

# 基于机器学习的股票预测方法研究

毛月月<sup>1</sup>,张秋悦<sup>2</sup>

(1. 贵州师范大学大数据与计算机科学学院, 贵阳 550025;

2. 贵州师范大学数学科学学院, 贵阳 550025)

## 摘要:

股票市场预测是金融界最热门、最有价值的研究领域之一,人们在早期仅使用结构化的数据即历史股票数据对股市进行预测。随着深度学习和自然语言处理技术的发展,研究者们开始探索金融新闻、社交软件上股民的评论等信息对股市价格的影响,从而使得预测的结果更好。从以上两种角度分别对各自使用的股市预测方法进行综述,最后对结果进行总结和展望。

## 关键词:

股票预测; 历史数据; 金融新闻; 机器学习

## 0 引言

股票市场是国家经济和金融活动的重要体现,在各国的经济发展中发挥着重要的作用,同时对于国家、企业、个人都有着重要的影响。如果能够提前掌握股市的动向,将会给投资机构和投资者带来收益。但是由于股票市场的不确定性和波动性,如果股民盲目投资,将会带来不可计量的损失。因此,股票的预测一直是一项具有挑战性和重要意义的任务。

股票预测,作为一个交叉的研究方向,数学家、经济学家、计算机领域的研究者们都在这个领域做了相当多的工作。这些工作主要分为两类,第一类方法是仅使用股票的历史数据<sup>[1-6]</sup>去预测股市,其忽略了影响股市的关键来源:金融新闻和社交平台的股民评论。近年来,大数据的计算技术、人工智能等技术的发展,促使人们不仅可以处理结构化的数据,对于非结构化的文本数据也能很好的处理。深度学习技术的广泛使用,使得自然语言处理的能力逐渐提升。越来越多的研究者开始使用数据挖掘技术从大量金融新闻、社交媒体中挖掘重要信息,研究其与股票历史数据的相关性,通过结合两者一起预测股市的波动情况,从而为股

民们提供建议。下面从以上两种方法对相关的工作进行介绍。

## 1 历史股票数据的股市预测

金融领域的许多分析和假设表明,股票市场是可预测的。研究者们通过对过去股票市场的数据进行研究来预测股市,下面介绍其预测使用的方法。

### 1.1 传统的机器学习方法

Tang C 等人<sup>[1-2]</sup>都使用分段线性表示(PLR)来确定历史数据的波谷或波峰。前者结合高斯过程分类来预测股票价格的转折点。后者结合随机森林去分类和预测转折点,最后使用深度递归神经网络(DRNN)设计投资决策模型。

传统的模型都是基于历史数据的统计特征和图像特征来分析数据,其只通过默认值固定每一个时间窗口来获取股票数据中不同指标的特征,这样预测的结果精度不高。RIT Ritzmann Junior 等人<sup>[3]</sup>提出使用遗传算法来最优化时间窗口的值,然后使用支持向量机(SVM)来决定未来交易是购买、持有还是出售。

传统的方法大多数使用线性模型来预测股票,例

如支持向量机、决策树、随机森林等。这种方法的优点是仅使用结构化的数据,数据的处理比较简单且运算速度快。这种线性统计模型的主要缺点是无法解释各种公司股票价格之间存在的潜在动态和关系,而且股票数据是带时间属性的,这种方法也不能捕获其时序特征。

## 1.2 深度学习的方法

CHEN Y 等人<sup>[4]</sup>提出深度学习的方法去预测股票的变化趋势和持续时间。使用卷积神经网络(CNN)去提取短期的特征,使用 PLR 对时间序列数据进行分割,提取长期的特征,最后通过双重注意力机制去关注长短期特征对股票预测的影响。

ZHANG K 等人<sup>[5]</sup>提出使用生成对抗网络来预测股票市场,其以多层感知机(MLP)为鉴别器,长短期记忆网络(LSTM)为生成器来预测收盘价,这是一种新方法的突破,但是其仅仅和基准的方法做了对比,值得继续深入和改进。

这种方法的优点是能够捕获到股票数据的时序特征,缺点是由于仅使用单一的历史股票数据,其预测结果精度不高,忽略了突发事件对股市造成的即时影响。

## 2 历史股票数据与带时间属性的多文档相结合的股市预测

股票市场中,股市的波动受国家宏观经济发展、法律法规,以及股民的情绪等因素的影响。因此,结合历史股票数据以及相应的金融新闻和股民的评论进行分析具有重要意义。基于对新闻文本的特征提取方式进行划分,使用的方法分为如下三个类型。

