DOI:10.16652/j.issn.1004-373x.2023.02.024
引用格式:文馨贤.基于深度强化学习的高频量化交易策略研究[J].现代电子技术,2023,46(2):125-131.

基于深度强化学习的高频量化交易策略研究

文馨贤
(昆明理工大学 信息工程与自动化学院,云南 昆明 650500)

摘 要: 当前国内金融市场的投资交易已从基于传统技术分析等方法的主观交易逐渐转向基于程序化的量化策略交易。股票市场已有大量量化策略的研究工作,但针对期货市场的量化交易策略的研究还不足,已有策略在日内高频交易中的投资回报和风险控制还有待优化。为提升期货高频量化策略的盈利和风控能力,文中设计一种期货交易环境,将1 min时间粒度的高频K线作为环境状态,针对期货交易中持仓状态和交易操作构建相应的动作空间及算法;采用基于LSTM的深度强化学习模型LSTM-Dueling DQN,使其更适用于处理序列输入的状态空间,并显著提升模型的学习速度。对DQN、Double DQN、基于全连接神经网络的Dueling DQN(FF-Dueling DQN)三个基准模型进行实验对比,得到文中构建的交易策略在四个黑色系商品期货交易中累计收益率最高达到43%,年化收益率达到153%,最大回撤控制在10.7%以内。实验结果表明,所提策略在震荡行情和趋势行情中都能实现超出业绩基准的超额收益。

关键词: 交易策略;深度强化学习;LSTM;Deep Q-Network;高频交易;期货;量化金融
中图分类号: TN911-34;TP3 **文献标识码:** A **文章编号:** 1004-373X(2023)02-0125-07

Research on high-frequency quantitative trading strategy based on deep reinforcement learning

WEN Xinxian
(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Investment trading in the domestic financial market has gradually shifted from subjective trading based on traditional technical analysis to quantitative trading based on programmability. There has been a considerable amount of research work on quantitative strategies in the current stock market, but there is not enough research on quantitative trading strategies for the futures market, and the investment returns and risk control of existing strategies in intraday high-frequency trading need to be further optimized. In order to improve the profitability and risk control of high-frequency quantitative strategies for futures, a futures trading environment is designed, which takes the 1-min time granularity high-frequency K-bar as the environment state, and corresponding action spaces and algorithms are constructed for position states and trading operations in futures trading. The deep reinforcement learning model LSTM-Dueling DQN based on LSTM is used to make it more suitable for processing the state space of sequence input, and significantly improve the learning speed of the model. Three benchmark models of DQN(Deep Q-Network), Double DQN, and Dueling DQN based on fully connected neural network (FF-Dueling DQN) are compared in this experiments. It is show that the trading strategy constructed in this paper has a maximum cumulative yield of 43%, an annualized yield of 153% and a maximum pullback of 10.7% in the four black commodity futures transactions. The experimental results show that the proposed strategy can achieve excess returns beyond the performance benchmark in both volatile and trend markets.

Keywords: trading strategy; deep reinforcement learning; LSTM; DQN; high-frequency trading; futures; quantitative finance

0 引 言

当前国内期货量化交易刚进入起步阶段,市场中的量化策略体量远不及国外,主流市场中仍有大量主观交易的机构和投资者。随着金融科技的发展,交易系统性能的提升,日内的分钟级别等高频交易获得了更多机构的青睐。市场行情瞬息万变,依靠人工完成高频交易是不现实的,通过量化策略自动完成择时、建仓、平仓等操作才是当前期货市场交易策略开发的主要方向。

随着深度学习在时间序列方面的研究发展,有大量

收稿日期:2022-06-18 **修回日期:**2022-07-27

研究人员[1-3]采用神经网络对标的价格进行预测，在此基础上获得交易信号，进而构建交易策略的方法。该类方法首先没有考虑到市场与投资者行为是相互交互的一个整体；其次，模型的训练数据标签的设置受人为因素影响较大；最后，深度学习中监督学习是即时反馈的，但实际交易策略的评价是延迟反馈的，深度学习没有考虑到策略长期投资回报的问题。

