

文章编号:1671-4229(2021)01-0044-12

基于 GA-Transformer 模型的多因子股票预测

陈诗乐^{a,b}, 王笑^b, 周昌军^{a*}

(浙江师范大学 a. 数学与计算机科学学院; b. 行知学院, 浙江 金华 321000)

摘要: 股票市场中拥有大量用于描述股票价格变化的财务指标, 这些指标为股票价格预测提供了良好的数据基础. 但由于股票数据存在高维相关性和时序性等特点, 导致精确预测股票价格存在困难. 为提高股票价格预测精度, 文章提出了基于 GA-Transformer 模型的多因子股票预测方法, 该方法使用遗传算法(GA)进行特征选择, 并结合 Transformer 模型进行股票预测, 提升了模型特征抽取能力. 实验结果表明, GA-Transformer 模型在包括建设银行和贵州茅台等六支股票数据集上的预测表现均优于股票预测主流模型.

关键词: 遗传算法; 多因子股票预测; Transformer 模型

中图分类号: F 832. 51; TP 181 **文献标志码:** A

Multi-factor stock forecasting method based on a GA-Transformer model

CHEN Shi-le^{a,b}, WANG Xiao^b, ZHOU Chang-jun^{a*}

(a. School of Mathematics and Computer Science; b. Xingzhi College, Zhejiang Normal University, Jinhua 321000, China)

Abstract: In the domestic stock market, there are a large number of financial indicators used to describe the change of stock prices, which provide a good data basis for stock price prediction. However, it is difficult to accurately predict stock prices because of a high dimensional correlation and the time sequence of stock data. In order to improve the accuracy of stock price prediction, this paper proposes a multi-factor stock prediction method based on a GA-Transformer model, which uses a genetic algorithm (GA) for feature selection and a Transformer model for stock prediction, and improves the ability of feature extraction. The experimental results show that the performance of the GA-Transformer model on six stock data sets, including CCB and Kweichow Moutai, is better than the mainstream models.

Key words: GA; multi-factor stock prediction; Transformer model

股票预测一直是社会关注的热点, 对于投资者而言, 对股票价格进行准确预测有利于对未来利润的获取和对风险的规避. 学者们致力于研究股票价格的变化规律, 提出了一些有效的股票预测方法, 并获得了一定的成果. 这些股票预测方法按先后发展顺序可划分为基于传统时序建模方法、基于机器学习方法和基于深度学习方法.

基于传统时序建模方法主要依靠统计学理论, 建立股票特征与股票价格之间的线性模型, 包

括早期的移动平均法、指数平均法和后来的自回归模型(AR)、滑动平均模型(MA)以及差分整合移动平均自回归模型(ARIMA). 这类方法对数据分布规则和完整性等方面要求非常严格, 并且认为引起股票价格变化的因素间仅存在线性关系. 然而事实证明, 股票价格变化往往与其影响因素之间存在复杂的非线性关系, 这也导致了传统时序建模方法最终预测结果精度不理想. 基于机器学习方法, 研究人员使用随机森林算法、xgboost 算

基金项目: 国家自然科学基金资助项目(62006106, 61672121)

作者简介: 陈诗乐(1994—), 女, 硕士研究生. E-mail: 1227437310@qq.com

* 通信作者. E-mail: zhou-chang231@163.com

引文格式: 陈诗乐, 王笑, 周昌军. 基于 GA-Transformer 模型的多因子股票预测[J]. 广州大学学报(自然科学版), 2021, 20(1): 44-55.

法和向量机算法等进行股票预测均取得了比 ARIMA 模型更好的效果,因为这些经典的机器学习算法均能够模拟股票特征与价格间复杂的非线性关系.本世纪初以来,随着计算机性能的提升和神经网络建模的日趋成熟,深度学习技术在特征抽取和构建复杂非线性模型上取得了巨大发展,研究人员运用深度神经网络(DNN)、卷积神经网络(CNN)以及长短期记忆神经网络(LSTM)等深度学习网络解决股票预测问题,取得了较好的效果,这也导致了深度学习理论在股票领域的相关研究越来越多.

1 股票预测研究现状

早期股票预测方法主要是基于统计学和经济学理论对股票价格进行分析,选取较为单一的股票特征数据进行股票预测,比较常用的有自回归模型(AR)、移动平均自回归模型(ARMA)及差分整合移动平均自回归模型(ARIMA)等.基于统计学的自回归模型,因其特征单一,在计算效率上有一定优势,但模型过于简单且输入维度低,限制了其对非线性非平稳金融时序序列的建模能力,以及对异常数据的抗干扰能力.

随着计算机技术的发展,研究人员逐渐开始将机器学习技术应用于股票预测问题当中,金融市场多年来累积的海量股票数据也为机器学习提供了充足的数据基础.Cao等^[1]考虑金融时间序列的非平稳性,建立了能够自适应优化参数的支持向量机预测模型.仿真结果表明,该模型能够降低陷入局部最优问题的可能性并提高模型泛化能力.Nair等^[2]提出了一种基于粗糙集的决策树系统,该方法以孟买证券交易所数据进行实验研究,使用C4.5决策树和粗糙集进行特征提取和规则总结.实验结果表明,该系统优于基于朴素贝叶斯的股票预测系统,但在处理噪声较大的数据时容易发生过拟合,影响股票预测精度.

从理论上讲,人工神经网络(ANN)可以学习任何非线性关系,且受噪声数据的干扰较小,因此,在时序预测领域得到了广泛的关注.1988年,White^[3]首次提出使用人工神经网络预测股票价格,但是由于梯度消失的问题,该方法并没有取得理想的效果.随着时间的推移,ANN在股票预测问题上的实用性被更多研究所证明,如Gen^[4]和

Zhang^[5]分别在各自的研究中,证实了基于ANN的非线性模型在股市预测问题上的表现优于线性预测模型.同时,ANN模型还可以与其他方法相结合来进行股票预测,如Mohammad等^[6]将ARMA模型、BP神经网络和多层感知机(MLP)等模型相结合进行股票预测研究,该混合模型取得了较好的预测效果.然而Wei等^[7]的研究表明,因为股票市场数据具有非线性、高维度和高噪声的特点,ANN在这些数据上表现出了浅层神经网络容易陷入局部最优解以及发生过拟合的问题.

