,它包含了大量的企业实体和企业关系,这可能对股票行情产生影响,因此我们第三种数据是来自 Wikidata 的关系数据。Wikidata 是一个免费的协作知识库,目前拥有 4800 万个项目(例如 Google Inc 等)和数亿个句子(如 Apple, founded by ,Steve Jobs)。受[28]工作的启发,我们通过最大只有 2 跳的元路径从[35]中分别提取了 9 种一阶关系和 62 种二阶关系,关于获取到的关系详细信息在文章后面附录部分。

#### 3.5.2 实验评估方法

为了比较提出方法和基准模型的性能,我们从模型分类和盈利能力分别选择常用指标去评估。股票趋势预测是一种典型的分类预测任务,因此我们首先选择了 2 个广泛使用于分类任务的评估指标,准确率和 F1 分数,计算公式如下:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3-15)$$

$$F1 = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (3-16)$$

式(3-16)中  $Recall = \frac{TP}{TP+FP}$ ;  $Precision = \frac{TP}{TP+FN}$ , 并且 TP、TN、FP 和 FN 分别表示真阳性,真阴性,假阳性和假阴性。

通过平均每个类别计算得到的 F1 分数,我们就可以得到宏观的 F1 值。

针对提出方法盈利能力的评估,我们使用以下两个指标来比较各个方法的盈利能力。

计算投资组合的回报公式为:

$$Return_i^t = \frac{1}{|F^{t-1}|} \sum_{i \in F^{t-1}} \frac{p_i^t - p_i^{t-1}}{p_i^{t-1}} \quad (3-17)$$

式(3-17)中  $F^{t-1}$  表示在时间 t-1 包含在投资组合中的一组股票,  $p_i^t$  表示股票 i 在时间 t 的价格,  $| \cdot |$  表示设定组合的项目数量。

夏普利率是一个同时对收益和风险综合考虑的指标,可以用来衡量投资风险与收益相比的表现。计算公式如下:

$$Sharpe_a = \frac{E[R_a - R_f]}{\sigma_p} \quad (3-18)$$

式(3-18)中  $R_a$  表示投资组合收益率,  $R_f$  表示无风险利率,  $\sigma_p$  是投资组合报酬率的标准差。

3.5.3 实验环境和参数设置

本仿真实验是使用 Python 编程语言，实现了基于 ML-GAT 的算法模型，本章实验涉及的软硬件环境如表 3-2 所示。

表 3-2 实验环境

名称	版本
操作系统	Windows 10
Python	3.9
Tensorflow	2.7.0
GPU	NVIDIA Tesla K80
编程工具	PyCharm
内存	16G
BERT 预训练模型	Uncased L-12 H-768 A-12

模型的实验参数和其他基准模型的实验参数设置如表 3-3 所示。

3.5.4 实验结果与对比分析

(1) 不同关系数据的效果分析

在股市趋势预测任务前，研究了不同类型的关系数据对股票预测的影响。本文使用了有效处理图结构数据的基础 GCN 模型，但其无法区分不同的关系类型，目的就是公平的衡量不同类型关系的不同影响程度，从而对现有的关系类型进行优化。我们使用具有两个卷积层和一个预测层的 GCN，其正向传播的定义如下：

$$Y^{GCN} = \text{softmax} \left( \hat{A} \text{ReLU} \left( \hat{A} \text{ReLU} \left( \hat{A} X' W^{(0)} \right) W^{(1)} \right) W^{(2)} \right) \quad (3-19)$$

式 (3-19) 中  $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ ，且  $\tilde{A} = A + I$  是邻接矩阵。改变关系类型会改变 GCN 输入的邻接矩阵。因此，这里将每种关系类型单独输入到 GCN 作为其邻接矩阵，最后在表 3-4 中列出了测试集上的 10 个最佳关系类型和 10 个最差关系类型及其 F1 分数。

从实验中发现，使用关系数据并不总能带来良好的效果，在最坏的情况下，关系数据的引入会显著降低股市预测的性能。由表可得，最佳关系性能比最差关系性能高 19.11%。因此，本章利用实验中获取的 10 种最佳关系为股票建立关系图进行预测。

表 3-3 实验参数设置

Methods	Parameters and Description	Methods	Parameters and Description
MLP	Hidden layers:128,64; Optimizer:Adam Learning rate:0.0001 Epochs:100	TGC	Hidden: LSTM layers:64,64 MLP layer:1 Optimizer:Adam Learning rate:0.001 Epochs:100
	Convolutional layer1:16filters 2*2 Convolutional layer2:32filters 2*2 Convolutional layer3:64filters 2*2 Max pooling layer:2*2 Fully connected layer: 500 Optimizer:rmsprop Learning rate:0.01 Epochs:100		Convolutional layer1:64filters 2*2 Convolutional layer2:64filters 2*2 Optimizer:Adam Learning rate:0.01 Epochs:100
LSTM	LSTM layers:60,50,50,50 Dropout layer:0.2 Optimizer:rmsprop Learning rate:0.01 Epochs:100	ML-GAT	The length of time series:50 Hidden: MLP layer:2 LSTM layer:128 Optimizer:Adam Learning rate:5e-4 Dropout layer:0.5 Activation function:Leaky_ReLU BERT: BATCH_SIZE=32 MAX_token_LENGTH=128 Epochs:100

表 3-4 不同关系的结果

Relation Type	F1-score
Industry-Legal form	0.4576
Parent organization-Owner of	0.4551
Industry-Product or material produced	0.455
Owned by-Subsidiary	0.4547
Founded by-Founded by	0.4543
Follows	0.4535
Parent organization	0.4521
Complies with-Complies with	0.4502
Subsidiary-Owner of	0.4491
Owner of-Parent organization	0.4484
...	
Worst 10	0.3112
Instance of-Legal form	0.3084
Location of formation-Country	0.3075
Stock Exchange	0.3053
Country-Location of formation	0.2952
Country of origin-Country	0.2948
Country-Board member	0.2886
Country-Country of origin	0.2851
Instance of-Instance of	0.2748
Stock Exchange-Stock Exchange	0.2665

## (2) 不同模型结果的对比分析

表3-5总结了不同指数股票基于不同方法在股票趋势预测实验中的分类精度结果。

表 3-5 不同模型的分类准确度

	F1-score					
	MLP	CNN	LSTM	GCN	TGC	ML-GAT
1	0.2462	0.2986	0.2873	0.2989	0.3337	<b>0.3983</b>
2	0.2331	0.2977	0.2821	0.3307	0.3528	<b>0.4224</b>
3	0.2728	0.2861	0.3155	0.3142	0.3476	<b>0.4035</b>
4	0.2661	0.3058	0.2958	0.3192	0.3352	<b>0.4183</b>
5	0.2749	0.2895	0.3129	0.3335	0.3449	<b>0.398</b>
6	0.2776	0.2984	0.2862	0.3011	0.3311	<b>0.4221</b>
7	0.2302	0.2892	0.3104	0.3308	0.3574	<b>0.4078</b>
8	0.2494	0.2733	0.3138	0.3317	0.3318	<b>0.4386</b>
9	0.2653	0.3012	0.312	0.2953	0.3289	<b>0.3941</b>
10	0.2525	0.2991	0.2882	0.2903	0.3484	<b>0.3985</b>
Average	0.25681	0.29389	0.30042	0.31457	0.34118	<b>0.41016</b>
	Accuracy					
	MLP	CNN	LSTM	GCN	TGC	ML-GAT
1	0.3267	0.2912	0.3467	0.3345	0.3779	<b>0.4172</b>
2	0.2711	0.3423	0.3365	0.3515	0.3674	<b>0.3972</b>
3	0.2759	0.3601	0.3212	0.3519	0.3443	<b>0.4094</b>
4	0.3017	0.3497	0.3498	0.3564	0.3498	<b>0.4036</b>
5	0.3181	0.3136	0.3317	0.3481	0.3569	<b>0.4184</b>
6	0.2875	0.3543	0.3616	0.3384	0.3473	<b>0.4069</b>
7	0.3095	0.2885	0.3412	0.3356	0.3694	<b>0.4174</b>
8	0.3073	0.3429	0.3345	0.3332	0.3665	<b>0.4193</b>
9	0.3039	0.2827	0.3574	0.3238	0.3769	<b>0.3983</b>
10	0.3199	0.3528	0.3474	0.348	0.3693	<b>0.3977</b>
Average	0.30216	0.32781	0.3428	0.34214	0.36257	<b>0.40854</b>



