

**课程报告**

**课程名称： 强化学习**

**项目名称： 强化学习在量化金融中的应用**

**小组成员： 钱以骞，马瑞廷，张文涛，丛宇恒**

**本人分工： 数据获取和预处理以及DDPG评估**

**学 院： 物理学院**

**学 号： D202180035**

**姓 名： 钱以骞**

**联系方式： 18687751766**

**报告日期： 2022年10月18日**

目 录

[强化学习在量化金融中的应用 2](#_Toc117082233)

[1 Background 2](#_Toc117082234)

[2 Dependencies 2](#_Toc117082235)

[2.1 FinRL（https://github.com/AI4Finance-Foundation/FinRL） 3](#_Toc117082236)

[2.2 Tushare（https://tushare.pro/） 3](#_Toc117082237)

[2.3 Pyfolio（https://github.com/quantopian/pyfolio） 4](#_Toc117082238)

[3 Data preparation and pre-processing 4](#_Toc117082239)

[3.1 Financial data downloading using Tushare 4](#_Toc117082240)

[4 RL environment setup 5](#_Toc117082241)

[5 Performance 5](#_Toc117082242)

[5.1 条目 5](#_Toc117082243)

[5.2 条目 5](#_Toc117082244)

[6 小组分工 5](#_Toc117082245)

[6.1 条目 5](#_Toc117082246)

[7 总结与展望 5](#_Toc117082247)

[7.1 条目 5](#_Toc117082248)

[7.2 条目 5](#_Toc117082249)

[参考文献 6](#_Toc117082250)

# 强化学习在量化金融中的应用

## Background

在本课题中，我们将从零开始探索强化学习在量化金融中的应用，包括金融数据的获取，数据的预处理，强化学习环境的搭建以及不同强化学习算法性能的评估。

我们将用四种不同的强化学习算法训练AI交易员（DDPG，SAC，PPO，A2C），并让他们自己找到他们所认为的最优投资策略，并和baseline进行比较，当然四个交易员之间也会进行对比，得出最优的投资策略。

由于本课题侧重于强化学习，所以我们将很少谈及量化策略，如在选股的部分我们将直接选择上证50中的成分股进行训练而非采用一些量化因子来进行选股。

## Dependencies

本课题依赖于FinRL，Tushare和Pyfolio以及其它一些常用的python库，如pandas，numpy等，这些库都很常用我们就不在这过多介绍，我们主要介绍一下FinRL，Tushare和Pyfolio三个在金融分析领域常用的库。

* 1. FinRL（<https://github.com/AI4Finance-Foundation/FinRL>）

FinRL是一个强化学习的金融分析库，通过它我们可以容易的仿真金融交易环境和模拟环境和agent之间的互动。它的框架大致分为3层，最上层的是应用层，中间的是agents，最底层的是环境，agents可以通过和环境作用来进行决策。

