doi:10.3969/j.issn.1001-4616.2025.01.011

# 基于深度强化学习的股票量化投资研究

宋　飞[1,2],田辰磊[2],冯传威[3]

(1.南京大学信息管理学院,江苏 南京 210023)
(2.南京林业大学理学院,江苏 南京 210037)
(3.南京大学数学学院,江苏 南京 210093)

[**摘要**]　针对股票量化投资,将深度强化学习中的 Deep Q-learning(DQN)模型应用于算法交易,构建端到端的算法交易系统. 首先,利用股票技术分析指标设计股票交易环境,从时间尺度扩充特征集;其次,定义智能体交易的奖励函数和动作空间;然后,设计 $Q$ 网络结构,将支持向量机和极致梯度提升法学习股票历史数据的涨跌信号加入强化学习中;最后,将算法交易系统应用于中国股票市场,并选择招商银行和泰和科技两支股票以及其余 4 支股票进行验证,从收益率、夏普比率和最大回撤率三方面评价投资绩效,结果表明该算法系统在收益率上有显著提升的同时,最大回撤率有所降低,模型的抗风险能力较高.

[**关键词**]　量化投资,DQN,强化学习,算法交易
[**中图分类号**]TP18　[**文献标志码**]A　[**文章编号**]1001-4616(2025)01-0085-08

# Research of Stock Quantitative Trading Algorithm Based on Deep Reinforcement Learning

Song Fei[1,2],Tian Chenlei[2],Feng Chuanwei[3]

(1.School of Information Management,Nanjing University,Nanjing 210023,China)
(2.College of Science,Nanjing Forestry University,Nanjing 210037,China)
(3.School of Mathematics,Nanjing University,Nanjing 210093,China)

**Abstract**:This paper focuses on quantitative stock investment and applies the Deep Q-learning(DQN)model from deep reinforcement learning to algorithmic trading,constructing an end-to-end algorithmic trading system. Firstly,the system utilizes stock technology analysis index to design a stock trading environment,expanding the feature set from a time scale;secondly,it defines the reward function and action space for intelligent agent transactions;then,it designs $Q$-network structure and incorporate Support Vector Machine and eXtreme Gradient Boosting method to learn the rise and fall signals of stock historical data into reinforcement learning;finally,the algorithmic trading system is applied to the Chinese stock market and China Merchants Bank and Taihe Technology,as well as the remaining four stocks are selected for validation. The investment performance is evaluated from three aspects:return rate,sharpe ratio,and maximum drawdown rate. The results are shown that the algorithmic system significantly improved the return rate while reducing the maximum drawdown rate,indicating that the model has a high risk resistance ability.

**Key words**:quantitative investment,DQN,reinforcement learning,algorithmic trading

量化投资是始于 20 世纪 70 年代的量化交易,近年来逐渐得到广大机构投资者的高度关注和应用. 量化投资是利用计算机技术对金融市场中大量的真实历史数据进行分析,进而建立科学模型挖掘投资机会和交易信号,从中学习高效且系统的交易策略.

我国量化交易发展的历史较短,首支量化基金(华安上证 180 指数增强型基金)诞生于 2002 年,当时国内量化投资仍处于摸索状态;2010 年首支股指期货沪深 300 股指期货上市,丰富了量化交易的可行策略. 人工智能技术的兴起使得量化交易策略有了新的应用[1-2]. 近年来,受益于计算机设备分析能力和机

**收稿日期**:2024-10-10.
**基金项目**:国家自然科学基金项目(12201303).
**通讯作者**:宋飞,博士,副教授,研究方向:深度学习. E-mail:songfei@njfu.edu.cn

器学习乃至整个人工智能领域的探索和发展,量化策略迎来了新的研究方向.

