Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2025, 14(1), 237-245
Published Online January 2025 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2025.141024

Hans汉斯

基于A2C算法的股票交易模型

肖　豪，柯宗武

湖北师范大学计算机与信息工程学院，湖北　黄石

收稿日期：2024年12月20日；录用日期：2025年1月20日；发布日期：2025年1月30日

摘　要

2024年9月中国A股市场大涨，再次点燃了全民的“炒股热”。然而，牵动股民心弦的股价涨跌——却跟许多因素息息相关。对于散户来说，除了筛选信息进行股票的买进卖出以外，通过算法模型预测也能够起到事半功倍的效果。上世纪六十年代初便有了通过计算机技术进行量化交易的雏形，随着技术的迭代，通过统计学和模型构建成为量化交易的主流选择。而本论文构建了一个使用A2C (优势行动－评论家)强化学习算法的股票交易模型。利用“gym-anytrading”库创建一个股票交易环境，并使用Stable-Baselines库训练一个策略网络来学习如何在该环境中进行交易以最大化收益。该模型的数据来源于Yahoo-Finance的阿里巴巴股票信息(2022年12月至2024年9月)，通过pandas-datareader库的接口获取。

关键词

量化交易，强化学习，A2C算法，Gym-Anytrading，Stable-Baselines

A Stock Trading Model Based on A2C Algorithm

Hao Xiao, Zongwu Ke

College of Computer and Information Engineering, Hubei Normal University, Huangshi Hubei

Received: Dec. 20th, 2024; accepted: Jan. 20th, 2025; published: Jan. 30th, 2025

Abstract

In September 2024, a significant surge in China’s A-share market reignited the public’s “stock trading frenzy”. However, the fluctuating stock prices that excited stock investors were closely related to many factors. For individual investors, in addition to screening information for buying and selling stocks, using an algorithm model to predict can also have a twice-as-effective effect. In the early 1960s, the embryo of quantitative trading using computer technology had appeared, and with the

文章引用: 肖豪, 柯宗武. 基于 A2C 算法的股票交易模型[J]. 人工智能与机器人研究, 2025, 14(1): 237-245.
DOI: 10.12677/airr.2025.141024

advancement of technology, quantitative trading based on statistics and model building became the mainstream choice. This paper constructs a stock trading model using the A2C (Advantage Actor-Critic) reinforcement learning algorithm. By using the "gym-anytrading" library to create a stock trading environment and training a policy network using the Stable-Baselines library to learn how to trade in this environment to maximize profits. The data source for the model comes from the stock information of Alibaba (2022 December to 2024 September) obtained through the interface of the pandas-datareader library.

Keywords

Quantitative Trading, Reinforcement Learning, A2C Algorithm, Gym-Anytrading, Stable-Baselines

1. 引言

随着各国经济联系日趋密切，全球的金融市场也呈现不同的态势，尤其是中国的股票市场，正以强劲的姿态发展。中国拥有着仅次于美国的股民数量，然而中国的股民素来被称为“韭菜”[1]。究其主因，股票行情信息纷繁复杂，往往散户会以远慢于金融市场震荡的速度，过早或过晚地抛售、买入股票，最终亏损被“割韭菜”或者接盘“垃圾股”成为冤大头[2]。如何对股票信息进行筛选和判断，提早发现单支股票的涨跌行情，对于散户或者投资机构来说十分重要。

股票量化交易早期主要依赖于简单的技术指标，例如移动平均线、相对强弱指标(RSI)、MACD 等[3]，一些统计模型计算能力有限，主要依靠人工进行数据分析和交易策略制定，策略简单，容易被市场噪音干扰，缺乏对市场风险的有效控制，交易频率较低[4]。

随着计算机技术和数据处理能力的提高，量化交易开始使用更复杂的统计模型，例如回归分析、因子模型等[5]。开始使用更高级的编程语言和数据库技术[6]。但是，随着高频交易开始出现，对数据质量和模型参数的依赖性较高，模型的有效性容易受到市场环境变化的影响，高频交易面临着技术风险和监管风险[7]。