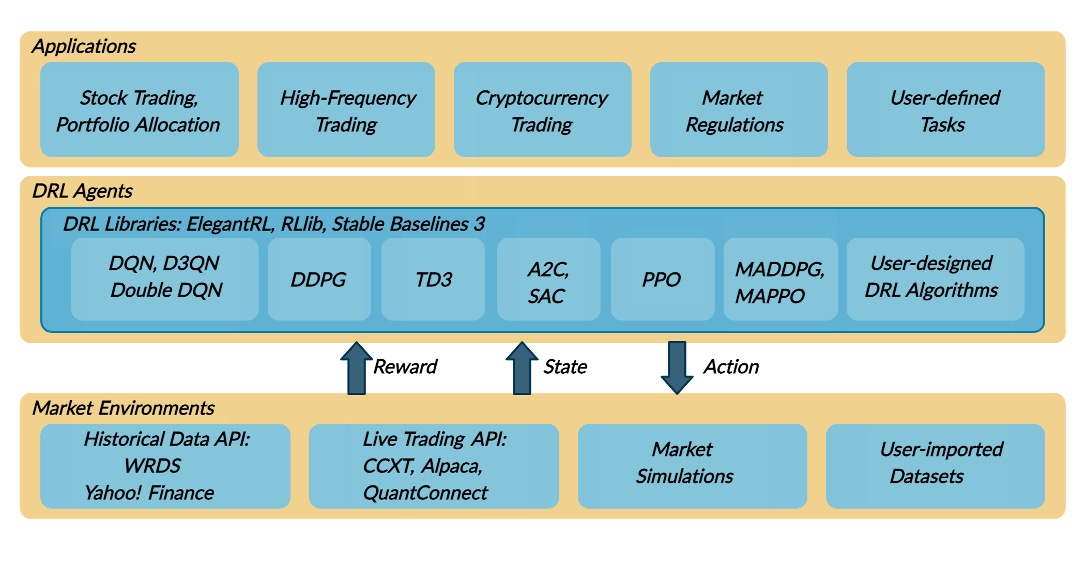


Figure 1 FinRL框架

* 1. Tushare（<https://tushare.pro/>）

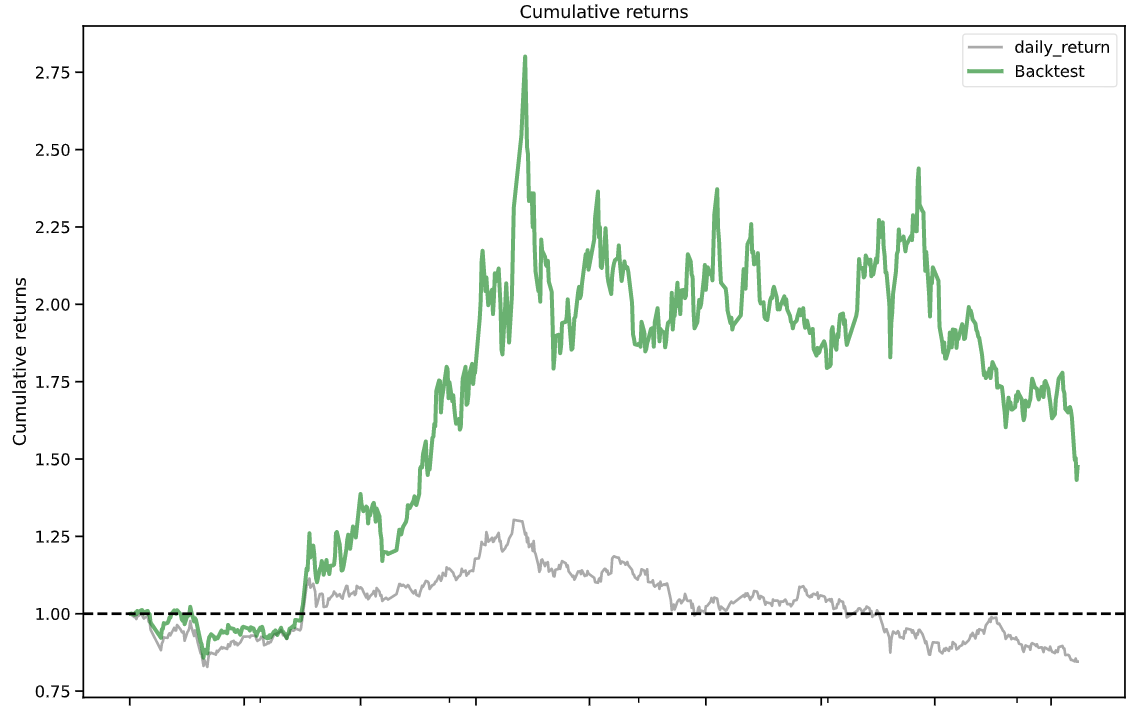
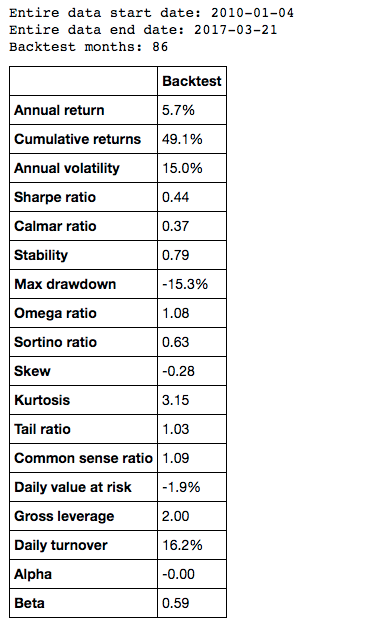
Tushare是一个开源的金融数据提供商，虽让一些数据需要权限，但是大部分数据可以免费的拿到。Tushare提供最高到tick级别的股票数据，包括开盘价，收盘价，最高和最低价以及成交量，还有其他一些指数和行情数据。