针对股票投资的研究,之前绝大多数是基于监督学习的价格走势进行预测. 由于股票历史数据具有嘈杂性和不稳定性,股市预测极具挑战性. 强化学习(Reinforcement Learning),尤其是 Q-learning,可以直接学习决策规则并获得合理的奖励,适用于学习交易策略. 强化学习在股票、期货等量化交易领域已经成为强化学习领域的研究热点[3]. Jiang 等[4]将 Deep Q-learning(DQN)用于数字货币投资;Xiong 等[5]将 DQN 算法应用于股票市场投资组合管理,选择了 30 只股票作为交易对象并将其每日价格作为市场环境,与道琼斯工业平均指数和传统最小方差投资组合分配策略进行了比较,其提出的深度强化学习智能体(Agent)在夏普比率和累计收益率方面均优于两个基准;韩道岐等[6]提出了基于深度强化学习的智能股市操盘手模型(ISTG),其加入历史行情数据、技术指标、宏观经济指标等数据,应用于中国股市 1 400 多支股票上的试验表明 ISTG 的总体收益率达到 13%,优于买入持有的表现;Azhikodan 等[7]提出利用深度强化学习实现自动交易,基于深度确定性策略梯度的神经网络训练模型,还使用了递归卷积神经网络建立情绪分析模型以预测股票趋势,证明了强化学习能够学习股票交易的技巧. Huang 等[8]提出了一种新的算法交易方法—CR-DQN,该方法将 DQN 与移动平均线(Moving Average,MA)和交易突破区间(Trading Range Break-out,TRB)结合在一起,并使用长短期记忆网络学习最优策略. Mahdi 等[9]开发了一种稳健 DQN 以预测波动性商品期货市场,使用门控循环单元学习策略. 深度强化学习结合了深度学习的感知能力和强化学习的自动决策能力,从而为股票自动化交易[10]提供了可行的解决方案. 研究者通常基于决策网络、奖励函数等模块进行改进,而鲜有针对股票多维度特征的改进,尽管目前的深度学习所给出的最优策略效果已经远超基准(Buy and Hold,BH),但是日益复杂的模型构造并未与多层感知机之类的简易网络全面拉开距离.

本文将针对股票市场建立深度强化学习模型进行端对端的自动交易. 基于 DQN 建立算法交易系统,进行股票日内交易,并从以下几个方面设计模型:首先,从时间尺度上丰富神经网络的输入特征集,选择不同时间分辨率的数据;其次,选择常用的股票技术分析指标加入环境特征. 此外,将 DQN 与监督学习结合,利用支持向量机(Support Vector Machine,SVM)和极致梯度提升(eXtreme Gradient Boosting,XGBoost)对原始数据进行学习,再将学习到的信号输入环境状态中.

本文设计的算法交易系统不仅拓宽了深度强化学习在金融领域的理论,也为机构投资者提供了有效的投资指导,丰富了深度强化学习在量化交易策略上的应用.

## 1　Deep Q-learning 模型

强化学习是机器学习的一个重要分支. 智能体(Agent)在与环境交互中学习动作策略 $\pi$,获得最大的累计奖励值. 具体来讲,如图 1 所示,在给定某时刻 $t$,智能体在状态空间 $S$ 中观测当前状态 $s_t$,在动作空间 $A$ 中选择动作 $a_t$ 与环境互动. 环境给予智能体奖励 $r_t$,并迁移到新的状态 $s_{t+1}$. 智能体根据 $r_t$ 逐步学习完善策略 $\pi$.

$Q$ 学习是一种基于价值的无模型强化学习方法. 通过动作—价值函数 $Q_\pi(s_t,a_t)$ 评估在当前状态 $s_t$ 下,智能体根据策略 $\pi$ 选择动作 $a_t$ 后可获得累计奖励 $G_t$ 的期望值. 动作—价值函数反映了智能体在某个状态选择一个动作,按照最优策略继续行动,最终能够获得的累计收益. 计算公式为

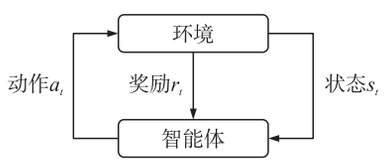


**图 1　强化学习方法的原理示意图**

**Fig. 1　Schematic diagram of the principle of reinforcement learning method**

$$Q_\pi(s_t,a_t)=E_{S_{t+1},A_{t+1},\cdots,S_n,A_n}[G_t\mid S_t=s_t,A_t=a_t], \tag{1}$$

其中 $G_t$ 为折扣回报,$\gamma\in[0,1]$ 为折扣因子,公式为

$$G_t=r_t+\gamma r_{t+1}+\gamma^2 r_{t+2}+\cdots+\gamma^{n-t}r_n. \tag{2}$$

由式(1)可知,$Q_\pi(s_t,a_t)$ 依赖于策略 $\pi$. 假设 $\pi^*$ 为最佳策略,则 $Q_*(s_t,a_t)$ 为最优动作—价值函数,此时 $Q_*(s_t,a_t)$ 只依赖于 $s_t$ 和 $a_t$,与 $\pi$ 不再相关,即