强化学习是通过智能体与环境不断交互学习，来优化策略从而实现目标。有研究通过基于值的强化学习来构建交易策略。M. Corazza等在意大利股票市场上针对每日股票数据，比较了SARSA和Q-Learning的效果，得出SARSA在短期收益上比Q-Learning稍好一些[4]。李静基于SARSA的强化学习算法，在6组期货品种上构建了配对交易策略进行每日交易，年化复合收益率为5.25%，相比传统套利策略在风险控制上有一定效果[5]。

为了解决SARSA和Q-Learning中Q表格的局限性和Q值估计问题，研究人员将深度学习引入强化学习中，用神经网络替代Q表格，提出了深度强化学习算法。文献[6]用每日回报作为状态空间，将累计利润作为奖励，使用Deep Q-Network（DQN）[7-8]构建了用于标普500股指期货连续合约的长期交易策略，年化收益率接近10%。Liu等人提出一个基于强化学习的股票量化交易框架——FINRL[9]，该框架构建了基础的股票交易环境，并集成了基础的强化学习方法，如Q-Learning、DQN、DDPG[10]等。FINRL框架主要针对股票交易，目前暂未支持期货交易。

当前期货量化交易策略研究存在以下挑战：期货日内交易活跃，分钟级别等高频量化交易策略有着很大的发展空间；已有研究的策略在收益和风险控制能力还有较大的优化空间。

因此，本文选取了我国期货交易所中流动性较好的黑色系板块品种的1 min交易数据构建数据集，构建了以K线作为输入的二维状态空间，设计了适用于期货交易的动作空间。考虑到状态空间为序列信息，LSTM擅长于处理序列任务，同时为了适用于高频的短时交易，采用Dueling DQN提升模型的学习效率，为Agent构建了LSTM-Dueling DQN模型。策略在实现自动交易的同时，在不同行情测试集上均获得了超额收益和较小的回撤。

1 模型构建

1.1 交易设置

期货交易中的交易成本主要为交易时产生的手续费，根据实际情况，本文设置仅在开仓交易时收取手续费，实验中设置手续费率为0.000 1。按照期货交易的保证金制度，本文实验中设置保证金比例为20%，交易初始资金设置为500 000元。本文假设Agent的交易行为对市场行情和其他投资者行为无影响，交易系统中不考虑滑点问题，委托下单的成交价默认为当前K线收盘价。

1.2 模型构建

模型由状态空间、动作空间、奖励函数、Agent构成，如图1所示，其目标是找到决策策略π来获得最大收益。状态和奖励作为环境的输出，输入到Agent中，Agent中基于LSTM的网络根据输入产生估计的Q值，通过Policy中Epsilon-Greedy探索来获得最终输出的交易动作，买卖或持有的交易动作为环境的输入，更新环境中的状态和奖励，环境和Agent不断相互交互直到回合结束或达到结束条件。

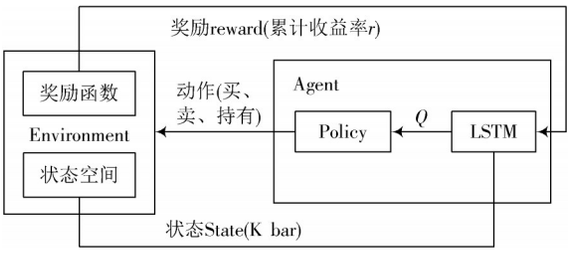

图1 模型架构

1.2.1 状态空间

本文中将固定滑动窗口长度的K线作为环境中的状态空间。每根K线包含7个特征：HC（最高价与收盘价之差）、HO（最高价与开盘价之差）、HL（最高价与最低价之差）、CL（收盘价与最低价之差）、OL（开盘价与最低价之差）、CO（收盘价与最低价之差）、MOV（收盘价与前一K线的收盘价之差）。将环境的状态空间定义为一个多维连续空间，滑动窗口固定长度设置为30，取1 min时间粒度的K线，即将过去30 min的交易K线特征作为环境中的状态空间。每个状态的维度为[30，7]。