综上所述，本文所使用的 A2C 算法能够自动学习最优的交易策略，而不需要人工设计复杂的规则[8]；能够适应市场环境的变化，因为它能够不断地学习和调整策略[9]；能够处理高维数据和时间序列数据，例如结合 LSTM 网络处理股票价格的时间序列特征[10]。

本文的组织结构如下：

第二节介绍了 A2C 算法模型的数据来源、环境与模型的结构、训练过程、模型的关键技术工作原理。第三节介绍了模型的训练，其中包括了：训练流程、评估指标、训练结果。最后在总结部分对本文的研究内容进行总结和模型改进的展望。

2. 基于 A2C 算法的交易模型创建及训练

2.1. 数据获取与预处理

本论文选择 2022 年 12 月 1 日至 2024 年 9 月 30 日阿里巴巴的股票信息作为数据研究对象。阿里巴巴作为中国互联网头部企业，其股票在美国纽交所上市。相比于 A 股市场的信息披露和监管力度，美国证券交易委员会(SEC)的监管力度更加严格和全面，吸引了全世界范围内的投资者，具有交易量大、流动

性更好的优势，作为投资者买卖其股票降低了交易成本，也减少了对于最终盈利核算的影响。如图 1 所示，为 2022-12-01 至 2024-9-30 阿里巴巴股票信息。

Price	Adj Close	Close	High	Low	Open	Volume
Ticker	BABA	BABA	BABA	BABA	BABA	BABA
Date						
2022-12-01 00:00:00+00:00	82.996269	85.940002	87.599998	84.260002	84.349998	20243000
2022-12-02 00:00:00+00:00	86.975143	90.059998	91.849998	86.050003	86.050003	35042600
2022-12-05 00:00:00+00:00	87.419380	90.519997	92.900002	89.629997	92.900002	30929800
2022-12-06 00:00:00+00:00	88.317520	91.449997	92.570000	89.199997	91.879997	26675700
2022-12-07 00:00:00+00:00	85.304398	88.330002	89.260002	86.620003	87.059998	19507400
...	...	...	...	...	...	...
2024-09-24 00:00:00+00:00	97.190002	97.190002	97.500000	94.400002	96.070000	47423200
2024-09-25 00:00:00+00:00	95.459999	95.459999	96.180000	94.059998	94.379997	18816600
2024-09-26 00:00:00+00:00	105.070000	105.070000	105.970001	101.760002	102.690002	67261600
2024-09-27 00:00:00+00:00	107.330002	107.330002	109.430000	105.730003	105.970001	50075200
2024-09-30 00:00:00+00:00	106.120003	106.120003	112.220001	106.105003	111.720001	58561500

Figure 1. 2022-12-01 to 2024-9-30 Alibaba stock information
图 1. 2022-12-01 至 2024-9-30 阿里巴巴股票信息

同时，选择了雅虎金融(Yahoo Finance)作为数据源。雅虎金融提供大量的金融数据包括：股票价格、交易量、财务报表数据等。相比较于其他数据提供商，雅虎金融大部分数据免费；数据通常以结构化的格式呈现——方便数据分析和处理；数据通常可以通过 API 接口获取。

使用“Pandas-Datareader”库的“Web.DataReader”函数获取在线数据，该函数将雅虎金融发送的 JSON 数据解析为 Python 对象，其交互过程如图 2 所示：

