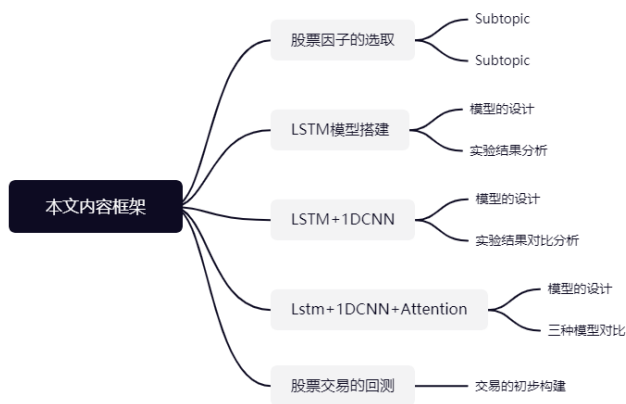


基于 CNN+LSTM+ATTENTION 的股票策略模型

蔡雨皓 202222011068 论文编辑器： \LaTeX Studio

了解项目更多请见 *Github*: [mygithub](#)

摘要—股票预测一直是金融量化的热门话题，由于股票价格指数的价格序列具有混沌性高、非线性程度较强的特点，用一般的统计学方法难以构建出模型预测股价。由于神经网络对预测序列的非线性关系具有突出优势，人们渐渐把目光转向了深度神经网络用于量化投资。Lstm 对序列数据处理能力非常可观，但是根据实验表明，层数到达一定程度后，Lstm 神经网络的深度对股票预测的效果提升并不明显，本文欲使用 1DCNN 帮助模型提取 20 种股票指标中的特征，再添加了注意力机制以更好地识别重要特征，对几种模型进行设计建模、预测对比、和回测验证，欲探究神经网络结构对股票预测效果的提升。



Index Terms—LSTM, 1DCNN, Attention, Keras, tensorflow, Ta-lib, BT

I. 股票因子的选取

引言：下面主要分享股票因子的选取，本文以沪深 300 成分股作为研究对象，利用 akshare 程序包获取了 2015 年 6 月到 2022 年 9 月的股票数据，利用 Ta-lib 程序包计算出与股票市场变动有很强相关性的常用指标技术指标。此内容不仅作为分享，也作为这次项目的学习记录。

股票量化分析分为基本面分析与技术分析两种，基本面分析又称基本分析，是以证券的内在价值为依据，着重于对影响证券价格及其走势的各项因素的分析，以此决定投资购买何种证券及何时购买。技术分析法是通过市场行为本身的分析来预测市场价格的变化方向，即主要是证券的日常交易状态，包括价格变动、交易量与持仓量的变化等资料，按照时间顺序绘制成图形或图表，或形成一定的指标系统，然后针对这些图

形、图表或指标系统进行分析研究，以预测证券价格走势的方法。用技术指标对股价进行预测时，认为技术分析可以获取超额收益，做出了如下假设：股票的价格未充分反映过去的交易量与价格信息，我们称之为无效市场或者称为未达到弱有效市场。

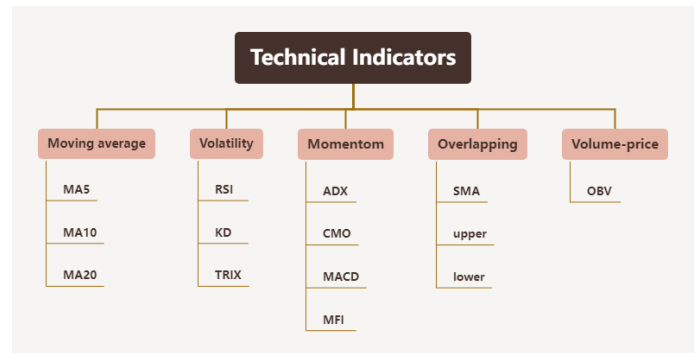


图 1. 技术指标汇总

众所周知，模型选用的因子对模型预测准确性、被解释变量的被解释程度具有密切关联，此模型根据文献[?]，选了五种技术指标以及五种基本股价指标作为股票预测因子，分别为动量指标、移动均线指标、量价指标、波动率指标、重叠研究指标。技术指标的分类以及具体计算方法见Akshare 技术文档。

