

基于 ATLG 混合模型的股票价格预测^①



王德广¹, 马恒锐¹, 梁叶²

¹(大连交通大学 软件学院, 大连 116052)

²(大连交通大学 计算机与通信工程学院, 大连 116052)

通信作者: 马恒锐, E-mail: 15941356654@163.com

摘要: 股市是金融市场的重要组成部分, 对股票价格预测有着重要的意义. 同时, 深度学习具有强大的数据处理能力, 可以解决金融时间序列的复杂性所带来的问题. 对此, 本文提出一种结合自注意力机制的混合神经网络模型 (ATLG). 该模型由长短期记忆网络 (LSTM)、门控递归单元 (GRU)、自注意力机制构建而成, 用于对股票价格的预测. 实验结果表明: (1) 与 LSTM、GRU、RNN-LSTM、RNN-GRU 等模型相比, ATLG 模型的准确率更高; (2) 引入自注意力机制使模型更能聚焦于重要时间点的股票特征信息; (3) 通过对比, 双层神经网络起到的效果更为明显. (4) 通过 MACD (moving average convergence and divergence) 指标进行回测检验, 获得了 53% 的收益, 高于同期沪深 300 的收益. 结果证明了该模型在股票价格预测中的有效性和实用性.

关键词: 股票预测; 长短期记忆网络; 门控递归单元; 自注意力机制; MACD 指标

引用格式: 王德广, 马恒锐, 梁叶. 基于 ATLG 混合模型的股票价格预测. 计算机系统应用, 2023, 32(3): 171-179. <http://www.c-s-a.org.cn/1003-3254/8964.html>

Stock Price Prediction Based on ATLG Hybrid Model

WANG De-Guang¹, MA Heng-Rui¹, LIANG Ye²

¹(Software Technology Institute, Dalian Jiaotong University, Dalian 116052, China)

²(School of Computer and Communication Engineering, Dalian Jiaotong University, Dalian 116052, China)

Abstract: The stock market is an important part of the financial market, and it is of great importance for stock price prediction. Meanwhile, deep learning has powerful data processing capability to solve the problems caused by the complexity of financial time series. In this regard, this study proposes a hybrid neural network model (ATLG) that combines a self-attention mechanism, a long short-term memory (LSTM) network, and a gated recurrent unit (GRU) for stock price prediction. The experimental results show the followings: (1) The ATLG model has higher accuracy than LSTM, GRU, RNN-LSTM, and RNN-GRU models. (2) The introduction of the self-attention mechanism makes the model more focused on the information of stock characteristics at important time points. (3) Comparison reveals that the two-layer neural network plays a more distinct role. (4) The backtesting with the moving average convergence and divergence (MACD) indicator achieves a 53% return, which is higher than the return of CSI 300 in the same period. The results prove the effectiveness and practicality of the model in stock price prediction.

Key words: stock prediction; long short-term memory (LSTM); gated recurrent unit (GRU); self-attention; moving average convergence and divergence (MACD) indicator

股票市场被认为是一个复杂的系统, 具有非线性特征, 受到经济、政治和心理因素的影响. 因此, 在这

样一个市场中, 对股票波动的准确预测是非常重要的^[1]. 机器学习是一种高度复杂的非线性人工智能系统,

^① 收稿时间: 2022-07-20; 修改时间: 2022-08-15; 采用时间: 2022-09-07; csa 在线出版时间: 2022-11-16
CNKI 网络首发时间: 2022-11-18

是对人脑抽象的人工模仿^[2]. 具有分布式存储、自组织、并行处理和自我调整能力. 机器学习的上述特点使其更适合处理具有多种影响因素、不稳定和随机类型的复杂非线性问题. 因此, 各国学者纷纷将机器学习用于股票市场的数据分析和预测, 并取得了一定的研究成果.

传统的机器学习方法包括自回归模型 (AR)^[3]、移动平均模型 (MA)^[4]、自回归移动平均模型 (ARMA)^[5] 和差分自回归移动平均模型 (ARIMA)^[6]. 但由于股票自身的非平稳与非线性特征, 这些方法不能在预测时达到较好的效果. 此外支持向量机^[7]、随机森林^[8]、遗传算法^[9]、集成学习^[10] 等技术也已成功应用于从原始金融数据中提取特征并进行预测. 调查表明, 在金融行为建模中使用的变量的数量和类型方面, 这些方法各不相同. 而且对于最适合使用哪些输入变量并没有达成共识. 此外, 需要注意的是, 在实际应用中使用时, 所提出的方法的利润评估通常被忽略.

近年来随着人工智能的不断发展以及计算机运算能力的提升, 机器学习中的深度学习技术受到广泛关注. Hajiabotorabi 等人^[11] 通过使用高效的离散小波变换对高频时间序列的预测进行了改进, 提出了 BSd-RNN 模型. 结果表明, BSd-RNN 模型优于其他常见的 DWT-RNN 模型. Cao 等人^[12] 将两种经验模态分解 (EMD) 与长短时间记忆网络模型 (long short-term memory, LSTM) 相结合, 形成两种混合预测模型分别对金融时间序列进行预测. Niu 等人^[13] 提出了一种基于变异模式分解 (VMD) 和门控递归单元 (gated recurrent unit, GRU) 的新型混合模型, 并通过注意力机制来提高股票价格指数预测的准确性.