$$\pi^*=\underset{\pi}{\operatorname{argmax}}Q_\pi(s_t,a_t), \tag{3}$$

$$Q_*(s_t,a_t)=\max\ Q_{\pi^*}(s_t,a_t). \tag{4}$$

最初 $Q$ 学习基于表格形式出现,适用于维度较低的环境. 文献[11]提出了 Deep Q-learning(DQN)模

型,使用神经网络 $Q(s_t,a_t;\theta)$ 估计复杂环境下的最优动作—价值函数,其中 $\theta$ 表示神经网络的参数. 智能体执行动作 $a_t$ 后的奖励 $r_t$ 为确定值,因此 $r_t+\gamma \max\limits_a Q(s_{t+1},a_{t+1};\theta)$ 作为 DQN 模型的训练目标. DQN 模型训练使用时间差分(Temporal Difference,TD)算法[12]. 时间差分法将一个连续的时间过程拆分成许多小的时间片段,然后在每个片段内近似计算系统的变化. 损失函数为

$$L(\theta)=\frac{1}{2}(Q(s_t,a_t;\theta)-(r_t+\gamma \max_a Q(s_{t+1},a_{t+1};\theta)))^2. \tag{5}$$

# 2　基于 DQN 的算法交易系统

## 2.1　股票模拟环境设置

由于股票交易数据具有时序性和互动性,本文提出的交易模型是基于马尔科夫决策过程,可以用四元组 $(S,A,P_a,R_a)$ 来描述股票交易过程,其中 $S$ 表示股票市场的环境状态;$A$ 表示在与市场互动中采取的行动;$P_a$ 代表动作对应的状态转移概率;$R_a$ 表示当前动作获得的回报;$a$ 表示当前动作.

### 2.1.1　状态(State)

智能体在与环境做互动的过程中需要接收环境提供的信息,这些信息就是状态. 状态一般是以向量或者矩阵形式表示的特征,受限于信息收集成本、状态空间的容量等原因,本文选择部分特征来描述环境信息.

股票价格会受到多种宏微观因素的影响,理论上来说股价是由股票的价值决定的,但是由于市场上有非常多的因素如股票供求关系、国家政策、投资者情绪、经济水平等都会对股价产生直接影响[13]. 另外,股票技术分析指标是投资者常用的分析工具,用来预测股价的趋势,一定程度上可以代表当前股市的状态,减少历史数据噪声和随机性. 本文选取了股票的每日数据,包括开盘价、收盘价、最高价、最低价;考虑到从短、中、长期不同的时间尺度来衡量股价的趋势,状态增加了 5 日、10 日、30 日、120 日移动平均线;同时,为了丰富股市环境特征,选择了一些与股票价格走势相关性较大的技术指标,将这些指标作为输入特征. 具体特征如表 1 所示.

**表 1　股票环境状态特征**

**Table 1　Characteristics of stock environment status**

| 一级指标 | 二级指标 | 含义 |
|---|---|---|
| 股价基本型 | 开盘价格 | 开盘时股票价格 |
| | 收盘价格 | 收盘时股票价格 |
| | 最高价格 | 当日股票最高价 |
| | 最低价格 | 当日股票最低价 |
| | 成交量 | 当日股票成交量 |
| 重叠指标 | MA5、10、30、120 | 股价 5、10、30、120 日移动平均线 |
| | EMA30 | 30 日指数移动平均数 |
| | BBANDS | 布林带 |
| 波动指标 | ATR | 平均真实波动幅度 |
| | ADX | 平均趋向指数 |
| 超买超卖指标 | CCI | 多空头市场判断信号 |
| | KDJ-K、KDJ-D、KDJ-J | 买入卖出信号 |
| | RSI6、12、24 | 6、12、24 日相对强弱指标 |

注:构建的深度强度交易模型的环境状态为一个 1×20 的二维矩阵,其中 20 个特征即为上述 15 项技术指标与 5 项股票交易数据构成.

### 2.1.2　动作(Action)

强化学习包含状态集、动作集、奖励集 3 个元素,动作集就是智能体与环境互动过程中采取的行动,动作空间的设置将直接关系到智能体对环境的影响,从而影响模型的学习效果. 本文只考虑对单一股票进行交易,只涉及建仓、持仓等待、平仓策略. 智能体执行的是日内交易,在第 $t$ 交易日的动作 $a_t\in$ {买,持有,卖} = {1,0,-1}. 具体来说,智能体在一个交易日内只能执行以下动作中的一个:

●买(1):看涨股票,购买股票并期望其升值;

●卖(-1):看跌股票,出售股票;

●持有(0):保持中立,不进行操作.

策略的目标是给定股票价格的趋势估计(价格上涨还是下跌),选择使得回报最大的动作. 如果以当前价格购入股票能产生长期持有效益时,就会采取买入操作;当智能体预测到该股票长期持有收益下降或卖出操作长期收益高时,就会采取卖出操作;此外的情况,智能体就不采取行动.