1.2.2 动作空间

期货市场中交易头寸即开仓买入或卖出期货合约的过程。与股票市场只能进行单边多头交易不同，期货市场可以进行多空两个方向的交易。期货交易中的Agent持仓状态有三种：多头、空头、无头寸。多头是指通过买入某项标的资产获得持仓，持仓一段时间后卖出该标的平仓。空头是指通过卖出某项标的资产获得持仓，持仓一段时间后买入该标的平仓。将Agent的持仓状态设为：

$$\text{postion} \in [0,1,2] \tag{1}$$

式中：0为多头（LONG）；1为空头（SHORT）；2为无头寸

(FLAT)。Agent作为交易者,可以进行买入、卖出、持有三种动作,构建一维离散动作空间为:

$$action\_sapce \in [0,1,2] \tag{2}$$

式中:0为买入(BUY);1为卖出(SELL);2为持有(HOLD)。本文实验设置在t时间步有多头持仓,在$t+1$时间步时,如果再次进行买入,将不改变持仓状态,此处的动作序列为持有(HOLD),空头头寸同理。图2为模型动作状态转移过程。

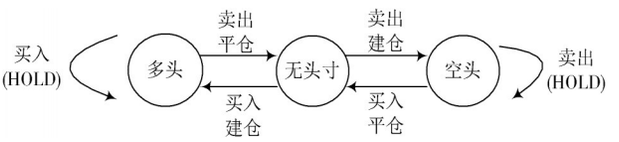

图2 动作状态转移过程

本文中构建的交易模型是针对固定标的物的买入卖出,每次建仓理论上使用全部资金。本文实验设置对单一标的物不能同时进行多头和空头的建仓,在每一时间步下,只存在单个方向的交易持仓。Agent在每个时间步进行一个动作,其算法流程如算法1所示。

算法1:Agent的step算法

Input:当前时间步t的action;
Output:状态、累计收益率、账户余额、当前Tick及收盘价;
Variables:
action:action ∈ [BUY,SELL,HOLD] = [买入,卖出,持有]
position:
postion ∈ [LONG,SHORT,FLAT] = [多头,空头,无持仓]
1. action = HOLD;
2. if action = BUY;
3. if position = FLAT;
4. position = LONG;
5.获取开仓价格为当前收盘价,记录action;
6. if position = SHORT;
7. position = FLAT;
8.获取平仓价格为当前收盘价,记录action;
9.计算平仓收益率,更新账户余额,空头数量+1,清空开仓价格;
10. if action = SELL;
11. if position = FLAT;
12. position = SHORT;
13. 获取开仓价格为当前收盘价,记录action;
14. if position = LONG;
15. position = FLAT;
16. 获取平仓价格为当前收盘价,记录action;
17. 计算平仓收益率,更新账户余额,空头数量+1,清空开仓价格

1.2.3 Agent构建

本文研究的期货交易问题,其动作空间是离散的,因此,采用基于值的强化学习算法。考虑到状态空间中是多维的序列数据,而LSTM擅长处理深度学习中的序列问题,同时,改进DQN的模型Dueling DQN能够提升模型学习效率,故实验采用基于LSTM的Dueling DQN来构建Agent。

Dueling DQN中预测以下两部分的值:价值函数与优势函数的输出,模型的Q函数如下:

$$Q(s,a,\omega,\alpha,\beta) = V(s,\omega,\beta) + A(s,a,\omega,\alpha) \tag{3}$$

式中:$V(s,\omega,\beta)$是价值函数,仅与状态s有关;$A(s,a,\omega,\alpha)$是优势函数,与状态s和动作a都有关;ω是整个网络的参数;β是价值函数部分的参数;α是优势函数的参数。对优势函数进行归一化后,最终模型的Q函数如下:

$$Q(s,a,\omega,\alpha,\beta) = V(s,\omega,\beta) + \left(A(s,a,\omega,\alpha) - \frac{1}{A}\sum_{a'\in A} A(s,a',\omega,\alpha)\right) \tag{4}$$