随着人工智能技术的发展,深度学习因其在机器翻译^[8]、语音情感识别^[9]、图像识别^[10]等方面的优异表现而受到广泛关注.与传统的统计模型相比,深度神经网络(DNN)可以通过分层特征表示来分析深度复杂的非线性关系,具有比浅层神经网络更好的特征抽取能力.研究表明,深度学习方法相比于过去的传统机器学习方法,在股票预测问题上有更高的准确度,如Abe等^[11]将深度学习网络模型与传统神经网络模型分别运用于股票预测,结果表明,深度学习网络模型能够取得更好的预测效果.Tsantekidis等^[12]提出了一种基于卷积神经网络(CNN)的股票预测模型,并与其他经典模型进行比较,验证了卷积模型在股票预测中的有效性.Fischer等^[13]通过LSTM模型建立股票预测模型,证明了该模型在股票预测问题上的表现优于随机森林、深度神经网络(DNN)和逻辑回归模型.Selvin等^[14]分别提出了3种基于CNN、循环神经网络(RNN)和LSTM的股票预测模型,并通过对外上市公司股票价格的预测,比较了3种模型的性能.最后得出结论:LSTM具有较好的记忆性,最适用于股票时序预测.近年来,注意力机制作为一种高效的权重优化算法常被用于深度学习模型的优化,如Qin等^[15]提出一种基于双阶段注意力机制的递归神经网络(DA-RNN)用于特征提取和序列预测,通过这种采用双阶段注意力机制方案,可以有效地预测股票价格.受注意力机制的启发,Ashish等^[16]构建了基于多头自注意力机制的Transformer模型,该模型在时序任务中取得了极大的成功.此后,一些学者开始将Transformer模型运用于股票预测问题的研究,如Liu等^[17]提出了一种基于Transformer模型的胶囊网络进行股票走势预测,其通过Transformer模型编码器提取社交媒体深层语义特征,然后使用胶囊网络捕捉

文本结构信息,结果表明,该方法能够有效预测股票价格走势. Molmann 等^[18]通过 Transformer 模型研究了未来货运协议与黄金海洋集团股票价格之间的预测关系,结果表明,该模型对黄金海洋集团股票价格具有一定预测能力,其预测精度超过了所有对比模型. Ding 等^[19]通过引入高斯先验、正交正则化和交易间隙分配器优化了 Transformer 股票预测模型,并对中国 A 股市场股票价格进行了有效预测. 基于上述研究成果,本文提出了基于遗传算法和 Transformer 模型的两阶段股票预测方法,该方法首先通过遗传算法根据因子重要性排名获得多个股票多因子集合,然后在不同股票多因子集合上分别使用 Transformer 模型进行股票预测,并通过实验对比获得了较好的预测结果.

2 GA-Transformer 多因子股票预测模型的构建

2.1 股票多因子分析及选取

股票价格的影响因素主要有公司内部因素、国家政策因素和市场因素^[20]. 公司内部数据通常只有在季度报表和年终总结才能够体现,国家政策也不是每天都会颁布,因此,它们都具有滞后性和间歇性,无法通过实时的量化数据进行反映,不适用于进行股票价格预测研究. 而股票市场每天都会进行交易,从而产生大量的数据,尤其是其中用于反映股票价格变化的各种财务指标,非常适用于作为股票预测研究的训练数据,是研究人员重点关注的对象^[21].

股票市场财务指标除了常见的开盘价、收盘价、最高价、最低价、成交量和成交金额等交易型指标以外,还有一些人工构建的更具建模意义的综合技术指标. 以下简要介绍几个常见的技术指标.

(1) 相对强弱指标(RSI): RSI 是一个信号指标,取值范围 0~100,两个临界点分别为 30 和 70. 当 RSI 小于 30,提示超卖信号,考虑买入;当 RSI 大于 70,提示超买信号,考虑卖出.

(2) 移动均线(MA): MA 是一种观察价格变动趋势的信号指标,所谓移动均线是把某段时间的股价加以平均,再依据这个平均值做出均线图像,将真实股价与移动均线进行对比,高于移动均线,考虑买入,低于移动均线,考虑卖出.

(3) 平滑异同移动均线(MACD): MACD 是一种趋势型指标,由一条 12 天的指数平滑快线和一条 26 天的指数平滑慢线之差,再以 9 天为窗口做指数平滑得到. 当 MACD 由负转正时,代表股价未来有可能上涨,考虑买入;反之,当 MACD 由正转负时,考虑卖出.

由于篇幅有限,没有对所有的股票技术指标进行详细介绍,为了对股票市场总体财务指标有更清晰的认识,将股票市场主流的技术指标大致分为 8 组并对其做出解释,描述如下:

(1) 趋势型指标

趋势型指标以两条线交叉为信号,能够反映股票价格的上涨或下降趋势,是投资者在研究股票价格变动时的第一选择,也是最容易让股民获利的指标. 趋势型指标主要包括震动升降指标(ASI)、指数均线(EXPMA)、佳庆指标(CHO)、简易波动指标(EMV)、移动方向指数(DMI)、区间震荡指标(DPO)、偏离移动均线(DMA)、变异平滑异同移动平均(QACD)和平滑异同移动均线(MACD)等.

(2) 能量型指标

能量型指标是股价热度的温度计,反映了股民情绪高亢或沮丧,指标数据过高代表高亢发烧,指标数据过低代表沮丧发冷. 能量型指标主要包括逆时钟线(PVC)、潜量消长指标(MAR)、梅斯线(MASS)、佳庆变异线(VCI)、心理线(Psy)、成交量变异率(VR)、人气指标(AR)、意愿指标(BR)和中间意愿指标(CR)等.