本文根据 $\varepsilon$-greedy 策略选择出当前动作 $a_t$,可表示为:

$$a_t=\begin{cases}\underset{a}{\operatorname{argmax}}Q(s_t,a_t;\theta), & \text{以概率}(1-\varepsilon),\\ \text{均匀抽取 } A \text{ 中的一个动作}, & \text{以概率 } \varepsilon.\end{cases}$$

2.1.3　奖励(Reward)

奖励函数是算法交易系统的又一个关键的组成部分,强化学习中价值函数学习就是学习一系列动作所对于的奖励值,策略优化过程中需要选择奖励值最大的行动,其目的是使得累积收益最大化. 奖励函数通常与当前状态,执行的动作和下一步状态有关. 以往大多数基于机器学习的研究通常使用及时奖励,如每日利润 $r_t=p_t-p_{t-1}$ 或 $r_t=p_t/p_{t-1}-1$,其中 $p_t$ 为第 $t$ 天的股价. 然而,这种瞬时奖励并不适合 DQN,因为瞬时奖励噪声很大,无法为模型训练提供可靠的监督,同时这种奖励与长期利润的目标不一致. 本文在尝试多种奖励函数及其对应的微调后,认为使用过去 $n$ 天内积累的收益减去交易成本作为奖励函数[14]可以达到较好的训练效果. 其公式如下:

$$r_t=\left(1+\operatorname{sgn}(a_t)\frac{p_t-p_{t-1}}{p_{t-1}}\right)\frac{p_{t-1}}{p_{t-n}}-c\,|a_t-a_{t-1}|, \tag{6}$$

其中,$a_t\in\{-1,0,1\}$,$p_t$ 为收盘价,sgn( )为符号函数;$c$ 为交易成本.

因为股票交易属于连续型任务,此时强化学习的长期收益选择折扣收益的形式,公式如式(2)所示.

## 2.2　智能体构建

智能体一次训练周期内,每步时间为 $t=0,\cdots,T$,在 $t$ 时刻 Agent 接收环境状态 $s_t$ 后,根据策略选择动作 $a_t$,该动作被环境接收后,环境状态会根据环境状态转移函数计算出下一步的状态 $s_{t+1}$,奖励函数根据 $(s_t,a_t)$ 计算出奖励 $r_t$,Agent 收到奖励后优化自身策略,根据 $s_{t+1}$ 和策略选择动作 $a_{t+1}$,循环往复. Agent 的目标是学习到最优策略,使得基于该策略做出的动作可以使得累计奖励 $G_t$ 最大. 图 2 描述了一次周期内,智能体与环境的互动流程,示意图与图 1 类似,只是智能体的目标变为了整体的累积收益,智能体需要考虑长期的收益,而不是短期内的.

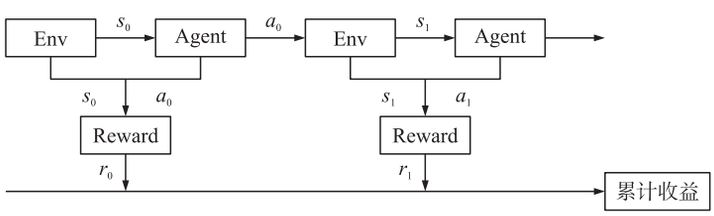


**图 2　一轮训练互动流程**

**Fig. 2　One round of training interactive process**

## 2.3　训练方式

本文采取经验放回机制进行模型的批次训练,将每个交易日获得的新样本放入历史交易数据,当达到足够数量后,随机选择一批数据作为 mini-batch 更新网络,其流程图如图 3 所示.

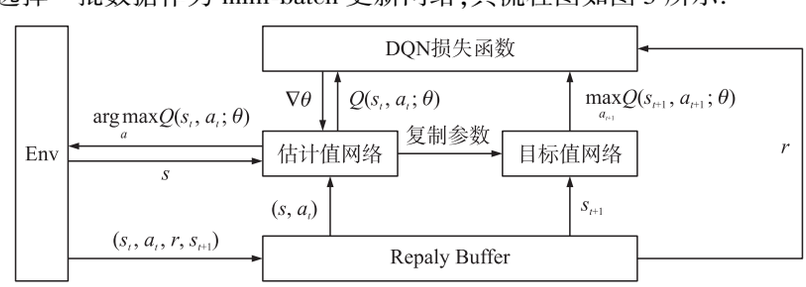


**图 3　DQN 流程图**

**Fig. 3　Flow chart of DQN**

在图 3 中,Repaly Buffer 是从历史数据中采样组成的样本集,每次从样本集中选取部分样本放入估计值网络中. 估计值网络是一个待训练的网络,其目的是拟合不同环境与行动下的收益,目标值网络是为了计算不同环境下的最优行动与对应收益.