注意力机制是在计算机视觉领域首次提出的. Google Mind 团队将注意力机制应用于循环神经网络模型 (recurrent neural network, RNN) 进行图像分类, 并取得了极高的准确率^[14]. 它模仿人脑的信号处理机制, 将计算资源分配给更重要的任务. Chen 等人^[15] 建立了一个基于注意力机制的 LSTM 模型并选择香港市场的股票来验证注意力机制的有效性. Qiu 等人^[16] 通过小波变换来处理股票数据, 使用基于注意力机制的 LSTM 神经网络来预测股票开盘价, 并与广泛使用的 LSTM、GRU 神经网络模型对比, 所提模型具有更好的拟合度. Chen 等人^[17] 在此基础上提出了一种新型混合深度学习模型, 它整合了注意力机制、多层感知器

和双向长短期记忆神经网络, 预测 4 种股票指数的收盘价.

与单个神经网络模型相比, 混合模型进行股票价格预测的例子相对较少. Lu 等人^[18] 提出了一种基于 CNN-LSTM 的股票价格预测方法, 采用卷积神经网络 (CNN) 从数据中提取有效特征, 来预测股票价格, 结果表明, 混合模型对股票价格预测是可行的, 并且预测精度较高. 这种预测方法不仅为股票预测提供了一种新的研究思路, 也为学者们研究金融时间序列数据提供了实践经验. Kumar 等人^[19] 提出了一种双混合方法, 即带有花粉算法的 RNN-LSTM 和带有粒子群优化的 RNN-LSTM, 以开发一个最佳的深度学习模型来提高当天内股票市场的预测.

在神经网络模型中, LSTM 网络因其独特的算法机制在时间序列预测领域占主导地位^[20]. 虽然 LSTM 网络克服了传统的循环神经网络中的梯度消失问题, 但它仍然是一种依赖顺序的计算方法. 此外, 单层 LSTM 模型在每个时刻都平等地对待所有股票状态^[21]. 基于以上文献讨论, 目前大部分对股票价格预测的模型都是基于单个神经网络, 部分混合模型也是单层相结合, 没有考虑过神经网络层数对预测结果的影响.

为了进一步减小股票价格的预测误差, 本文进行多次实验, 结合每种神经网络的优势, 对比挑选出由双层长短期记忆网络 (LSTM) 和双层门控递归单元 (GRU), 并结合自注意力机制组成的新型混合神经网络模型 (ATLG), 来对股票价格进行精确预测, 然后应用股票基本指标和技术指标来优化预测结果. 最后通过 MACD 指标进行回测, 将收益与同期沪深 300 作对比. 通过对比实验得出的结论, 本文所提出的模型在股票价格预测中取得了良好的成果.

1 股票价格预测基本原理与实现

1.1 LSTM、GRU 神经网络模型

LSTM 是 RNN 的改进版, 广泛应用在时间序列分析、文件分类、声音和语音识别等领域. 它使记忆过去的数据变得更加容易, 同时解决了梯度消失的问题. LSTM 的核心在于细胞状态以及“门”结构, 这样的设置有助于更新或删除信息. 其计算过程如下:

$$f_t = \text{sigm}(W_f x_t + U_f h_{t-1} + b_f) \quad (1)$$

$$i_t = \text{sigm}(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (3)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (4)$$

$$o_t = \text{sigm}(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(c_t) \quad (6)$$

GRU 是 LSTM 的简化变体, 它有控制单元内部信息流的门控机制, 而没有单独的存储单元. 所以它跟 LSTM 有相同的优点, 但是比 LSTM 的计算更简单, 更节省时间和算力. GRU 的设计将遗忘门和输入门合并为一个单元, 命名为“更新门”, 并有一个额外的“重置门”来处理单元内的信息流. 其计算过程如下:

$$z_t = \text{sigm}(W_z x_t + U_z h_{t-1} + b_z) \quad (7)$$

$$r_t = \text{sigm}(W_r x_t + U_r h_{t-1} + b_r) \quad (8)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \times h_{t-1}) + b_h) \quad (9)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (10)$$

根据式 (1)–式 (6) 中遗忘门 (f_t)、输入门 (i_t)、细胞状态 (c_t)、输出门 (o_t) 各部分功能结构, 推导出 LSTM 的简写模型, 如式 (11) 所示. 根据式 (7)–式 (10) 中更新门 (z_t)、重置门 (r_t) 各部分功能结构, 推导出 GRU 的简写模型, 如式 (12) 所示. 其中 (x_t) 为输入向量, (h_{t-1}) 为前一单元的输出, W 、 U 、 b 中的权重矩阵和偏执向量是在实验中逐步训练得出.

$$(h_t, c_t) = LSTM(x_t, h_{t-1}, c_{t-1}, W, U, b) \quad (11)$$

$$(h_t, z_t) = GRU(x_t, h_{t-1}, z_{t-1}, W, U, b) \quad (12)$$

1.2 自注意力机制 (self-attention)

注意力机制是一种通过加权方式来消除无用信息, 更有效地完成任务的方法. 在分析股票价格走势时, 利用注意力机制来计算不同时间段的股票价格数据的权重, 以便更好地预测股票价格的走势. 如图 1(a) 所示.