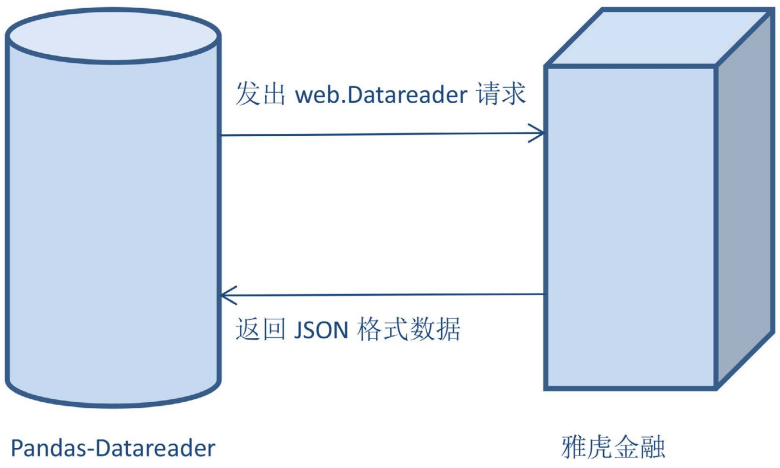

Figure 2. Data streaming acquisition process
图 2. 数据流获取过程

2.2. 环境与模型的构建

本论文通过使用 Python 的 Gym 库为模型构建一个股票交易环境。Gym 库的核心函数是 gym.make

(stocks-v0)。其中“stocks-v0”是“gym-anytrading”库注册的环境 ID，表示要创建的是一个股票交易环境。创建好的环境会接收已经打包完成的股票数据集，该数据集包括了开盘价、最高价、最低价、收盘价、成交量等信息。数据传输完成后，环境进行规范化预处理，以便输入模型训练。该环境使用指定的股票数据(阿里巴巴)、数据范围(2022-12-01~2024-9.30)和窗口大小(每一次循环中模型训练的数据规模)，如图 3 所示：

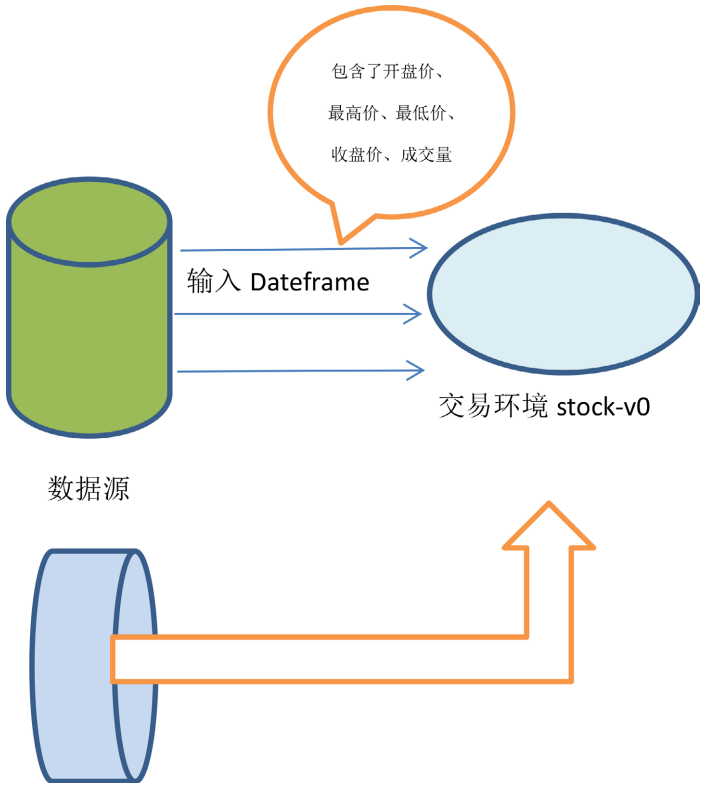

Figure 3. Environment creation
图 3. 环境的创建

而模型的构建主要工作是对环境预处理好的数据进行接收、设置好训练的次数、选取合适算法和策略网络、确定好训练流程。

模型选取了 Stable Baselines3 库的 A2C 算法作为模型训练的核心算法，并辅助使用了 MlpLstmPolicy 网络结构进行策略的更新。

其中 A2C 算法指的是 Actor-Critic，一种基于策略梯度(Policy Gradient)和价值函数(Value Function)的强化学习方法。它通过产生交易动作提前完成对股票的买入或者卖出，通过与实际股票的涨跌情况作对比，评判本次交易的结果是盈利还是亏损，并根据奖励(判断的结果)更新下一个阶段的行为，也就是下一阶段买入还是卖出，以得到更好的资金回报。

