

基于情感分析大模型的股票预测:结合 GRU 和 ALBERT 的预测模型

崔 婷¹, 黄斐然²

(1. 广东财经大学 经济学院, 广东 广州 510320; 2. 暨南大学 网络空间安全学院, 广东 广州 510632)

[摘 要]自股票市场诞生以来,对股票价格的预测一直是研究的热点。已有的研究主要根据股票基本面数据、股价时间序列等结构化数据,利用数学、统计、深度学习等多种方法进行预测;或者基于新闻报道、社交网络文本等非结构化数据,采用文本挖掘、情感分析等技术获取情绪指标进而实现股价预测。然而,将以上两种方法进行结合的研究目前并不多见。本文提出了一种新的股票价格预测方法,该方法利用 GRU 深度学习网络,根据股票历史数据进行股价的预测,并构建出基于 GRU 框架的预测模型。同时,采用轻量级大模型 ALBERT,对媒体新闻进行情感提取,获取情感特征。最后,将情感特征输入至于 GRU 预测模型中,并适应性调整 GRU 的输出结果,以获得最终的股票预测。本文最终选择了 GRU-attention 与 Google Chinese ALBERT-Large 结合的方案,该方案充分利用了客观特征以及情感特征包含的信息,使得输入模型的特征更加全面,并且在预测精度上相对其他方案有明显的优势。从实际应用上来看,本文搭建的股票预测模型能够为投资者提供更为准确的股票价格预测结果,帮助投资者制订更加合理的投资决策,并可以预警金融市场中可能发生的风险,帮助投资者规避损失。

[关键词]股票预测;文本情感分析;自然语言处理;门控循环单元;神经循环网络

[中图分类号] C812, O212

[文献标识码] A

[文章编号] 1003-8353(2024)02-0113-11

引 言

股票预测一直是金融领域广泛关注的问题,一般通过企业的基本面、股价的历史数据及其技术指标来实现对未来股价波动的预测。随着大数据以及人工智能的发展,传统的时间序列预测模型已经被逐渐替代,机器学习尤其是深度学习方法凭借其处理非线性数据的能力,成为股票预测的主流方法。

本文研究的重点在于探索一个结合时间序列预测和文本情感分析的股票预测模型。主要根据数据特性开展不同工作:对于历史数据以及技术指标的时间序列,构建 GRU 网络框架,并引入小波去噪、注意力机制等特征提取手段强化预测精度;对于情感特征,选取目标股票的相关新闻报道,利用一种轻量级的大模型 ALBERT 进行情感提取;最后根据新闻的时间戳,利用提取的情感特征对 GRU 框架的输出结果进行适应性的调整并得到最终的预测。相较于该领域的现有研究,本文的结合方法优势如下:首先,充分利用了客观特征以及情感特征包含的信息,使得输入模型的特征更加全面,有利于预测精度的提升。其次,目前大多数金融领域的文本情感分析采用的还是传统的机器学习分类方法,在分析非结构化的文本数据时优势不足^①,本文采用轻量级大模型 ALBERT,对文本信息具有更强的处理能力。最后,本文利用情感特征以调整 GRU 的输出结果,为克服调整幅度受一天内新闻数目差异的影响,构建了四

[基金项目]国家自然科学基金资助项目(项目编号:62002068)。

[作者简介]崔婷(1983—),女,广东财经大学经济学院副教授,博士;黄斐然(1988—),男,暨南大学网络空间安全学院副教授。

^①Fataliyev, K., Chivukula A., Prasad M., et al., "Stock Market Analysis with Text Data: A Review", *Preprint Arxiv*, 2021.

种方案实现情感综合,利用总情感调整输出结果。根据本文提供的方法,可以获得更精准的预测结果,从而为投资者决策提供有效参考,同时能够对突发新闻事件作出反应,及时预警可能发生的金融风险,帮助投资者规避股价波动带来的损失,因此具备充分的现实意义。

简而言之,本文的主要贡献如下:1. 通过收集和整理大量的金融股票相关新闻文本,构建了一个适用于金融股票领域的文本数据集。该数据集可用于金融领域情感分析研究,是一个全面且多样化的数据集,为研究人员提供了丰富的文本材料进行情感分析;2. 提出了一种将情感分析与股票预测相融合的双路模型。将股票预测模型与情感分析模型相结合,从而能够更全面地考虑股票市场的情感因素,提高股票预测的准确性。3. 本研究进行了大量实验,分别在股票预测模型和情感分析模型的分支上进行了探索,并筛选出了最优的模型。在股票预测模型方面,尝试了多种神经网络模型,并创新性地对它们进行有效组合,通过性能的比较最终选择了最佳的预测模型,以提高预测的准确性和稳定性;实验结果显示,在当前场景下,GRU 模型的表现优于 LSTM 模型,这与预期相符。此外,在情感分析模型方面,也进行了大量实验,筛选出了在金融领域表现最好的情感分析大模型,以确保其结果能够更准确地反映股票市场的情绪变化。最后,提出了一种将两个模型结合的方法,以进一步提高股票预测的准确性和可靠性。

本文的结构安排如下:第一部分综述支撑本研究的关键文献。第二部分描述模型构建、数据处理方法以及实验结果的数据分析和可视化呈现。第三部分讨论模型的内在机制和理论意义。第四部分对研究进行总结并展望未来研究方向。

一、相关工作及文献回顾

与一般预测模型类似,股票预测同样主要从预测模型以及特征两个方面入手实现预测精度的提升。对于预测模型,考虑到股价作为时间序列,具有非平稳非线性和高度的顺序依赖性的特点,循环神经网络(RNN, Recurrent Neural Network)及其变种,如长短期记忆网络(LSTM, Long Short-Term Memory)以及门控循环单元(GRU, Gate Recurrent Unit)成为合适的选择。特别地,针对于股价序列, Bhavani 等^①对比研究了分别使用 GRU 和 LSTM 方法预测股价的效果差异,虽然 GRU 有着更简单的网络结构,但其预测性能并不弱于 LSTM。在此基础上, Zou 等^②进一步优化模型,利用堆叠 LSTM 以及基于注意力的 LSTM 预测股价并取得了不错的效果。