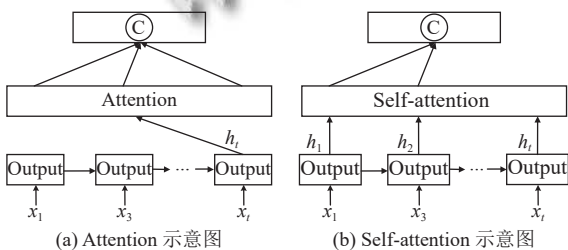


图 1 注意力机制示意图

本文所用的自注意力机制是注意力机制的变体, 其减少了对外部信息的依赖, 更擅长捕捉数据的内部

相关性. 如图 1(b) 所示, 其本质思想如式 (13) 所示:

$$Attention(Q, S) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (13)$$

2 ATLG 股票价格预测模型

本文在结合 LSTM 和 GRU 这两种神经网络模型的基础上, 加入了 self-attention 机制, 使混合后的模型可以更好地从股票数据中提取有效信息, 抓住核心关键点. 所提出的 ATLG 混合神经网络模型融合了每种模型的优点, 即 LSTM 良好的顺序数据处理能力, GRU 计算速度更快, 参数更少, 以及 self-attention 强大的获取长期依赖的能力. 为了建立一个稳定的数据处理模型, 采用了单层 self-attention 和双层 LSTM 以及双层 GRU. 高级模型的结构如图 2 所示.

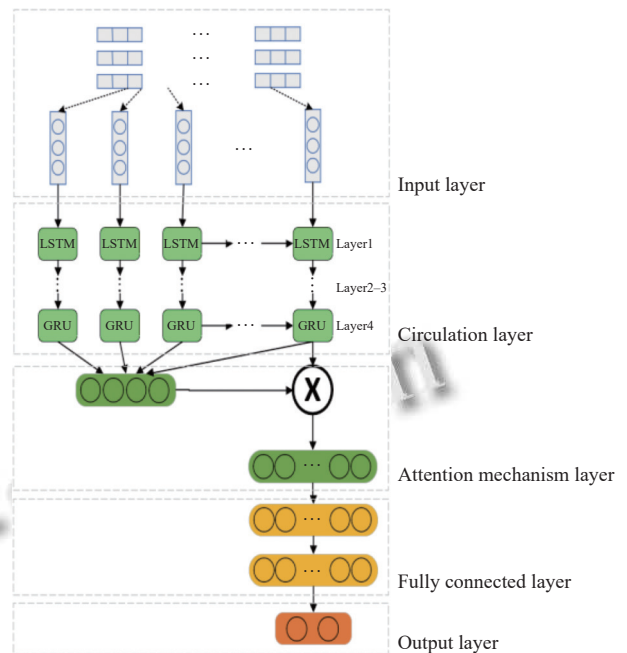


图 2 ATLG 结构示意图

输入层: 本研究使用前 T 个交易日的数据来预测第 $(T+1)$ 天的股价变化, 股票数据特征集的维度为 15, 12 个基本市场指标加上 3 类股票技术指标.

循环层: 循环层采用双层 LSTM 以及双层 GRU 模型的结构. 运用控制变量法, 在训练的过程中不断调整参数, 直至模型的预测效果最佳. 经过多次调参及优化, 最终确定, LSTM 层搭建具有 80 个记忆体的循环层, GRU 层设置为 100 个记忆体. 堆叠式的优点是在不同的时间步长内学习原始时间数据的特征, 而且可

以减少内存容量,使模型可以加速收敛.堆叠的实验模型可以描述如下:

$$(h_t^1, c_t^1) = LSTM1(x_t, h_{t-1}^1, c_{t-1}^1, W^1, U^1, b^1) \quad (14)$$

$$(h_t^2, c_t^2) = LSTM2(h_t^1, h_{t-1}^2, c_{t-1}^2, W^2, U^2, b^2) \quad (15)$$

$$(h_t^3, z_t^1) = GRU1(h_t^2, h_{t-1}^3, z_{t-1}^1, W^3, U^3, b^3) \quad (16)$$

$$(h_t^4, z_t^2) = GRU2(h_t^3, h_{t-1}^4, z_{t-1}^2, W^4, U^4, b^4) \quad (17)$$

自注意力机制层: RNN、LSTM、GRU 和其他 RNN 在预测的每一步都使用相同的固定状态向量 C . C 是一个压缩的向量,没有办法表达完整的序列信息,之前输入的信息也会被后来的信息所覆盖.引入自注意力机制可以更好地从股票数据中提取有效的信息,抓住核心关键点.

输出层:将通过第 4 个 GRU 层的输出数据 h_t^4 送入一组全连接层进行预测,通过 self-attention 层生成编码向量后,并通过 Softmax 输出类别概率. ATLG 模型整体输出被表示为:

$$y_t = w \times h_t^4 + b \quad (18)$$

其中, w 是隐藏层到输出层的权重矩阵, b 是偏置向量, y_t 是一个介于 0-1 之间的预测值.另外,为了防止模型过拟合,在各层之间进行 Dropout 操作.

本文主要是针对股票价格的预测,输出变量线性依赖于其先前的值,因此是一个回归性问题.实验使用大量数据来训练模型并估计回归方差的未知系数,通过训练后的模型来预测股票的未来价格,在实际价格和预测价格之间进行比较,选用回归评价指标来量化模型的性能,同时也将准确率作为比较模型的辅助指标.