从结果表 3-5 中,我们可以看出,在没有考虑对股票间关系建模的 3 个基准模型中, LSTM 的 F1-score 和 accuracy 都要比其他模型表现的更好。因此,在和无关系建模模块模型比较时,只需将我们模型的结果和 LSTM 模型的结果进行比较。在 F1-score 方面,所有具有关系建模模块的模型比 LSTM 表现的更好。然而并不是所有考虑对股票间关系建模的模型的准确性都比 LSTM 的准确性更高, GCN 在 accuracy 方面就比 LSTM 表现略差。特别注意的是,模型 ML-GAT 在重复进行的 10 次实验中,实验结果表现通常都显著优于其他模型。

我们根据上节的交易策略计算投资组合的每日收益,不同模型的盈利能力结果汇总如表 3-6。平均情况来看, ML-GAT 和 TGC 获得了比较高的日均回报,其中 ML-GAT 日均回报的表现最具有竞争力,远大于 TGC 和其他基准模型。这里比较特殊的是, GCN 在 F1-score 上要优于没有关系建模模块的模型,但 GCN 的夏普利率远低于 LSTM 等无关系建模模块基准模型,但是 TGC 的夏普利率高于除了 ML-GAT 之外的基准模型。从多次实验结果方差来看,几组基准模型的夏普利率方差都远远大于我们的模型,进一步验证我们模型在盈利能力测试中具有更稳定的特性。总之, ML-GAT 模型在预期日均收益和夏普利率两个指标上都取得了显著效果。

表 3-6 不同模型的盈利能力对比

	Average Daily Return					
	MLP	CNN	LSTM	GCN	TGC	ML-GAT
1	-0.0279	0.0804	0.0756	0.0744	<b>0.1532</b>	0.1152
2	0.0161	0.0851	0.0557	-0.0557	<b>0.1457</b>	0.1114
3	-0.0605	-0.0627	-0.1037	0.0907	0.1588	<b>0.2130</b>
4	-0.0538	<b>0.0784</b>	0.0580	0.0531	-0.1295	0.0268
5	-0.0025	0.0408	0.0883	0.0852	<b>0.1420</b>	0.1155
6	0.0630	-0.0212	0.0731	0.0450	<b>0.0923</b>	0.0382
7	-0.0501	0.0493	-0.0914	0.0861	0.0774	<b>0.1051</b>
8	<b>0.0862</b>	-0.0540	0.0445	-0.1064	-0.0587	0.0242
9	-0.0111	0.0479	-0.0296	0.1165	0.0427	<b>0.3991</b>
10	0.0344	0.0722	<b>0.1171</b>	0.0973	0.0948	0.0440
Average	-0.0006	0.0316	0.0288	0.0486	0.0719	<b>0.1193</b>

	Sharpe Ratio					
	MLP	CNN	LSTM	GCN	TGC	ML-GAT
1	1.6805	-1.5475	1.4447	<b>2.3555</b>	0.4076	1.3997
2	-0.2416	0.3258	-2.0754	1.0749	<b>1.8283</b>	1.4691
3	-1.7833	1.2976	<b>1.9577</b>	-0.2295	0.7839	1.2263
4	1.3725	-0.6959	1.0672	-0.2419	-2.0261	<b>2.4552</b>
5	-0.4729	<b>2.1807</b>	1.8222	0.4541	1.3664	2.0245
6	-1.8405	0.3249	1.4248	-0.9291	<b>1.5023</b>	1.1347
7	1.9891	-1.6977	0.4555	-1.2779	0.2599	<b>2.2177</b>
8	-0.1580	0.9360	0.2072	-2.0270	0.7183	<b>1.9186</b>
9	0.0119	0.8290	-1.4811	0.7792	0.6135	<b>3.0144</b>
10	1.5192	-0.6809	1.0610	1.1681	1.0720	<b>2.0282</b>
Average	0.2077	0.1272	0.5884	0.1126	0.6526	<b>1.8889</b>

此外，我们假设资产价值从 1000 开始，不同指数股票基于不同预测模型在模拟交易中资产价值变化如图 3-5 所示。在 2018 年底到 2019 年初，我们的模型 ML-GAT 预测的资产价值出现了大跌的情况，而其他方法的资产价值似乎没有发生很大变化。事实上，2018 年末美股主要因土耳其金融危机发酵和特朗普大力度对华制裁政策而集体大跌；加上 2019 年新冠疫情的爆发，美股也受到很大的影响。这也验证了我们方法的优越性，通过 BERT 捕获金融领域的时事新闻的特征，在经历两次股市突如其来的黑天鹅事件，通过我们的方法和交易策略起到了止损的重要作用，并获得了较好的收益。

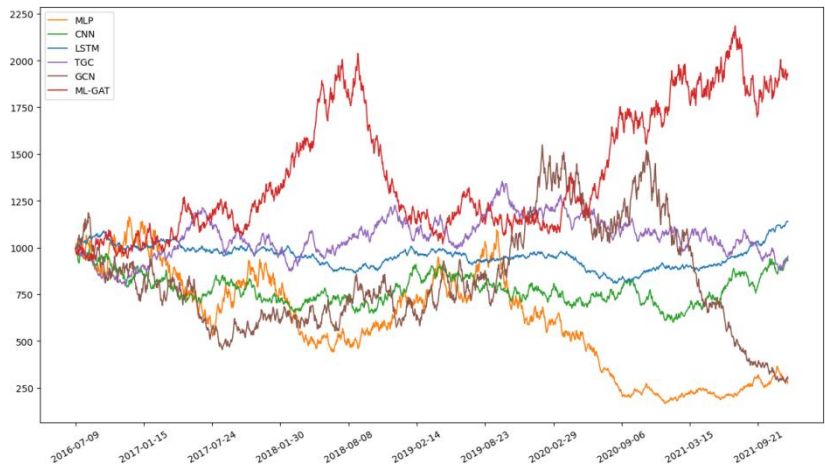


图 3-5 不同预测模型及其资产价值变化的对比。假设资产价值从 1000 开始。

最后，我们对以上实验结果和观察结果进行分析，得出以下几点结论：

1)通过对比有关系建模模块和无关系建模模块模型来看，股票间关系的引入通常可以对趋势预测任务产生积极的影响。基于模型实验结果表现，我们提出的方法显著优于其他基线。

2)提高模型预测的准确性，不仅仅要考虑对股票关系建模，还要有选择性地聚合不同类型关系信息。我们在实验过程中发现，GCN 在对关系建模时，公平地考虑了所有关系信息，所以导致 GCN 获得更高的 F1-score，但准确性比 LSTM 更低。而 ML-GAT 通过多层注意力机制根据节点的当前状态给不同关系类型分配不同权重，因此取得更好的效果。

3)通过从金融新闻中用 BERT 提取的向量表示，然后嵌入 ML-GAT 中进行预测。我们的模型可以从新闻事件中找出股票走势，进一步验证了金融新闻在股票趋势预测中的良好的表现，具体见下文。

### 3.6 本章小结

在本章研究中，提出了用于股票趋势预测的多层图形注意力神经网络模型。研究将金融市场，新闻时讯和公司关系等数据通过特定的特征提取模块纳入基于图神经注意力网络的模型中。ML-GAT 旨在通过多层不同级别的注意力机制，选择性地筛选不同类型的信息形成聚合图，以学习对预测任务有用的节点的特征表示，希望通过基于图的学习使预测精度达到更高。我们将 ML-GAT 与常见基准模型在公开数据集进行实验对比，评估了本文提出方法的有效性。结果还证明了使用金融新闻和关系数据的重要性，并表明了关系数据不同的聚合方法会带来不同的预测精度。

## 第 4 章 双图注意力和多源特征融合的股票预测模型

目前市场信息与股票关系建模的方法大多使用预定义行业关系构建股票关系图，例如前面第三章是基于 Wikidata（预定义关系）去研究股票间的关系。但是，人为预定义的行业关系不是充分的，存在局限性，不可避免的会忽略股票间潜在的相互作用。为此，在前章的研究内容基础之上，上本文设计新的多源信息融合的双图注意力模型（MF-DAT），弥补前文预定义行业关系表示股票间关系的局限性。MF-DAT 模型主要包括三个步骤：首先通过 LMF 模块融合提取到股票的不同特征，捕获不同市场信息间的相关性；其次学习股票的加权长短期状态特征，通过引入注意力机制使我们模型能够选择性地叠加不同时刻的特征信息，避免学习数据中过多的冗余特征而导致的不良效果；最后，通过图注意力层对股票集群间关系特征进行学习，可以对不同集群间关系赋予不同的权重，并选择性地前面捕获到的节点表示聚合到股票图网络中。我们的模型同时考虑了股票时间序列信息和股票潜在关系。