### 2.4　加入 SVM 和 XGBoost

SVM 模型经常应用于股票市场的预测,它的收敛速度快,精度高,能够很好地预测股票涨跌趋势[15]. XGBoost 具有计算复杂度低、运行速度快、准确率高等特点,使用 XGBoost 分析时间序列数据,能够快速且精确地学习到股票趋势[16].

本文将股票交易数据作为输入变量,股票价格的涨跌信号作为标签,使用 SVM 和 XGBoost 对股票涨跌进行预测. 其中标签分为 1 和-1,若今日收盘价减去昨日收盘价大于 0 则为上涨,标为 1;若差值小于 0 则为下降,标为-1. 本次学习使用滑动窗口法,即预测某一天的涨跌信号,使用前 30 d 的股票基础数据作为训练集进行训练. SVM 采用 sklearn 中的线性二分类支持向量机 LinearSVC,其使用 $L^2$ 惩罚,损失函数定义为 squared_hinge,错误项的惩罚系数 $C$ 为 1. XGBoost 采用 XGBClassifier,其使用 gbtree 树模型作为基分类器,sample_weight 设置为 weight=np.exp(0.08 * np.arange(1,1+len(y_train))).

## 3　模拟试验

本次试验仅考虑单一资产的交易任务,在每个交易日只能选择进行一个行动,行动有 3 种可供选择:买入(1),卖出(-1),中立(0),且模型允许做空. 试验基于 Python 的 pytorch 库进行. 本文选择中国股票市场上市的招商银行、泰和科技两只股票. 招商银行使用的数据时间范围为 2008 年 1 月 2 日到 2024 年 1 月 2 日,泰和科技使用数据时间范围为 2019 年 11 月 28 日到 2024 年 5 月 31 日. 数据通过 Tushare 读取,包括股票收盘价、开盘价、最高价、最低价、成交量. 本文涉及的股票技术指标均通过 TA-Lib 来计算.

此外,本文使用买入持有(Buy and Hold,BH)策略来与基于 DQN 的算法交易系统进行比较,BH 策略是一种被动的投资策略,即在合适的时间点买入资产,无论市场如何变动都不会卖出,该策略在多数文献中用作评估其他投资策略盈利能力的基准,可以作为本次试验的比较基准.

### 3.1　训练与测试区间划分

本文的试验将数据分为训练集和测试集,表 2 为训练集与测试集的划分,图 4 和图 5 展示了两只股票在对应测试集上的走势图.

**表 2　交易股票及时间段**

**Table 2　Trading stocks and time periods**

| 股票代码 | 股票名称 | 训练集 | 测试集 |
|---|---|---|---|
| 600036.SH | 招商银行 | 2008-01-02 至 2018-01-03 | 2018-01-03 至 2024-01-02 |
| 300801.SZ | 泰和科技 | 2019-11-28 至 2022-01-04 | 2022-01-04 至 2024-01-02 |