Figure 2 Tushare官网

* 1. Pyfolio（<https://github.com/quantopian/pyfolio>）

Pyfolio是一个金融投资表现和风险评估的库，它可以通过自动回测来评估一个投资算法（策略）的表现和风险程度，并且通过可视化来直观的展示策略的性



**Figure 3** **Pyfolio可视化**

能。

## Data preparation and pre-processing

我们通过tushare的API接口可以获取到股票的交易数据，考虑到我们的计算资源和时间关系，我们选择上证50作为我们的baseline，并且选择日线数据作为我们的数据库。我们选取了上证50中的成分股从2012年1月1日到2022年9月30日总共10年多的日线数据：

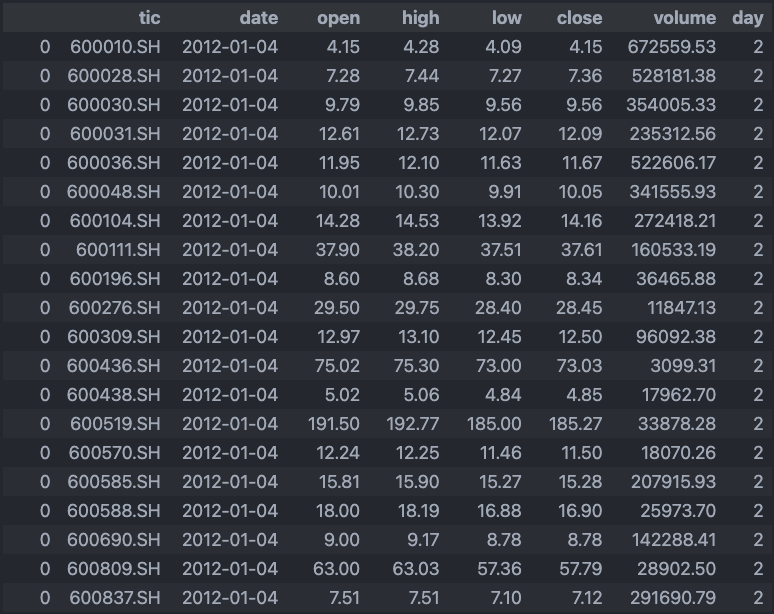


Figure 3 部分上证50成份股数据

接下来我们对数据进行了清洗和缺失值处理，去除了缺失值多于整个数据长度10%的成分股，最后剩下了38支满足要求的股票。紧接着我们给每只股票增加了一些金融上常用的技术指标如MACD，SMA和布林线等。