所提出的 ATLG 模型的新颖性如下: (1) ATLG 模型是一个深度的 5 层神经网络模型,它弥补了传统神经网络中层数的不足,提高了模型的准确性. (2) 由于是对整个数据求 self-attention 的,所以能抓取到当前数据和该股票所有数据的依赖关系强度.这方面的能力显然比 RNN 的获取长依赖的能力强大得多. (3) ATLG 模型结合了 LSTM 的良好数据呈现能力和 GRU 的易收敛能力,因此克服了传统 RNN 的局限性,提高了模型的性能.

3 实验

3.1 股票数据集的选择

从股价反映市场的角度来看,筛选出了能够代表

市场的个股,形成一个金融时间序列数据集.为了选择代表市场的优质股票,本文从 tushare 上获取了“上证 50 指数”成分加权表,其中部分股票如表 1 所示.本文选择了权重较高的前 30 只股票组成股票数据集.股票数据集包括贵州茅台、招商银行、中国平安、兴业银行、上海浦东发展银行等.

表 1 上证 50 指数的部分权重表

Date	Stock code	Stock name	Weight (%)
2022/01/28	600519.SH	Moutai	15.83
2022/01/28	600036.SH	CMB	8.135
2022/01/28	601318.SH	PingAn	7.23
2022/01/28	601166.SH	CIB	4.056
2022/01/28	601012.SH	LONGi	4.055
2022/01/28	600900.SH	CYPC	3.329
...

3.2 股票市场基本数据

在筛选出要选择的股票后,下一步是获得这 30 只股票的基本市场数据.本文从 tushare 获取所有股票的基本数据,时间是从 2010 年 1 月-2019 年 12 月.输入特征包含两类,基本指标和技术指标.

基本指标包括:开盘价、最高价、最低价、收盘价、涨跌额、涨跌幅、成交量、成交额、换手率、量比、市盈率 PE、市净率 PB 等.

技术指标包括: (1) 指数移动平均 EMA: 价格跟随指标,能准确反映近期价格的变化.选择了 5 天、10 天、20 天、30 天、60 天的 EMA 值作为输入特征. (2) 平滑异同移动平均线 MACD: 表征当前的多空状态和股价可能的发展变化趋势.指标包括 MACD 快速线 (DIF)、慢线 (DEA) 和 MACD 柱. (3) 个股资金流向: 小、中、大、特大单的买入卖出量以及买入卖出额,以及净流入量和净流入额.

3.3 股票数据预处理

(1) 数据集划分

本文选取 2010 年 1 月-2019 年 12 月的 30 只股票组成原始数据集.2010-2017 年的数据用于训练,2018 年和 2019 年的数据用于测试.这包括 12 个基本的市场数据和 3 类技术指标数据.

(2) 归一化处理

由于将股票的价格数据、交易量数据和资金流向数据等同时作为输入的特征参数,而数值差异巨大,为了消除指标间的量纲影响,提升模型精度和收敛速度,本文对数据进行了归一化处理.由于数据集的不同项

目处于不同的数量级,因此需要进行数据规范化操作,否则大数量级的项目会掩盖小数量级的情况,导致小数量级项目的信息损失问题。

(3) 涨跌标签

“涨幅”为用来训练的标签,定义为:(当日收盘价-昨日收盘价)/昨日收盘价;“涨幅正负”为用来验证准确率和精度的标签.设定为当涨幅为正时,则判定该股票的趋势为涨(label: 1);当涨幅不为正时,则判定该股票的趋势为跌(label: 0)。

3.4 评价指标

为了评估模型对股票价格的预测性能,通过均方根误差(RMSE)、平均绝对误差(MAE)和决定系数(R^2)这3种回归评价指标来量化模型的性能. RMSE和MAE越小,预测值越接近真实值;系数 R^2 越接近1,模型的拟合度越高,4种指标总结为式(19)~式(21):

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - y_i| \quad (19)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N ((x_i - y_i)^2)}{\sum_{i=1}^N ((x_i - \bar{y})^2)} \quad (21)$$

根据样本的真实结果和模型预测结果,将样本分为4类: true positive (TP), 样本是正样本并且模型预测结果也是正样本的个数; false negative (FN) 样本是正样本但模型预测结果是负样本的个数; false positive (FP) 样本是负样本但模型预测结果是正样本的个数; true negative (TN) 样本是负样本并且模型预测结果也是负样本的个数. 准确率用式(22)表示:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (22)$$

3.5 回测指标

移动平均线收敛和背离(MACD)指标是股票交易中的一个常见技术分析工具. MACD指标的设计是基于移动平均线的原理,它是对收盘价进行平滑处理(计算加权平均)后的价格指标. MACD指标由一系列曲线和垂直条组成,反映了股票波动周期的特点,这些特点能够钻研出买卖股票的最佳时机. MACD的优势在于它消除了移动平均线(MA)所传递的频繁买入和卖出信号的缺点,这增加了对信号传输的要求和限制.它在实战中的使用比MA更稳定. 指数移动均线(EMA)

是当天的收盘价减去昨天的EMA,乘以一个平滑系数,然后再乘以昨天的EMA. 计算公式为:

$$EMA = \frac{2}{N+1}C + \left(1 - \frac{2}{N+1}\right)EMA' \quad (23)$$

$$DIF = EMA(12) - EMA(26) \quad (24)$$

$$DEA(N) = \frac{2}{N+1}DIF + \frac{N-1}{N+1}DEA' \quad (25)$$

$$MACD = 2 \times (DIF - DEA) \quad (26)$$

其中, N 为所计算的天数, C 为今日收盘价, EMA' 表示昨日的EMA, DIF 为离差值, DEA 为离差平均值, DEA' 为昨日DEA.

