实验成果汇报

基于LSTM-DoubleDQN深度强化学习的量化交易策略

研究对象

- 在沪深300期权上做一种对冲交易模式[1]，当沪深300指数出现较大幅度的波动时可实现盈利，而出现亏损情况则是沪深300长时间波动较小。这样的交易模式是否盈利则与沪深300指数的涨跌无关。

注：[1]这种对冲交易模式是同时买入一定数量的看涨期权和看跌期权，期权是一种亏损有限，但盈利上限极高的金融衍生工具

数据介绍

- 实验收集了沪深300从2018年到2022年的15分钟k线数据，每条数据包含了最高价，最低价，开盘价，收盘价，成交均价，成交量，成交额，时间戳

code	high	low	open	close	avg	vol	amount	time
300	4148.634	4143.402	4148.293	4145.557	4145.012666666666	7738892.0	10007790473.0	2018-01-05 13:30:00
300	4145.594	4142.048	4145.557	4142.451	4143.393	7430118.0	9302780359.0	2018-01-05 13:45:00
300	4142.904	4133.627	4142.451	4135.955	4135.932	6472777.0	8702622069.0	2018-01-05 14:00:00
300	4143.103	4135.761	4135.955	4142.248	4139.490066666666	5251690.0	7015551585.0	2018-01-05 14:15:00

- 将2018-2021年四年数据作为训练集，2022年作为测试集

数据预处理

- 金融数据的各个属性量纲不同，波动程度也不一样，这会导致模型一开始会受到较强干扰。为了统一量纲，让模型关注数据的相对波动，做出了对数化处理。

- $\overline{p}_t = ln(\frac{p_t}{p_{t-1}})$ 当前价格与前一刻的价格做除法后取对数

- $\overline{vol_t} = ln\frac{vol_t}{\sum_{i=1}^{n} vol_{t-16i} / n}$ 成交量与过去n天的同一时刻的均值做比后取对数