在针对股票预测已有的研究中,分析师通常会考虑公司财务指标、宏观经济指标,并结合历史数据和实时信息来分析和预测未来股票价格走势。这些预测指标和数据不仅为投资者解读市场趋势和指导交易决策提供了参考价值,而且被证实对股票收益有一定程度的预测能力。如 Chemmanur 和 Yan^③发现广告支出的增加对股价的短期变动有正向影响。李昌明^④指出企业高管的大幅减持行为会使股价走势达到相应的顶部,但是减持行为往往会领先 10—15 天。张琢^⑤通过研究人民币汇率波动与中国 A 股的关系,提出短期人民币汇率波动与股票价格反相关,而长期人民币汇率波动与股票价格正相关的观点。徐加根和王波^⑥研究了交易量和涨停板作为投资者关注度的间接指标对中国 A 股市场的影响。

这类结构化数据大部分属于时间序列,而 RNN 由于能够存储过去的信息,正适合于学习序列预测

①Bhavani A., Ramana A. V., Chakravarthy A. S. N., “Comparative Analysis Between LSTM and GRU in Stock Price Prediction”, 2022 International Conference on Edge Computing and Applications (ICECAA) IEEE, 2022, pp. 532–537.

②Zou Z., Qu Z., “Using LSTM in Stock Prediction and Quantitative Trading”, CS230: Deep Learning, Winter, 2020, pp. 1–6.

③Chemmanur T. J., Yan A., Chhaochharia V., et al., “Advertising, Investor Recognition, and Stock Returns”, AFA 2011 Denver Meetings Paper, 2010.

④李昌明:《上市公司高管的关联交易对股价影响的实证分析:以创业板高管的增持与减持为例》,复旦大学硕士学位论文,2014 年。

⑤张琢:《人民币汇率波动与中国 A 股价格的关系》,暨南大学硕士学位论文,2010 年。

⑥徐加根,王波:《投资者关注:涨停板还是大交易量——来自中国 A 股市场的实证研究》,《投资研究》,2013 年第 4 期。

问题中的顺序依赖性,因此 RNN 非常适合用于预测股票价格。对于深度学习来说,使用 LSTM^{①②} 进行股票预测成为了一种主流的方法。Gunduz^③ 等人将 LSTM 结合 LightGBM 预测了八家银行股票的每小时走势,表现出了非常具有竞争力的效果,同时使用变分自编码器将 LSTM 生成的股票数据进行降维处理也可达到相似的结果。Zou 等^④采用堆叠 LSTM 与基于注意力的 LSTM 的模型来预测第二天的股价,结果表明基于注意力机制的 LSTM 在预测误差方面优于所有其他模型,并且在交易策略中显示出比其他模型更高的回报。除了 LSTM 之外,GRU 也是一种常用的循环神经网络,它较 LSTM 相比具有结构简单、运行速度快的优点。Bhavani 等^⑤对比研究了使用 GRU 和 LSTM 方法来预测股价,实验结果表明尽管 GRU 具有更少的特征,但其准确度是有竞争力的,同时 GRU 在性能方面优于 LSTM。

关于模型输入特征的考量,又可以进一步分为特征选择和特征提取两个方向。特征选择即选择哪些指标作为预测股价的依据,常见的有历史数据和基本面数据。然而随着大数据技术的发展,大数据为我们提供了更多特征来源,比如新闻报道传递的情感信息,网络评论反映的投资者情绪等。这些特征的可靠性也在一些研究中被佐证,例如,Hillert 等^⑥验证了主流媒体报道会影响到投资者的情绪,其中媒体关注度高的公司股票收益可预测性更显著。相应的文本情感分析方法也被用于包含该类特征的股价预测^⑦。特征提取主要是对已有特征的进一步处理,如数据平滑或去噪、降维等。例如,Bao 等^⑧将小波去噪后的多个序列使用堆叠自编码器提取深层特征并输入 LSTM,在多个市场指数的数值实验中验证了模型的有效性。

在有效市场假说^⑨成立的条件下,一切有价值的信息都将反映在股价走势中,因此诸如上市公司信息披露信息、财经媒体新闻报道、社交网络文本这类非结构化文本数据同样具有价值。俞庆进和张兵^⑩通过研究百度搜索指数发现投资者关注度与股票短期波动之间存在相关性。胡国强、黄波^{⑪⑫}基于中国股市经验证明了媒体情绪会影响到投资情绪,导致资本市场错误定价。孙书娜和孙谦^⑬基于“雪球网”论坛数据指出,股票短期价格和交易量受到投资者关注度影响。Frank & Sanati^⑭研究发现,股票市场对于新闻态度有差异性反应,正面新闻的冲击强于负面新闻。

而随着互联网和大数据技术的发展,非结构化文本数据(如上文提到的投资者情绪、媒体关注度

①Hochreiter S., Schmidhuber J., “Long Short-Term Memory”, *Neural Computation*, 1997, 9(8), pp. 1735–1780.

②Greff K., Srivastava R. K., Koutnik, J., et al., “LSTM: A Search Space Odyssey”, *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(10), pp. 2222–2232.

③Gunduz H., “An Efficient Stock Market Prediction Model Using Hybrid Feature Reduction Method Based on Variational Autoencoders and Recursive Feature Elimination”, *Financial Innovation*, 2021, 7(1).

④Zou Z., Qu Z., “Using LSTM in Stock Prediction and Quantitative Trading”, *CS230: Deep Learning*, Winter, 2020, pp. 1–6.

⑤Bhavani A., Ramana A. V., Chakravarthy A. S. N., “Comparative Analysis Between LSTM and GRU in Stock Price Prediction”, 2022 International Conference on Edge Computing and Applications (ICECAA) IEEE, 2022, pp. 532–537.

⑥Hillert A., Jacobs H., Müller S., “Media Makes Momentum”, *The Review of Financial Studies*, 2014, 27(12), pp. 3467–3501.