3.6 实验对比

实验1. 在其他条件相同的情况下,探索基础模型RNN、LSTM和GRU之间的性能差异.

实验2. 探讨自注意力机制对股票价格预测的影响. 分析和比较基础模型RNN、LSTM和GRU在加入自注意力机制后各项指标的变化.

实验3. 研究探讨神经网络层数对AT-RNN、AT-LSTM和AT-GRU模型的影响. 基于第2个实验基准模型,循环层分别从1层升级为2层、3层或4层. 观察实验结果,比较不同深度的循环层对实验结果的影响.

实验4. 基于第3个实验得出的结论,得到了对模型影响的最佳层数. 在最佳层数的基础上对模型两两结合,分别于基本模型、加入自注意力机制的模型做出对比,评选出对股票预测的最优模型.

4 实验结果和分析

4.1 实验环境

实验环境: Windows 10 操作系统 16 GB 内存; 开发工具: PyCharm 2020 TensorFlow 2.0; 显卡: NVIDIA GeForce MX450; CPU: Intel(R) Core(TM) i5-1135G7 @2.40 GHz.

4.2 探索RNN、LSTM、GRU之间的性能差异

分别比较单层的RNN、LSTM、GRU模型对股票价格短期预测的效果. 使用Adam优化器,损失函数定为均方差(mean squared error, MSE),批处理大小(batch_size)设置为64. 对每个模型分别进行多次实验,取多次结果的平均值,得到最终的实验结果,如图3和表2所示.

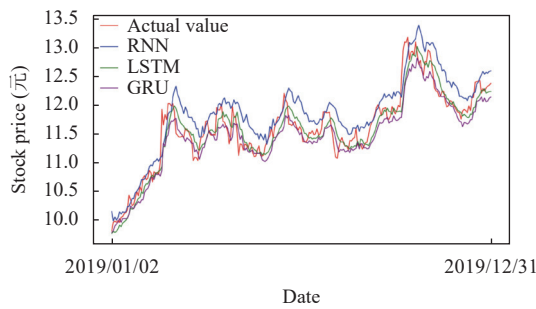


图3 基础模型预测效果

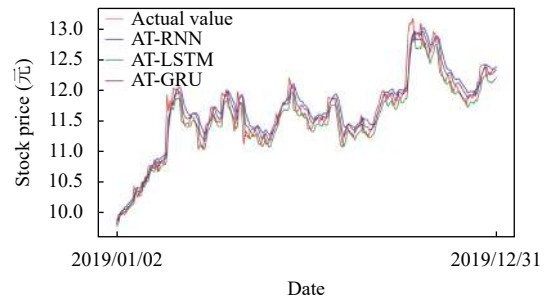


图4 加入 self-attention 后模型预测效果

表2 基础模型的实验结果比较

Model	RMSE	MAE	R ²	Acc (%)
RNN	4.7197	3.3256	0.8145	50.54
LSTM	4.5256	3.1475	0.8583	52.93
GRU	4.2527	2.9903	0.8647	53.31

表3 加入注意力机制后的模型实验结果比较

Model	RMSE	MAE	R ²	Acc (%)
AT-RNN	0.2225	0.1718	0.8806	54.36
AT-LSTM	0.2002	0.1502	0.9062	56.43
AT-GRU	0.1931	0.1387	0.9052	57.25

从表2可以看出GRU的误差最低,其次是LSTM,而表现最差的是RNN.就准确率而言,GRU的准确率最高为53.31%,其次是LSTM达到52.93%,RNN最低只有50.54%.这也验证了之前的分析,RNN难用梯度下降法来训练,而且存在梯度消失和爆炸问题.而LSTM和GRU很好地解决了这一问题.并且GRU模型将LSTM中的遗忘门和更新门优化合并为重置门,需要训练的参数更少且可以防止过拟合.实验证明在相同参数时,GRU模型效果普遍好于RNN、LSTM模型.

4.3 探索加入自注意力机制对模型的影响

如图4和表3所示,可以看出AT-GRU的误差普遍小于AT-LSTM和AT-RNN.在准确率方面,AT-GRU最好,达到57.25%,其次是AT-LSTM,达到56.43%,而准确率最低的是AT-RNN.相对于基础模型,加入self-attention机制的预测模型效果更优.原因主要有两个:第一,未加入self-attention机制时,模型对不同时间点编码向量关注度相同,未能对重点信息的影响特征赋予更多注意力.第二,由于依靠self-attention机制进行加权,可以适用不断变化股市波动规律,泛化能力较强,并且模型用自注意力机制来处理特殊时间段特征信息时,所以可以关注重要时间点位置信息,从而减小实验的误差.

4.4 探讨循环层的层数对预测模型的影响

在加入self-attention机制的模型的基础上,循环层分别增加到2层、3层和4层.分析了循环层的层数对股票预测模型的影响.对性能进行分析和评估后,各个模型的预测结果的准确性如图5所示.