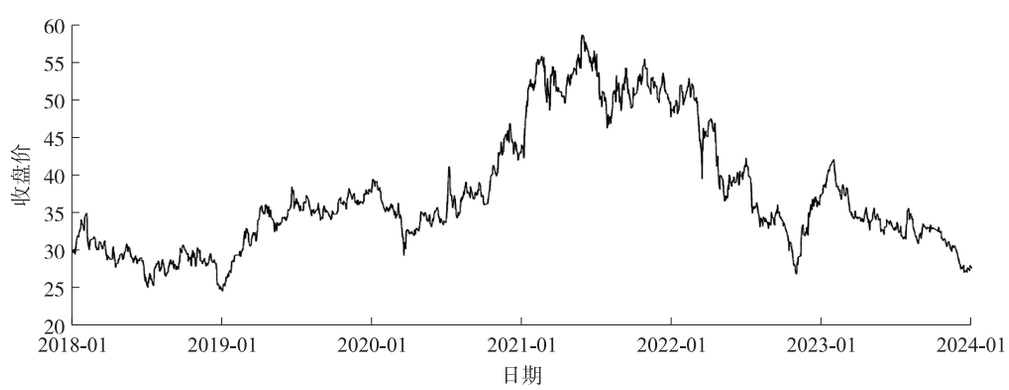


**图 4　招商银行 2018—2024 走势图**

**Fig. 4　China Merchants Bank 2018-2024 trend chart**

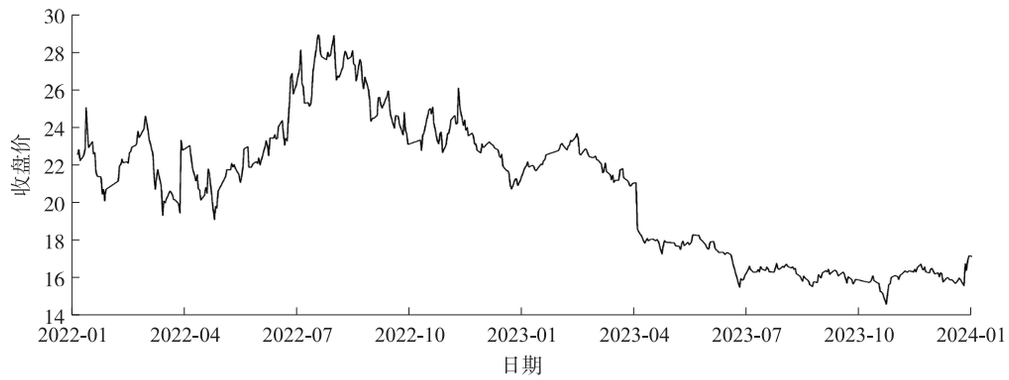


**图 5　泰和科技 2022—2024 走势图**

**Fig. 5　Taihe technology 2022-2024 trend chart**

### 3.2　试验参数设定

本次试验部署于服务器上,硬件配置为显存 24 GB 的 NVIDIA RTX 4090、64 核 vCPU 及 100GB 内存. 编程语言为 Python3.8.所使用的深度学习框架为 Pytorch2.0,相关环境为 CUDA11.8.

试验设定初始金额 1,折扣因子 $\gamma=0.98$,交易成本为交易金额的 0.000 1. 神经网络的输入特征分成 3 类,第一类仅包含基本的股票信息;第二类额外添加股票技术分析指标;第三类额外添加 SVM 与 XGBoost 的预测涨跌信号. 输出为 3 种对应行为产生的 $Q$ 值. 具体如表 3 所示.

**表 3　对照试验输入特征**

**Table 3　Control experiment input characteristics**

| 环境状态特征 | 区别 |
|---|---|
| 算法交易系统(简称系统) | 仅包含过去 30 日收盘价格数据 |
| 算法交易系统 1.0(简称系统 1.0) | 额外加入技术指标(MA、EMA、CCI、RSI、ATR、ADX 等) |
| 算法交易系统 2.0(简称系统 2.0) | 额外加入 SVM、XGBoost 涨跌信号 |

其余参数设置如表 4 所示.

**表 4　网络参数及结构设置**

**Table 4　Network parameters and structure settings**

| 训练轮次 | Batchsize | 激活函数 | 优化器 | $\varepsilon$-greedy | 网络结构 |
|---|---|---|---|---|---|
| 200 | 64 | ReLU | Adam(lr=1E-5) | $0.01+(0.9-0.01)\exp(-x/500)$ | [nn.Linear(state_dim,128),nn.ReLU()<br>nn.Linear(128,128),nn.ReLU()<br>nn.Linear(128,3)] |

### 3.3　试验结果

图 6 和图 7 分别是招商银行和泰和银行的累计利润图,初始资金记为 1,曲线代表累计的利润值. 从招商银行的累计利润上看,系统 2.0 大于系统 1.0,且均大于系统. 而在泰和科技上,系统 2.0 和系统 1.0 累计利润相近.