⑦Kalyani J., Bharathi H. N., Jyothi R., “Stock Trend Prediction Using News Sentiment Analysis”, *International Journal of Computer Science & Information Technology*, 2016, 8(3): pp. 67–76.

⑧Bao W., Yue J., Rao Y., “A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and Long-Short Term Memory”, *PLoS ONE*, 2017, 12(7).

⑨Malkiel B. G., “Efficient Market Hypothesis”, *Finance*, London: Palgrave Macmillan UK, 1989, pp. 127–134.

⑩俞庆进, 张兵:《投资者有限关注与股票收益——以百度指数作为关注度的一项实证研究》,《金融研究》,2012年第8期。

⑪胡国强, 肖志超:《媒体关注、资本市场错误定价与企业投资》,《广东财经大学学报》,2019年第2期。

⑫黄波, 姜华东:《情绪引致的股价噪声会增进投资—估值敏感性吗?——源自中国股市的经验证据》,《广东财经大学学报》,2021年第6期。

⑬孙书娜, 孙谦:《投资者关注和股市表现——基于雪球关注度的研究》,《管理科学学报》,2018年第6期。

⑭Frank M. Z., Sanati A., “How Does the Stock Market Absorb Shocks?”, *Journal of Financial Economics*, 2018, 129(1), pp. 136–153.

等)的收集和分析成为可能。在金融市场中,互联网每天会产生大量的金融文本信息,其中包含着股票、机构和市场的新闻,提取这些新闻中情感倾向不仅可以监测市场稳定^①,而且有助于预测股票走势^②。Kalyani^③等人通过朴素贝叶斯分类新闻的正负情感,预测股票的涨跌趋势,结果表明与新闻随机标签相比,模型的准确度有大幅提升。因此,分析这些庞大的文本数据信息可以为市场参与者提供获利的优势。然而,大多数的研究方法都是基于传统的方法,在分析非结构化的文本数据时优势不足^④,同时在对文本情感进行处理时,基于机器学习的传统方法仍是主流。此外,基于卷积神经网络或循环神经网络的情感分析模型均暴露出了一些问题:卷积神经网络无法获取文本的远距离依赖信息不适用于处理自然语言,而循环神经网络无法同时获取双向语义信息且并行能力较差,参数规模较小无法较好拟合大规模数据。而近年来,随着大语言模型的发展(LLM, Large Language Model)的发展,基于Transformer^⑤的模型,如BERT^⑥、ChatGPT,表现出了对文本信息极强的处理能力,因此本文使用一种轻量级的大模型ALBERT^⑦来对新闻数据进行情感提取。

综合以上观点,使用深度学习方法进行股票市场预测不仅需要考虑数据特性和选择适当的模型方法,还要全面考虑影响股价波动的各种因素。特别是在当前容易获取客观特征数据的情况下,通过探索不同渠道的文本情感特征可能会带来更有价值的信息。然而,目前的深度学习方法在股票预测研究中通常只注重模型或特征的单一应用,鲜有结合两者的研究。为此,本文提出了一种新的股票预测模型,充分利用大型模型在非结构化数据特征提取方面的强大能力,同时结合传统深度神经网络中的循环神经网络(GRU)对时间序列数据预测的准确性。具体而言,本文采用了ALBERT模型进行文本情感分析,以更好地理解市场参与者的情感状态。同时,引入GRU作为时间序列数据的处理部分,以提高对股票价格变动预测的准确性。这种双管齐下的策略使得模型在处理多源信息时能够更全面、精准地进行预测。通过实验证明,本文提出的股票预测模型在综合考虑大型模型和传统深度神经网络的优势上,取得了较好的预测性能。这为未来深度学习方法在股票市场预测中的应用提供了一种新的思路和方法。

二、模型搭建及指标设定

(一) 算法流程和模型架构

整体实验的流程分为三大步,首先是股票模型的训练和预测,之后是金融文本情感多分类模型的训练与预测,最后是二者融合。具体流程见图1。

本文提出的股票预测模型总体结构如图2所示。它主要由GRU-attention模型和ALBERT-large模型两部分组成。

在GRU-attention模型中,股票数据首先按照(二)2节中的方式进行划分,作为GRU的输入,之后将每一个GRU单元的输出进行注意力计算,取其中的最后一个输出经过全连接层映射为一个预测值。

①许伟,梁循,杨小平:《金融数据挖掘:基于大数据视角的展望》,知识产权出版社,2013年版。

②Fung G. P. C., Yu J. X., Lam W., "News Sensitive Stock Trend Prediction", Pacific-Asia Conference on Knowledge Discovery and Data Mining, Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 481-493.

③Kalyani J., Bharathi H. N., Jyothi R., "Stock Trend Prediction Using News Sentiment Analysis", *International Journal of Computer Science & Information Technology*, 2016, 8(3), pp. 67-76.

④Fataliyev, K., Chivukula A., Prasad M., et al., "Stock Market Analysis with Text Data: A Review", *Preprint Arxiv*, 2021.

⑤Vaswani A., Shazeer N., Parmar N., et al., "Attention is All You Need", *Advances in Neural Information Processing Systems*, 2017, 30, pp. 5998-6008.

⑥Devlin J., Chang M. W., Lee K., et al., "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding", *The North American Chapter of the Association for Computational Linguistics*, 2019, pp. 4171-4186.

⑦Lan Z., Chen M., Goodman S., et al., "Albert: A Lite Bert for Self-Supervised Learning of Language Representations", *International Conference on Learning Representations*, 2020.