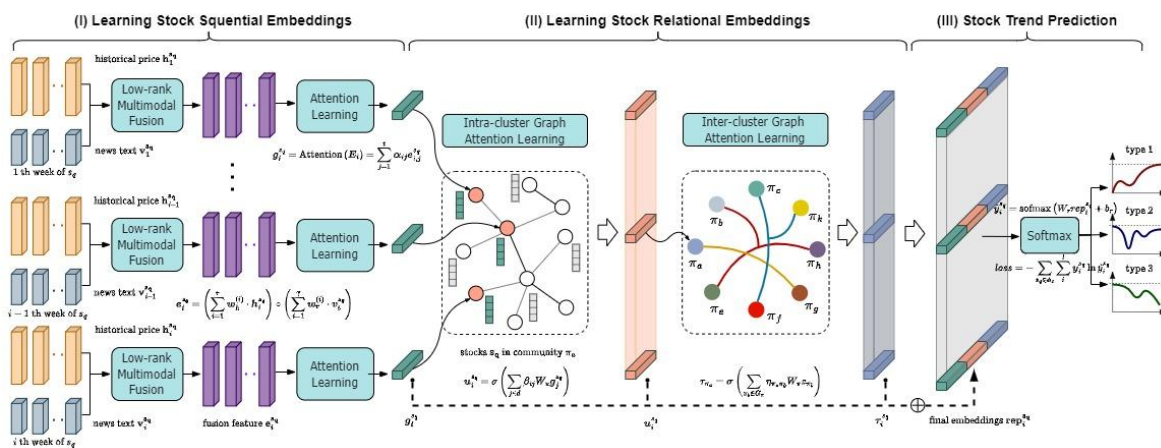


图 4-1 MF-DAT 模型框架图

### 4.1 问题描述

在前章的研究中，采用基于 ML-GAT 模型的方法有效解决了目前研究中存在的几个问题：一是如何通过适合不同数据特征的提取模型实现高效挖掘金融市场多种来源信息，二是如何基于 Wikidata 构建有效的股票关系网络，三是如何通过多级注意力机制权衡各种特征对股价波动不同的影响。但在后面的实际研究

过程中发现, Wikidata 库里的实体关系存在一定局限性, 并且多种来源信息特征之间存在一定交互特征。于是在前章研究内容基础上出现了新的问题:

(1) 如何高效融合多种模态特征向量? 股票市场具有复杂而又多变的特点, 其价格受到多种因素的影响, 比如公司基本面、金融新闻、政策等等。因此, 在进行股票的预测时, 单一的信息源难以提供足够的准确性。而多模态信息的融合可以最大限度地利用不同信息源之间的交互特征, 进而提高预测准确性, 为投资者提供更为科学、精准的决策支持。同时, 高效地融合不同模态特征向量也可以大大加快预测的速度和效率, 提高预测的实时性和可靠性, 对投资者进行交易决策和风险管理提供更为可靠的依据。

(2) 如何突破预定义行业关系对股票关系建模? 前章的股票关系建模采用固定的预定义关系来描述股票之间的关系网络, 这种做法的局限性在于预定义的关系是有限的, 无法对金融市场中复杂而多变的关系进行完整的建模。尤其是在股票市场波动性较大、非线性关系较多的情况下, 预定义关系的建模方法往往难以达到良好的预测效果。因此, 突破预定义关系对股票关系建模成为了必要的趋势。

## 4.2 特征提取

### 4.2.1 股票特征

本章研究内容是基于第三章扩展研究内容, 因此股票的历史价格数据特征和金融新闻特征的提取方法和第三章一样, 具体可见第三章。

### 4.2.2 关系特征

股票  $s_q$  的收益率为  $R^{s_q} = (p_1^{s_q}, \dots, p_i^{s_q})$ , 对应的收益率区间可定义为  $[\min p_i^{s_q}, \max p_i^{s_q}]$ , 然后该区间被平均划分为  $m$  个子区间。计算股票  $s_q$  的收益率  $p_i^{s_q}$  落在第  $m$  个子区间的近似概率为:

$$p_m^{s_q} \approx \frac{f_m^{s_q}}{d} \quad (4-1)$$

式 (4-1) 中  $f_m^{s_q}$  为股票  $s_q$  落在第  $m$  个区间的频数,  $d$  为样本容量 (股票数量)。

因此股票  $s_q$  的收益熵为:

$$H(R^{s_q}) = -\sum_{m=1}^m p_m^{s_q} \log_2 p_m^{s_q} \quad (4-2)$$

定义股票  $s_s$  的收益率为  $R^{s_s}$ , 则其对应的收益率区间为  $[\min p_i^{s_s}, \max p_i^{s_s}]$ 。

然后将  $[\min p_i^{s_q}, \max p_i^{s_q}] \times [\min p_i^{s_s}, \max p_i^{s_s}]$  划分为  $M \times M$  个子区间。计算股票

$s_q$  和股票  $s_s$  的联合收益率  $(p_i^{s_q}, p_i^{s_s})$  落在子区间  $(k,l)$  的近似概率为:

$$p_{k,l}^{s_q,s_s} \approx \frac{f_{k,l}^{s_q,s_s}}{d} \quad (4-3)$$

于是联合收益率的联合熵为:

$$H(R^{s_q}, R^{s_s}) = -\sum_{k=1}^m \sum_{l=1}^m p_{k,l}^{s_q,s_s} \log_2 p_{k,l}^{s_q,s_s} \quad (4-4)$$

最后, 得到股票  $s_q$  和股票  $s_s$  的相互信息表示:

$$I(R^{s_q}, R^{s_s}) = H(R^{s_q}) + H(R^{s_s}) - H(R^{s_q}, R^{s_s}) \quad (4-5)$$

为了方便比较互信息, 一般会对互信息进行标准化处理, 标准化互信息越大, 表示股票间波动相关性越强。

$$NMI(R^{s_q}, R^{s_s}) = \frac{2I(R^{s_q}, R^{s_s})}{H(R^{s_q}) + H(R^{s_s})} \quad (4-6)$$

通过计算股票间互信息我们得到了一个全连接网络, 此时网络中还存在冗余信息, 因此本文选择阈值法[33]对网络进行初步过滤, 筛选其中关键信息。阈值法是一种常见的对股票网络信息进行过滤的方法, 主要基于阈值筛选的思想。通过设置一个预先定义的阈值, 只有当股票之间的关联度超过这个阈值时, 这些关联才会被认定为有意义, 并被加入到股票网络中。对于如何设置阈值, 一般需要根据具体的预测目标和股票市场情况来进行调整, 同时也需要考虑到投资者对风险的容忍度和收益目标。具体而言, 阈值的设置通常可以基于相关系数、方差、互信息、协方差等指标进行计算和比较, 以确定是否满足设定的阈值条件。

同时还需要注意到, 设置阈值时不宜过高或过低。因为过高的阈值会使得很多关联被排除, 导致网络出现不完整的情况, 从而影响预测精度。而过低的阈值则会使得网络中包含大量的冗余信息, 从而加大了预测的计算开销和复杂度, 同时也容易受到数据噪声和干扰的影响。因此, 需要考虑多方面的因素来决定阈值,

以实现更为准确的股票趋势预测。

通过调整阈值从而控制对网络信息的过滤后,最后通过 Louvain 算法对股票网络进行社区检测,根据股票社区结构挖掘新的股票关系。Louvain 算法是一种基于模块化的社区检测算法,可以对股票网络中的股票进行社区划分,即将股票按照其相似性聚集到不同的社区中。通过社区检测,我们可以发现股票与股票之间的潜在关系,并根据股票所属的社区结构挖掘新的股票关系。

Louvain 算法的核心思想是将网络划分为多个社区,每个社区中的股票之间具有较强的相似性,而不同社区之间的相似性较小。具体而言,算法分两个阶段。在第一阶段中,将每个节点看作一个单独的社区,然后通过社区合并的方式来最大化整个网络的模块度(即社区内部的紧密度与社区之间的稀疏度之间的平衡),从而实现网络划分。在第二阶段中,将第一阶段得到的社区看做节点,再次进行社区合并,以获得更加合适的社区划分结果。

通过 Louvain 算法对股票网络进行社区检测,我们可以发现潜在的股票之间的关系,包括股票之间的相似性、信息共享等。同时,我们也可以根据股票所属的社区结构来发现新的股票关系,例如同一社区内的股票之间具有较高的联系和相关性,可以提供更加有价值的股票投资策略。此外,对于股票组合的构建和会补偿策略的制定,股票社区结构也可以提供重要的参考信息。