特征工程

- 阻力位区域信号：当价格运动到阻力位区域附近时，发出阻力位信号1，否则为0。

阻力位的定义：阻力位由前期相对高点(或低点)组成，是一个相对区域，实验中设定高点附近正负0.3%都是阻力位区域

特征工程

- 近n日波动率：衡量近一段时间内行情波动程度，使用年化收益率方差衡量。

$$volatility = \sqrt{\frac{F}{N-1}\sum_{i=1}^{N} ln(\frac{C_i}{C_{i-1}})}$$

F是一年可以获取的样本量，N是实际计算式近n日选取的样本量，Ci是第i根k线的收盘价

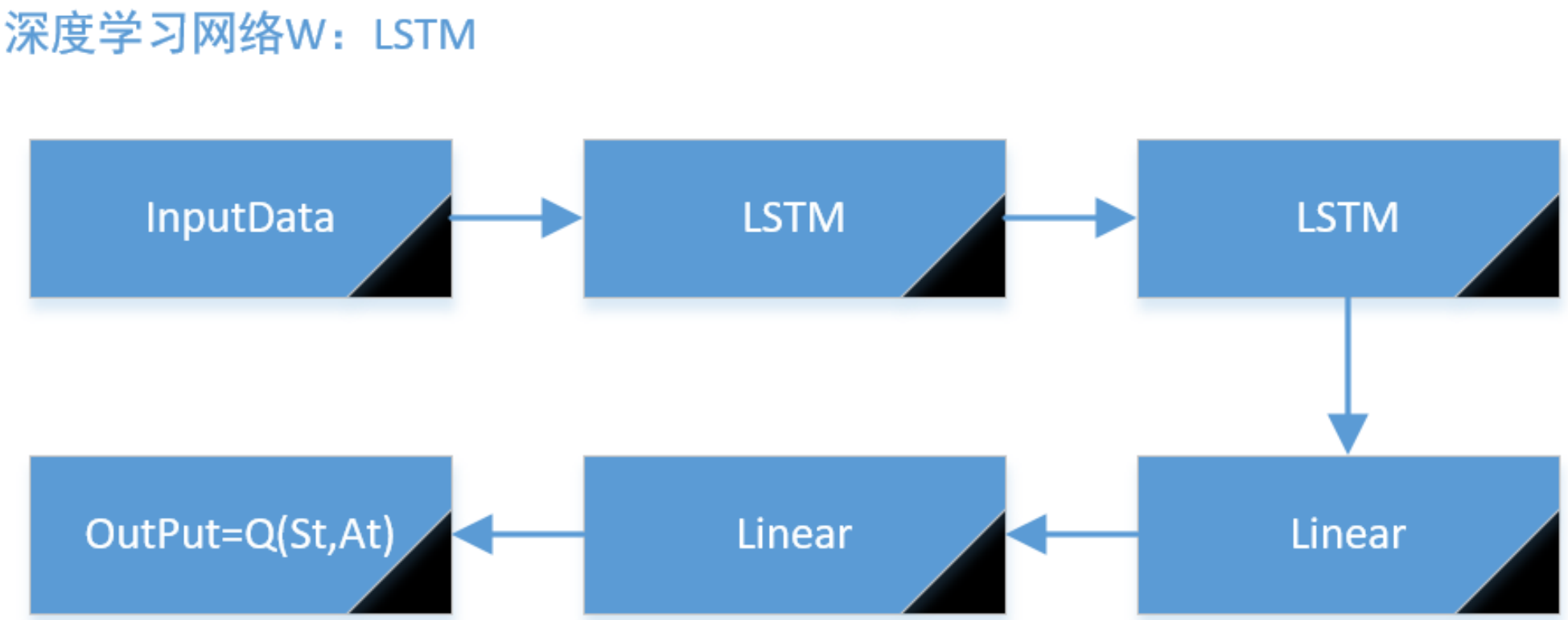
智能体的环境设置

- 环境状态定义St：过去n天的15分钟经过预处理的k线数据（最高价，最低价，开盘价，收盘价，收盘价，成交量，成交额），波动率，单笔交易浮盈浮亏，距离下一个交易日的天数，是否到达阻力位区域，持有时间。 $s_t \in S_t$
- 动作空间At：动作只有两个：开仓和平仓，1代表持仓状态，0代表空仓状态。由0到1代表开仓动作，由1到0为平仓动作。 $a_t \in A_t$

LSTM-DoubleDQN模型结构

- 深度学习网络A由两层LSTM加两层全连接层构成，来学习在状态St下对应各个动作的估值Q(St,At)
- 样本输入：当前时刻的状态数据St，在St状态下做出的动作At，在当前状态下做出动作后产生的奖惩Rt，下一个状态数据S（t+1），回合结束标志done。在实验中，模型完成空仓等待，开仓，持仓，平仓这一流程动作后视为一个回合。
- 输出：当前状态下输出持仓的估值评分(Q(St,1))和空仓动作的估值(Q(St,0))，以评分较大者作为交易行为。

LSTM-DoubleDQN的深度学习部分结构



LSTM-DoubleDQN算法流程

- 1、初始化两个LSTM神经网络W1，W2，分别为选择最优动作网络和计算估值评分（Q值）网络
- 2、输入状态st，使用epsion-gredy策略，有epsion概率随机选择一个动作，1- epsion选择Q(st,a)最大的动作：

$$a_t = argmax_{a \in A_t} Q(s_t, a)$$

- 3、环境返回奖励rt和下一个状态s(t+1)
- 4、将这个训练样本放入经验回放池中（DataBuffer）
- 5、经验回放池数据超过一定数量后，随机取出一定数量样本训练

LSTM-DoubleDQN训练流程

- 1、从经验回放池DataBuffer中随机采样，一个样本中含有st, rt, s(t+1), done
- 2、输入st和at，用W1计算当前状态下的估值eval_q=Q(st,at,W1)
- 3、输入s(t+1)，用W1获在下一个状态中估值最大的动作:

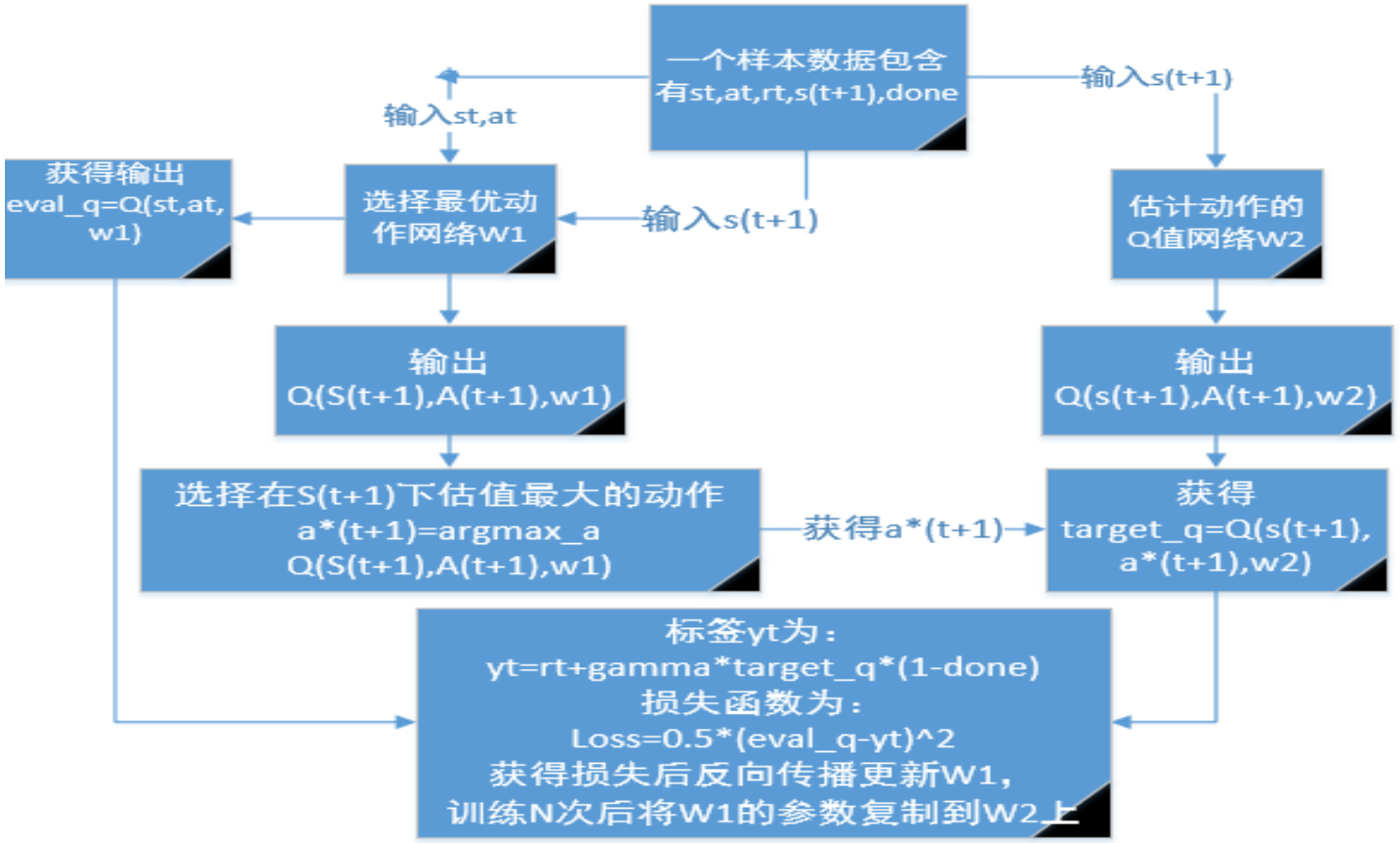
$$a_{t+1}^{*} = argmax_{a \in A} Q(s_{t+1}, a, W1)$$

- 4、目标值y由W2计算得到:

$$y = r_t + gamma * Q(s_{t+1}, a_{t+1}^{*}, W2))(1 - done)$$

- 5、获得损失Loss=0.5(eval_q-y)^2
- 6、获得参数梯度，反向传播更新W1参数。W1训练若干次后将W1的参数复制到W2上

LSTM-DoubleDQN强化学习部分结构



深度强化学习网络算法框架Double-DQN，其中W1和W2是第9页ppt采用的LSTM深度学习神经网络

激励函数设定(Reward Function)