在 ALBERT 模型中,文本数据首先被转换为词向量,并且将一个无语义的 CLS 符号插入至文本前,将其作为模型的输入,在经过 ALBERT 模型的运算后,取 CLS 对应位置的值作为分类的标签,得到模型的预测输出。最后将 GRU-attention 模型的输出与 ALBERT 模型的输出进行整合,得到整体模型的输出。

(二) 数据预处理

1. 数据集选择

对于股票数据,本文通过 Tushare^① 大数据开放社区提供的 api 接口,爬取了宁德时代(300750.sz)自 2021 年 7 月 1 日至 2022 年 8 月 3 日的股票数据信息,包括开盘价(open)、收盘价(close)、最高价(high)、最低价(low)等,在本实验中,取收盘价作为训练与预测数据来源。

对于金融新闻文本数据,本文首先爬取了云财经 2022 年 2 月 5 日至 2022 年 12 月 20 日共计 3200 条金融新闻数据,用作金融文本情感分类的模型训练;之后爬取了云财经、新浪财经、东方财富、同花顺自 2022 年 5 月 18 日至 2022 年 8 月 3 日发布的全部金融新闻,在其中筛选出包含“宁德时代”的新闻条目共计 318 条,作为后期对预测进行调整的数据来源。

2. 股票数据处理

实验首先将股票数据样本进行归一化处理,归一化区间为-1 至 1。之后选择 Daubechies8 小波进行小波去噪处理。随后将每 20 个数据分为一组,每组用前 19 个数据作为模型输入 x ,第 20 个数据作为预期结果 y 。再将所有的数据按照 8:2 划分为训练集与测试集,即可以得到训练集: x_{train} (模型输入)和 y_{train} (模型输出),测试集: x_{test} (模型输入), y_{test} (模型输出)。

3. 金融文本数据处理

首先对 3200 条金融新闻进行人工标注,标准为:对经济积极标注为 2,对经济消极标注为 0,中立标注为 1。之后取出其中的 800 条数据作为测试集,剩余 2400 条数据作为训练集,318 条包含“宁德时代”的金融新闻作为待预测文本数据。

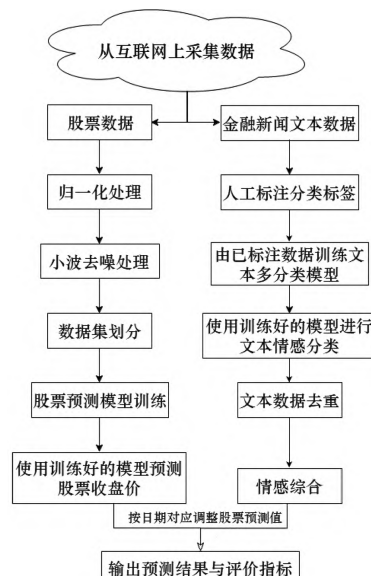


图1 算法流程图

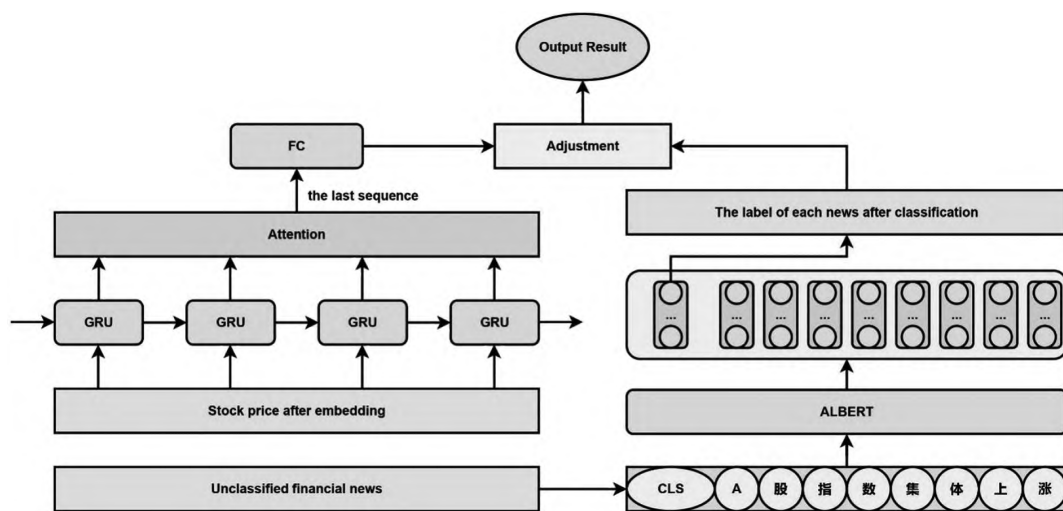


图2 模型结构

(三) 评价指标

本文提出的模型结构(图2)中,左路为传统股票预测模型,右路为金融文本情感分析模型,以下为

^①Tushare 大数据开放社区网址:https://www.tushare.pro/。

针对左右两路的模型分别制定的评价指标。

1. 股票预测模型评价指标

股票预测模型的评价指标分为两部分:值准确率和趋势准确率。

考虑到若使用常规的值准确率评价指标——即预测值必须与真实值完全相等才算做预测正确,将会使得准确率非常低,几乎没有参考价值,因此本文采取预测值在真实值上下一定范围内即算作预测正确的方式,避免准确率偏低。在下文中,统一称被算作预测正确的范围为准确范围。

对于准确范围的界定,出于对不同的股票数据集有较好的泛化能力的考虑,本文不直接指定准确范围的具体数值,而使用股票收盘价最大值与最小值之差的 7% 作为准确范围。公式表示为(1)、(2)。

$$prediction \in [actuality - 7\%(\max - \min), actuality + 7\%(\max - \min)] \quad (1)$$

$$accuracy_value = \frac{prediction_correct}{prediction_all} \quad (2)$$

其中, $prediction$ 表示预测值, $actuality$ 表示真实值, \max 表示最大值, \min 表示最小值, $accuracy_value$ 表示值准确率, $prediction_correct$ 表示预测值在准确范围内的数目, $prediction_all$ 表示测试集的值的总数目。

为了计算趋势准确率,需要计算涨幅。我们定义今日收盘价涨幅为 α , 今日收盘价为 $closing_t$, 昨日收盘价为 $closing_y$, 计算方式见公式(3)。

$$\alpha = \frac{closing_t - closing_y}{closing_y} \quad (3)$$

若 α 为正,则判断为股票上涨,设置标签为 1, 否则为股票下跌,设置标签为 0。根据真实值和预测值的涨幅,可以将样本划分为 TP (真实值与预测值均为上涨的数目)、 TN (真实值与预测值均为下跌的数目)、 FP (真实值为下跌且预测值为上涨的数目)、 FN (真实值为上涨且预测值为下跌的数目)四类,上述划分方式如表 1 混淆矩阵所示。进而可以按照公式(4)、(5)、(6)、(7)计算 $accuracy$ 、 $precision$ 、 $recall$ 、 f_1_trend 。其中,准确率 $accuracy$ 指正确预测的股票涨跌的比例;精确率 $precision$ 指模型在预测涨的情况下,实际上股票真的上涨的比例,精确率关注模型在预测为涨的情况下的准确性,避免过多的假阳