II. LSTM 模型搭建

A. 长短期记忆神经网络

虽然循环神经网络的目的是学习长期的依赖性，但理论和经验的证据表明很难学习并长期保存信息，长期依赖性成了需要解决的问题。九十年代中期，德国学者 Sepp Hochreiter 和 Juergen Schmidhuber¹提出这种循环网络的变体，带有所谓长短期记忆单元，或称 LSTM，同时 LSTM[1], [2] 还能解决损失函数梯度消失的问题。带有长短期记忆单元类似累加器和门控神经元：它在下

¹两位德国学者在 LSTM 论文中引入了 CEC 单元，解决了 BPTT 带来的梯度消失和梯度爆炸问题，被称为递归神经网络之父

一个时间步长将拥有一个权值并连接到自身，拷贝自身状态的真实值和累积的外部信号，但这种自联接是由另一个单元学习并决定何时清除记忆内容的乘法门控制的。[3]

遗忘门：遗忘门是以上一单元的输出 $h_{(t-1)}$ 和本单元的输入 $x_{(t)}$ 作为输入的 sigmoid 函数，为 $C_{(t-1)}$ 中的每一项产生一个 $[0,1]$ 内的值，来控制上一单元被遗忘的程度。

$$f_{(t)} = \delta(w_f[h_{(t-1)}, x_{(t)}] + b_f) \quad (1)$$

输入门：输入门和一个 tanh 函数配合控制有哪些新信息被加入。tanh 函数产生一个新的候选向量 \tilde{C}_t ，输入门 \tilde{C}_t 的每一项产生一个在 $[0,1]$ 内的值，控制新信息被加入的数量。

$$i_{(t)} = \delta(w_i[h_{(t-1)}, x_{(t)}] + b_i) \quad (2)$$

$$\tilde{C}_t = \delta(w_c[h_{(t-1)}, x_{(t)}] + b_c) \quad (3)$$

$$C_t = i_{(t)} * \tilde{C}_t + f_{(t)} * C_{(t-1)} \quad (4)$$

输出门：输出门用来控制当前的单元状态有多少被过滤掉。先将单元状态激活，输出门为其中每一项产生一个在 $[0,1]$ 内的值，控制单元状态被过滤的程度决定什么样的信息要输出。

$$o_{(t)} = \delta(w_o[h_{(t-1)}, x_{(t)}] + b_o) \quad (5)$$

$$h_{(t)} = o_{(t)} * \tanh(C_{(t)}) \quad (6)$$

B. LSTM 模型结构

LSTM 模型结构主要由 LSTM 层，dropout 层，和 dense 层构成，用 keras.layers 的 API 进行模型的构建，由于计算机的计算能力有限，只用了三层 LSTM，没有对层数进行比较分析。将模型的所有参数存入 json 文本中，方便对模型进行调参以及之后加入 CNN 层和 Attention 层。模型加入了 early stopping 的机制，当预测效果没有显著提升后，将终止 epoch 循环，为防止模型开始过拟合从而失去泛化能力。

输入特征的维数为 20，用前 85% 的数据作为训练集。Lstm 选取时间步长为 19。模型采用了滑动窗口法，滑动窗口法 [4] 是利用一个窗口里的所有数据来预测下一个时间点的被预测值。对窗口大小从 5-20 进行参数对比，也就是从一周的交易日到一个月的交易日进行搜索对比，最后选取了 19 作为窗口大小。需要注意的是，前两层的 lstm 中的 return 是隐藏状态的所有值，是序列形式，而最后一层 lstm 只返回时间步长中最后一个输出值，也就是隐藏状态中最后一个值。