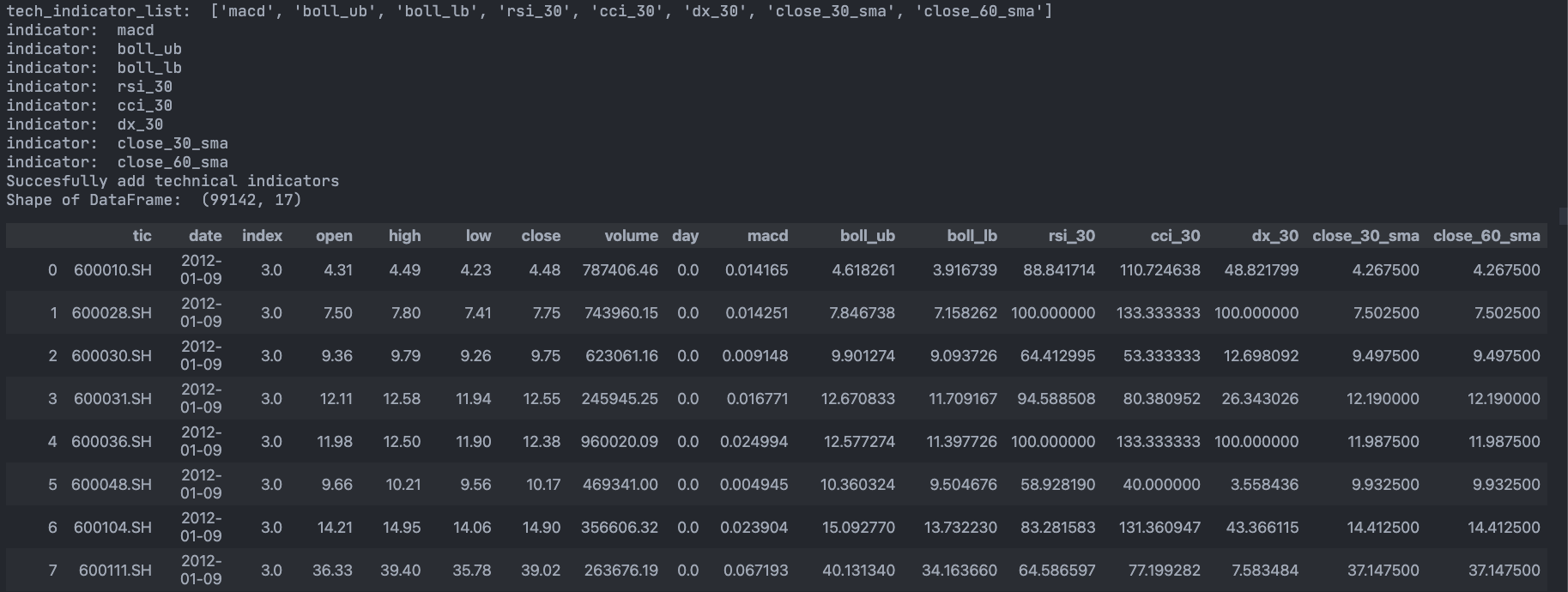


Figure 4 增加技术指标

至此数据已经准备就绪，接下来我们把数据分为训练集（training）和测试集（trading）。训练集包括2012年1月1日至2020年1月1日总共8年的数据，我们用近两年2020年1月1日至2022年9月30日的数据来进行测试。

## RL environment setup

* 1. Market environment

我们通过FinRL来构建一个类似OpenAI Gym-style的金融仿真环境，RL agents可以通过和这个环境进行互动，包括股价的监测，动作的采取和奖励的计算，最终学会一个投资策略来使得奖励最大化。



Figure 5 Market environment setup

在这里我们简要说明一下我们的市场环境，动作和奖励的定义：

1. 动作：买，卖和持有，通过{-1,0,1}来表示，1代表买，-1代表卖，0代表持有。假设我们持有k\*100股（A股规定交易必须是100的整数倍）某股票，那么动作空间为{-k,…,0,…,k}, 这里的k\*100≤hmax，即k≤10，那么在本课题中，动作空间大致为10^38
2. 奖励：R(a,s,s’) = v’-v, v’和v分别为状态s’和s的账户总资产，账户总资产=股票价值+现金
3. 状态：在本课题中，状态为每个股票的技术指标加上收盘价和持有股数再加上现金，写成数学表达式为s = (技术指标数量+1+1)\*股票数量+1=381

在训练环境中，我们设置每只股票单次最大交易1000股（hmax），初始资金有1000万。设置好金融环境后就可以开始训练我们的AI交易员了。

* 1. Agent training

FinRL也为我们提供了一些常用的RL agents，我们选取了其中四种：DDPG，SAC，PPO和A2C。



Figure 6 RL agents以及各自的超参数

在训练的时候发现DDPG和SAC的训练时间要比PPO和A2C短很多，但是表现要弱于PPO和A2C，在接下来的部分我们会着重分析各个agent的性能表现。

## Performance

在经过一段时间的训练之后，我们的AI交易员就准备完毕了。接下来将在交易环境中对他们所学到的策略进行测试。交易环境的设置大致上和训练环境相一致，只是把初始资金降到了100万。