(3) 超买超卖型指标

超买超卖型指标是一种专门用于研究股票指数走势的中长期技术分析指标,通过计算一定时期内市场涨跌股票数量之间的相关差异性,了解整个市场买卖气势之强弱,以及未来大势走向如何. 超买超卖型指标主要包括商品路径指标(CCI)、随机指标(KDJ)、资金流量指标(MFI)、动量线(MOM)、变动速率线(OSC)、变动速率指标(ROC)和相对强弱指标(RSI)等.

(4) 均线型指标

均线型指标是用于反映价格运行趋势的重要指标,其运行趋势一旦形成,将在一段时间内继续保持,为投资者提供买入或卖出的有利时机. 均线型指标包括移动均线(MA)、变异均线(VMA)、高价均线(HMA)和低价均线(LMA)等.

(5)成交量型指标

成交量型指标是证券市场技术分析中用成交量数据作为主要参数的大类指标.成交量型指标主要包括离散量(ADV)、负量指标(NVI)、累计能量线(OBV)、正量指标(PVI)、波幅价量(PVT)及邱氏量法(CHIU)等.

(6)选股型指标

选股型指标是一种帮助投资者筛选高报酬个股的技术指标,指标数值越高,表示其投资价值越高.选股型指标包括股票选择指标(CSI)、趋向比例(DX)、现值率(TAPI)、威力雷达(RAD)以及邱氏曲腰量(SV)等.

(7)大势型指标

大势型指标是一种反应大盘走向与趋势的技术指标,其不对个股的涨跌提供信号,投资者可以

通过其了解股市人气状况,探测股市内在的资金动能是强势还是弱势,判断股市未来动向.大势型指标主要包括绝对广量指标(ABI)、腾落指标(ADL)、涨跌比率(ADR)、阿姆氏指标(ARMS)、广量动力指标(BTI)、钱龙中线指标(CA)、估波指标(COP)及麦克连指标(MCL)等.

(8)压力支撑型指标

压力支撑型指标是一种反应股价上限和下限的技术指标,上限代表压力,下限代表支撑,股价触碰上线会回落,股价触碰下限会反弹.压力支撑型指标主要包括布林线(BOLL)、轨道线(ENV)及风林火山(FIRE)等.

选取国内股票市场中较为典型的 40 个财务指标作为原始特征用于后续模型研究,所有指标汇总如表 1 所示.

表 1 股票多因子选取汇总

Table 1 Summary of stock multi-factor selection

| 因子符号 | 因子释义 | 因子符号 | 因子释义 | 因子符号 | 因子释义 | 因子符号 | 因子释义 |
|-------|---------|-------|---------|------|--------|-------|---------|
| EMA | 指数移动平均值 | EMV | 简易波动指标 | RSI | 相对强弱指标 | ROC | 变动速率指标 |
| HMA | 高价平均线 | MACD | 异同移动平均线 | ATR | 均幅指标 | SSL | 压力支撑线 |
| DKX | 多空线 | CCI | 商品路径指标 | AMO | 成交金额 | PVT | 波幅价量 |
| Low | 最低价 | OBV | 累计能量线 | CR | 中间意愿指标 | WVAD | 威廉变异离散量 |
| OSC | 变动速率线 | DMI | 移动方向指数 | MA | 移动平均线 | KBJ | 随机指标 |
| MAUDL | 平均引力线指标 | BBI | 多空指标 | AR | 人气指标 | BB | 布林极限指标 |
| CHO | 佳庆指标 | High | 最高价 | BR | 意愿指标 | PSY | 心理线 |
| BIAS | 乘离率 | MAFSL | 平均分水岭线 | MAR | 潜量消长指标 | MACHO | 平均佳庆指标 |
| UDL | 引力线 | TRIX | 三重指数平均数 | MOM | 动量线 | MASS | 梅斯线 |
| FLS | 分水岭线 | ACCER | 幅度涨速 | ABI | 绝对广量指标 | Open | 开盘价 |

2.2 GA 特征组合构建

遗传算法是一种基于自然选择和遗传进化思想的自适应启发式搜索算法,广泛用于寻找搜索空间大的优化问题的近似最优解,可以有效地用于优化特征的选择.它将一个问题的潜在解决方案编码成个体,即染色体,所有个体组成的集合称为种群,遗传算法的优化过程就是在种群中进行.染色体是遗传物质的主要载体,代表某些基因的组合,其决定了个体的外在表现,如黑头发的特征是由控制这一特征的染色体上某些基因组合决定的.遗传算法在初始化阶段,需要将问题的解进行编码形成表现型到基因型的映射,较为常用的编码方式为二进制编码.在种群初始化之后,遗传算法根据问题域内个体的适应度进行个体选择,并利用遗传算子结合交叉和变异生成代表新解集的种群,这个过程会导致种群进化,类似于生物

种群的进化,每一代新种群中的个体将比上一代种群的个体更适应于外在环境.解码后,最后一代种群中的最优个体可以作为问题的近似最优解.

对选取的 40 维因子展开研究,从中获得最优的股票多因子特征组合,达到提高股票预测精度以及加快训练速率的目的.具体做法是使用遗传算法对 40 维多因子进行重要性排名,按排名顺序取不同数量的股票因子组合作为训练模型的输入,进行股票预测分析.

图 1 说明了通过遗传算法进行特征选择的整个过程.遗传算法可分为编码、初始化、适应度评估、终止条件检查、选择、交叉和变异 7 个阶段^[22]. $[\alpha_1, \alpha_2 \dots \alpha_n]$ 代表原始的特征集, β 代表随机产生的一条染色体.使用遗传算法获得因子重要性排名进行特征选择的具体步骤如下.