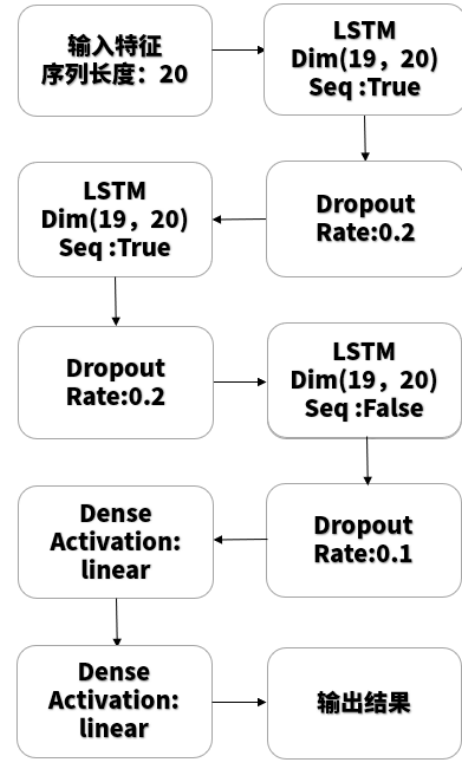


图 2. LSTM 结构图

由于股票的走势不尽相同，所以本次策略欲对沪深 300 跑出不同模型，选取正确率大于直观预测（50%）的股票作为回测的股票池，以下展示预测准确率等结果，均以 600048 为例。

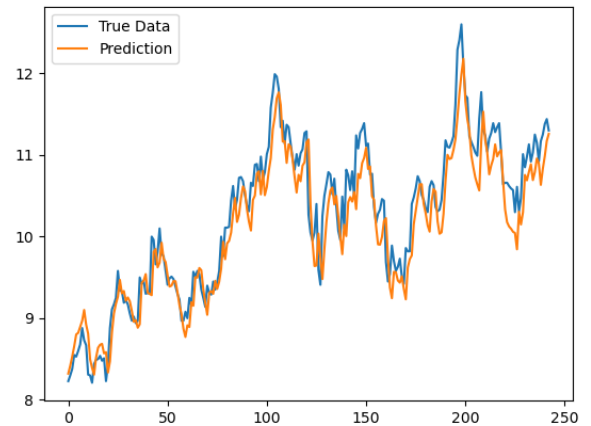


图 3. LSTM 模型预测

III. CNN+LSTM 模型

A. 一维卷积神经网络

一维卷积神经网络 [5], [6], [7] 通常用于对时间序列等进行处理, 其网络结构与二维卷积神经网络类似, 输入序列首先经过卷积层, 然后经过池化层, 最后通过全连接层。二维卷积神经网络是采用一个二维矩阵的卷积核对图像的像素特征进行提取, 类似地, 一维卷积神经网络是采用一个一维的小序列在时间序列上滑动, 从而对整个时间序列的特征进行提取。下图展示了一维卷积神经网络卷积计算的过程, 图中的三色线分别代表着大小为 3 的卷积核的元素值。与二维卷积的计算过程相比, 一维卷积在序列横向滑动的卷积计算过程只有横向移动。

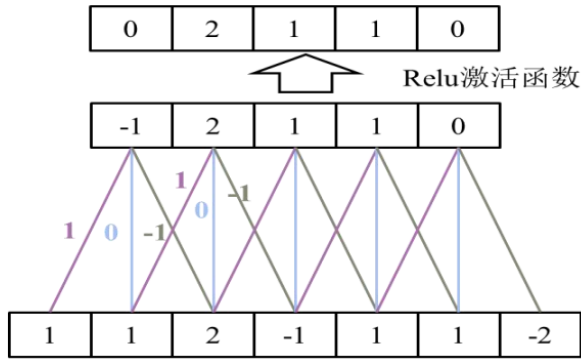


图 4. 一维卷积核计算示意图

B. LSTM+CNN 模型结构

在 LSTM 层之前加入 keras.layers 中的 Conv1D, 卷积核个数为 50, padding=same, 即不改变输入层尺寸, 输入的尺寸为 (19,20), 以便于之后连接 LSTM 层。Conv1D 层之后加入池化层, 使用最大池化方法。加入了 CNN 层后, 会发现模型的预测效果变差, 但同时泛化能力变强, 一定程度上可以改善 LSTM 模型的特征提取能力弱、具有表示瓶颈以及特征识别能力弱的问题, 得到一个更为准确的股价指数预测模型。