A2C 算法会创建多个进程与环境进行交互，以收集更多的交易经验，这对于优化每一次的动作都能起到正向效果。如图 4 所示。

MlpLstmPolicy 策略网络是多层感知器 MLP 和长短期记忆网 LSTM。MLP 能够辅助处理交易量、技术指标等不随时间变化的数据；LSTM 则辅助处理价格序列等随时间而变动的因素，捕捉股票价格变化过程中的依赖关系，通过与 MLP 相结合形成 MlpLstmPolicy 策略网络，能够更好地为 A2C 算法更新策

略，以得到更好的预期收益。

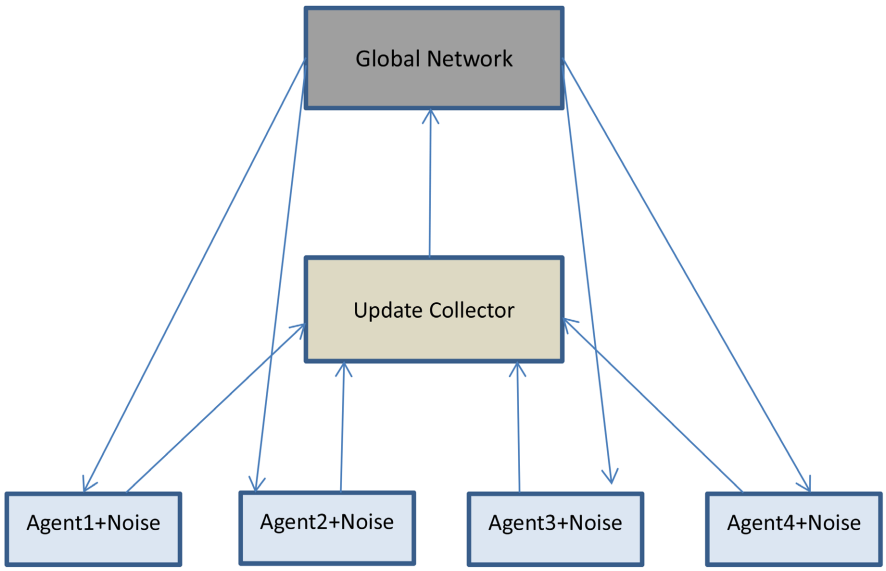

Figure 4. The working principle of A2C algorithm in multi-threading
图 4. A2C 算法多线程工作原理

3. 模型的训练

3.1. 训练流程

模型训练开始前，必须对模型的训练参数进行设置，主要是循环训练的次数(根据实际数据源的数量进行估计)、获取的窗口大小(一次抓取的数据规模)。

模型训练第一步：初始化。这个过程由 Gym 环境处理完成。这个处理过程隐含在“gym-anytrading”库中。

模型训练第二步：数据获取。前文所述，由环境会自动将打包完成的数据输入给模型。

模型训练第三步：决策网络。该步骤负责生成一个交易策略，以 A2C 算法为核心辅以 MlpLstmPolicy 策略网络生成一个决策。该决策是动态存在，在设定的最大训练次数之前，每一次完整的训练都会根据奖励更新策略。

模型训练第四步：行动。即根据策略完成一次完整的股票买入或者卖出。在预期中该行动应该是股票的高点卖出，低点买入。

模型训练第五步：奖励。奖励本质是上一次交易动作产生的后果，后果可能为正向或负向。根据一次决策完成一次行动，将行动结果与股票的数据进行比对，即可知道该行为产生的后果。