### 4.3 双图注意力机制

#### 4.3.1 多模态特征融合层

股票趋势受到多种模态信息的影响,不只是股票预测领域,包括信息安全领域[42]基于多特征融合的方法具有更高的匹配精度,因此如何高效融合多种模态特征向量十分重要的。传统的融合方式将特征向量直接拼接会忽略模态间的相关性,而[9]和[34]中运用的 TFN 会大大的增加特征维度。为了更高效从历史价格和文本数据中提取高级特征,同时保留模态间的相关性,本文采用用 LMF 模块[35]通过利用低秩权重张量和输入张量的并行分解来计算基于张量的融合。股票的融合特征向量的计算公式如下:

$$e_i^{s_q} = \left( \sum_{i=1}^r w_h^{(i)} \otimes w_v^{(i)} \right) Z = \left( \sum_{i=1}^r w_h^{(i)} \cdot h_i^{s_q} \right) \circ \left( \sum_{i=1}^r w_v^{(i)} \cdot v_i^{s_q} \right) \quad (4-7)$$

式(4-7)中  $Z$  为输入特征张量,  $r$  为股票数量。  $w_h^{(i)} \in \mathbb{R}^{h \times l}$  和  $w_v^{(i)} \in \mathbb{R}^{l \times l}$  分别为

价格模态和文本模态各自对应的低阶因子。 $\otimes$ 表示在输入的两种特征向量上提供的外张量积， $\circ$ 表示元素的乘积。

#### 4.3.2 序列特征嵌入层

针对股票市场数据具有可变性高、时间跨度大和高度非线性特点，数据输入过长会使得朴素的特征提取模型无法捕获某些时间点更准确的向量表示，从而会导致下游预测任务的准确性较低。通过引入注意力机制使我们的模型能够选择性地叠加不同时刻的特征信息，避免学习数据中过多的冗余特征而导致的不良性能。

给定  $E_i = \{e_{i,1}^{s_q}, \dots, e_{i,t-1}^{s_q}, e_{i,t}^{s_q}\}$  为注意力学习模块的输入，我们可以得到股票  $s_q$  的融合特征在第  $i$  周的加权短期状态特征向量表示  $g_i^{s_q}$ ：

$$g_i^{s_q} = \text{Attention}(E_i) = \sum_{j=1}^t \alpha_{ij} e_{i,j}^{s_q} \quad (4-8)$$

$$\alpha_{ij} = \frac{\exp(W_g e_{i,j}^{s_q})}{\sum_{k \in t} \exp(W_g e_{i,k}^{s_q})} \quad (4-9)$$

式 (4-9) 中  $W_g$  为可学习的参数矩阵。

#### 4.3.3 关系特征嵌入层

学习股票关系模块的最终目的是更新由社区检测识别的股票社区网络中的节点特征。而节点特征描述的准确性是图预测任务的关键，因此模型需要学习到更多有效的关系信息进行聚合。为此，我们提出了一种通过两层不同类型的图注意力层，使得我们的模型在不同类型的信息下更好的收集到不同类型的关系信息，过滤掉与下游任务无关的冗余特征。具体来说，我们设计了类内图注意学习层和类间图注意学习层，对股票长期状态下不同节点之间的权重和股票各社区间的联动效应进行建模分析。

##### (1) 集群内图注意力学习

在真实的股票市场中，股票价格不仅会受到短期状态的影响，还会受到其长期走势的影响。因此模型有必要学习股票长期状态下不同时间点之间不同的权重。首先将股票  $s_q$  基于注意力机制学习到的短期（周水平）特征  $g_i^{s_q}$  嵌入节点进行聚合，我们的目标是根据短期特征通过社区内注意力学习层学习到股票的加权长期状态特征向量表示。通过嵌入第  $i$  周过去  $x$  周的短期特征向量  $g_i^{s_q}$  可以得到股票  $s_q$  未加权的长期状态序列表示  $G^{s_q} = \{g_{i-x}^{s_q}, g_{i-x-1}^{s_q}, \dots, g_{i-1}^{s_q}\}$ 。第一层图注



注意力网络就是用来学习同一社区内股票各短期状态间的注意力权重，并根据注意力系数更新每个节点的特征信息。计算目标周期  $i$  的状态与目标周期  $j$  的状态之间的注意力系数  $\beta_{ij}$  公式为：

$$\beta_{ij} = \frac{\exp \left( \text{Leaky Re LU} \left( r_u^T \left[ W_u g_i^{sq} \parallel W_u g_j^{sq} \right] \right) \right)}{\sum_{k \in \mathcal{X}} \exp \left( \text{Leaky Re LU} \left( r_u^T \left[ W_u g_i^{sq} \parallel W_u g_k^{sq} \right] \right) \right)} \quad (4-10)$$

式 (4-10) 中  $W_u$  为可学习权重矩阵。

基于求得的周期状态注意力系数  $\beta_{ij}$ ，我们可以得到加权的长期状态特征：

$$u_i^{sq} = GAT(G^{sq}) = \sigma \left( \sum_{j \in \mathcal{d}} \beta_{ij} W_u g_j^{sq} \right) \quad (4-11)$$

## (2) 集群间图注意力学习

通过基于注意力机制层和社区内图注意学习层可分别学习到股票加权短期状态特征向量和加权长期状态特征向量，但是社区内部股票周期状态间关系只是股票间关系一部分，股票网络社区间的复杂关系也会影响到股票价格走势。在社区间图注意学习层，就是应用注意力机制去学习社区间不同的权重，避免忽略股票间其他复杂关系。与先前的工作 [1, 4, 7, 8] 有所不同，我们首先通过 Louvain 算法对股票网络进行社区检测，将其中某个股票社区作为一个节点  $\pi_a$ ，节点  $\pi_a$  中包含此社区的所有股票，而不是将每个股票作为一个节点，然后构建一个社区全连接网络  $G_{\pi_a}$ 。

在计算社区间注意力系数前，我们需要通过图池化操作去基于社区内股票的加权长期特征和社区结构特征去生成一个社区特征向量嵌入，这里受到前人工作的启发，选择用图池化的操作去实现。定义社区  $\pi_a$  的社区内图为  $G_{\pi_a} = (M_{\pi_a}, E_{\pi_a})$ ， $M_{\pi_a}$  是属于社区  $\pi_a$  的股票集合， $E_{\pi_a}$  为股票  $s_q$  和  $s_s$  之间的边集合， $s_q, s_s \in M_{\pi_a}$ 。将股票的加权长期状态特征向量  $u_i^{sq}$  和学习到的社区结构特征向量  $e^{sq}$  连接起来并表示为  $\tau_i^{sq}$ ，我们通过元素级最大池化操作生成社区  $\pi_a$  的嵌入向量  $z_{\pi_a}$ ：

$$z_{\pi_a} = \text{MaxPool} \left( \{ \tau_i^{sq} \mid \forall s_q \in M_{\pi_a} \} \right) \quad (4-12)$$

最终，我们可以得到所有社区的特征向量表示序列  $Z_{\pi} = \{z_{\pi_a}, z_{\pi_b}, \dots, z_{\pi_k}\}$ ， $k$  为股票网络社区数量。

因为每个股票社区之间的关系强度可能会有所不同，因此我们采用图注意网络去学习社区间不同的关系权重，目标股票社区  $\pi_a$  与目标股票社区  $\pi_b$  之间的关系

系数通过以下公式计算：

$$\eta_{\pi_a \pi_b} = \frac{\exp \left( \text{Leaky Re LU} \left( r_{\pi}^T \left[ W_{\pi} z_{\pi_a} \| W_{\pi} z_{\pi_b} \right] \right) \right)}{\sum_{\pi_k \in G_{\pi}} \exp \left( \text{Leaky Re LU} \left( r_{\pi}^T \left[ W_{\pi} z_{\pi_a} \| W_{\pi} z_{\pi_k} \right] \right) \right)} \quad (4-13)$$

最后可以将聚合特征计算为具有 sigmoid 函数的源隐藏特征的加权平均值，公式为：

$$\tau_{\pi_a} = GAT(z_{\pi}) = \sigma \left( \sum_{\pi_b \in G_{\pi}} \eta_{\pi_a \pi_b} W_{\pi} z_{\pi_b} \right) \quad (4-14)$$