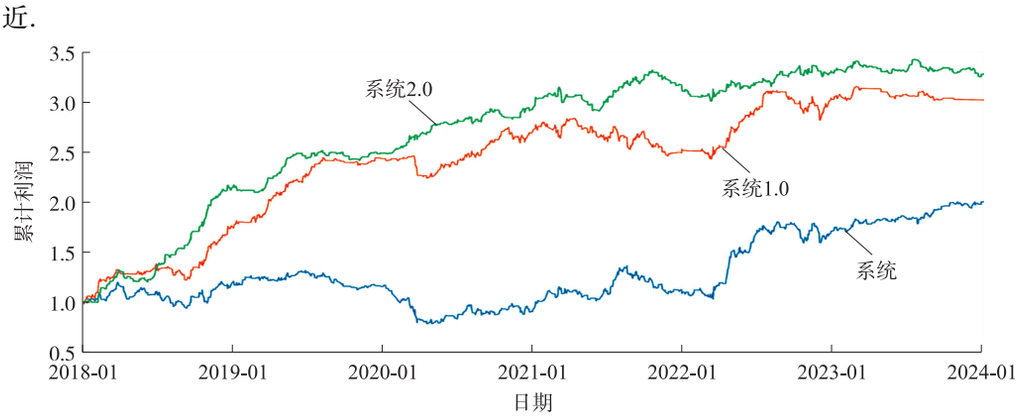


**图 6　招商银行利润累计图**

**Fig. 6　Cumulative profit chart of China Merchants Bank**

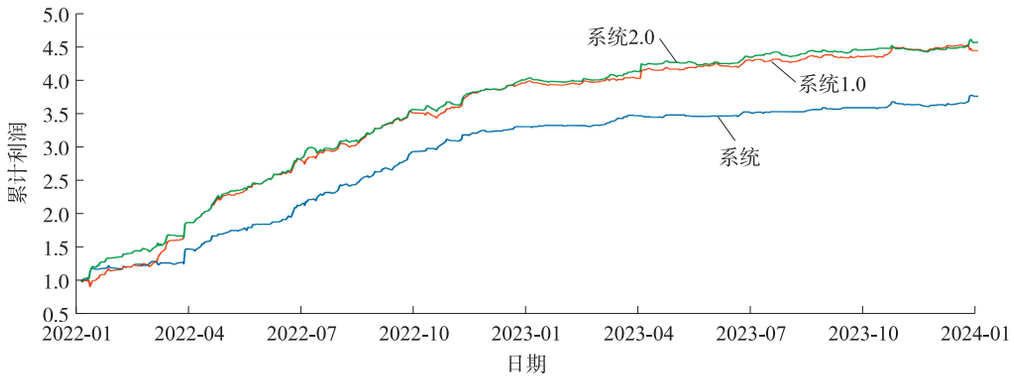


**图 7　泰和科技利润累计图**

**Fig. 7　Cumulative profit chart of Taihe technology**

表 5 更详细的展示了股票交易一些衡量指标的比较结果,包括最终金额、收益率、夏普比率、最大回撤率.

**表 5　试验衡量指标**

**Table 5　Experimental measurement indicators**

| 股票 | 模型 | 最终金额 | 收益率/% | 夏普比率 | 最大回撤率 |
|---|---|---|---|---|---|
| 招商银行 | BH 策略 | 1.15 | 15 | 0.03 | 0.397 |
| | 系统 | 2.01 | 101 | 0.60 | 0.400 |
| | 系统 1.0 | 3.02 | 201 | 0.91 | 0.140 |
| | 系统 2.0 | 3.28 | 228 | 0.99 | 0.090 |
| 泰和科技 | BH 策略 | 0.92 | −8 | −0.19 | 0.460 |
| | 系统 | 3.76 | 275 | 1.41 | 0.042 |
| | 系统 1.0 | 4.44 | 344 | 1.34 | 0.096 |
| | 系统 2.0 | 4.56 | 356 | 1.41 | 0.032 |

从表 5 可以看出,系统 2.0 和系统 1.0 均优于 BH 策略与系统,仅在泰和科技上,系统 1.0 的夏普比率略高于系统,承受的风险有所提高. 系统 2.0 交易策略能够在不同的市场上采取不同的行动. 例如,在下跌的市场中,智能体预测到卖出或做空,这种对市场的敏感度很大程度上归因于深度神经网络强大的发掘市场状态的能力. 此外,考虑到了股票市场的多样性和复杂性,本文额外选择了几支股票以提高结果的可靠性与普遍性,其结果如表 6 所示.

**表 6　测试样本指标**

**Table 6　Test sample indicators**

| 股票 | 模型 | 最终金额 | 收益率/% | 夏普比率 | 最大回撤率 |
|---|---|---|---|---|---|
| 000858.SZ | BH 策略 | 1.21 | 21 | 0.03 | 0.11 |
| | 系统 | 2.33 | 133 | 0.60 | 0.082 |
| | 系统 1.0 | 3.24 | 224 | 0.68 | 0.074 |
| | 系统 2.0 | 4.33 | 333 | 0.77 | 0.056 |
| 002594.SZ | BH 策略 | 0.88 | −12 | −0.27 | 0.11 |
| | 系统 | 1.75 | 75 | 0.72 | 0.055 |
| | 系统 1.0 | 3.48 | 248 | 0.95 | 0.082 |
| | 系统 2.0 | 3.62 | 262 | 0.64 | 0.076 |
| 600519.SH | BH 策略 | 1.01 | 1 | 0.05 | 0.11 |
| | 系统 | 3.88 | 288 | 1.09 | 0.034 |
| | 系统 1.0 | 3.9 | 290 | 1.21 | 0.031 |
| | 系统 2.0 | 4.08 | 308 | 1.36 | 0.016 |
| 601988.SH | BH 策略 | 1.32 | 32 | 0.26 | 0.15 |
| | 系统 | 1.46 | 46 | 0.21 | 0.11 |
| | 系统 1.0 | 2.01 | 101 | 0.66 | 0.088 |
| | 系统 2.0 | 3.22 | 222 | 0.37 | 0.097 |