模型训练第六步：更新策略。由第五步的后果，回到第三步，调整策略，该过程由算法自主学习完成。更新策略后使用新策略继续完成第三步到第六步。

模型在执行循环往复的训练次数后结束。其流程图如图 5 所示。

3.2. 评估指标

股票交易模型的最终评估指标以收益率为准。本模型引入 MATLAB 以构建可视化的图表，最终的收益率和决策过程会生成散点图，散点会落在股票的实际涨跌线上，以进行对比。

如图 6 所示，在没有任何算法训练的情况下，Gym 环境内自行交易的结果。

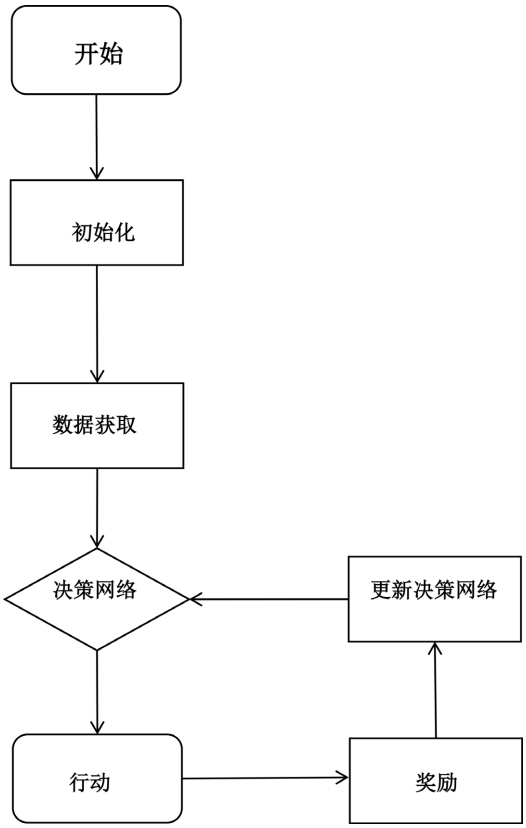

Figure 5. Training flowchart
图 5. 训练流程图

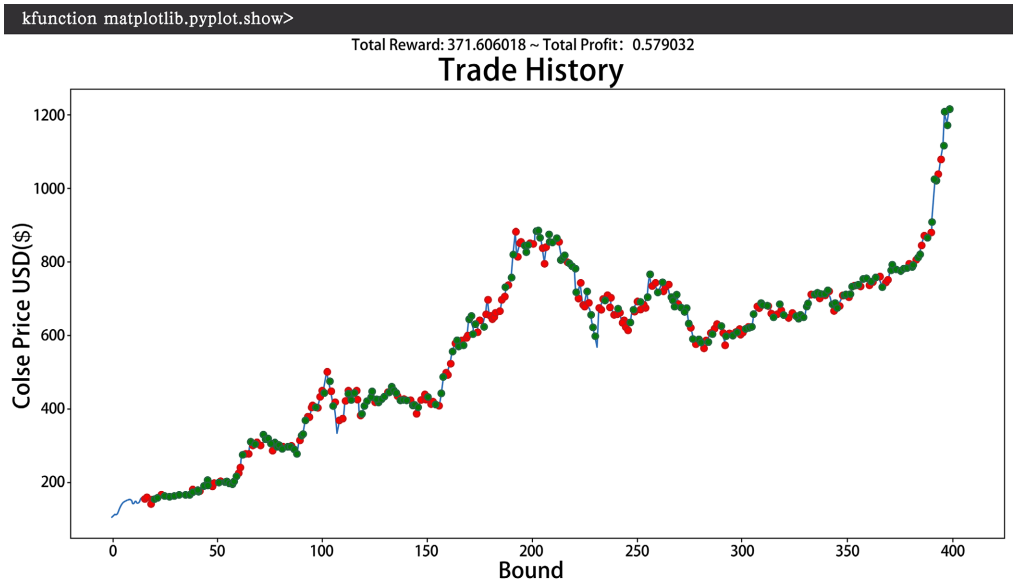

Figure 6. Chart of transactions in the system without training
图 6. 无训练状态下系统的交易图

在该图 6 中，红点表示卖出行为，绿点表示买入行为。由于第一次进行数据训练不会产生买入和和卖出行为，因此最开始的一段曲线上并无散点。

分析图中的散点图可知，在没有算法的情况下，自主交易存在混乱的情形，在持续走高的趋势中连续买入，在持续走低的趋势里卖出的行为却寥寥无几。根据股票常识易知，假如阿里巴巴股票在 2021-12-01 至 2024-09-30 两年间的股票走势总体降低，那么连续买入的操作(图 6 中绿点占据大部分比重)最终结果将是亏损，同时根据图中最终的收益显示为：“Total Profit” = 0.579032 可知，其收益率小于 1，即最终余额约只有本金的 58%。