式（4-14）中  $W_{\pi}$  为可学习参数。

#### 4.4 分类预测及模型训练

模型最后一层主要是将前面注意力层学习到的股票表示，聚合生成股票最终特征向量，最终的股票表示通过浅层神经网络和 softmax 函数进行股票趋势预测，如图 1-(III)所示。

考虑到股票长短期状态和股票社区间关系的影响，我们将加权短期状态特征向量表示  $g_i^{sq}$ 、加权的长期状态特征  $u_i^{sq}$  和社区间关系特征  $\tau_{\pi_a}$  连接为最终特征向量表示  $rep_i^{sq}$ ：

$$rep_i^{sq} = [g_i^{sq} \oplus u_i^{sq} \oplus \tau_{\pi_a}]^T \quad (4-15)$$

我们把股票趋势预测问题作为一个分类任务，即预测的结果会表现为  $\{\text{up, neutral, down}\}$ ，通过完全连接层去预测结果标签：

$$\hat{y}_i^{sq} = \text{softmax} (W_r rep_i^{sq} + b_r) \quad (4-16)$$

式（4-16）中  $W_r$  为可训练权重矩阵， $b_r$  是偏差向量。

股票趋势预测是一个多分类问题，因此我们将最小化交叉熵定义为模型的损失函数：

$$loss = - \sum_{s_q \in \phi_s} \sum_i^l y_i^{s_q} \ln \hat{y}_i^{s_q} \quad (4-17)$$

式（4-17）中  $y_i^{s_q}$  为股票在第  $i$  周的真实值标签， $\phi_s$  定义为数据集中的所有股票集合。

最后，我们详细介绍提出模型 MF-DAT 的完整训练过程，解释了输入变量，特征的提取和融合，模型的优化等过程。在下面伪代码中，步骤 2 体现了股票社区网络的建立过程，步骤 9-11 是算法的关键部分。步骤 13 和 14 描述了模型的优化过程。

图 4-2 MF-DAT 算法伪代码

**Algorithm 1:MF-DAT 算法**


---

**输入:** 每日级别的历史价格变动率  $\mathbf{P}$ , 股票新闻文本  $\mathbf{Q}$   
**输出:** 模型参数  $\Theta$

```

1  while 没有达到停止条件时 do
2    for 所有  $\mathcal{D}^{\text{batch}} \leftarrow \mathcal{D}^{\text{train}}$  do
3      for 所有股票  $\phi_s \leftarrow s_q$  do
4        通过等式 3-5 对历史价格进行嵌入编码, 得到价格特征  $h^{sq}$ ;
5        通过等式 3-6 对新闻文本进行嵌入编码, 得到文本特征  $v^{sq}$ ;
6        通过等式 4-7 得到多模态融合状态表示  $e^{sq}$ ;
7        通过等式 4-8 得到注意力机制生成加权的短期状态表示  $g^{sq}$ ;
8        通过等式 4-11 生成加权的长期状态表示  $u^{sq}$ ;
9        通过等式 4-14 生成聚合关系特征  $\tau_\pi$ ;
10       end
11       对于每个  $\mathcal{D}^{\text{batch}}$ , 通过等式 4-17 计算损失  $loss$ 
12       进行反向传播并更新参数  $\Theta: \Theta^{(\text{new})} = \Theta^{(\text{old})} - \lambda \frac{\partial loss}{\partial \Theta^{(\text{old})}}$ 
13     end
14   end
15 返回  $\Theta$ 

```

---

## 4.5 实验设计和结果分析

在本节中, 我们进行了广泛的实验来研究提出方法的有效性和局限性, 并且将我们的方法和近几年金融领域流行的几种方法进行了比较。我们先介绍实验数据来源和实验设置, 然后依次汇报每个实验的结果, 并进行对比分析。

### 4.5.1 数据集

本章实验采用的数据集为第三章介绍的数据集, 详情见第三章数据集说明。

### 4.5.2 评价指标

一般来说, 提出新模型去预测股票趋势的最终目的是有利可图, 为了衡量我们提出方法的盈利能力, 我们使用常见的交易策略去模拟股票交易活动。具体来说, 就是将模型预测的向量设置为三维的, 每个维度分别代表着上涨, 持平和下降的预测概率。当上涨概率最高时, 则以当天收盘价买入股票; 如果下降概率最高时, 则以当天收盘价卖出股票, 否则, 不采取任何交易活动。

为了评估提出方法的盈利能力, 我们通过累积投资收益率和夏普利率两个指标去衡量模型的盈利水平。累积投资收益率的计算公式如下:

$$IRR_i^t = \sum_{i \in F^{t-1}} \frac{p_i^t - p_i^{t-1}}{p_i^{t-1}} \quad (4-18)$$

式 (4-18) 中  $F^{t-1}$  表示在时间  $t-1$  交易的一组股票,  $p_i^t$  表示在时间  $t$  股票  $i$  的价格。

而夏普利率是一项综合考虑交易收益和风险的指标，可以用来衡量投资风险与收益相比的表现。夏普利率的计算公式如下：

$$SR_a = \frac{E[R_a - R_f]}{std[R_a - R_f]} \quad (4-19)$$

式（4-19）中  $R_a$  表示投资回报， $R_f$  表示无风险利率。

为了对比基准模型和我们模型的分类预测的效果，我们选取了广泛应用与分类任务的评价指标：准确率和 F1-Score。他们的计算公式为：

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (4-20)$$

$$F1 = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (4-21)$$

$$Recall = \frac{TP}{TP+FP}, Precision = \frac{TP}{TP+FN} \quad (4-22)$$

式（4-22）中 TP、TN、FP 和 FN 分别表示真阳性，真阴性，假阳性和假阴性。

### 4.5.3 实验环境和参数设置

#### 1、实验环境

本节实验环境和前面第三节保持一致。

#### 2、参数设置

为了验证我们提出的模型的有效性，我们选择了几个经典的时间序列预测模型作为基准模型，它们是没有考虑股票关系对股票趋势的影响的。除此之外，我们还选择了考虑股票间复杂关系的 SOTA 模型。我们将结果与以下这些模型进行了对比。

不考虑股票关系的方法：

（1）**MLP** 多层感知机是时间序列预测领域最广泛使用的模型之一，在本文中，我们使用一个简单的多层感知机模型，它具有四层，包括 128 维度和 64 维度的两层隐藏层，学习率设置为 0.0001。

（2）**CNN** 卷积神经网络建模速度很快因此也被广泛应用与股票预测领域。我们使用了具有 3 层卷积层和 1 层全连接层的 CNN 网络，并选择 rmsprop 作为优化器，学习率设置为 0.01。

（3）**LSTM** 作为时间序列预测领域广泛使用的神经网络模型，在股票预测任务中具有优越的性能。在我们的基准模型中，我们使用一个具有 2 层的

LSTM 模型对历史价格数据进行特征学习最后预测。

考虑股票关系的方法：

(4) **GCN** 使用具有 2 个卷积层的 GCN 模型，将历史价格数据作为节点的输入，对包含目标公司关系的历史信息进行重新构建股票图，并利用 Adam 作为优化器。

(5) **TGC Feng** 等人提出的时态图卷积模块，被用于股票关系建模。本文采取了原文章一样的参数设置。

(6)**AD-GAT** 一个通过特征交互进一步对市场信息进行建模分析的 SOTA 模型，提高了股票走势预测任务的性能。使用 Glorot 对所有参数进行初始化，选择 Adam 作为优化器，初始学习率设置为 0.0005。

(7) **MF-DAT** 在 NVIDIA Tesla K80 GPU 上使用了 Adam 优化器，调整参数 100 epochs，初始学习率设置为 5e-4，权重衰减为 5e-5，批量大小为 32。为了减轻过度拟合，我们在每一层的末端设置 dropout=0.5。

4.5.4 实验结果与分析

在这一节中，我们将从一系列实验结果的表现来评估提出模型的性能。我们不仅和经典模型进行比较，还与当前先进的 SOTA 模型进行比较。进一步地，本文还通过消融实验探究不同的组件和超参数对 MF-DAT 性能的产生不同的影响。

(1) 不同阈值的股票集群结果分析

为了探究阈值对模型预测任务的影响，我们通过社区检测基于 5 组不同的阈值挖掘股票不同的集群结构，最后基于不同阈值下的股票关系图去更新股票特征表示，模型的准确度结果如下表。

表 4-1 不同阈值情况下模型的表现

threshold	S&P 500			
	edges	cluster	Accuracy	F1
0.90	13706	6	51.37%	50.59%
0.92	10717	7	53.79%	52.48%
0.94	8019	8	55.61%	54.34%
0.96	5133	8	52.85%	51.65%
0.98	2624	9	50.83%	49.56%