IV. CNN+LSTM+ATTENTION 模型

A. 注意力机制

人的眼睛往往会首先注意到物体最重要的特征, 并依赖这些重要特征进行物体的识别, 研究者从这一现象中得到灵感, 提出了注意力机制 [8](Attention Mechanism, AM)。注意力机制可以通过对重要的信息赋予较高的权重, 提高信息的利用程度, 因此具有更高的可扩展性和鲁棒性 [9]。它克服了传统神经网络中的一些局限, 如随着输入长度增加系统的性能下降、输入顺

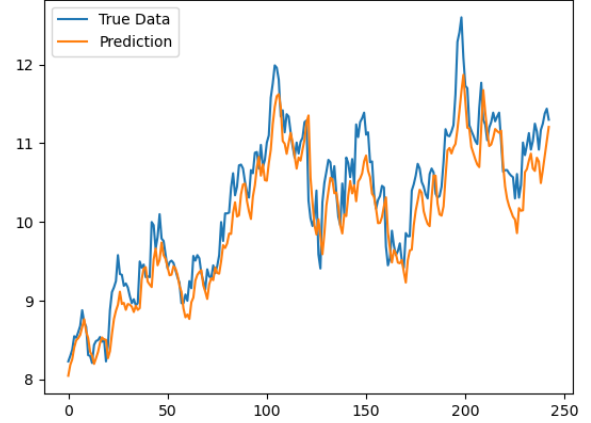


图 5. LSTM+CNN 模型预测

序不合理导致系统的计算效率低下、系统缺乏对特征的提取和强化等。

注意力机制从一开始就因其独特的思想深受广大学者的喜爱, 通过实验研究将其进行拓展应用于多种情景。注意力机制与传统算法的简单结合就可以提高系统的性能, 因此注意力机制的提出对深度学习许多结构都有着性能提高的作用。而在 2017 年, Vaswani 提出了 Transformer 模型 [10], 更是将注意力机制推向了诸多应用方向的热潮。

B. 模型对比分析

为了可以精确地比较本章所建模型之间的回归预测能力, 根需要选取一些客观且能反映模型预测性能优劣的指标 [11], 从而使得模型之间具有可比性。本章选取了四个指标来评价所提模型以及对照模型的回归预测能力, 分别是平均绝对误差、均方根误差、平均绝对百分误差以及涨跌准确率。前三个指标是用来评价模型回归预测结果与真实值的差别, 而最后一个指标是用来衡量模型在回归预测过程中捕捉股票价格指数收盘价走势的能力。

1) 平均绝对误差: 平均绝对误差 (Mean Absolute Error, MAE), 用以计算预测值与真实值的误差绝对值的平均值, 其公式如下:

$$MAE = \frac{1}{n} \sum_{k=1}^n |(predict_k - actual_k)| \quad (7)$$

其中, predict_k 为第 k 个预测值, actual_k 为第 k 个预测值对应的真实值。MAE 值可以衡量预测值与真实值的差距, MAE 值越大, 模型预测结果越偏离真实值。

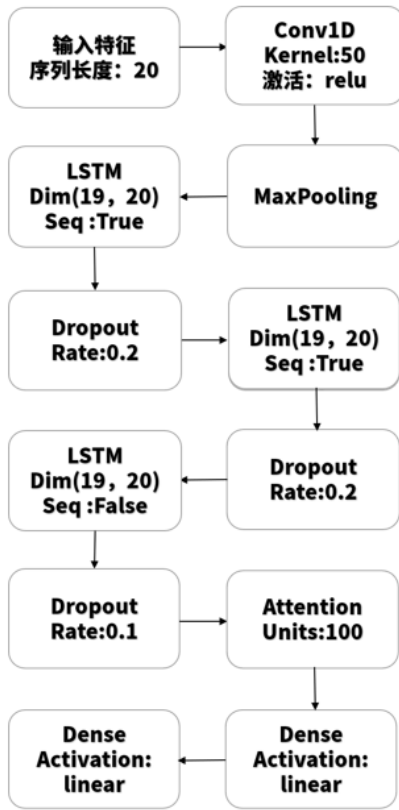


图 6. LSTM+CNN+ATTENTION 结构图

2) 均方误差:

$$MSE = \frac{1}{n} \sum_{k=1}^n (predict_k - actual_k)^2 \quad (8)$$

3) 平均绝对百分误差:

$$MAPE = \frac{100\%}{n} \sum_{k=1}^n \left| \frac{predict_k - actual_k}{actual_k} \right| \quad (9)$$