将价值函数和优势函数两部分合并后即为Q网络的估计。Q网络结构如图3所示。

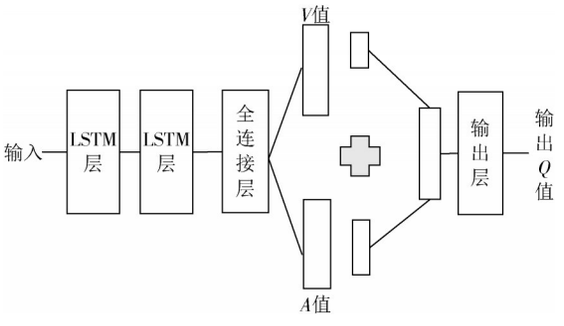

图3 Q网络结构

Q网络的输入是标准化后的环境状态和Agent的动作。原始Dueling DQN[11]中的Q网络采用的是卷积神经网络架构,由于本文中状态输入为期货的相关价格序列,考虑到LSTM在时间序列问题上的特征提取能力,本文中的Q网络采用两个LSTM层和一个全连接层来处理和学习数据,输出分上路和下路,分别为价值函数和优势函数的计算,上下两路聚合后经过全连接层输出获得最后的Q值。

2 实验分析

2.1 实验配置

2.1.1 实验环境

实验采用的深度学习框架为TensorFlow和Keras,TensorFlow版本为1.14.0 ,Keras版本为2.1.2。采用基于Keras_RL框架[12]构建深度强化学习模型,以Open AI的Gym模块[13]为基础来构建交易环境。

2.1.2 数据集

黑色系板块的期货品种在整个商品期货市场交易中有着重要地位，实验选择其中4个品种的主力合约交易，分别为焦炭、焦煤、铁矿石和螺纹钢，对应的交易代码为J、JM、I、RB。数据来源为万得金融数据库，数据的内容包括开盘价、最高价、最低价、收盘价，时间粒度为1 min。数据集具体内容如表1所示，4个品种的收盘价序列如图4所示。实验将4个品种的K线数据进行标准化、滑动窗口机制预处理后，作为模型的状态空间输入。

表1 数据集描述

数据集	主力合约	训练数据集	测试数据集
焦炭J	5月、9月合约	2019-08-06—2019-12-10	2019-12-10—2020-03-30
焦煤JM	5月、9月合约	2019-08-16—2019-12-18	2019-12-18—2020-04-02
铁矿石I	5月、9月合约	2019-07-30—2019-12-05	2019-12-05—2020-03-30
螺纹钢RB	5月、10月合约	2019-12-04—2020-03-19	2020-03-20—2020-06-10

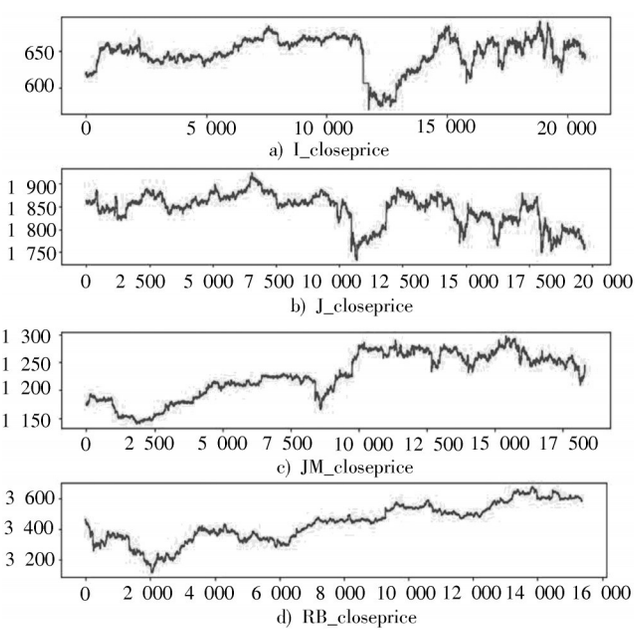

图4 收盘价序列

2.1.3 评价指标

1）累计收益率：是指在某个周期内完成多头、空头平仓交易的收益率之和，用于衡量策略的整体收益情况。单次平仓收益率为r_i，则累计收益率r_{total}为：

$$r_{\text{total}} = \sum_{i=1}^{n} r_i \tag{5}$$

2）年化收益率：为评估策略的长期收益能力，使用年化的收益率作为评价指标之一。不考虑复利影响，n为累计收益率天数，年化收益率r为：

$$r = r_{\text{total}} \cdot \frac{250}{n} \tag{6}$$

3）最大回撤：是衡量策略风险的重要指标，反映了策略可能产生的最大损失。当策略无回撤时，最大回撤值为0，最大回撤的绝对值越大说明策略的风险控制能力越差。设某个周期内的收益率序列为$T=\{r_1,r_2,\cdots,r_n\}$，最大回撤为：

$$\text{MaximumDrawDown} = \max_i \left(r_i - \min_{1\leqslant i\leqslant j\leqslant n} r_j \right) \tag{7}$$