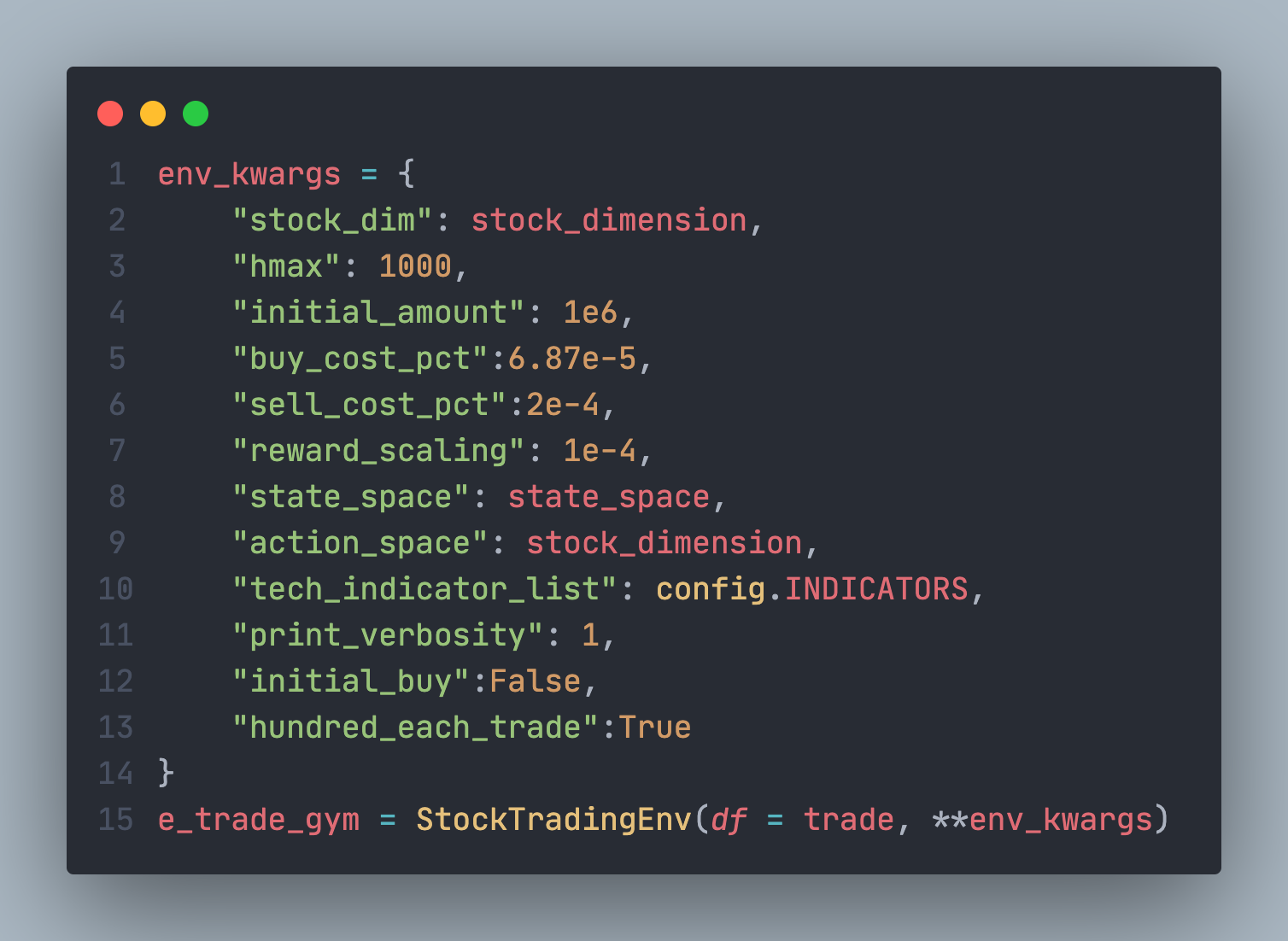


Figure 7 Trade environment setup

接下来我们让AI交易员在此环境中进行交易，最后我们通过Pyfolio给我们输出各个交易员的性能评估。

* 1. DDPG

通过查看DDPG的动作，我们发现DDPG倾向于在交易初期选择几只股票买入然后一直持有，期间不做任何交易。通过Pyfolio进行回测后，发现它的年化收益为9.2%，累计收益为26.2%，最大回撤为43.5%。与此同时，baseline，即上证50的年化收益为-6%，累计收益为-15.5%，最大回撤为35.2%，可见DDPG的表现是要优于上证50的。