MAPE 指标在一定程度上消除了量纲的影响，其可以用于比较模型预测不同股票时的性能。MAPE 指标也是一个负向指标，其值越小说明模型的预测性能越好。

4) 预测准确率:

$$ACC = \frac{1}{n} \sum_{k=1}^n I(predict_k = actual_k) \quad (10)$$

其中 I 为指示函数， $predict_k = actual_k$ 代表预测涨跌和实际涨跌相同。ACC 可以描述模型捕捉涨跌信号的能力，ACC 越大，模型捕捉涨跌信号能力越强，模型越优秀。

对于传统 LSTM 模型，LSTM+CNN 模型以及 LSTM+CNN+ATTENTION[11], [12] 模型，我们使用沪深 300 中的成分股 600048 进行演示对比。根据图 3，

表 1
TREE DIFFERENT MODEL COMPARISON

Model	MAE	MSE	MAPE(%)	ACC(%)
LSTM	0.2805	0.1305	2.673	0.5165
LSTM+CNN	0.3257	0.1658	3.1172	0.5023
LSTM+CNN+ATT	0.215	0.0815	2.0838	0.5413

图 5，图 7 对比我们可以发现，传统 LSTM 模型捕捉涨跌信号的能力与加入 CNN 接近，但是均低于 LSTM+CNN+ATTENTION 模型，再加入注意力机制之前，会发现 CNN 模型的引入反而将模型的预测精度降低，这可能是 CNN 提供的特征模型无法很好分析。

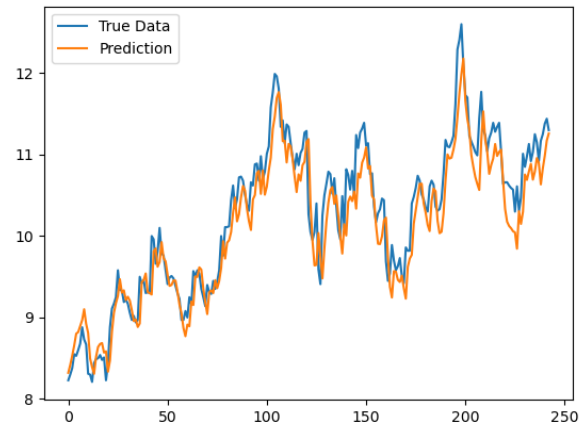


图 7. LSTM+CNN+ATTENTION 模型预测

由表格 1 可知，加入了注意力机制后，可以发现模型不管从拟合的精度还是涨跌准确率都优于其他两个模型，LSTM+CNN+ATT 模型 MSE 在 LSTM 模型的基础上降低了 40% 左右，这说明加入了 CNN 以及注意力机制之后，能实质性地改进模型的预测能力。所以我们得到的结论是，最好的模型是 LSTM+CNN+ATT 模型，次优的是 LSTM 模型，表现最差的是 CNN+LSTM 模型。

V. 股票交易的回测

A. 模型回测

回测 (Back Test) 是根据历史数据来验证交易策略的可行性和有效性的过程。做回测是希望可以用回测后的表现来评估未来实盘表现。我们假设如果回测结果好，实盘结果也不会太差，也就是说我们假设回测时的市场历史表现会在未来重演。现在提供回测的策略平台有很多，比如国泰安，BigQuant，掘金量化等，但是根据策略的不同，假设的不同，或者股票指数的不同，回测