2.1.4 比较基准

本文中选择了三个基于值的深度强化学习模型作为比较基准，同时为了更好地评估模型策略的收益能力和风险情况，选择了市场上管理期货策略的主流业绩基准——南华商品指数收益率作为本实验中策略的业绩基准。

1）DQN：基于值的深度强化学习算法，在Q-Learning方法基础上引入神经网络来对Q函数进行近似，获得Q值。

2）Double DQN：为解决Q值存在过高估计的问题，文献[14]中提出改进的DQN方法，利用原始DQN中的两个神经网络，通过交换的方式结合两个网络来获得最终Q值。

3）FF-Dueling DQN：基于全连接神经网络的Dueling DQN。在原始Dueling DQN框架中，使用一个前馈神经网络和一个全连接层构建神经网络。

南华商品指数是国内主流的商品指数之一，是通过某种价格编制规则形成的一种指数，主要反映了国内期货市场整体行情。

2.2 参数设置

实验中为实现针对期货单一标的的自动交易，采用经过固定滑动窗口为30个时间步，处理1 min时间粒度的K线数据作为环境状态输入。Agent使用LSTM-Dueling DQN，其中神经网络包含两个LSTM层和一个全连接层，Agent中采用Epsilon-Greedy策略选择动作。具体参数设置如表2所示。

2.3 实验结果对比与分析

2.3.1 模型性能对比

本文构建的期货交易策略模型和三个基准模型分别在4个数据集上进行对比实验。实验初始资金均设置为50万元，输入的K线固定滑动窗口步长为30步，每一步为1 min。在每个时间步，针对单一标的进行一次买卖交易或不操作，单次交易使用当前账户的全部可用资金。在对应的测试周期内计算同周期下南华商品指数的累计收益率、年化收益率、最大回撤、账户收益作为市场业绩基准参考，具体如表3所示。

表2　参数设置

超参数	取值
滑动时间窗口长度T	30
开始学习时间步长	200
目标网络的更新步数	10 000
手续费率	0.000 1
保证金比率	0.2
初始资金	500 000
循环步数	5 000
一次学习周期中最大步数	10 000
折扣衰减因子Gamma	0.99
LSTM的网络层数	2
每层LSTM的神经元个数	64
全连接层神经元个数	32
学习率	0.001
损失函数	MSE
Batch_size	32
优化器	Adam

从表3的4个不同品种测试集的回测数据可以看出,基于DQN 和Double DQN的策略在所有测试集上的收益率表现接近,区间累计收益率相差最小0.7个点。主要原因是Double DQN没有从根本架构上改进DQN,所以两个模型的效果接近。本文的LSTM-Dueling DQN模型策略,累计收益率、年化收益率、账户收益均超过其他基准模型,区间累计收益率最大超过43%,年化收益率超过153%,在焦煤和螺纹钢测试集中的最大回撤均为最小,不超过4%,在铁矿石和焦炭测试集中最大回撤不超过11%。

从市场整体情况看,在铁矿石、焦炭、焦煤测试数据周期内,同周期南华商品指数呈趋势性下跌;螺纹钢测试数据周期内,同周期南华商品指数短期震荡后呈趋势性地上涨,市场整体回暖。在市场整体下跌的周期中,四个模型策略最终都获得正收益,均超过了业绩基准。而在市场整体上涨的周期中,LSTM-Dueling DQN模型策略累计收益率为18.050 6%,超过同周期南华商品指数收益率11.504 0%,取得了接近7个点的超额收益。而DQN、Double DQN和FF-Dueling DQN最终累计收益率均低于业绩基准,说明这三个模型策略没有跑赢市场。