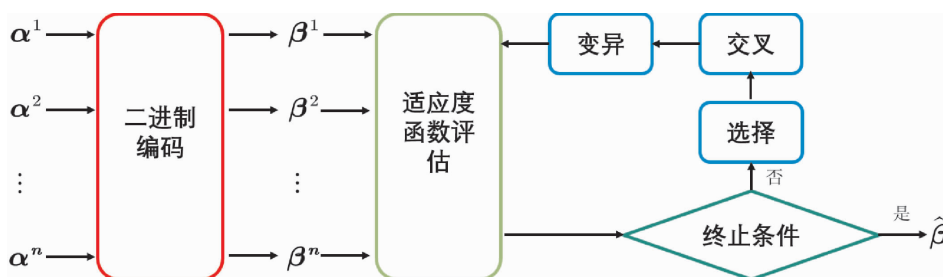


图 1 遗传算法特征选择流程图

Fig. 1 Flow chart of feature selection based on genetic algorithm

(1) 对染色体进行二进制编码, 随机初始化种群, 种群用以下矩阵表示:

$$POP = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,k} \end{bmatrix},$$

在矩阵中, 每一行代表一条染色体(或一个特征选择集合), 染色体长度为 k 代表基因总数, m 代表种群数量, $a_{i,j}$ 取值为 0 或 1, 0 代表该特征不被选中, 1 则代表该特征被选中。

(2) 采用轮盘赌方法进行选择操作. 设定个体适应度函数, 计算种群中每个个体的适应度, 个体被选择的概率与该条染色体的适应度成正比, 所有个体被选择的概率之和为 1, 每一轮迭代过程更新一次种群。

(3) 采用多点交叉法进行交叉操作. 设置交叉概率, 根据概率将两个个体之间的部分染色体进行交换, 每一轮迭代对每一条染色体进行一次交叉操作. 具体方法是使用 Python 随机生成一个 0 到 1 之间的概率, 如果随机生成的概率小于交叉概率, 则进行染色体交换, 否则不进行交换。

(4) 采用单点变异方法进行变异操作. 设置变异概率, 根据概率对某个个体的基因进行单点变异, 每一轮迭代对每一条染色体进行一次变异操作. 具体方法是使用 Python 每次随机生成一个 0 到 1 之间的概率, 如果该概率小于变异率, 则进行染色体变异, 否则不进行染色体变异。

循环步骤(2)到(4), 直到迭代 100 次. 最后, 遗传算法将生成一个接近最优解的最优种群. 将各因子在种群中出现的总次数统计排名作为因子重要性排名, 因子在种群中出现的次数越多, 说明它越重要。

特别指出, 适应度函数是用来评价染色体性

能的一个评价指标, 在遗传算法中, 适应度函数的定义是影响性能的关键因素^[23]. 计算适应度函数的过程将用于保留优秀的解, 以便进一步复制, 适应度高的染色体更有可能被多次选择, 而适应度低的染色体则更有可能被淘汰. 遗传算法的适应度函数设置为决定系数, 其反映了特征变量 X 对 Y 目标值的解释程度. 其定义如公式(1)所示.

$$r^2 = 1 - \frac{\sum (y - y')^2}{\sum (y - \bar{y})^2} \quad (1)$$

其中, 决定系数用 r^2 表示, y' 为标签值, y 为预测值, \bar{y} 为平均值, r^2 取值范围为 $[0, 1]$. 染色体表现越好, X 解释 Y 的能力就越强, 该染色体也就越有可能遗传给下一代。

遗传算法参数设置见表 2. 染色体长度设为 40, 即全部多因子种类数, 种群大小设为 100, 说明种群中随机存在 100 种多因子组合, 交叉率设为 0.8, 变异率设为 0.003, 算法总共迭代 100 次。

表 2 遗传算法参数设置

Table 2 Parameter setting of genetic algorithm

| 参数名 | 参数 |
|-------|-------|
| 染色体长度 | 40 |
| 种群大小 | 100 |
| 交叉率 | 0.8 |
| 变异率 | 0.003 |
| 迭代次数 | 100 |

2.3 Transformer 模型构建

Transformer 模型构建由位置编码、多头注意力机制以及残差连接和归一化等组成, 以下将对 Transformer 模型构建技术细节进行详细阐述。

(1) 位置编码

由于 Transformer 模型没有循环神经网络的迭

代操作,模型中没有天然存在输入数据间的时序关系,使用频率不同的正弦函数和余弦函数对 Transformer 模型的输入进行位置编码,位置编码公式原理如下:

$$PE(t, 2i) = \sin(t/10\,000^{2i/d_{model}}) \quad (2)$$

$$PE(t, 2i+1) = \cos(t/10\,000^{2i/d_{model}}) \quad (3)$$

其中, t 表示时间步, d_{model} 表示输入特征的维度总数, i 表示输入特征的维度编号, i 的取值范围是 $[0, d_{model}/2]$.

位置编码的每一个维度对应一个正弦信号或余弦信号,当 i 在规定范围内取不同的值进行位置编码时,上式中的 \sin 函数和 \cos 函数分别对应一组奇数和偶数的特征序号维度,例如 0 和 1 为一组,2 和 3 为一组,分别用 \sin 函数和 \cos 函数进行处理,从而产生不同的周期变化.位置编码随着维度序号的增大,周期从 2π 到 $10\,000 * 2\pi$ 变化,每个位置维度在位置编码中都会得到不同周期的 \sin 和 \cos 函数的取值组合,从而产生唯一的纹理位置信息,最终使得模型学习到位置之间的依赖关系和输入数据的时序特性.

图 2 对位置编码进行可视化,横轴为特征的维度,纵轴为时序的维度,图中每一行代表一个时刻的输入.纵向观察,随着特征维度的序号增大,位置嵌入函数的周期变化越来越平缓.

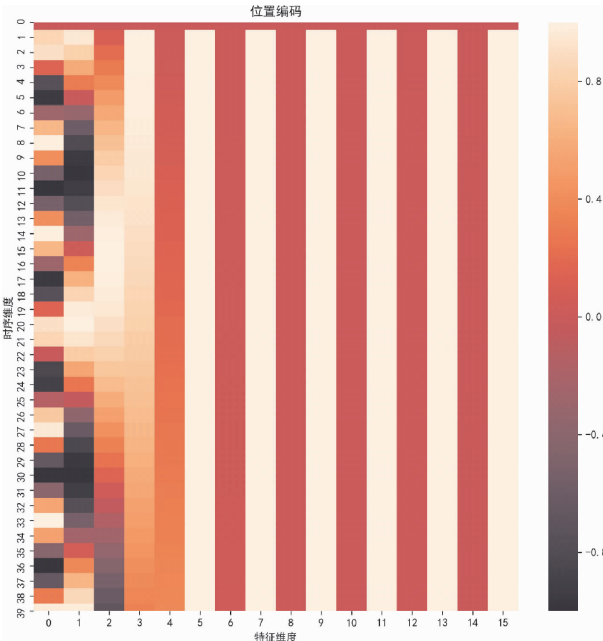


图 2 位置编码可视化

Fig. 2 Visualization of location coding

使用 \sin 和 \cos 函数进行位置编码的优势在于任意位置的 PE_{t+k} 都可以被 PE_t 的线性函数表示,从而使得两个位置之间存在某种位置依赖关系.原理如下:

$$\begin{cases} \cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \\ \sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \sin(\beta)\cos(\alpha) \end{cases} \quad (4)$$

最后,将位置编码信息添加到输入,从而产生拥有位置信息的输入.