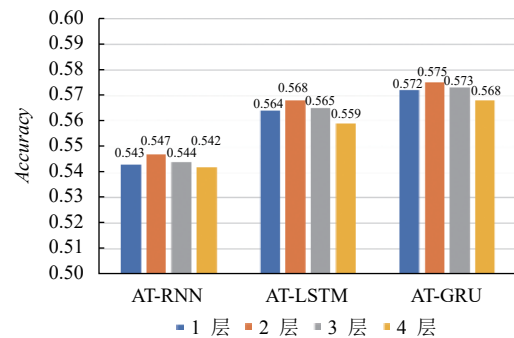


图5 不同层数对实验结果的影响

就准确性而言AT-GRU的两层循环层模型结果最优,达到57.5%,而表现最差的是有着4层循环层的RNN,准确率为54.2%.普遍来看,2层循环层表现得结果最为突出,准确率普遍高于3层和1层,4层循环层得准确率最低.从实验结果中可以看出,增加神经网络的层数在一定程度上可以增加模型的精度,但是会出现过拟合的情况,因此,增加模型的层数,要避免过拟合现象的产生,并不是一味地增加网络层数就可以提高模型的精度.

4.5 探究混合模型对实验结果的影响

第3个实验中,得到了循环层数对预测模型的影响,在其他条件相同的情况下,2层循环层表现得效果最佳.所以在最优层数的基础上对已有自注意力机制的模型两两结合,搭建5层神经网络模型,依次为自注意力机制层加双层RNN加双层LSTM(ATRL),自注意力机制层加双层RNN加双层GRU(ATRG),以及自注意力机制层加双层LSTM加双层GRU(ATLG).进

行多次对比实验, 评选出对股票价格预测的最优模型. 各个模型的预测结果的准确性如图 6 和表 4 所示, 通过实验结果可知自注意力机制层加双层 LSTM 加双层 GRU 的混合模型表现最好, 均方根误差为 0.1515、平均绝对误差为 0.1057、决定系数为 0.9407. 就准确率而言 ATLG 的准确率达到 62.84%, 优于其他两种组合模型, 实验并给出了 ATLG 的损失函数, 如图 7 所示.

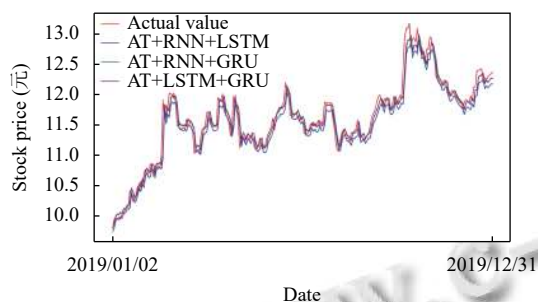


图 6 混合模型预测效果

表 4 混合模型实验结果

Model	RMSE	MAE	R^2	Acc (%)
ATRL	0.1812	0.1319	0.9136	59.68
ATRG	0.1541	0.1084	0.9392	61.25
ATLG	0.1515	0.1057	0.9407	62.84

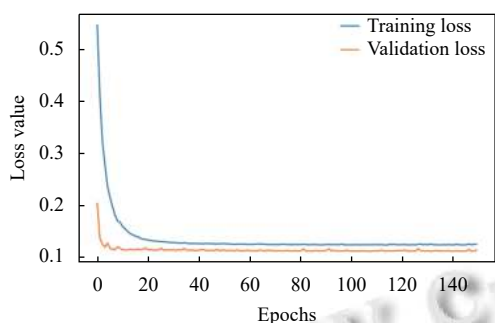


图 7 ATLG 的损失函数

4.6 实验结果分析

为了验证本模型的先进性, 选取了一些最近流行的实验模型进行对比, 分别是, 2022 年严冬梅等人^[22]提出的一种结合自注意力机制和残差网络的生成式对抗神经网络模型 (SAR-GAN); 2022 年刘茜阳等人^[23]提出的一种基于模糊 K 线的长短期记忆 (LSTM) 网络和支持向量回归多阶段混合模型 (FCLSTM-vSVR); 2021 年, 郑树挺等人^[24]提出改进的 self-attention 模型; 以及 2021 年蒙懿等人^[25]构建的基于 CNN-BiLSTM 和注意力机制的股价预测模型. 如表 5 所示.

表 5 实验结果分析

Model	RMSE	MAE	R^2	Acc (%)
SAR-GAN ^[22]	1.5191	1.1844	—	—
FCLSTM-vSVR ^[23]	0.1700	0.1210	0.9280	—
改进self-attention ^[24]	0.1546	—	—	61.32
CNN-BiLSTM ^[25]	—	0.1666	0.9550	57.58
ATLG	0.1515	0.1057	0.9407	62.84

通过实验结果分析可知, 本文提出的 ATLG 模型在均方根误差、平均绝对误差、决定系数和准确率方面均优于其他对比模型, 仅在决定系数上略差于 CNN-BiLSTM 模型, 这表明该模型在股票价格预测上具有一定的有效性和实用性.

4.7 消融实验

为了验证所提出的方法的每个组成部分对股票价格预测的有效性, 现删除模型中的一个组成部分, 探究每个组成部分对模型的影响. 结果如表 6 所示.

表 6 消融实验结果分析

ATLG model	RMSE	MAE	R^2	Acc (%)
(1) without LSTM	0.1783	0.1147	0.9108	58.46
(2) without GRU	0.1942	0.1472	0.9094	57.63
(3) without self-attention	0.2193	0.1573	0.8301	54.81
ATLG	0.1515	0.1057	0.9407	62.84

从实验结果可以看出, 加入双层 LSTM 结构和双层 GRU 结构后, 模型准确率分别提升了 4.38% 和 5.21%, 充分证明了 LSTM 模型可以很好地解决股票预测系统中股票数据难以预测及股票数据过长的问题, 以及 GRU 模型计算更简单, 更节省时间和算力的优点. 加入自注意力机制后, 模型准确率提升了 8.03%, 同样可以验证自注意力机制能更好地从股票数据中提取有效的信息, 抓住核心关键点.