### 2.1 新闻事件的结构化表示

Xiao Ding 等人<sup>[6]</sup>证明了使用结构化的事件去预测股票的有效性。其总结了前人的不足,即以前关于新闻驱动的股票市场预测的工作通过词袋模型、命名实体识别和名词短语的方式提取文本特征,得到的仅是文本浅层的特征,这些特征不能捕获结构化的实体关系信息。其提出采用事件的结构化表示即:  $E = (O_1, P, O_2, T)$ , 其中  $O_1$  代表参与者,  $P$  代表行动,  $O_2$  代表执行动作的对象,  $T$  代表时间戳,用于将股票数据与新闻数据对齐。以此来实现股票涨、跌的预测。

Xiao Ding 等人<sup>[7]</sup>在 2015 年提出一个基于深度学习的事件驱动型股票预测模型。首先,将大型的财

经新闻语料转换为词向量的形式。其次,不同于文献[6],这里使用神经张量网络去获取新闻的事件表示,然后将提取到的具有时间顺序的向量表示作为输入,采用卷积神经网络(CNN)去提取长、中、短期的特征,长期使用 30 天的数据,一天的事件通过取平均值获得。将其按照时间顺序排列,设置一个滑动窗口,同一个窗口内的事件进行卷积操作。之后,将所有卷积层输出的向量进行最大池化(获得局部的最大特征),得到最终的长期事件向量。中期时间定义为过去的一周,方法和上面类似,短期事件使直接将前一天的所有事件求平均。最终得到长、中、短期的三个事件向量,将三者拼接为一个向量表示,联系这个特征向量和股票价格,得到最后的分类结果即股票价格的涨跌。

### 2.2 新闻的情感极性分析法

Dev Shah 等人<sup>[8-12]</sup>都提出基于词典的新闻情绪分析模型来研究新闻文章中情感表达对股票的影响。首先需要建立一个情感词典,然后对新闻文本进行情感极性分析,最后结合新闻的情感极性、股票历史数据以及相邻日收盘价的方差来进行股市的预测。

Manoj S Hegde 等人<sup>[13]</sup>提出利用公司历史股价和金融新闻及社交媒体内容的情绪得分进行预测。与上述方法不同的是,其使用了 LSTM 模型去捕获数据中的时间序列信息,同时还提出了基于 CNN 分类器的推荐模块,最后采用背包投资模型来使投资者收益最大化,投资最小化。

Luca Cagliero 等人<sup>[14]</sup>提出了一种将新闻信息、历史股票价格相结合的趋势逆转方法来量化股票交易。文章中主要研究三个问题,在基于分类的趋势逆转方法中新闻信息是否值得考虑? 新闻信息对于股票预测有什么影响? 什么特征应该结合新闻情感分析去最大化趋势逆转的平均回报率? 其使用方法也是对新闻进行情感分析,不同的是其提出了基于股票预测分类的趋势反转策略。其监控每只股票的每日收盘价格变化,以检测用户在持续时间的上升趋势或下降趋势,最后针对上一步检测到的每一个趋势,训练分类器以便于预测未来五天的收盘价格变化。

### 2.3 新闻文本向量的直接表示结合深度学习方法进行特征提取

Ryo Akita 等人<sup>[15]</sup>提出将每一篇新闻文档表示成一个固定长度的向量,然后将所有文章的向量表示连接

起来,同时与经过标准化处理的股票价格一同送入长短期记忆网络(LSTM)中捕获文本和股票的时序信息,最后通过回归输出预测值。LSTM 捕获文本和股票的时序信息时,LSTM 的输入是文本序列和股票价格序列的连接,其存在一个问题,即文本信息的表示维度远大于股票数值表示的维度,其容易造成文本信息对于预测的影响更大,导致预测精度的不平衡。为了解决这个问题,文章提出对向量的大小进行缩放,以使每个向量具有相同的维数。

Che-Yu Lee 等人<sup>[16]</sup>提出基于递归神经网络的股票预测模型,采用 Word2Vec 将新闻标题转化为词向量的形式,然后使用 CNN 提取文本的局部特征,将提取的文本特征向量和过去的股票价格送入到 LSTM 中,其可以捕获金融新闻和价格的长期时间依赖性,最后预测输出股票类别,分别是上升、下降、持平。