(2) 多头自注意力机制

Transformer 模型主体部分的构建是通过多头自注意力机制实现的.为便于理解,先介绍单头自注意力机制.单头注意力机制是将输入矩阵经过三次不同的线性变换映射到三组空间向量矩阵,分别记为查询矩阵 Q 、键矩阵 K 及值矩阵 V . Q, K, V 矩阵每一行维度均设为 d_k ,单头自注意力机制计算公式为

$$Q = \text{Linear}(X) = XW^Q \quad (5)$$

$$K = \text{Linear}(X) = XW^K \quad (6)$$

$$V = \text{Linear}(X) = XW^V \quad (7)$$

$$\text{SelfAttention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

其中, X 为输入矩阵, W^Q, W^K, W^V 均为线性变换权重矩阵, softmax 归一化函数用于计算注意力权重,计算公式在前文中已经介绍.

相较于只进行一次自注意力机制运算,如果对于 K, Q, V 矩阵实行多次不同的线性变换,各自进行自注意力机制运算后,将多个输出拼接起来进行最终的输出,能够使模型关注到不同方面的信息并进行整合,起到增强模型性能的作用,因此,使用多头自注意力机制构建 Transformer 模型.多头自注意力机制表达式为

$$\text{head}_i = \text{SelfAttention}(QW_i^Q, KW_i^K, VW_i^V) \quad (9)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \quad (10)$$

其中, W_i^Q, W_i^K, W_i^V 分别是 Q, K, V 第 i 头的线性变换权重矩阵, W^O 为多头注意力矩阵拼接以后的线性变换权重矩阵.

(3) 残差连接和归一化

Transformer 模型子模块主要由一个多头注意力机制层和一个全连接层构成,注意力层和全连接层之间使用残差连接和归一化对数据进行处理.

输入数据经过多头注意力机制运算以后的输出数据维度与输入保持一致,以便进行残差连接,残差连接能够解决梯度消失和权重矩阵的退化问题.残差连接具体做法很简单,即对输入和多头注意力层的输出进行叠加,公式如下:

$$X = X_{input} + SelfAttention(Q, K, V) \quad (11)$$

归一化的作用是把模型的隐藏变量归一化为标准正态分布,以起到加速收敛的作用.具体计算过程如式(12)~(14)所示:

$$\mu_j = \frac{1}{m} \sum_{i=1}^m x_{ij} \quad (12)$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^m (x_{ij} - \mu_j)^2 \quad (13)$$

$$LayerNorm(x) = \frac{x_{ij} - \mu_j}{\sqrt{\sigma_j^2 + \epsilon}} \quad (14)$$

其中, x_{ij} 为输出矩阵第 i 行第 j 列的值, μ_j 为输出矩阵第 j 列的均值, σ_j^2 为输出矩阵第 j 列的方差.然后用每一列的元素减去这列的均值,再除以这一列的标准差,从而得到归一化后的数值,为了防止分母为 0,式(14)分母加上了不为 0 的常数 ϵ .

下面对 Transformer 模型子模块进行整合,构建完整的 Transformer 模型,如图 3 所示.

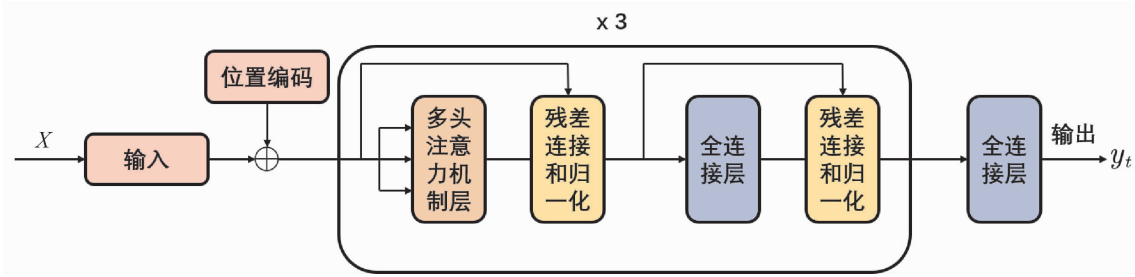


图 3 Transformer 股票预测模型^[16]

Fig. 3 Stock forecasting model based on Transformer

由图 3 可知,Transformer 模型由 3 个相同子模块堆叠而成,每个子模块主要包含多头注意力机制层和全连接层.多头注意力机制层和全连接层的输出都需要进行残差连接和归一化操作,数据输入以后,首先进行位置编码,然后经过 3 个 Transformer 子模块的特征抽取,最后外接一个全连接层,将数据维度转化到 1 维进行股票价格输出,其中, x 代表输入, y_t 代表 t 时刻的股票价格预测结果. Transformer 模型参数设置见表 3.