4.8 模型回测及投资收益

通过实验得出对股票预测得最优模型, 将模型对真实股票数据 (以浦发银行为例) 进行买卖点以及收益率的回测, 计算出收益并且和同期沪深 300 进行对比. 买卖的策略为, 假设初始资金为 100 000 元, 当 MACD 从负数转向正数, 模型提示买入. 当 MACD 从正数转向负数, 模型提示卖出. 当 MACD 以大角度变化, 表示快的移动平均线和慢的移动平均线的差距非常迅速地拉开, 代表了一个市场大趋势的转变. 图 8 展示了浦发银行的部分 K 线及买卖点的提示图, 图中提示的买卖时间点是每日的刚开盘时间 (即 9:30). 为显

示策略的可行性,选取同期沪深300指数进行对比,在所有买卖点之后,浦发银行总收益为53146.687元,收益率为53%,同期沪深300总收益为142621.410元,收益率为42%,图9展示了浦发银行和沪深300的收益对比。可以发现所提的模型可以提高收益率,证明了所提出的ATLG模型可以用于真实股票预测。

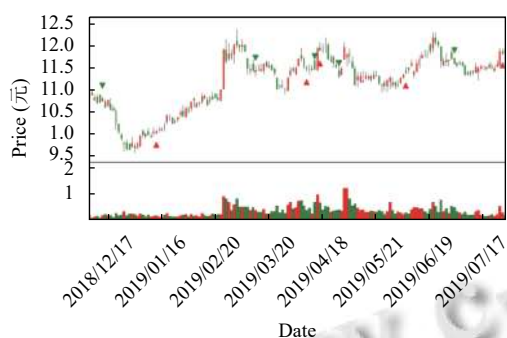


图8 浦发银行K线及买卖提示图

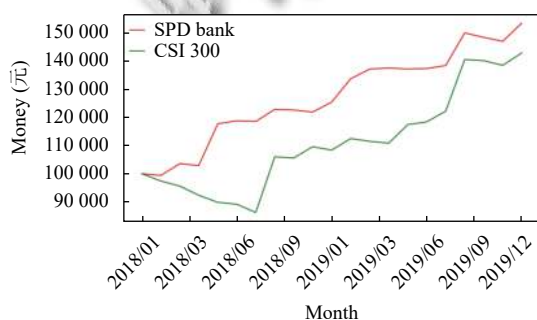


图9 浦发银行和沪深300的收益对比

5 总结

本文提出了一个由双层LSTM、双层GRU和自注意力机制组成的混合ATLG神经网络模型,用于股票价格的预测。为了提高性能,逐步探讨了加入自注意力机制后对基本模型的误差和准确率的影响,探讨了神经网络中层数对模型的影响。并且在得出最优层数的结论后,将模型两两组合最终得出对股票价格预测的最优模型。对比其他文献在各项指标上,所提模型均有提高,实验结果反映出ATLG模型可以适用于市场的股票价格预测。

此外,本文进行了额外的基金模拟,提出的框架可以产生可观的利润,年化收益明显增加,与市场表现相比,这是一个显著的改进。由于股价预测任务的不确定性和难度,以及相当大的交易成本使得轻微的价格波

动毫无意义,因此,关注更有可能成为高利润机会的重要交易点,而不是预测每个时间点的趋势,可以带来更多的利润。在未来的工作中,将继续研究深度学习投资组合构建技术,更加关注中长期的预测,还将通过纳入自然语言处理NLP技术来扩展预处理部分,以捕捉社交媒体、新闻和谣言对股票市场价格的影响。

参考文献

- 1 Wang L, Ma F, Liu J, *et al.* Forecasting stock price volatility: New evidence from the GARCH-MIDAS model. *International Journal of Forecasting*, 2020, 36(2): 684–694. [doi: 10.1016/j.ijforecast.2019.08.005]
- 2 Freeborough W, van Zyl T. Investigating explainability methods in recurrent neural network architectures for financial time series data. *Applied Sciences*, 2022, 12(3): 1427. [doi: 10.3390/AP12031427]
- 3 Taylor JW, Yu KM. Using auto-regressive logit models to forecast the exceedance probability for financial risk management. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2016, 179(4): 1069–1092. [doi: 10.1111/rssa.12176]
- 4 Pradhan R, Bharadwaj K, Saxena A, *et al.* Stock trend prediction and analysis using LSTM neural network and dual moving average crossover algorithm. *IOP Conference Series: Materials Science and Engineering*, 2021, 1131: 012005. [doi: 10.1088/1757-899X/1131/1/012005]
- 5 Chen B, Buesser B, DiPietro K. Recurrent neural networks for autoregressive moving average model selection. *IEEE International Conference on Big Knowledge*. Beijing: IEEE, 2019. 33–41. [doi: 10.1109/ICBK.2019.00013]
- 6 Bukhari AH, Raja MAZ, Sulaiman M, *et al.* Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. *IEEE Access*, 2020, 8: 71326–71338. [doi: 10.1109/ACCESS.2020.2985763]
- 7 赵澄, 叶耀威, 姚明海. 基于金融文本情感的股票波动预测. *计算机科学*, 2020, 47(5): 79–83. [doi: 10.11896/jsjcx.190400145]
- 8 Khaidem L, Saha S, Dey SR. Predicting the direction of stock market prices using random forest. *arXiv:1605.00003*, 2016.
- 9 Hu Y, Liu K, Zhang XZ, *et al.* Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review. *Applied Soft Computing*, 2015, 36: 534–551. [doi: 10.1016/j.asoc.2015.07.008]
- 10 谢琪, 程耕国, 徐旭. 基于神经网络集成学习股票预测模型