根据表 4-1 的结果，可以看到随着阈值的增加，连接股票的边越来越少，集群数量却随之增加。通过对比分析不同阈值情况下模型的表现，可以得出在阈值为 0.94 的情况下，模型表现出了最好的性能。这是因为，阈值太小时对应的边就过多，股票社区检测过粗容易引入过多的不显著相关的冗余信息；而阈值过大时对应的边就过少，检测过细容易忽略显著相关股票间的相关信息。阈值过大或者过小都无法准确通过图注意力神经网络学习股票间相互关系，从而影响预测任务的效果。虽然不同阈值下模型的性能表现存在一定差异，但是我们模型在大部分阈值下的表现要优于基于预定义行业关系。

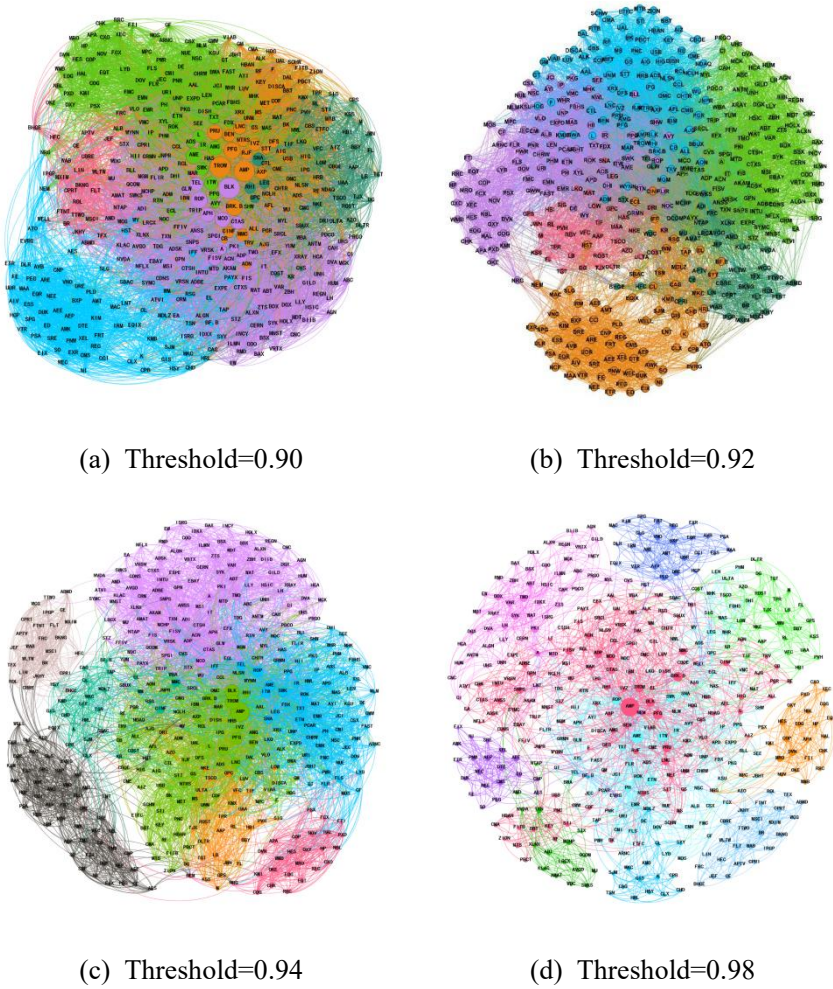


图 4-3 不同阈值对应不同的股票集群结构

(2) 不同模型实验结果的对比分析

1) 分类准确度对比分析

表 4-2 总结了基准模型和我们方法在分类任务中的准确度得分。LSTM 和 CNN 等经典模型只考虑了时间序列数据，表现相对于基于图的模型差很多。实验表明，股票关系因素的介入有利于股票趋势预测。而对比我们提出方法与其他基于图的模型，会发现我们的模型 MF-DAT 表现要比其他模型都要好。虽然 GCN 和 TGC 等基于图的方法都考虑了股票间关系会影响股票走势，但是他们是基于预定义的行业关系或者市场先验信息去建模分析股票间相互作用，比如 TGC[7] 方法就是通过提取 Wikidata 中的行业关系挖掘股票间关系。虽然 Wikidata 有丰富的关系数据，但是预定义式行业关系的局限必然会导致预测的偏差。我们认为这就是我们模型取得更好的效果的原因。MF-DAT 模型的提出就是为了避免先验信息匮乏带来的误导，从而提高预测效果。

总的来说，我们提出的模型相比于其他基准模型拥有更准确和更稳定的预测性能。在 S&P 500 数据集中，准确度和 F1 分数分别比第二名高出约 4.56%和 4.03%；在 NASDAQ 数据集中，分别超出第二名 4.75%和 5.02%。

表 4-2 不同模型预测的分类准确度对比

方法	S&P 500		NASDAQ	
	Accuracy	F1	Accuracy	F1
MLP	35.68%	30.22%	36.68%	34.59%
CNN	39.39%	32.78%	38.39%	36.87%
LSTM	41.55%	40.28%	41.24%	40.39%
GCN	45.08%	44.21%	46.37%	45.28%
TGC	46.74%	46.26%	48.56%	47.57%
AD-GAT	51.05%	50.34%	51.38%	50.67%
MF-DAT	<b>55.61%</b>	<b>54.57%</b>	<b>56.13%</b>	<b>55.69%</b>

2) 盈利能力对比分析

为了验证模型的盈利能力表现，我们在测试期间对股票的交易进行了回测，并且进行了模拟交易。表 4-3 汇总了不同模型在保持交易策略不变的情况下的盈利能力结果。图 4-4 展示了不同方法在模拟交易中资产价值变化情况。在回测期间内，S&P 500 指数和 NASDAQ 指数分别上涨了 10.47%和 15.82%（雅虎财经

显示:S&P 500 指数从 2615.75 增长到 2889.75，NASDAQ 指数从 6561.25 增长到 7599.25）。从表 4-3 和图 4-4 可以得出，相对于其他模型，我们的模型的盈利能力更加稳定。从 IRR 的结果来看，MF-DAT 模型的累积收益率分别为 13.69%和 16.79%高于同时期指数的涨幅，也高于其他基准模型；从 SR 的结果来看，MF-DAT 的夏普比率为正值且更大，说明期间股票收益增长率高于无风险利率且获得的风险回报越高。

表 4-3 不同模型回测的盈利能力对比

方法	S&P 500		NASDAQ	
	IRR	SR	IRR	SR
MLP	-5.01%	0.4243	-3.28%	0.5443
CNN	-2.12%	0.4272	-1.87%	0.5873
LSTM	3.28%	0.7884	3.88%	0.7642
GCN	4.50%	1.1681	5.69%	0.9643
TGC	6.87%	1.2599	7.24%	1.1354
AD-GAT	10.51%	1.9186	11.88%	1.8936
MF-DAT	<b>13.69%</b>	<b>2.7890</b>	<b>16.79%</b>	<b>2.6793</b>

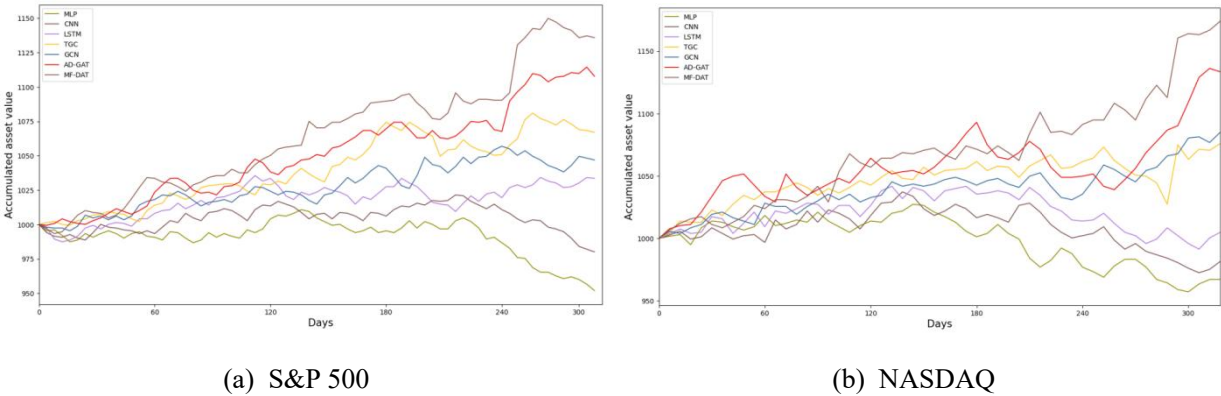


图 4-4 不同模型模拟交易中资产变化对比。假设资产价值从 1000 开始。

(3) 消融实验分析对比

为了检验 MF-DAT 模型各组成模块的有效性，我们在两个数据集上进行了相关的消融实验，我们通过剔除替换某个模块设计了 4 个变体：



1) w/o LMF:MF-DAT 没有通过 LMF 模块对价格数据和新闻文本进行特征融合,而是直接将两者串联拼接。

2) w/o Intra-community Graph Attention Learning (w/o intra):MF-DAT 剔除了股票长期状态学习模块,直接将前面学习到的股票短期状态嵌入股票社区图结构中。