表3　模型性能对比

数据集	评价指标	LSTM-Dueling DQN	FF-Dueling DQN	Double DQN	DQN	南华商品指数
铁矿石I	累计收益率/%	27.108 8	9.574 1	13.220 1	12.200 3	−14.783 5
	年化收益率/%	89.173 8	31.493 8	43.487 1	40.132 5	−48.629 9
	最大回撤/%	−10.677 6	−7.225 7	−8.465 66	−14.681 9	−18.600 8
	账户收益	646 137	549 261.01	555 061.60	553 471.59	426 082.50
焦炭J	累计收益率/%	13.667 7	5.626 85	5.909 8	8.696 3	−16.231 9
	年化收益率/%	46.807 2	19.270 0	20.239 0	29.782 0	−55.588 7
	最大回撤/%	−7.282 8	−7.573 3	−6.935 7	−6.590 3	−18.600 8
	账户收益	565 399.61	526 378.04	528 609.08	545 039.38	418 840.50
焦煤JM	累计收益率/%	43.000 2	7.808 4	23.160 4	23.941 5	−14.338 0
	年化收益率/%	153.572 3	27.887 4	82.715 7	85.505 4	−51.207 3
	最大回撤/%	−2.927 3	−4.341 8	−6.120 4	−3.647 2	−16.691 5
	账户收益	765 707.53	533 721.95	615 802.00	619 707.50	428 310.00
螺纹钢RB	累计收益率/%	18.050 6	3.844 0	3.853 2	3.181 8	11.504 0
	年化收益率/%	82.048 2	17.473 1	17.514 9	14.463 0	52.291 0
	最大回撤/%	−3.755 8	−10.160 3	−6.468 4	−9.082 6	−0.875 5
	账户收益	596 806.84	517 751.50	518 180.10	514 412.71	557 520.00

2.3.2　收益及回撤情况分析

图5~图8分别是四个模型在铁矿石、焦炭、焦煤、螺纹钢四个期货品种测试集上的累计收益率曲线,曲线上的黑点是最大回撤开始产生和回撤修复的对应节点。

铁矿石的前期行情波动小,各模型收益率曲线均稳定上涨,中期出现急跌行情,LSTM-Dueling DQN和DQN模型策略都在该行情下进行了正确操作,收益率短期涨幅接近5个点。焦炭为持续震荡行情,4个模型收益率都呈现震荡波动,最大回撤发生在各个不同的时期。焦煤中期出现了急跌行情,只有LSTM-Dueling DQN模型策略在该行情下进行了正确操作,收益率短期涨幅超过3个点,而最大回撤发生的周期短,说明策略的风险控制有效果。螺纹钢行情呈现一定趋势,只有LSTM-Dueling DQN模型策略抓住了中长期趋势,实现收益率

的持续稳定上涨。总体上，LSTM-Dueling DQN在波动率较大的震荡行情下能够识别极端行情并进行正确方向的交易操作，在整体上涨的趋势行情下能够抓住中长期趋势实现平稳的收益，并且能够实现一定程度的风险控制。