- 定义在t时刻的持仓市值为MarketValue_t，定义开仓成本为Cost，则这笔交易在t时刻的对数收益率为return_t=ln(MarketValue_t/Cost)，设定止损线stop(stop<0)
- 若a(t-1)->at为0->1，是开仓动作，reward_t=0

- 若a(t-1)->at为1->1，是持仓动作，分有两种情况：

1、若return_t>stop，reward_t=0

2、若return_t<stop，reward_t=exp(return_t)-1(转换为普通收益率)

激励函数设定(Reward Function)

- 若a(t-1)->at为1->0，是平仓动作，分三种情况：

1、在止损线下平仓，reward_t=a (a>0)，因为止损是正确行为

2、在非止损情况下， reward_t=exp(return_t)-1

3、若平仓时止盈时的点位相比于开仓点位偏离幅度达到1.5%以上，给予双倍奖励

- 若a(t-1)->at为0->0，是空仓动作，reward_t=0

智能体环境参数

- 模型回看历史数据长度为20天，一个交易日有16根15m级别的k线，所以ModelWindow = 20*16=320
- 波动率选择近5日波动率n=5
- 手续费：期权一张15元
- 账户初始资金InitCash=1000000
- 开仓限制比例：开仓时市值不能超过总资金20%（交易所限制）
- 随机选择动作概率Epsilon=0.1