表 3 Transformer 模型参数设置

Table 3 Parameter setting of Transformer model

| 参数名 | 参数 | 参数名 | 参数 |
|--------------|-----|---------|-----------|
| 多头数 h | 5 | 损失函数 | MSE |
| 权重矩阵维度 d_k | 400 | 激活函数 | $sigmoid$ |
| 子模块数 | 3 | 优化器 | $Adam$ |
| $dropout$ | 0.2 | 训练轮数 | 100 |
| 全连接层节点数 | 128 | 批量 | 512 |
| 时间步 | 5 | 训练集:测试集 | 8:2 |

表 3 显示了 Transformer 模型参数.自注意力机制多头数 h 设为 5,多头自注意力机制权重矩阵

K, Q, V 维度 d_k 设为 400,Transformer 子模块数设为 3,全连接层节点数设为 128, $dropout$ 参数设为 0.2,用于随机移除一部分神经元,以防止全连接层发生过拟合,LSTM 的时间步设置为 5,即通过过去 5 天的历史股票数据预测后一天的股票价格,模型训练轮数设为 100 轮,使用 Adam 优化器作为训练优化器,全连接层激活函数使用 $sigmoid$ 函数,训练批量设为 512,实验将数据按照 8:2 划分,即前 80% 的数据作为训练集进行模型训练,后 20% 的数据作为验证集来验证模型训练结果,使用均方差 MSE 作为模型损失函数.

2.4 GA-Transformer 模型构建

本文结合遗传算法和 Transformer 模型构建 GA-Transformer 混合模型,进行多因子股票价格预测研究.首先采用遗传算法对初始多因子组合进行因子重要性排名,然后根据因子排名顺序分别取排名前 5,10,20,30,40 的因子组合作为输入特征输入 Transformer 模型进行股票预测,最后通过实验结果对比分析模型性能,并获得最优因子组合. GA-Transformer 模型框架如图 4 所示.

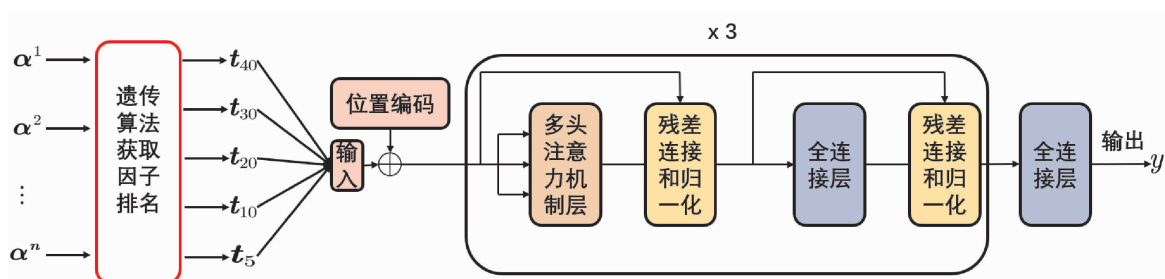


图 4 GA-Transformer 模型框架^[16]

Fig. 4 The framework of GA-Transformer model

其中, $[\alpha^1, \alpha^2, \dots, \alpha^n]$ 代表原始特征集, $t_5, t_{10}, t_{20}, t_{30}, t_{40}$ 分别代表原始特征按重要性排名前 5, 10, 20, 30, 40 的多因子组合, y_t 代表 t 时刻的预测输出结果。

典型的两支股票进行举证和分析, 这两支股票分别是建设银行和贵州茅台。股票数据时间跨度为 10 年, 即 2010 年 1 月 1 日至 2020 年 4 月 1 日, 共有 2 490 条日数据, 每一条数据的维度均为 40 维。建设银行和贵州茅台的历史股价走势如图 5 所示。

3 实验及结果分析

3.1 数据获取及预处理

为了实现普适性, 选取来自金融、餐饮、医药、房地产、通信及航空等不同行业的 6 支股票进行实证研究, 所选股票均为该行业较为典型的股票, 数据汇总如表 4 所示。

鉴于选取了股票市场上各个不同行业中具有代表性的 6 支股票作为股票价格预测研究对象, 如果一一具体介绍过分冗余, 因此, 仅对其中比较

表 4 股票数据汇总

Table 4 Stock summary

| 股票名称 | 股票代码 | 股票行业 |
|------|------------|------|
| 建设银行 | 601939. SH | 金融 |
| 贵州茅台 | 600519. SH | 餐饮 |
| 中国医药 | 600056. SH | 医药 |
| 保利地产 | 600048. SH | 房地产 |
| 中国联通 | 600050. SH | 通信 |
| 东方航空 | 600115. SH | 航空 |

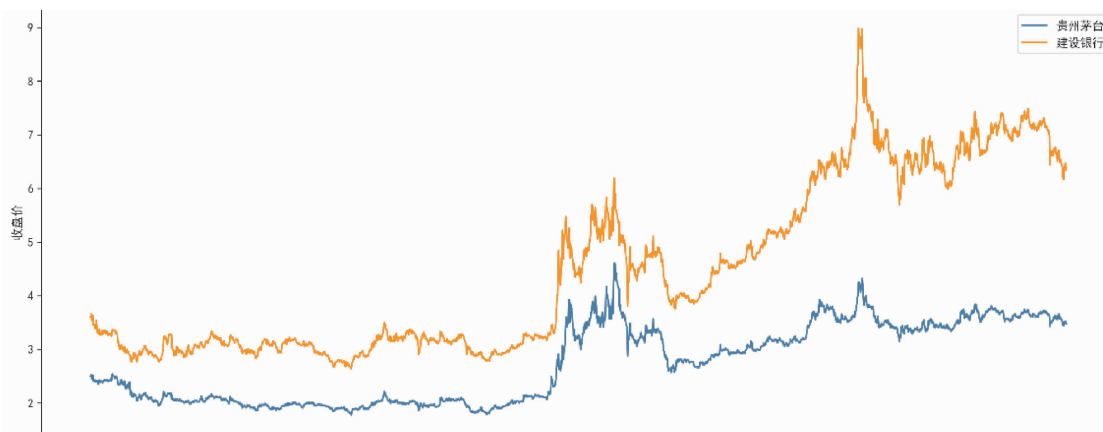


图 5 建设银行和贵州茅台 10 年股票价格趋势图

Fig. 5 Ten year stock price trend chart of China Construction Bank and Guizhou Maotai

由于股票的原始数据存在缺失和量纲现象, 容易影响股票预测精度, 因此, 对原始数据进行数据清洗是必要的, 具体做法包括缺失值处理和数据标准化。

(1) 缺失值处理

由于样本丢失及统计不全等原因, 样本存在着数据缺失的问题, 而不同的样本数据缺失情况造成数据的破坏程度不同, 其处理方式也不完全

预测精度进行评估,模型预测结果如表 5 所示.