表 1 预测与真实情况类型表

	真实值上涨	真实值下跌
预测值上涨	TP	FP
预测值下跌	FN	TN

性,即误以为股票上涨;召回率 $recall$ 指实际股票上涨的情况下,模型成功预测为涨的比例,召回率关注模型对实际涨的情况的捕捉能力,避免过多的假阴性,即误以为股票跌了;而 f_1_trend 在平衡精确率和召回率之间取得折衷,对于在涨跌预测中寻求平衡的情况更为有用,避免过于偏向某个指标。

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$precision = \frac{TP}{TP + FP} \quad (5)$$

$$recall = \frac{TP}{TP + FN} \quad (6)$$

$$f_1_trend = \frac{2 \times precision \times recall}{precision + recall} \quad (7)$$

根据股票预测的实际价值,本文将两种指标的结果进行加权平均,得到最终指标 $final_score$, 值准确率的权值为 $weight_value$, 趋势准确率的权值为 $weight_trend$ 。计算公式见公式(8)。出于对实际行情研究重要程度的考虑,最终确定 $weight_value$ 为 0.7, $weight_trend$ 为 0.3。

$$final_score = weight_value \times accuracy_value + weight_trend \times f_1_trend \quad (8)$$

2. 文本情感分类模型评价指标

文本情感分类模型的评价指标为经典的文本情感多分类评价指标,即对每一种分类进行单独评估,对得到的二分类评估值以宏平均方式统计得到多分类的评估值。

(四) 两路模型训练

本文分别对左右两路的股票预测模型和金融文本情感分析模型进行了大量实验进行训练、筛选并分析原理,以下是模型选择过程和实验结果。

1. 股票模型选择

实验选择了、ResNet34、ResNet101、ResNet152、DNN、CNN、LSTM、LSTM-attention、GRU、GRU-attention、GRU-attention-sum、GRU-multi-head-attention 模型分别进行训练,通过调整模型参数得到效果最好的参数组合如表 2 所示。

DNN 为包含三个隐藏层的深度神经网络,CNN 为包含两个 3 核 Conv1d 和两个 2 核 AvgPool1d 的卷积神经网络^①,LSTM^②、GRU^③ 为经典的 RNN(循环神经网络)的改进版本,LSTM 包括遗忘门、输入门和输出门三个门结构,而 GRU 只有更新门和重置门两个门结构。

LSTM-attention、GRU-attention、GRU-attention-sum、GRU-multi-head-attention 均为与注意力机制^④相结合的 RNN 模型。LSTM-attention、GRU-attention 与 GRU-attention-sum 中的 Query 初始化方式为以 -0.1 至 0.1 上的均匀分布随机生成;GRU-attention-sum 在最后使用注意力机制中的加权求和,而 LSTM-attention 与 GRU-attention 则只加权,不求和;GRU-multi-head-attention 中的 Query 初始化方式为与 Key 相等。

2. 股票模型预测结果

上述每种模型的预测结果评价指标如表 3 所示。

表 2 效果最好的股票预测模型参数

model	learning-rate	epochs
ResNet34	0.01	300
ResNet101	0.01	100
ResNet152	0.01	300
DNN	0.01	100
CNN	0.001	500
LSTM	0.01	100
LSTM-attention	0.01	200
GRU	0.01	100
GRU-attention	0.01	100
GRU-attention-sum	0.01	100
GRU-multi-head-attention	0.01	500

表 3 不同模型的预测结果评估指标

model	accuracy_value	f1_trend	final_score
ResNet34	0.7551	0.5306	0.6878
ResNet101	0.6122	0.5306	0.5878
ResNet152	0.6531	0.5769	0.6302
DNN	0.6327	0.5098	0.5958
CNN	0.6531	0.5098	0.6101
LSTM	0.8163	0.5417	0.7339
LSTM-attention	0.5918	0.6415	0.6067
GRU	0.7959	0.4800	0.7011
GRU-attention	0.8776	0.5385	0.7758
GRU-attention-sum	0.7551	0.4348	0.6590
GRU-multi-head-attention	0.7755	0.5882	0.7193

由于篇幅原因,图 3 只展示出 CNN、LSTM、GRU、GRU-attention 的拟合图像。

其中,DNN 通过多层感知机结构进行训练,笼统地将定长序列输入模型,进而运算得出预测值,不能有效利用输入的时序信息,因此效果不好。

CNN 主要面向二维的数据样本,通过卷积计算来提取局部特征,虽然可以学习到部分时序信息,但其忽略了股票时序序列的单向性,且其并不能避免深度增加造成的梯度问题,因此效果依旧不佳。可以看到折线图中,CNN 在训练集和测试集的预测值曲线基本不会出现远大于或远小于真实值曲线的点,但预测值曲线的涨跌趋势与真实值曲线差异较大。

ResNet 系列作为针对梯度问题与网络退化问题的改进网络,通过残差连接的方式,将网络深度从几

①Lecun Y.,Bottou L.,Bengio Y.,et al.,“Gradient-Based Learning Applied to Document Recognition”,*Proceedings of the IEEE*,1998,86(11),pp.2278-2324.

②Hochreiter S.,Schmidhuber J.,“Long Short-Term Memory”,*Neural Computation*,1997,9(8),pp.1735-1780.

③Chung J.,Gulcehre C.,Cho K.,et al.,“Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling”,*Preprint Arxiv*,2014.

④Vaswani A.,Shazeer N.,Parmar N.,et al.,“Attention is All You Need”,*Advances in Neural Information Processing Systems*,2017,30,pp.5998-6008.