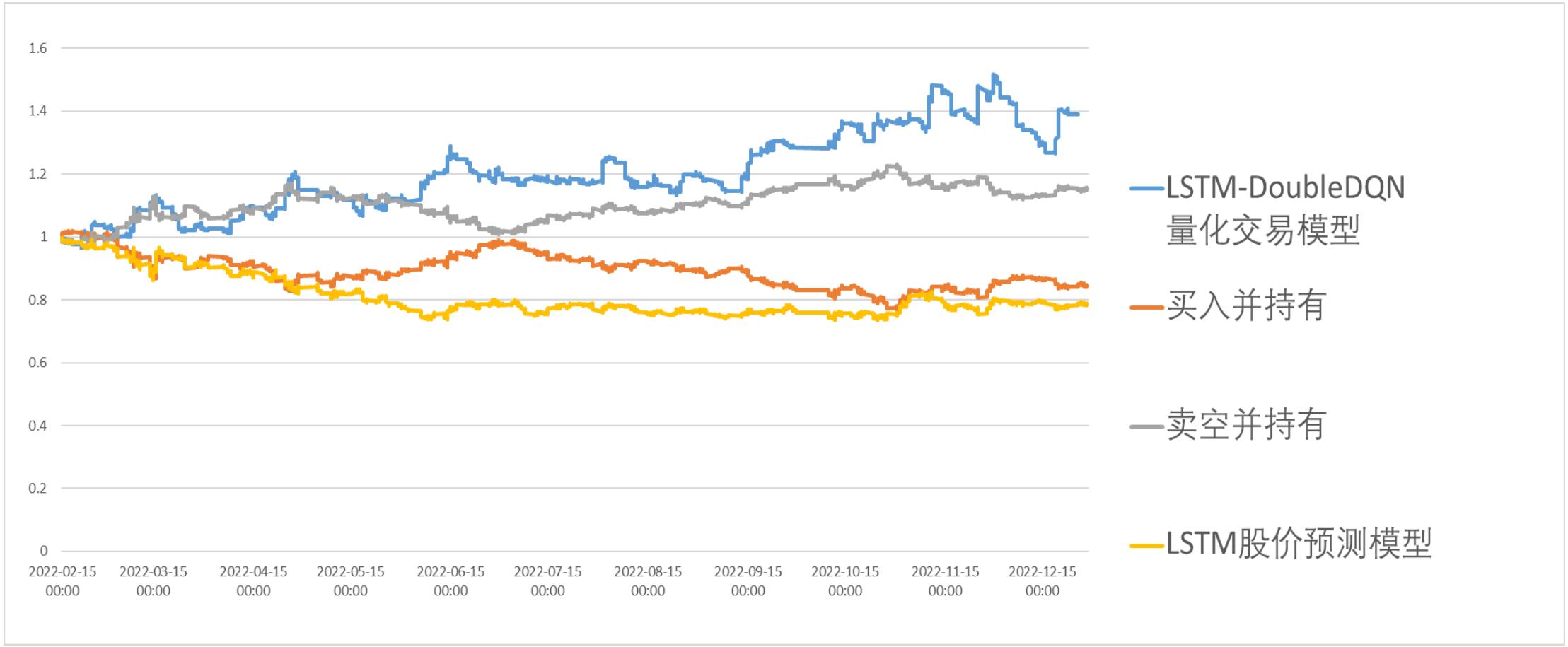
神经网络训练参数设置

- 优化器学习率LR=0.01
- 奖励折扣系数gamma=0.99
- 经验回放池大小MEMORY_SIZE = 15000
- 训练开始所需样本MEMORY_THRESHOLD=5000
- BatchSize=128
- Epoch=100
- 更新目标网络（W2）参数频率 UPDATE_TIME=100，即每训练100次更新一次目标网络参数

Baseline选择

- 1、账户长期满仓持有沪深300ETF多头
- 2、账户长期满仓融券沪深300ETF空头
- 3、使用LSTM网络每天进行股价预测，预测下一天股价上涨则满仓买入沪深300ETF做多，预测下一天股价下跌则满仓融券卖空沪深300ETF做空

测试集实验结果（以初始净值为1计算）



评价指标

模型\指标	夏普比率	到期总对数收益率	最大回撤（对数形式）
LSTM-DoubleDQN	1.1931	0.3176	-0.1818
LSTM	-1.258	-0.240	-0.1874
长期持有多头	-0.872	-0.166	-0.1854
长期持有空头	0.9147	0.1411	-0.1811

注：指标解释
夏普比率：$$SharpeRatio = \frac{E(R_P) - R_f}{\sigma_P}$$

E(Rp)：平均年化收益率
Rf：年化无风险利率
σp：年化收益率的标准差
夏普比率衡量的是投资者每承受一单位总风险，会产生多少的超额回报

到期总对数收益率：Rt=ln(pt/p0)即期末总资产除以期初总资产后取对数。

最大回撤率（对数形式）：是指在选定周期内任一历史时点往后推，产品净值走到最低点时的收益率回撤幅度的最大值。如果一只基金的最高净值是10元，最低净值是5元，那么它的最大回撤率就是ln(5/10)

附注：阻力位模型构建（支撑与压力）

- 1、定义一个固定的时间长度d
- 2、定义在时刻t时，指数点位是Pt，在时间长度d下，指数运动的平均速度为

$$Vt=(Pt-P(t-d))/d$$

3、若出现$Vt*V(t-1)<0$，则说明在时间区间d之内出现了一个反转点，若$Vt>0$则代表指数见底回升，是一个相对低点，是一个可能的支撑位，$Vt<0$则代表指数见顶回落，是一个相对高点，是一个可能的压力位。

3、若相对高点比前一个支撑点涨幅超过1.5%，或突破前一个压力位0.5%，记为一个压力点位。若相对低点比前一个压力位跌幅超过1.5%，或跌破前一个支撑位0.5%，记为一个支撑点位

实盘交割单

期权名称	浮动盈亏	持仓/可用	均价/最新
权利 300ETF购2月3329A 沪	-3,576.22	0 0	0.0000 0.0298
期 2024-02-28 剩余29日到期			
权利 300ETF沽2月3329A 沪	5,242.54	0 0	0.0000 0.1189
期 2024-02-28 剩余29日到期			

没有更多数据了

期权名称	浮动盈亏	持仓/可用	均价/最新
权利 300ETF购3月3400 沪	4,676.40	0 0	0.0000 0.1307
期 2024-03-27 剩余35日到期			
权利 300ETF沽3月3400 沪	-2,612.00	0 0	0.0000 0.0438
期 2024-03-27 剩余35日到期			

没有更多数据了