而阿里巴巴股票的总体趋势图也验证了这一次结果必然是亏损的，如图 7 所示：

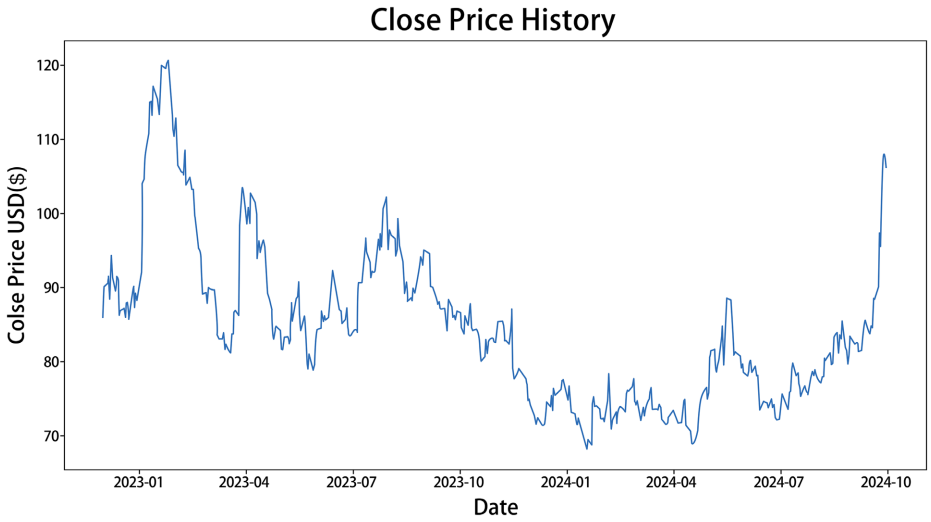

Figure 7. Alibaba stock trend chart
图 7. 阿里巴巴股票涨跌趋势图

3.3. 训练结果

由 3.2 可知，一个无算法的股票交易模型最终的结果是亏损的，下文将展现运用算法模型的训练结果。

在经过 3.1 的训练步骤之后，该模型的训练结果比较于无算法状态下的收益率有了较为明显的进步。如图 8 所示：

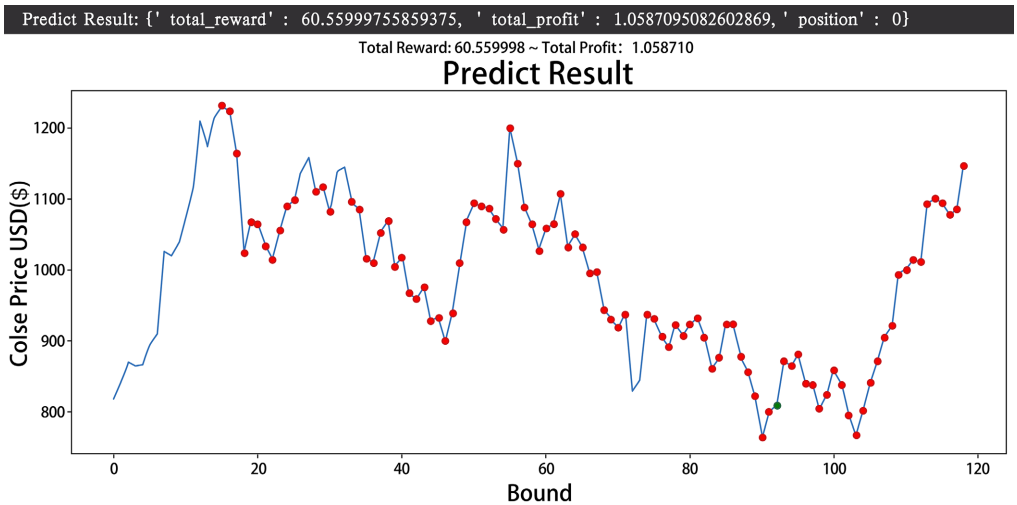

Figure 8. Initial training behavior trajectory and return rate
图 8. 初步训练行为轨迹与收益率

在图 8 中，收益率为 1.058710，大于 1，交易模型有了预期的效果。但是约 5.87%的盈利率有待提高。实际的股票交易中还有手续费等相关费用，所以现实的收益率依然大打折扣。