十层提高至几百层甚至上千层,但其并没有解决时序序列特征学习的本质问题,因此它的效果较 DNN 与 CNN 更加优秀,但不及 RNN。

LSTM 作为专门为时序序列预测任务而诞生的 RNN 的改进版,它通过遗忘门、输入门、输出门三个门结构来控制信息的保留程度,能够有效克服 RNN 中存在的梯度问题,尤其在长距离依赖的任务中的表现远优于 RNN,梯度反向传播过程中不会再受到梯度消失问题的困扰,可以对存在短期或者长期依赖的数据进行精确的建模^①,故其效果较 ResNet 有了较大提升。从图中可以看到 LSTM 的预测值曲线与真实值曲线几乎完全重合。

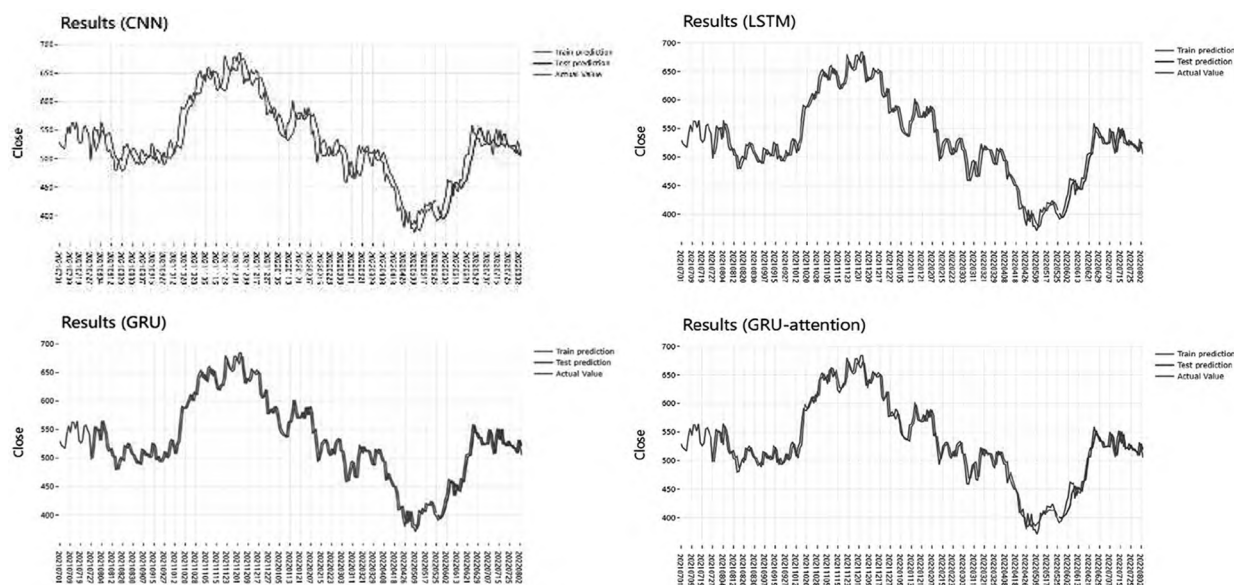


图3 CNN、LSTM、GRU、GRU-attention 的拟合图像

GRU 作为 LSTM 的改进简化版,它将输入门与遗忘门合二为一,比 LSTM 减少了一个门结构,因此参数量较少,训练速度很快,在数据规模较小时可以取得比 LSTM 更好的效果,但在本次股票预测任务中表现较 LSTM 稍差一些。GRU 的预测值曲线与 LSTM 的差别很小,都与真实值曲线几乎完全重合。

在引入注意力机制(attention)后,进一步加强了模型对于全体时序序列的特征提取与集中能力,但 LSTM 的效果有所下降,而 GRU 模型的效果明显提升,说明对于稍复杂的 LSTM 来说,引入注意力机制导致了对于数据的特征拟合过度,而稍稍增加 GRU 的特征提取能力则恰到好处。

对于 GRU 模型,实验又测试了引入注意力机制加权后是否求和对于效果的影响,结果表明,加权求和后的效果较差,远不及不求和。这是由于单步时序预测只取模型的最后一个输出作为输出预测,强行将其与之前的输出进行求和,会影响模型的判断;若继续增强特征提取能力,引入多头注意力机制(multi-head-attention),则可以发现 GRU 模型的效果有微小提升,但不如简单的注意力机制效果更好,因此可以更加确定,简单的注意力机制与 GRU 组合是非常适合的。

3. 金融文本分类模型选择与实验结果

实验使用 UER-py 工具^②,对其提供的 HFL Chinese BERT-wwm^③、Google Chinese BERT-Base^④、Google

①杨丽,吴雨茜,王俊丽:《循环神经网络研究综述》,《计算机应用》,2018 年第 7 期。

②Zhao Z., Chen H., Zhang J., et al., "UER: An Open-Source Toolkit for Pre-Training Models", Conference on Empirical Methods in Natural Language Processing, 2019, pp. 241-246.

③Cui Y., Che W., Liu T., et al., "Pre-Training with Whole Word Masking for Chinese Bert", IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29, pp. 3504-3514.

④Devlin J., Chang M. W., Lee K., et al., "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding", The North American Chapter of the Association for Computational Linguistics, 2019, 1, pp. 4171-4186.