- 的研究. 计算机工程与应用, 2019, 55(8): 238–243. [doi: [10.3778/j.issn.1002-8331.1807-0294](https://doi.org/10.3778/j.issn.1002-8331.1807-0294)]
- 11 Hajiabotorabi Z, Kazemi A, Samavati FF, *et al.* Improving DWT-RNN model via B-spline wavelet multiresolution to forecast a high-frequency time series. Expert Systems with Applications, 2019, 138: 112842. [doi: [10.1016/j.eswa.2019.112842](https://doi.org/10.1016/j.eswa.2019.112842)]
- 12 Cao J, Li Z, Li J. Financial time series forecasting model based on CEEMDAN and LSTM. Physica A: Statistical Mechanics and Its Applications, 2019, 519: 127–139. [doi: [10.1016/j.physa.2018.11.061](https://doi.org/10.1016/j.physa.2018.11.061)]
- 13 Niu HL, Xu KL. A hybrid model combining variational mode decomposition and an attention-GRU network for stock price index forecasting. Mathematical Biosciences and Engineering, 2020, 17(6): 7151–7166. [doi: [10.3934/mbe.2020.367](https://doi.org/10.3934/mbe.2020.367)]
- 14 Mnih V, Heess N, Graves A, *et al.* Recurrent models of visual attention. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 2204–2212.
- 15 Chen S, Ge L. Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. Quantitative Finance, 2019, 19(9): 1507–1515. [doi: [10.1080/14697688.2019.1622287](https://doi.org/10.1080/14697688.2019.1622287)]
- 16 Qiu JY, Wang B, Zhou CJ. Forecasting stock prices with long-short term memory neural network based on attention mechanism. PLoS One, 2020, 15(1): e0227222. [doi: [10.1371/journal.pone.0227222](https://doi.org/10.1371/journal.pone.0227222)]
- 17 Chen Q, Zhang WY, Lou Y. Forecasting stock prices using a hybrid deep learning model integrating attention mechanism, multi-layer perceptron, and bidirectional long-short term memory neural network. IEEE Access, 2020, 8: 117365–117376. [doi: [10.1109/ACCESS.2020.3004284](https://doi.org/10.1109/ACCESS.2020.3004284)]
- 18 Lu WJ, Li JZ, Li YF, *et al.* A CNN-LSTM-based model to forecast stock prices. Complexity, 2020, 2020: 6622927. [doi: [10.1155/2020/6622927](https://doi.org/10.1155/2020/6622927)]
- 19 Kumar K, Haider MTU. Enhanced prediction of intra-day stock market using metaheuristic optimization on RNN-LSTM network. New Generation Computing, 2021, 39(1): 231–272. [doi: [10.1007/s00354-020-00104-0](https://doi.org/10.1007/s00354-020-00104-0)]
- 20 Torres JF, Hadjout D, Sebaa A, *et al.* Deep learning for time series forecasting: A survey. Big Data, 2021, 9(1): 3–21. [doi: [10.1089/big.2020.0159](https://doi.org/10.1089/big.2020.0159)]
- 21 Pang XW, Zhou YQ, Wang P, *et al.* An innovative neural network approach for stock market prediction. The Journal of Supercomputing, 2020, 76(3): 2098–2118. [doi: [10.1007/s11227-017-2228-y](https://doi.org/10.1007/s11227-017-2228-y)]
- 22 严冬梅, 李斌. 基于生成式对抗神经网络的股票预测研究. 计算机工程与应用, 2022, 58(13): 185–194. [doi: [10.3778/j.issn.1002-8331.2112-0005](https://doi.org/10.3778/j.issn.1002-8331.2112-0005)]
- 23 刘茜阳, 宋燕, 张亚萌. 基于模糊 K 线的 FCLSTM-vSVR 模型的股票价格预测. 智能计算机与应用, 2022, 12(4): 54–60, 69. [doi: [10.3969/j.issn.2095-2163.2022.04.010](https://doi.org/10.3969/j.issn.2095-2163.2022.04.010)]
- 24 郑树挺, 徐菲菲. 基于改进 Self-Attention 的股价趋势预测. 计算机技术与发展, 2021, 31(3): 33–38. [doi: [10.3969/j.issn.1673-629X.2021.03.006](https://doi.org/10.3969/j.issn.1673-629X.2021.03.006)]
- 25 蒙懿, 徐庆娟. 基于 CNN-BiLSTM 和注意力机制的股票预测. 南宁师范大学学报(自然科学版), 2021, 38(4): 70–77. [doi: [10.16601/j.cnki.issn2096-7300.2021.04.010](https://doi.org/10.16601/j.cnki.issn2096-7300.2021.04.010)]

(校对责编: 牛欣悦)