且观察其行为轨迹能看出模型的行为较为单一保守，几乎所有行为都是卖出股票，受益于阿里巴巴的股票整体呈下跌趋势，因此总体比较什么也不做或者连续买入是获利的。

考虑到模型的优越性与训练次数(时间)和观察窗口有关，提高训练次数并扩大观察窗口，得到了更加理想的结果。如图 9 所示：

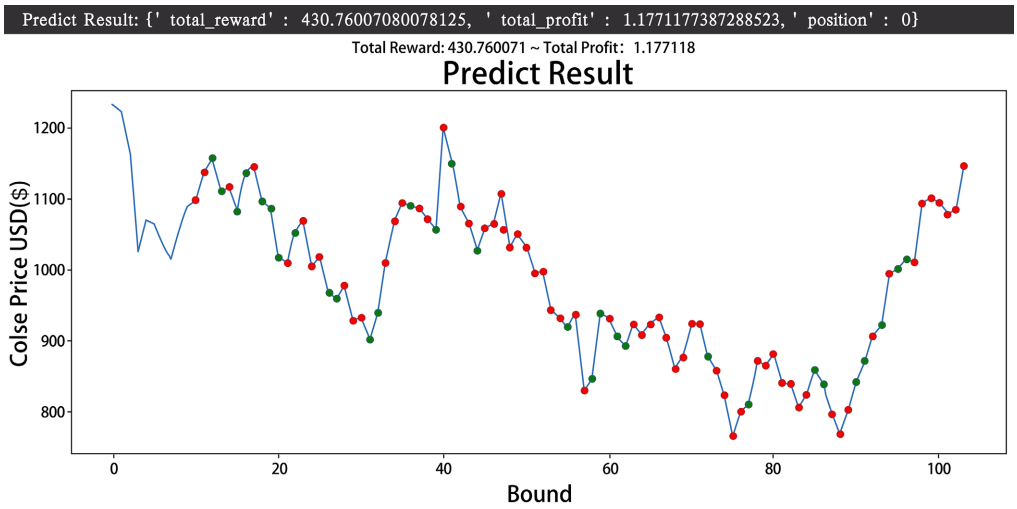

Figure 9. Adjusted parameters behavior trajectory and return rate
图 9. 调整参数后的行为轨迹与收益率

观察图 9 可知收益率为 1.177118，最终盈利率率约达到了 17.71%。根据图中的散点轨迹能够发现，买入和卖出的操作在未知实际涨跌的情况下跟预期(高点卖出，低点买入)大致吻合。这说明调整参数起到了正向的效果，最终是盈利的。

同时，17.1%的盈利率远高于美联储的存储利率，意味着投资者运用了本模型，将本金用于阿里巴巴股票的量化交易远大于储蓄利息，运用了 A2C 算法的交易模型训练成功。