Chinese ALBERT-Large^①、HFL Chinese RoBERTa-wwm-ext^②、HFL Chinese RoBERTa-wwm-large-ext^③、HFL Chinese BERT-wwm-ext^④、Google Chinese ALBERT-Base^⑤共计七个预训练模型进行增量预训练和下游任务微调,得到的结果如表 4 所示。

实验选用上述模型中效果最好的 Google Chinese ALBERT-Large 模型进行预测,得到包含预测标注的 318 条金融新闻。部分预测结果展示如表 5 所示。

表 4 不同文本分类模型的训练效果

model	accuracy	f ₁ -trend
HFL Chinese BERT-wwm	0.7025	0.6403
Google Chinese BERT-Base	0.7113	0.6536
Google Chinese ALBERT-Large	0.7313	0.6724
HFL Chinese RoBERTa-wwm-ext	0.7038	0.6483
HFL Chinese RoBERTa-wwm-large-ext	0.7013	0.6444
HFL Chinese BERT-wwm-ext	0.7063	0.6427
Google Chinese ALBERT-Base	0.6775	0.5859

表 5 部分金融新闻文本情感分类预测结果

date	text	label
20220609	云财经讯,创业板指跌幅扩大至 2%,沪指跌 0.5%,深成指跌 1.32%,两市超 3900 股飘绿,宁德时代跌超 5%。	0
20220701	宁德时代:公司有超快充技术,最快 5 分钟充至 80%电量。	1
20220713	创业板指强势拉升,涨幅超 1%,此前一度跌超 1%,宁德时代涨超 3%。	2

(五) 文本分类模型的特殊处理

1. 金融文本数据去重

考虑到不同媒体可能重复多次报导相同的新闻事件,导致 318 条包含“宁德时代”的金融新闻有对相同事件的重复描述,因此实验中根据新闻语句的相似度,进行了去重处理。限于篇幅原因,以下仅阐述去重算法的核心思想:

首先使用 jieba 分词将新闻语句拆分为词语序列,之后统计给定的两条语句中相同的词语个数。我们定义相似度为 β ,相同的词语个数为 $word_i$,语句 1 的词语总数为 $word_1$,语句 2 的词语总数为 $word_2$ 。通过公式(9)计算得到两条语句的相似度,相似度的取值范围为 $[0,1]$,越接近 1 说明两条语句越相似,反之则说明不相似。通过计算相似度,将金融新闻语句划分为不同的群组。在每个群组中,选择一条新闻作为代表,将其余与该代表相似度超过阈值的新闻视为重复,从中筛选出每个群组的唯一代表,形成去重后的数据集。通过这个算法,相似度高于阈值的新闻被认为是重复的,最终只保留每组相似新闻中的一条。

$$\beta = \frac{2 \times word_i}{word_1 + word_2} \quad (9)$$

2. 情感综合

考虑到若某一天的新闻总量很多,则会导致对那一天的股票数据调整幅度很大,我们不希望报导的数目成为调整幅度的决定性因素,因此需要对每一天的多条新闻进行情感综合,得到当天的总情感,再进行股票数据调整。对此我们提出了四种方案用于情感综合。

方案一:分别统计当天 0 类、1 类与 2 类的数量,取数量最多的作为当天的总情感;若 0 类与 1 类数量相等则归为 0 类,若 0 类与 2 类数量相等则归为 1 类,若 1 类与 2 类数量相等则归为 2 类。

方案二:对 0 类赋予数值 0,1 类赋予数值 1,2 类赋予数值 2,将当天的所有数值求和后除以当天的新闻数量,将结果四舍五入得到 0、1 或 2 中的一个整数,再将得到的整数重新对应映射回 0 类、1 类或 2 类,得到当天的总情感。

方案三:对 0 类赋予数值 0,1 类赋予数值 1,2 类赋予数值 2,将当天的所有数值求和后除以当天的新闻数量,结果在 $[0, \frac{2}{3}]$ 范围内映射为 0 类,结果在 $(\frac{2}{3}, \frac{4}{3})$ 范围内映射为 1 类,结果在 $[\frac{4}{3}, 2]$ 范围内映射为 2 类,得到当天的总情感。

^{①⑤}Lan Z., Chen M., Goodman S., et al., “Albert: A Lite Bert for Self-Supervised Learning of Language Representations”, International Conference on Learning Representations, 2020.

^{②③④}Cui Y., Che W., Liu T., et al., “Pre-Training with Whole Word Masking for Chinese Bert”, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, 29, pp. 3504–3514.

方案四:对 0 类赋予数值-1,1 类赋予数值 0,2 类赋予数值 1,将当天的所有数值求和后除以当天的新闻数量,得到一个介于[-1,1]范围内的小数,将其作为当天的总情感。

对于方案一、方案二与方案三,若当天的总情感为 0 类,则调整表达式为公式(10);若当天的总情感为 2 类,则调整表达式为公式(11);若当天的总情感为 1 类,则不作调整;对于方案四,调整表达式为公式(12)。

$$y_final=y_pred-y_pred\times weight \tag{10}$$

$$y_final=y_pred+y_pred\times weight \tag{11}$$

$$y_final=y_pred+y_pred\times weight\times emotion \tag{12}$$

其中, y_final 表示经过调整后的最终股票预测值, y_pred 表示未经过调整的股票预测值, $weight$ 表示调整权值,为可调整参数, $emotion$ 表示方案四中得到的总情感。

(六) 整体模型综合结果

实验使用对于股票预测结果最好的模型 GRU-attention 得到的结果作为 y_pred ,结合已经做好分类工作、去重工作和情感综合工作的金融新闻情感数据,分别将既不使用情感综合方案也不使用数据去重(记为 A)、使用不同情感综合方案但不使用数据去重(记为 B)、同时使用情感综合方案与数据去重(记为 C)这三种方式通过调整去重阈值、调整权值两个参数,得到的最好结果列举如表 6。

表 6 综合结果

ways	accuracy_value	f1_trend	final_score	RMSE	情感综合方案	去重阈值	weight
A	0.8776	0.5385	0.7758	14.75			
B	0.9388	0.5490	0.8218	14.31	方案一		0.005
B	0.9388	0.5490	0.8218	14.38	方案二		0.005
B	0.9388	0.5490	0.8218	14.30	方案三		0.005
B	0.9388	0.5385	0.8187	14.41	方案四		0.005
C	0.9388	0.5490	0.8218	14.27	方案一	0.7	0.005
C	0.9388	0.5490	0.8218	14.35	方案二	0.7	0.005
C	0.9388	0.5769	0.8302	14.15	方案三	0.7	0.005
C	0.9388	0.5490	0.8187	14.39	方案四	0.7	0.005