## 4 结论

本文基于 DQN 构建算法交易系统应用于中国股市,分别在招商银行和泰和科技两支股票以及其余 4 支股票上进行模拟交易,并以买入持有策略为基准进行对比,从收益率、夏普比率、最大回撤率三方面进行对比验证模型的有效性. 经过对照试验综合对比分析后,本文得出以下结论:

(1)算法交易系统在解决股票日内交易问题时是有效的,其能够利用神经网络挖掘到股票市场中有效的特征信息,同时利用强化学习使得智能体能够进行自主决策,可以进行在线学习实时更新,追踪最新市场趋势,并获得可观的收益,因此该模型可以应用到股票量化交易中.

(2)算法交易系统不仅在收益率上优于买入持有策略,并且最大回撤率较低,模型的抗击风险能力较好,说明本文对智能体交易环境的构建是有效的,与监督学习的结合也显著地提升了模型的稳定性.

(3)该方法目前只针对单支股票的过去信息进行学习,无法充分考虑整个金融市场的宏观信息、行业数据信息,以及互联网上相关舆论风向等信息,可进一步考虑丰富环境特征,转化、统合、筛选并提取信息,以提高模型的决策能力.

[参考文献]

[1] 姜宇薇. 金融市场量化交易的国际经验[J]. 中国货币市场,2022,246(4):22-25.
[2] 张晓燕,张远远. 量化投资在中国的发展及影响分析[J]. 清华金融评论,2022(1):44-45.
[3] 梁天新,杨小平,王良,等. 基于强化学习的金融交易系统研究与发展[J]. 软件学报,2019,30(3):20.
[4] JIANG Z,XU D,LIANG J. A deep reinforcement learning framework for the financial portfolio management problem[J/OL]. arXiv Preprint arXiv:1706.10059,2017.
[5] XIONG Z,LIU X Y,SHAN Z,et al. Practical deep reinforcement learning approach for stock trading[J/OL]. arXiv Preprint arXiv:1811.07522,2018.
[6] 韩道岐,张钧垚,周玉航,等. 基于深度强化学习的股市操盘手模型研究[J]. 计算机工程与应用,2020,56(21):145-153.
[7] AZHIKODAN A R,BHAT A,JADHAV M V. Stock trading bot using deep reinforcement learning[M]. Singapore:Innovations in Computer Science and Engineering,2019:41-49.
[8] HUANG Z,LIN N,MEI W L,et al. Algorithmic trading using combinational rule vector and deep reinforcement learning[J]. Applied soft computing,2023,147:110802.
[9] MAHDI M,MAHOOTCHI M. A deep Q-learning based algorithmic trading system for commodity futures markets[J]. Expert systems with applications,2024,237:121711.
[10] TRELEAVEN P,GALAS M,LALCHAND V. Algorithmic trading review[J]. Communications of the ACM,2013,56(11):76-85.
[11] MNIH V,KAVUKCUOGLU K,SILVER D,et al. Playing atari with deep reinforcement learning[EB/OL](2013-12-19)[2023-01-08].
[12] BANOTHS P R,DONTA P K,AMGOTH T. Dynamic mobile charger scheduling with partial charging strategy for WSNs using deep-Q-networks[J]. Neural computing and applications,2021,33(22):15267-15279.
[13] 石泽宇. 我国股票市价变动影响因素的实证分析[J]. 北方经贸,2020(9):125-127.
[14] WANG Y,WANG D,ZHANG S,et al. Deep Q-trading[R]. CSLT Technical Report-20160036,2017.
[15] 李嘉浩. 基于支持向量机的股票预测与分析[J]. 经济研究导刊,2021(32):107-110.
[16] 葛橹漠,周显. 基于 XGBoost 的多因子选股模型[J]. 信息技术与标准化,2020(5):36-41.

[责任编辑:陆炳新]