结合非结构化的文本数据(金融新闻、社交媒体数据)和结构化的历史数据(股票数值数据)来对股市进行预测时,如何从众多文本数据中挖掘出对于股票价

格影响的信息十分重要。同时,如何把非结构化的文本数据表示成轻量、精简而不失其语义信息的抽象特征十分重要。

### 3 结语

以上就是对近年来股票预测方法的总结,我们可以看出使用股票的历史数据去预测股市,其忽略了市场波动中金融新闻和公众评论所造成的影响。结合历史数据和金融新闻和公众评论去预测股票时,如何准确、快速地从文本中挖掘出关键信息十分重要。基于深度学习的预测方法能更好地提取文本和股票数据的特征,特别是循环神经网络其能够捕获数据的时序特征。

股票预测领域大多数是确定并跟踪持续的价格趋势,这个趋势在未来几天将会持续。相反,利用机器学习算法预测股市趋势的逆转的研究相对较少,未来可以在这个方向上深入研究。

#### 参考文献:

- [1]李丰,高峰,寇鹏.基于分段线性表示和高斯过程分类的股票转折点概率预测[J].计算机应用,2015,35(08):2397-2403.
- [2]TANG C,ZHENG X,Yu X,et al. Design and Research of Intelligent Quantitative Investment Model Based on PLR-IRF and DRNN Algorithm,2018[C]. IEEE,2018.
- [3]RIT Ritzmann Junior, Norberto&Nievola, Julio. A Generalized Financial Time Series Forecasting Model Based on Automatic Feature Engineering Using Genetic Algorithms and Support Vector Machine,2018.
- [4]CHEN Y, LIN W, WANG J Z. A Dual-Attention-Based Stock Price Trend Prediction Model with Dual Features[M]. DOI10.1109/ACCESS.2019.2946223,IEEE Access,2019.
- [5]ZHANG K, ZHONG G, DONG J, et al. Stock Market Prediction Based on Generative Adversarial Network[J]. Procedia Computer Science,2019,147:400-406.
- [6]DING X,ZHANG Y. Using Structured Events to Predict Stock Price Movement:An Empirical Investigation[M]. EMNLP,2014.
- [7]DING X,ZHANG Y. Deep Learning for Event-Driven Stock Prediction[M]. IJCAI,2015.
- [8]SHAH D, ISAH H, Zulkernine F. Predicting the Effects of News Sentiments on the Stock Market. IEEE,2018.
- [9]Ramalingam V V, Pandian A, DWIVEDI S. Analysing News for Stock Market Prediction[M]. J.Phys.:Conf.Ser.1000 012026,2018.
- [10]KIM J, SEO J, LEE M. Stock Price Prediction Through the Sentimental Analysis of News Articles[M]. IEEE,2019.
- [11]Shankar P, Arora M A. Stock Price Forecasting by Combining News Mining and Time Series Analysis. IEEE Computer Society,2019.
- [12]Hegde M S, Krishna G, Srinath R. An Ensemble Stock Predictor and Recommender System. IEEE,2018.
- [13]Kalra S, Prasad J S. Efficacy of News Sentiment for Stock Market Prediction:International Conference on Machine Learning, Big Data, Cloud and Parallel Computing[C],2019.
- [14]Attanasio G, Cagliero L, Garza P, et al. Combining News Sentiment and Technical Analysis to Predict Stock Trend Reversal. IEEE,

2019.

[15]Akita R, Matsubara T. Deep Learning for Stock Prediction Using Numerical and Textual Information[M]. IEEE, 2016.

[16]Lee C, Soo V. Predict Stock Price with Financial News Based on Recurrent Convolutional Neural Networks[M]. Conference on Technologies and Applications of Artificial Intelligence (TAAI). IEEE, 2017.

#### 作者简介:

毛月月(1995-),女,贵州贵阳人,研究生,研究方向为自然语言处理

张秋悦(1995-),女,重庆永川人,研究生,研究方向计算数学、自然语言处理

收稿时间:2020-04-14 修稿时间:2020-05-20

## Research on Stock Prediction Method Based on Machine Learning

MAO Yue-yue<sup>1</sup>, ZHANG Qiu-yue<sup>2</sup>

(1. School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550025;

2. School of Mathematical Sciences, Guizhou Normal University, Guiyang 550025)

#### Abstract:

Stock market forecasting is one of the most popular and valuable research areas in the financial world. In the early days, people used only structured data, historical stock data, to predict the stock market. With the development of deep learning and natural language processing technology, researchers have begun to explore the impact of financial news, social commentary on social software and other information on stock market prices, In order to making predictions better. The article summarizes the respective stock market forecasting methods used from the above two perspectives, and finally summarizes and prospects the results.

#### Keywords:

Stock Prediction; Historical Data; Financial News; Machine Learning