表 5 GA-Transformer 模型股票预测结果分析

Table 5 Analysis of stock forecasting results based on GA-Transformer model

| 多因子组合总因子数目 | MSE | |
|------------|---------|---------|
| | 建设银行 | 贵州茅台 |
| 5 | 0.001 9 | 0.001 8 |
| 10 | 0.001 4 | 0.001 3 |
| 20 | 0.001 7 | 0.001 9 |
| 30 | 0.002 1 | 0.002 2 |
| 40 | 0.002 4 | 0.002 5 |

当建设银行和贵州茅台两支股票均取因子排名前 10 的多因子组合作为输入特征时,股票预测结果均方误差达到最小值分别为 0.001 4 和 0.001 3. 由此可以得到,建设银行和贵州茅台两只股票的最优特征集均为各自因子排名前 10 的多因子组合.

图 8 和图 9 分别显示了 GA-Transformer 模型在建设银行和贵州茅台数据测试集上进行股票预测的可视化结果. 可以直观看到,预测数据和真实数据偏差很小,两条数据线几乎重合,说明 GA-Transformer 模型具有很强的模型拟合能力.

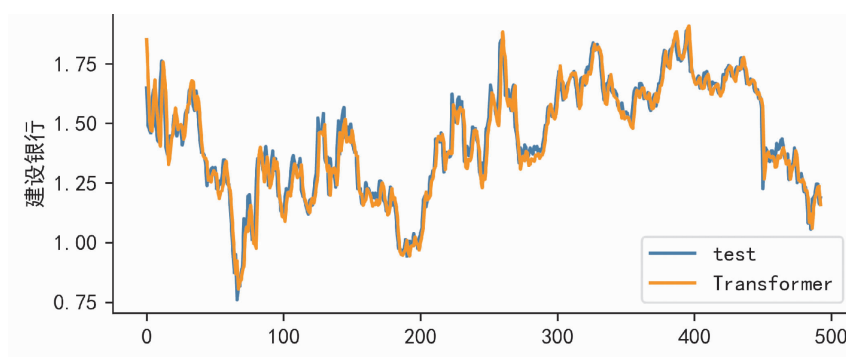


图 8 GA-Transformer 模型建设银行测试集预测结果

Fig. 8 Prediction results of test set of China Construction Bank based on GA-Transformer model

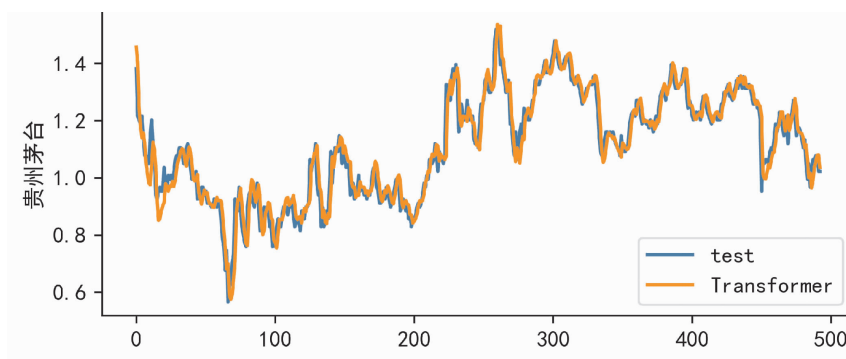


图 9 GA-Transformer 模型贵州茅台测试集预测结果

Fig. 9 Prediction results of test set in Guizhou Maotai based on GA-Transformer model

(2) 探究超参数 h 取值对模型性能的影响

GA-Transformer 模型引入了多头自注意力机制进行模型构建,引入多头的概念是为了增强模型性能. 本章为探究多头超参数 h 对 GA-Transformer 模型性能的影响,将超参数 h 分别取值为 3, 5 和 10,在建设银行和贵州茅台数据集上取排名前 10 的特征集进行实验研究. 表 6 显示了 GA-Transformer 模型分别在建设银行和贵州茅台股票数据测试集上使用不同超参数 h 的预测结果.

表 6 超参数 h 对 GA-Transformer 模型预测结果的影响

Table 6 Influence of super parameter h on prediction results of GA-Transformer model

| 超参数 h | MSE | |
|---------|---------|---------|
| | 建设银行 | 贵州茅台 |
| 3 | 0.001 8 | 0.001 9 |
| 5 | 0.001 4 | 0.001 3 |
| 10 | 0.005 4 | 0.004 9 |

由表 6 可知,当超参数 h 取值为 3 和 5 时,模型预测结果均方差 MSE 较为接近,并且当 h 取 5 时,GA-Transformer 模型在建设银行和贵州茅台股票数据测试集上的预测结果均方差达到最低;当超参数 h 取值为 10 时,GA-Transformer 模型预测性能出现了较大的降低,其原因是当超参数 h 取值过大时,GA-Transformer 模型结构过于复杂,使得训练出现了过拟合,可以通过降低模型复杂程度或者增大训练数据来解决该问题.

(3) 模型性能横向对比

为了进一步评估 GA-Transformer 模型在股票价格预测问题上的性能表现,分别选取基于 MLP^[24]、RNN^[25]、随机森林^[26]、LSTM^[27]、支持向量机^[28]和 DA-RNN^[15]等 6 种经典机器学习算法的股票预测模型进行横向比较.训练完成后,实验结果统计了各模型在 6 支股票数据测试集上均方差的均值,如表 7 所示.