3) w/o Inter-community Graph Attention Learning (w/o inter):MF-DAT 剔除了股票社区间关系状态学习模块,直接将前面学习到的股票长短期状态馈送训练层进行预测。

4) w/o community detection (w/o com):MF-DAT 不再通过社区检测挖掘股票社区结构构建关系图,而是通过传统方法提取 Wikidata 中的预定义股票关系生成股票关系图。

表 4-4 不同变体的结果对比

变体	S&P 500		NASDAQ	
	准确度	F1	准确度	F1
w/o LMF	53.86%	52.43%	54.37%	52.89%
w/o intra	52.25%	51.17%	53.63%	52.32%
w/o inter	51.27%	50.23%	52.29%	51.48%
w/o com	50.49%	49.85%	51.84%	50.39%
MF-DAT	55.61%	54.57%	56.13%	55.69%

由表 4-4 的结果可得,完整的 MF-DAT 模型才会产生更好的性能,剔除其中任何一个模块都会导致更坏的结果,这样的结果证明了模型中每个组件的有效性。在两个不同的数据集中,各模块发挥的作用不一样,但都有助于提高模型的预测性能。通过对这四个变体模型的实验研究,与其他三种情况比较,取缔社区检测挖掘股票社区结构去构建股票关系图会导致模型预测性能最大程度的下降。这表明,社区检测挖掘的股票社区结构之间的隐性关系可以弥补目前先验信息不足的问题,对模型预测有着显著的效果。相比之下,剔除用于多特征融合的 LMF 模块,带来的性能下降最小。尽管如此,以上结果还是证明所提出模型在预测任务中的表现更好。

## 4.6 本章小结

在这项研究中,为了预测股票未来走势,我们提出了新的多特征融合的双图注意力网络。我们将股票历史价格特征和金融新闻特征通过 LMF 模块融合生成新的股票特征,弥补了直接拼接多特征时无法捕获市场信息间相互作用的问题。MF-DAT 通过双层图注意力网络能够从市场信息和股票关系中学习股票多个状态特征。特别地, MF-DAT 模型在学习股票间关系时,不需要预定义关系,而此前大多数研究都是基于先验信息完成的。为了评估提出方法的先进性,我们在公开数据集 S&P 500 和 NASDAQ 上进行了实验。实验结果表明, MF-DAT 具有更好的准确性和更高的投资回报率。

将来,我们的研究可以从以下几个方向深入研究: 1)虽然本文避免了根据股票预定义关系去生成股票关系图,而是通过股票波动相关性去构建股票网络,但是没有考虑股票间相互作用会随时间的变化而变化,因此股票社区结构也是动态变化的。在将来的研究中,我们应该考虑股票关系的动态化对预测任务的影响。 2)本文通过 Louvain 算法对股票社区结构进行检测,只能检测到非重叠社区结构,不能挖掘到重叠社区结构。我们将考虑应用更加先进的社团检测方法对股票重叠社区结构进行挖掘。 3)在将来的研究中,我们考虑将模型应用到其他领域,比如网站流量预测等。

## 第五章 结论与展望

本文研究的内容是基于图注意力神经网络和多特征融合的股票趋势预测,重点介绍了本篇提出的两个预测模型,对其模型架构和工作机制进行了详细阐述,并对两个模型的预测效果和性能提升进行了验证。接下来在本章中,将对全文研究内容进行总结,并且基于当前模型的不足,对未来可能解决问题的方案进行探索和展望。

### 5.1 全文总结

股市价格趋势的反映对行业和整个国家的经济状况都具有重要意义。为准确预测股价趋势,过去人们通常采用单一的股票历史价格数据,但近年来,越来越多的研究开始融合多种数据特征,如股票历史价格和金融新闻等,以提高预测的准确性。尽管取得了不俗的成果,但这些研究中尚未充分挖掘利好利空特征对应的新闻信息,也未考虑股票间复杂关系联合作用对股价的影响。此外,需要探究不同特征对股价趋势的影响程度、特征之间的交互特征等因素。因此,本研究结合图注意力神经网络和多特征融合技术,深入探索股票预测算法。本研究的研究成果如下:

(1) 通过有效的特征提取模型,高效挖掘股票市场多种信息特征,以准确预测股票趋势。然而,金融市场的信息众多且多样化,包括股票价格历史数据和金融新闻等,其中并非所有信息都是有效的,可能存在重复和虚假的数据。因此,需要不断更新和优化挖掘方法,以获取更有效的信息特征,并提高预测模型的准确性。

(2) 利用元路径方法构建有力的股票关系网络,以融合不同股票之间的关系。股票关系网络指通过连接同板块、同产业、同消息面等关系,将不同股票之间的联系建立起来。通过构建股票关系网络,可以更好地了解不同股票之间的关联关系,进而有助于预测股票的走势。

(3) 引入股票关系建模模型 ML-GAT,以权衡不同特征对股价波动的影响。不同信息特征和不同类别关系对股票趋势具有不同的重要性,因此需要给予股票不同特征分配不同的信息权重。通过 ML-GAT 模型,可以有效地建模不同特征

之间的关系，并为股票关系赋予更合适的权重，以提高预测准确性。

(4) 采用新颖的多特征融合方法，应用 LMF 模块捕获股票市场多种信息模态间的交互特征。股票市场价格受多种因素影响，如公司基本面、金融新闻、政策等，因此，单一信息源的预测难以提供足够精准性。相反，多模态信息融合可以最大限度利用不同信息之间的交互特征，提高预测准确性，并为投资者提供更科学、精准的决策支持。

(5) 在 ML-GAT 模型基础上，提出了新的股票关系挖掘模型 MF-DAT。该模型通过股票价格收益波动相关性和阈值法建立新的股票网络。然后，利用社区检测算法从股票网络中挖掘新的股票关系。实验证明，相对于预定义行业关系表示股票关系方法，利用股票收益波动相关性来挖掘股票关系的方法具有更优秀的性能和有效性。

## 5.2 未来展望

在未来的研究中，可以从以下几个方向深入探索：

(1) 探索适用于金融领域不同特征的编码工具。例如，可以采用更先进的自然语言处理技术处理金融新闻文本，以捕捉更准确的股票变化信息的多个维度。然后，对获取到的多种特征进行因子优化分析，选择最重要的几个技术指标来预测股票趋势，以提高预测的准确性。

(2) 虽然本文避免了根据预定义的股票关系生成股票关系图，而是通过股票波动相关性构建股票网络，但未考虑股票之间的相互作用会随时间变化而变化，实际上股票社区结构也是动态变化的。因此，在未来的研究中，应考虑股票关系的动态化对预测任务的影响。

(3) 本文使用 Louvain 算法检测股票社区结构，只能检测非重叠的社区结构，而无法挖掘重叠的社区结构。因此，将考虑应用更先进的社团检测方法来挖掘股票的重叠社区结构。

(4) 在未来的研究中，可以考虑将该模型应用于其他领域，例如网站流量预测等，以探索其在不同领域的适用性。

## 参 考 文 献

- [1] Ahmadi E, Jasemi M, Monplaisir L, et al. New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the Support Vector Machine and Heuristic Algorithms of Imperialist Competition and Genetic[J]. Expert Systems with Applications, 2018, 94: 21-31.
- [2] Rather A M, Agarwal A, Sastry V N. Recurrent neural network and a hybrid model for prediction of stock returns[J]. Expert Systems with Applications, 2015, 42(6): 3234-3241.
- [3] Patel J, Shah S, Thakkar P, et al. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques[J]. Expert systems with applications, 2015, 42(1): 259-268.
- [4] Zhou X, Pan Z, Hu G, et al. Stock market prediction on high-frequency data using generative adversarial nets[J]. Mathematical Problems in Engineering, 2018: 1-11.
- [5] Fama E F. The behavior of stock-market prices[J]. The journal of Business, 1965, 38(1): 34-105.
- [6] Malkiel, B. G. The efficient market hypothesis and its critics.[J] Journal of Economic Perspectives, 2003,17, 59–82.
- [7] 陈擎霄. 基于深度学习的股票走势分析系统的研究与实现[D].北京邮电大学,2021.DOI:10.26969/d.cnki.gbydu.2021.001796.
- [8] Huang K, Li X, Liu F, et al. ML-GAT: A Multilevel Graph Attention Model for Stock Prediction[J]. IEEE Access, 2022, 10: 86408-86422.
- [9] Wang G, Cao L, Zhao H, et al. Coupling macro-sector-micro financial indicators for learning stock representations with less uncertainty[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(5): 4418-4426.
- [10] 程孟菲,高淑萍.基于深度迁移学习的多尺度股票预测[J].计算机工程与应用,2022,58(12):249-259.
- [11] Bollerslev, T. Generalized autoregressive conditional heteroskedasticity.[J]

Journal of econometrics, 1986,31, 307–327.