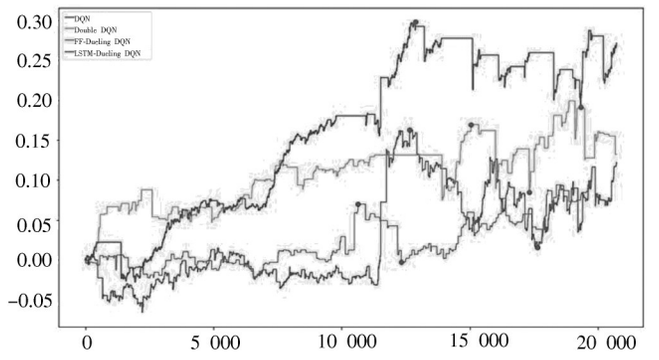

图5 铁矿石I收益率曲线

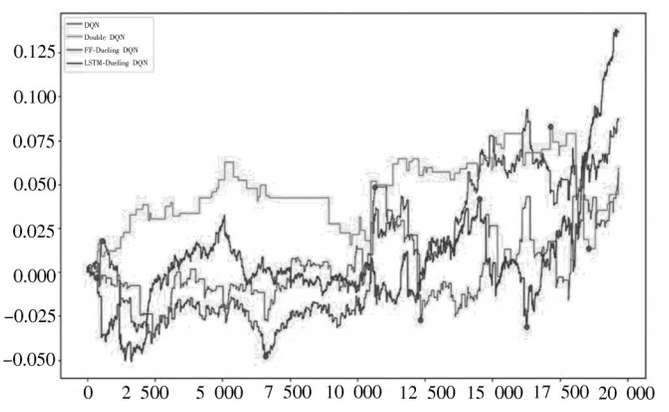

图6 焦炭J收益率曲线

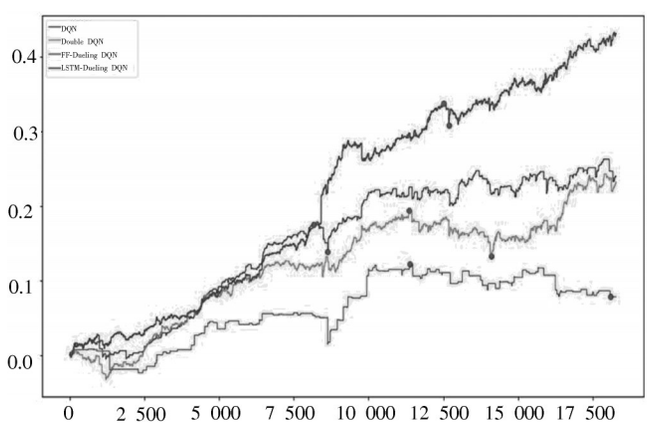

图7 焦煤JM收益率曲线

2.3.3 模型开平仓情况分析

对策略的开仓情况进一步分析，针对焦煤测试集上均有频繁操作的两个模型进行分析，图9和图10分别为Double DQN、LSTM-Dueling DQN在焦煤出现急跌行情周期的开平仓图。由图可以看出：Double DQN模型策略避开了极端行情，交易操作在行情平稳时十分频繁，这种过于频繁的交易会通过手续费率对投资收益产生影响；LSTM-Dueling DQN模型策略能够明确抓住价格波动进行正确的开平仓，尤其是充分利用极端行情进行交易操作。