表 7 各模型股票预测结果横向对比

Table 7 Comparison of stock forecast results of various models

| 模型 | 平均 MSE |
|----------------|----------|
| 支持向量机 | 0.008 0 |
| 随机森林 | 0.007 5 |
| MLP | 0.005 9 |
| RNN | 0.005 1 |
| LSTM | 0.004 5 |
| DA-RNN | 0.002 0 |
| GA-Transformer | 0.001 4 |

从各模型预测结果来看,MLP 和 RNN 这两种神经网络预测模型在股票价格预测精度上比支持向量机和随机森林两种机器学习预测模型略有提升;LSTM 作为 RNN 的改进神经网络模型,在股票预测效果上相比于 RNN 得到了一定提升,这也证明了 LSTM 模型在股票时序数据的预测中能够取得良好的表现;DA-RNN 股票预测模型相较于 LSTM 模型在预测精度上提升了 0.002 5,证明了该模型在解决股票价格预测问题上拥有一定的优越性;GA-Transformer 模型预测结果达到最低为 0.001 4,相较于 DA-RNN 模型在预测精度上提升了 0.000 6,优化效果明显.

4 结 语

本文提出了一种基于 GA-Transformer 模型的股票预测方法,与传统的股票预测方法相比,该模型通过引入自注意力机制,能够快速获取全局时序信息并且实现并行计算,在增加计算效率的同时,可以有效解决长距离依赖问题,拥有更强的特征抽取能力.实验表明,该方法能够有效预测股票价格,在预测精度上优于其他对比方法.同时,通过实验结果可见,Transforme 模型参数量大,对数据的依赖性强,在大数据集中能够发挥更好的效果.因此,如何在保证模型效果的前提下,减少模型的参数,提高运算效率将是未来的主要研究方向.

参考文献:

- [1] Cao L, Tay F. Support vector machine with adaptive parameters in financial time series forecasting[J]. IEEE Transactions on Neural Networks, 2003, 14(6): 1506-1518.
- [2] Nair B, Mohandas V, Sakthivel N R. A decision tree-rough set hybrid system for stock market trend prediction[J]. International Journal of Computer Applications, 2010, 6(9): 1-6.
- [3] White H. Economic prediction using neural networks: The case of IBM daily stock returns[J]. Neural Networks IEEE International Conference, 1988, 8(5): 409-422.
- [4] Gen R. Non-linear prediction of security returns with moving average rules[J]. Journal of Forecasting, 1996, 15(3): 165-174.
- [5] Zhang P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Computer Science, 2008, 13(1): 123-126.
- [6] Mohammad Y, Mohammad S. Application of combined Arma-neural network models to predict stock prices[C]//Proceedings of the 3rd Multidisciplinary International Social Networks Conference on Social Informatics. New York: ACM, 2016, 40(3): 1-5.
- [7] Wei S, Guo X, Chao W, et al. Forecasting stock indices using radial basis function neural networks optimized by artificial

- fish swarm algorithm[J]. Knowledge-based Systems, 2011, 24(3): 378-385.
- [8] Costa J, Marta R, Allauzen A, et al. Introduction to the special issue on deep learning approaches for machine translation [J]. Computer Speech & Language, 2017, 46(3): 367-373.
- [9] Haytham M, Margaret L, Lawrence C. Evaluating deep learning architectures for speech emotion recognition[J]. Neural Networks, 2017, 13(2): 92-99.
- [10] Xing J. Diagnosing deep learning models for high accuracy age estimation from a single image[J]. Pattern Recognition the Journal of the Pattern Recognition Society, 2017, 45(1): 34-37.
- [11] Abe M, Nakayama H. Deep learning for forecasting stock returns in the cross-section[J]. Papers, 2018, 12(5): 273-284.
- [12] Tsantekidis A, Passalis N, Tefas A, et al. Forecasting stock prices from the limit order book using convolutional neural networks[C]//2017 IEEE 19th Conference on Business Informatics (CBI). Piscataway: IEEE, 2017: 123-128.
- [13] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions[J]. European Journal of Operational Research, 2017, 2(3): 45-48.
- [14] Selvin S, Sreelekshmy A. Stock price prediction using LSTM, RNN and CNN-sliding window model[J]. International Conference on Advances in Computing IEEE, 2017, 13(21): 453-456.
- [15] Qin Y, Song D, Chen H, et al. A dual-stage attention-based recurrent neural network for time series prediction[C]//Twenty-sixth International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2017:2627-2633.
- [16] Vaswani A, Shazcer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-06-12)[2021-01-15]. <https://arxiv.org/pdf/1706.03762v1.pdf>.
- [17] Liu J, Liu X, Lin H, et al. Proceedings of the first workshop on financial technology and natural language processing[J]. Macao: IJCAI, 2019:66-73.
- [18] Molmann L, Aasen U. Transforming data into profit: Building a transformer neural network to predict golden ocean stock price based on forward freight agreements[J]. NHH Brage, 2020, 11(1): 23-29.
- [19] Ding Q, Wu S, Sun H, et al. Hierarchical multi-scale Gaussian transformer for stock movement prediction[C]//Proceedings of the International Joint Conference on Artificial Intelligence. Freiburg: IJCAI, 2020:4640-4646.
- [20] Fama E, French K. Common risk factors in the returns on stocks and bonds[J]. Journal of Financial Economics, 1993, 33(1): 3-56.
- [21] 丁鹏. 量化投资:策略与技术[M]. 北京:电子工业出版社, 2012.
- [22] Goldberg E. Genetic algorithm in search[J]. Journal of Time, 1989, 31(2): 138-145.
- [23] Jadhav K, Swati H. Information gain directed genetic algorithm wrapper feature selection for credit rating[J]. Applied Soft Computing, 2018, 34(5): 13-19.
- [24] 杨杰, 贾树文, 徐霁亨, 等. 基于 BP 神经网络的股票价格预测研究[C]//2020 万知科学发展论坛论文集. 重庆:中国智慧工程研究会智能学习与创新研究工作委员会, 2020: 9-15.
- [25] 陈佳. RNN 神经网络在股指预测中的应用研究[D]. 天津:天津科技大学, 2019.
- [26] Deng J, Li L. Application of parameter optimization stochastic forest in stock prediction [J]. Software, 2020, 41(1): 178-182.
- [27] 隋金城. 基于 LSTM 神经网络的股票预测研究[D]. 青岛:青岛科技大学, 2020.
- [28] 孙若愚. 基于支持向量机的股票价格预测[D]. 大连:大连理工大学, 2019.

【责任编辑:卓祯雨】