[12] 万建强, 文洲. ARIMA 模型与 ARCH 模型在香港股指预测方面的应用比较[J]. 数理统计与管理, 2001, 20(6): 1-4.

[13] Li W, Bao R, Harimoto K, et al. Modeling the stock relation with graph network for overnight stock movement prediction[C]//Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021: 4541-4547.

[14] 程孟菲,高淑萍.基于深度迁移学习的多尺度股票预测[J].计算机工程与应用,2022,58(12):249-259.

[15] 严冬梅,李斌.基于生成式对抗神经网络的股票预测研究[J].计算机工程与应用,2022,58(13):185-194.

[16] Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction[J]. Neural Computing and Applications, 2021, 33(10): 4741-4753.

[17] Nayak R K, Mishra D, Rath A K. A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices[J]. Applied Soft Computing, 2015, 35: 670-680.

[18] Lohrmann C, Luukka P. Classification of intraday S&P500 returns with a Random Forest[J]. International Journal of Forecasting, 2019, 35(1): 390-407.

[19] Hoseinzade E, Haratizadeh S. CNNpred: CNN-based stock market prediction using a diverse set of variables[J]. Expert Systems with Applications, 2019, 129: 273-285.

[20] Zhao J, Zeng D, Liang S, et al. Prediction model for stock price trend based on recurrent neural network[J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(1): 745-753.

[21] Ding G, Qin L. Study on the prediction of stock price based on the associated network model of LSTM[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(6): 1307-1317.

[22] Ding X, Zhang Y, Liu T, et al. Deep learning for event-driven stock prediction[C]//Twenty-fourth international joint conference on artificial intelligence. 2015.

- [23] Wang H, Wang T, Li Y. Incorporating expert-based investment opinion signals in stock prediction: A deep learning framework[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(01): 971-978.
- [24] Dash R K, Nguyen T N, Cengiz K, et al. Fine-tuned support vector regression model for stock predictions[J]. Neural Computing and Applications, 2021: 1-15.
- [25] Chen W, Zhang H, Mehlawat M K, et al. Mean-variance portfolio optimization using machine learning-based stock price prediction[J]. Applied Soft Computing, 2021, 100: 106943.
- [26] Zhang Q, Yang L, Zhou F. Attention enhanced long short-term memory network with multi-source heterogeneous information fusion: An application to BGI Genomics[J]. Information Sciences, 2021, 553: 305-330.
- [27] Chen Y, Wei Z, Huang X. Incorporating corporation relationship via graph convolutional neural networks for stock price prediction[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018: 1655-1658.
- [28] Li W, Bao R, Harimoto K, et al. Modeling the stock relation with graph network for overnight stock movement prediction[C]//Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021: 4541-4547.
- [29] Wang H, Li S, Wang T, et al. Hierarchical Adaptive Temporal-Relational Modeling for Stock Trend Prediction[C]//IJCAI. 2021: 3691-3698.
- [30] Matsunaga D, Suzumura T, Takahashi T. Exploring graph neural networks for stock market predictions with rolling window analysis[J]. arXiv preprint arXiv:1909.10660, 2019.
- [31] Chen W, Jiang M, Zhang W G, et al. A novel graph convolutional feature based convolutional neural network for stock trend prediction[J]. Information Sciences, 2021, 556: 67-94.
- [32] Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.
- [33] Dou Y, Liu Z, Sun L, et al. Enhancing graph neural network-based fraud

detectors against camouflaged fraudsters[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020: 315-324.

[34] Gurcan C, Yitao L, Gel Yulia R, et al. BitcoinHeist: Topological data analysis for ransomware detection on the bitcoin blockchain[J]. arXiv preprint arXiv: 1906.07852, 2019, 2019.

[35] Feng F, He X, Wang X, et al. Temporal relational ranking for stock prediction[J]. ACM Transactions on Information Systems (TOIS), 2019, 37(2): 1-30.

[36] Cheng D, Yang F, Xiang S, et al. Financial time series forecasting with multi-modality graph neural network[J]. Pattern Recognition, 2022, 121: 108218.

[37] Feng S, Xu C, Zuo Y, et al. Relation-aware dynamic attributed graph attention network for stocks recommendation[J]. Pattern Recognition, 2022, 121: 108119.

[38] Ali U, Hirshleifer D. Shared analyst coverage: Unifying momentum spillover effects[J]. Journal of Financial Economics, 2020, 136(3): 649-675.

[39] 乔扬,戴洛特,朱宏泉.A+H 交叉上市股票涨跌幅度的溢出效应[J].金融论坛,2017,22(03):66-80.

[40] 段丙蕾,汪荣飞,张然.南橘北枳: A 股市场的经济关联与股票回报[J].金融研究,2022(02):171-188.

[41] Xu C, Huang H, Ying X, et al. HGNN: Hierarchical graph neural network for predicting the classification of price-limit-hitting stocks[J]. Information Sciences, 2022, 607: 783-798.

[42] Cheng R, Li Q. Modeling the Momentum Spillover Effect for Stock Prediction via Attribute-Driven Graph Attention Networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(1): 55-62.

[43] Zhong T, Peng Q, Wang X, et al. Novel indexes based on network structure to indicate financial market[J]. Physica A: Statistical Mechanics and its Applications, 2016, 443: 583-594.

[44] MacMahon M, Garlaschelli D. Community detection for correlation matrices[J]. arXiv preprint arXiv:1311.1924, 2013.

[45] Wang G J, Xie C, Stanley H E. Correlation structure and evolution of world



- stock markets: Evidence from Pearson and partial correlation-based networks[J]. Computational Economics, 2018, 51(3): 607-635.
- [46] Mantegna R N. Hierarchical structure in financial markets[J]. The European Physical Journal B-Condensed Matter and Complex Systems, 1999, 11(1): 193-197.
- [47] Onnela J P, Chakraborti A, Kaski K, et al. Dynamics of market correlations: Taxonomy and portfolio analysis[J]. Physical Review E, 2003, 68(5): 056110.
- [48] 黄玮强,庄新田,姚爽.中国股票关联网络拓扑性质与聚类结构分析[J].管理科学,2008(03):94-103.
- [49] de Pontes L S, Rêgo L C. Impact of macroeconomic variables on the topological structure of the Brazilian stock market: A complex network approach[J]. Physica A: Statistical Mechanics and its Applications, 2022: 127660.
- [50] Li H, An H, Fang W, et al. Global energy investment structure from the energy stock market perspective based on a Heterogeneous Complex Network Model[J]. Applied Energy, 2017, 194: 648-657.
- [51] Chmielewski L, Amin R, Wannaphaschaiyong A, et al. Network analysis of technology stocks using market correlation[C]//2020 IEEE International Conference on Knowledge Graph (ICKG). IEEE, 2020: 267-274.
- [52] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [53] Kenton J D M W C, Toutanova L K. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of naacL-HLT. 2019, 1: 2.
- [54] Zhang K, Zhong G, Dong J, et al. Stock market prediction based on generative adversarial network[J]. Procedia computer science, 2019, 147: 400-406.
- [55] Thakkar A, Chaudhari K. A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions[J]. Expert Systems with Applications, 2021, 177: 114800.
- [56] J. Tang, C. Deng, G.-B. Huang, Extreme learning machine for multilayer perceptron, IEEE transactions on neural networks and learning systems 27 (4) (2015) 809 – 821.
- [57] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances

in neural information processing systems, 2017, 30.

[58] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.