4. 总结和展望

本论文通过 Python 库工具包抓取阿里巴巴两年内的股票数据，使用 Gym 库创建了一个股票交易的环境，将数据封装进算法模型训练，训练需要一定的 GPU 资源，本项目通过谷歌的 COLAB 笔记提供的 GPU 完成训练，最终得到了一个基于 A2C 算法的股票交易模型。该模型利用 A2C 算法的行为-评价体系，对过往行为结果进行反复的评价后对策略进行往复的调整，优化行为，最终根据训练结果验证了该模型的有效性。本论文存在一些可以改进的地方，例如：能够尝试在数据处理环节，运用对数据优化的算法，对数据进行筛选，提高模型的运行速度；实际上股民的手里不可能单一持股，因此模型可以为用户添加搜索指定股票并下载其数据的功能，用户能够对多支的股票进行训练，得到针对不同股票的不同模型。

展望本论文提出的算法模型，本文的提高工作可以着眼于对收益率的提高和行为的稳定性。第一：由于影响股票涨跌存在诸多因素，因此可以根据本论文内容设置两个股票模型。其中，A 模型对股票的数据进行训练，B 模型以 A 模型的行为和收益率为数据基础，再结合影响股票的宏观政策、突发事件、市场变动等外界要素融合进行训练，最终得到一个更加贴合现实情况的交易模型。第二：考虑到两年内

的数据并不具备代表性，对于一支股票而言，全面综合地考察其表现需要追溯到上市伊始的数据，因此增加数据量和扩大每一次的数据观察窗口也是提高模型性能的途径，当然这也对训练的硬件和时间提出了更严格的要求。第三：本论文训练的是上市于美股的阿里巴巴股票，不同股市的股票具备不同的涨跌规律，为了更加全面和客观，亦可挑选中国的 A 股及港股股票进行训练。

参考文献

[1] 牛晓健, 侯启明. 基于 CNN-LSTM 模型的中国股票价格预测与量化策略研究[J/OL]. 贵州省党校学报, 2024: 1-18. https://doi.org/10.16436/j.cnki.52-5023/d.20241128.005, 2024-12-19.

[2] Han, A. and Munan, L. (2024) Exploiting the Potential of a Directional Changes-Based Trading Algorithm in Stock Market. *Finance Research Letters*, **60**, Article ID: 104936.

[3] 吴灿柳, 陈小英. 基于邻近森林的量化交易系统[J]. 软件导刊, 2024, 23(10): 82-87.

[4] Dumiter, F.C., Turcaș, F., Nicoară, Ș.A., Bențe, C. and Boiță, M. (2023) The Impact of Sentiment Indices on the Stock Exchange—The Connections between Quantitative Sentiment Indicators, Technical Analysis, and Stock Market. *Mathematics*, **11**, Article No. 3128. https://doi.org/10.3390/math11143128

[5] 雷鹏. 基于 A2C 算法考虑投资风险的股票交易策略研究[D]: [硕士学位论文]. 成都: 西南财经大学, 2024.

[6] 蔡云龙. 大数据驱动的湿法冶金全流程优化控制模型及实证研究[J]. 湿法冶金, 2024: 1-10.

[7] 何杉杉, 周雅兰, 郭宇阳. 基于 LSTM 和 DDPG 的股票交易决策算法[J]. 南京大学学报(自然科学), 2024, 60(6): 940-953.

[8] Wu, Y., Fu, Z., *et al*. (2023) A Hybrid Stock Market Prediction Model Based on GNG and Reinforcement Learning. *Expert Systems with Applications*, **228**, Article ID: 120474. https://doi.org/10.1016/j.eswa.2023.120474

[9] 陈家祥. 基于强化学习的股票交易自适应辅助决策系统的设计与实现[D]: [硕士学位论文]. 北京: 北京邮电大学, 2024.

[10] 徐智钊. 量化交易策略问题中的深度强化学习算法研究[D]: [硕士学位论文]. 济南: 山东师范大学, 2024.