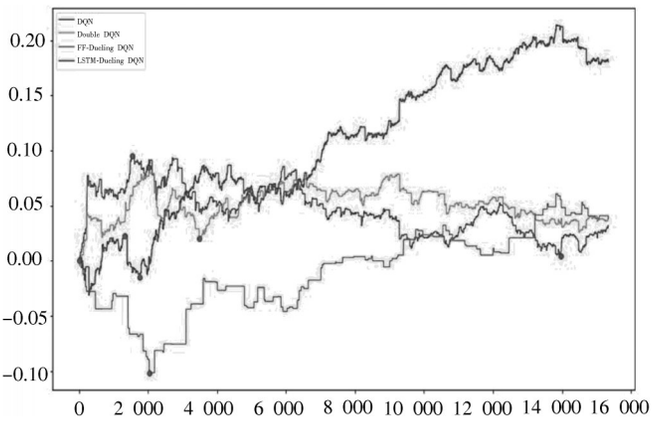

图8 螺纹钢RB收益率曲线

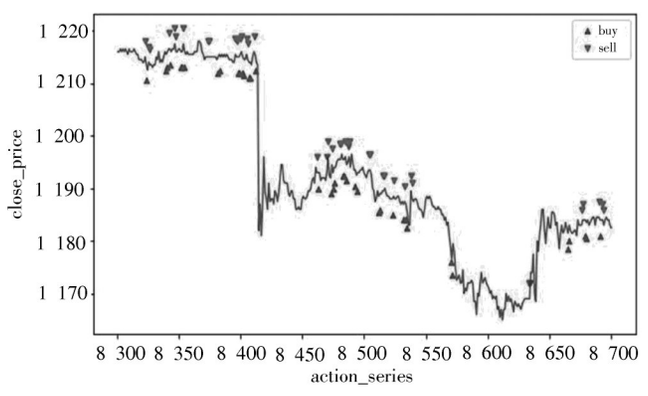

图9 Double DQN的开平仓情况

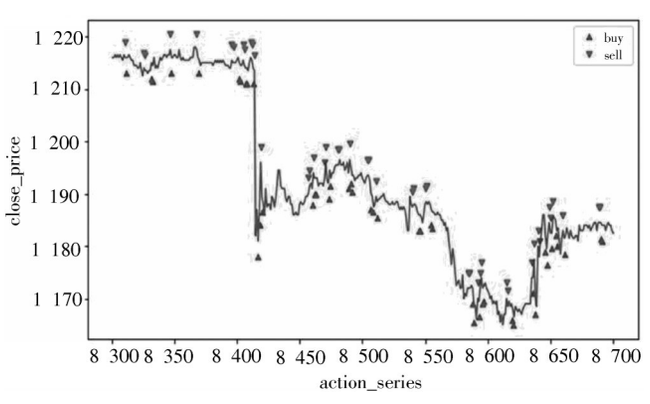

图10 LSTM_Dueling DQN的开平仓情况

3 结 语

本文将期货交易过程构建为深度强化学习模型，实现了在高频行情下的自动交易策略。在黑色系板块的4个期货品种上，模型策略取得了超出所有基准的最大超额收益，并实现了相对较小的策略回撤，实验验证了

其在不同行情下均能实现超额收益,并且在极端行情下能够做出正确的交易操作。

本文在实验交易环境中设置了一些理想条件,如交易不考虑滑点问题、成交价即为当前时间步K线的收盘价。在实际市场中,尤其是高频交易中,往往会因为交易系统、交易网络等问题出现滑点,使得策略不能以理想的价格成交。金融市场情况复杂,行情瞬息万变,本文提出的量化策略在用于实际交易前还需进一步优化。另外,当前的实验考虑的是对单一品种的交易,后续也可以根据需要扩展到两两品种间的配对交易和多品种间的投资组合交易策略中。

参 考 文 献

[1] 李亚峰,王洪波,李晨,等.融合注意力机制的LSTM期货投资策略[J].计算机系统应用,2021,30(8):22-30.

[2] LI Y, WANG S, WEI Y, et al. A new hybrid VMD-ICSS-BiGRU approach for gold futures price forecasting and algorithmic trading [J]. IEEE transactions on computational social systems, 2021, 8(6): 1357-1368.

[3] GU Q, LU N, LIU L. A novel recurrent neural network algorithm with long short-term memory model for futures trading [J]. Journal of intelligent & fuzzy systems, 2019, 37(4): 4477-4484.

[4] CORAZZA M, SANGALLI A. Q_Learning and SARSA: a comparison between two intelligent stochastic control approaches for financial trading [D]. Venice: Ca'Foscari University of Venice, 2015.

[5] 李静.基于强化学习算法的商品期货配对交易策略设计[D].上海:上海师范大学,2021.

[6] HIRSA A, HADJI MISHEVA B, OSTERRIEDER J, et al. Deep reinforcement learning on a multi-asset environment for trading [EB/OL]. [2021-01-21]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3867800.

[7] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning [EB/OL]. [2020-07-17]. https://www.xueshufan.com/publication/1757796397.

[8] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529-533.

[9] LIU X Y, YANG H Y, CHEN Q, et al. FinRL: a deep reinforcement learning library for automated stock trading in quantitative finance [EB/OL]. [2020-12-07]. https://www.xueshufan.com/publication/3162607076.

[10] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning [J]. Computer science, 2015(10): 67722-67724.

[11] WANG Z, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning [C]// International Conference on Machine Learning. Jeju Island: PMLR, 2016: 1995-2003.

[12] MATTHIAS P. Keras-RL [EB/OL]. [2016-11-21]. https://github.com/keras-rl/keras-rl, 2016.

[13] GREG B, VICKI C, LUDWIG P, et al. OpenAIGym [EB/OL]. [2020-02-15]. https://www.xueshufan.com/publication/3037207827.

[14] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q_Learning [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Washington: ACM, 2016: 14-20.

作者简介:文馨贤(1994—),女,硕士研究生,主要研究方向为数据挖掘、时间序列。