的逻辑结构也不尽相同，如果没有能力搭建自己的回测算法，那说明还没有进入股票量化门槛。

本文根据以上对股票时间序列的预测模型，搭建了结合止盈止损策略的较为浅显回测算法，由于是第一次做股票量化，所以此算法只能提供部分验证性。回测结果的好坏可以很大程度反应运用该模型的历史收益率，是在进行了较多的理想假设的基础上实现的：

label=hypothesis

- 1) 忽略操作员的交易时需要的时间以及失误
- 2) 操作员或者策略者是完全理性的，是策略至上主义者
- 3) 认为数据都是可以及时得到的，也有充足的时间训练模型
- 4) 模型对股票的适用性在短时间内不发生变化

B. 回测算法的搭建

本文采用较为简单的固定止损 (需要金融学知识较少)，每次进场都带有固定点数的止损进场，而固定的点数也是根据自己策略的级别和资金的规模决定的，止盈也是如此，虽然看起来有些机械化但是至少能够保护账户的安全。该模型的止损率设置为-0.05(非常保守)，止盈率设置为 0.10，初始资金为 500,000 元，股票最大持有天数为 5 天，即一周的交易日，这也是短线策略较常用的参数。

回测算法包括两大模块，分别为 Create_Pool 和 Account，Create_Pool 的功能是 load 那些事先训练好的进入初始股票池的模型，根据预测下一天的股票价格计算预测当天收益率，筛选出预测收益率为正的股票并根据收益率排序作为第二天的股票池。Account 模块是对个人账户的模拟，会根据止盈止损策略进行股票买卖，并进行必要信息的记录。以下以沪深 300 中的 30 支成分股作为初始股票池，对于 2022 年 11 月 1 日中的交易日进行回测。

2022-11-15 00:00:00	止盈止损 (002864)	1200股, 股价: 81.61, 收入: 97932.0, 手续费: 127.31, 剩余现金: 99079.22, 最终盈利: 1211.72
2022-11-15 00:00:00	止损止损 (002850)	400股, 股价: 215.17, 收入: 86068.0, 手续费: 111.89, 剩余现金: 185035.33, 最终亏损: -1894.24
2022-11-15 00:00:00	止盈止损 (002849)	180股, 股价: 628.7, 收入: 62878.0, 手续费: 81.73, 剩余现金: 247823.6, 最终盈利: -91.59
2022-11-15 00:00:00	止盈止损 (002807)	180股, 股价: 548.82, 收入: 54882.0, 手续费: 71.35, 剩余现金: 302634.25, 最终盈利: 1904.79
2022-11-15 00:00:00	止盈止损 (002236)	180股, 股价: 540.91, 收入: 54091.0, 手续费: 70.32, 剩余现金: 356654.93, 最终盈利: 833.73
2022-11-15 00:00:00	止盈止损 (600745)	600股, 股价: 150.35, 收入: 90210.0, 手续费: 117.27, 剩余现金: 446747.66, 最终盈利: 677.91
2022-11-15 00:00:00	止盈止损 (600837)	500股, 股价: 170.54, 收入: 85270.0, 手续费: 110.85, 剩余现金: 531906.81, 最终盈利: -66.41
2022-11-15 00:00:00	买入 002050 (002050)	400股, 股价: 215.17, 收入: 86068.0, 手续费: 25.82, 剩余现金: 445812.99
2022-11-15 00:00:00	买入 601111 (601111)	9700股, 股价: 10.9, 收入: 105730.0, 手续费: 31.72, 剩余现金: 340051.27
2022-11-15 00:00:00	买入 002044 (002044)	1200股, 股价: 81.61, 收入: 97932.0, 手续费: 29.38, 剩余现金: 242089.89
2022-11-15 00:00:00	买入 002049 (002049)	100股, 股价: 628.7, 收入: 62870.0, 手续费: 18.86, 剩余现金: 179201.03
2022-11-15 00:00:00	买入 600115 (600115)	19700股, 股价: 5.35, 收入: 105395.0, 手续费: 31.62, 剩余现金: 73774.41
2022-11-15 00:00:00	买入 600010 (600010)	5000股, 股价: 14.7, 收入: 73500.0, 手续费: 22.05, 剩余现金: 252.36