从结果可以看出,在使用情感综合方案后,最终的评价指标 $final_score$ 较既不使用情感综合方案也不使用数据去重提高了 5.53%至 5.93%,而在引入数据去重后,方案三进一步有所提升,较不使用数据去重提高了 1.02%。虽然在 $final_score$ 上来看,引入数据去重后的提升不明显,只有方案三有小幅度的提升,但从均方误差 $RMSE$ 上来看,引入数据去重后,所有方案的 $RMSE$ 均有所下降,下降幅度在 0.14%至 1.05%。

(七) 模型在不同股票数据集的综合表现

为了验证模型在不同股票数据集上的综合性能,实验又选取了贵州茅台、工商银行、招商银行、农业银行、中国石油、中国人寿、中国银行、中国平安、中国中免和海天味业共 10 支股票作为

表 7 贵州茅台测试结果

model	accuracy_value	f1_trend	final_score
ResNet34	0.9722	0.4167	0.8056
ResNet101	0.9167	0.3478	0.7460
ResNet152	0.9907	0.5000	0.8435
DNN	1.0000	0.4565	0.8370
CNN	0.9630	0.3404	0.7762
LSTM	0.9907	0.4742	0.8358
LSTM-attention	0.9537	0.3958	0.7863
GRU	0.9907	0.4211	0.8198
GRU-attention	1.0000	0.5161	0.8548
GRU-attention-sum	0.9907	0.4632	0.8325
GRU-multi-head-attention	0.9901	0.5051	0.8450

测试对象,分别使用表 7 中的 11 个模型进行训练与预测,并将得到的预测结果进行评估。表 7 列出了在贵州茅台股票数据上进行测试的结果,其他股票的测试结果形式与“贵州茅台”类似,因此不再分别列出。

将不同的模型按照 $final_score$ 由高到低排序后,分别给第一名至第五名赋予分数 $score_5$ 、4、3、2、1,其余的赋予分数 0,综合宁德时代与上述 10 支股票共计 11 支股票的预测情况,按照公式 13 的计算方式,得到每一个模型的综合得分 $total_score$ 。其中, $score_i$ 表示在第 i 支股票上该模型的得分。

$$total_score=\sum_{i=0}^{10} score_i \tag{13}$$

之后,我们对 $total_score$ 排名前 5 的模型在宁德时代股票数据上进行金融新闻综合调整,得到的最

优结果如表 8 所示。由表 8 可以看出,GRU-attention 结合金融新闻综合调整的表现依然是最好的。并且,在最优情况下,使用金融新闻情感综合方案三的数量是最多的,这说明方案三较其他方案更优。因此,对单支股票与多支股票的综合分析结果充分说明,GRU-attention 与金融新闻情感综合方案三的结合,具有最优的预测效果。

表 8 最终结果

model	score	accuracy_value	f1_trend	final_score	情感综合方案	去重阈值	weight
GRU-attention	34	0.9388	0.5769	0.8302	方案三	0.7	0.005
GRU-attention-sum	31	0.8776	0.5306	0.7735	方案三	0.6	0.015
LSTM	25	0.8980	0.5882	0.8050	方案三	0.5	0.022
CNN	18	0.6939	0.5306	0.6449	方案一	0.6	0.025
GRU-multi-head-attention	16	0.7959	0.6275	0.7454	方案一	0.6	0.015

三、讨 论

本文分析了为什么结合 GRU 和 ALBERT 的预测模型在股票预测问题上可以表现得较其他模型优秀,对该模型的作用解读如下。

GRU 模型以其出色的序列建模能力在股票预测中扮演着关键角色。其能够有效捕捉股价时间序列中的非线性模式和长期依赖关系。考虑到股市波动的复杂性,采用 GRU 模型来学习股价变化的动态过程。GRU 模型的门控机制通过更新和重置门,有助于解决长期依赖关系,使模型更好地适应市场的长期趋势,从而提高了预测准确性。与主流时间序列模型 LSTM 相比,GRU 的门控结构相对简单,参数较少,减少了过拟合风险。在较小的数据集中,GRU 更容易泛化到未见过的数据。在股票预测任务中,很多时候短期信息更为关键,GRU 的简化结构和较少的参数量使其更容易捕捉隐藏的短期趋势和模式。

在对股票相关的新闻文本进行分析时,我们使用基于 Transformer 的大语言模型。相较于传统方法,它已被证明在情感信息提取方面更为强大。该模型通过深入挖掘文本的语义关系,能够更准确地评估新闻对市场的实际影响。模型理解文本中的隐含信息和语境,进而辨别对市场情绪有重要影响的关键信息。这种综合分析使得模型能更准确地量化新闻对股票价格的潜在影响。

总体而言,基于新闻市场的情感分析使股票预测模型更全面、深入地考虑市场的客观趋势和主观情感因素。情感分析提高了模型的信息涵盖范围,缓解了股票预测中的滞后性问题,同时提供了更及时、灵敏的信息,有助于更早地捕捉市场的变化信号。模型还具有动态调整权重的机制,使其更灵活地适应不同市场环境的变化。

四、结论与展望

本文综合考虑了模型和特征两个方面,提出了一种融合客观特征和情感特征的股票预测模型,相较于传统方案更为丰富。我们采用了以股价序列为输入的循环神经网络(RNN)框架作为主体,结合媒体新闻提取的情感得分,通过自适应调整 RNN 输出结果,从而得到更准确的股票预测结果。

实验以“宁德时代”股票为例,并与主流预测模型进行对比。结果显示,在只考虑客观特征时,引入注意力机制的门控循环单元(GRU)模型表现最佳。通过引入情感特征的调整,我们进一步提升了模型的预测性能。将该框架应用于其他股票的测试中,结果表明其在各类股票预测中均表现出色,验证了其普适性和高预测性能。然而,模型仍未完全解决滞后性问题。未来的研究方向将集中在进一步开发情感特征,探索更多文本情感来源,并设计更精密的情感分析方法,以更好地克服滞后性。

总体而言,本文提出的模型框架不仅在实证检验中表现出卓越的预测性能,而且具备较高的通用性。未来的工作将进一步深化对情感特征的研究,并探索其他可能影响股票走势的因素,以使模型更加全面、准确地预测股市动态。

[责任编辑:王成利]