图 8. 11-15 的交易详情

由图 9 可知，模型在 11 月内一月的收益达到了 23658 元，说明模型在股市中获利的能力是非常不错的，但是也可以发现的问题是，模型的表现随着时间的推移逐渐下降，到 17 号之后开始收益率为负了。这是因为

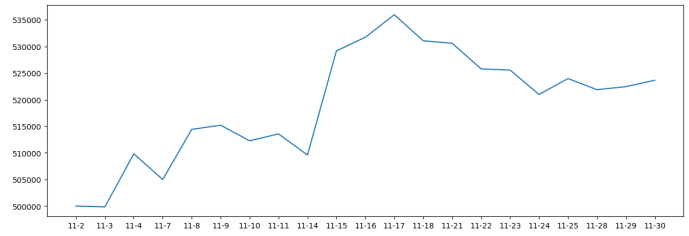


图 9. 11 月模型回测总资金情况

由于计算机算力限制，我们用的模型是不随着时间更新的，由于股市的波动是非常大的，再加上模型预测 n 天后的误差不是线性上升而是指数上升。这为后续的研究提供了思路，可以制作随着时间不断更新的策略模型，编写更有效率的代码来改进该策略。

致谢

本文不仅是一份项目汇报，更是一份学习的记录，感谢传龙和童行伟两位教授的教学与指导，没有他们为学生制定的学习计划，作者不可能开始去接触金融量化，并产生了极大的学习兴趣。同时需要感谢 keras-team(git), 唐宇迪博士(git), Gambler_Evolution(git)这些开发者在 github 上提供了开源学习代码，为此次项目提供了许多的借鉴意义和学习支持。最后要感谢 IEEE²提供的 Latex 模板，为编写论文提供了“方便”。

参考文献

- [1] Y. Yu, X. Si, C. Hu, and J. Zhang, “A review of recurrent neural networks: Lstm cells and network architectures,” *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
- [2] Z. Huang, W. Xu, and K. Yu, “Bidirectional lstm-crf models for sequence tagging,” *arXiv preprint arXiv:1508.01991*, 2015.
- [3] 柏万宽, “Rnn 神经网络在股票指数价格预测模型的研究与应用,” Master’s thesis, 重庆大学, 2018.
- [4] 任君, “基于 svm 和 lstm 网络预测的股票量化投资模型,” Master’s thesis, 武汉理工大学, 2019.
- [5] 周飞燕, 金林鹏, 董军 *et al.*, “卷积神经网络研究综述,” *计算机学报*, vol. 40, no. 6, pp. 1229–1251, 2017.
- [6] 王莉静, 戴远东, 赵海盟, and 李雅婷, “基于一维卷积神经网络客户需求分类模型,” *天津城建大学学报*, 2022.
- [7] 王宇轩, “基于卷积神经网络的股票预测,” Master’s thesis, 天津工业大学, 2019.
- [8] 任欢 and 王旭光, “注意力机制综述,” *计算机应用*, vol. 41, no. S01, pp. 1–6, 2021.
- [9] S. Hao, D.-H. Lee, and D. Zhao, “Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system,” *Transportation Research Part C: Emerging Technologies*, vol. 107, pp. 287–300, 2019.

²电气和电子工程师协会 (IEEE, 全称 Institute of Electrical and Electronics Engineers) 是一个美国的电子技术与信息科学工程师的协会，是目前世界上最大的非营利性专业技术学会，其会员人数超过 40 万人，遍布 160 多个国家。

- [10] A. Katharopoulos, A. Vyas, N. Pappas, and F. Fleuret, “Transformers are rnns: Fast autoregressive transformers with linear attention,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 5156–5165.
- [11] 田志宏, “基于 cnn-rlstm-atten 模型的股票价格指数预测研究,” Master’s thesis, 哈尔滨工业大学, 2021.
- [12] 李晨阳, “基于 cnn-lstm 的股票价格预测及量化选股研究,” Master’s thesis, 西北大学, 2021.