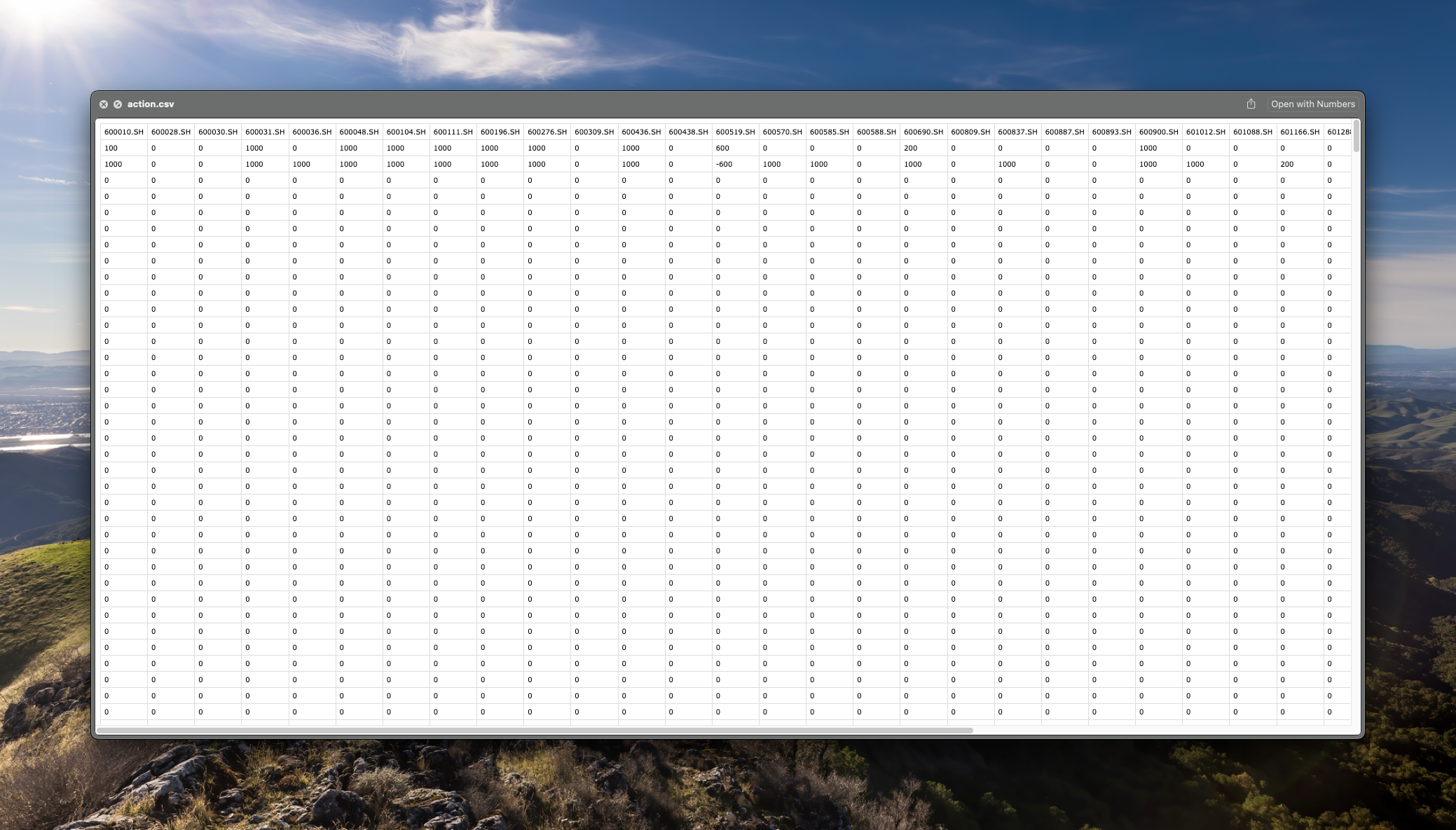
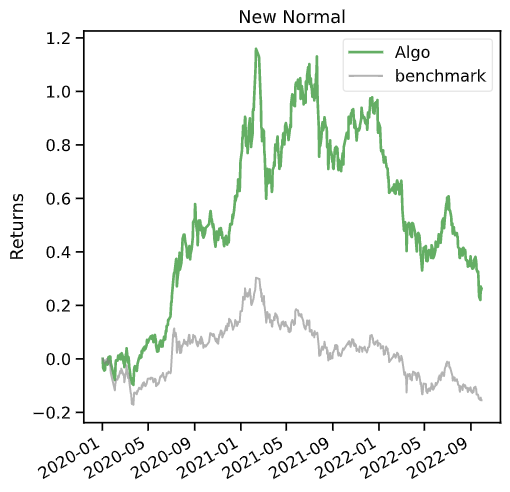
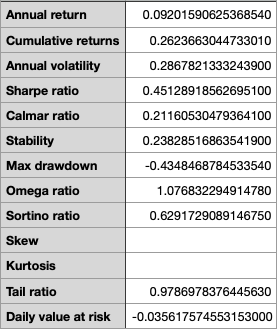
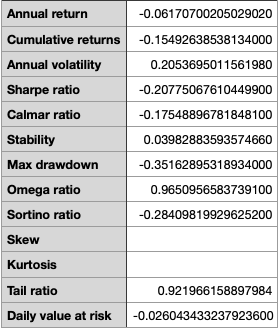


Figure 8 DDPG actions



**Figure 9 DDPG performance: 从左到右为baseline，DDPG以及累计收益对比**

* 1. SAC

SAC的表现与DDPG类似，也是选取几只股买入后一直持有，但是选取的个股不同，收益也不同：年化收益11.6%，累计收益33.8%，最大回撤29.6%，

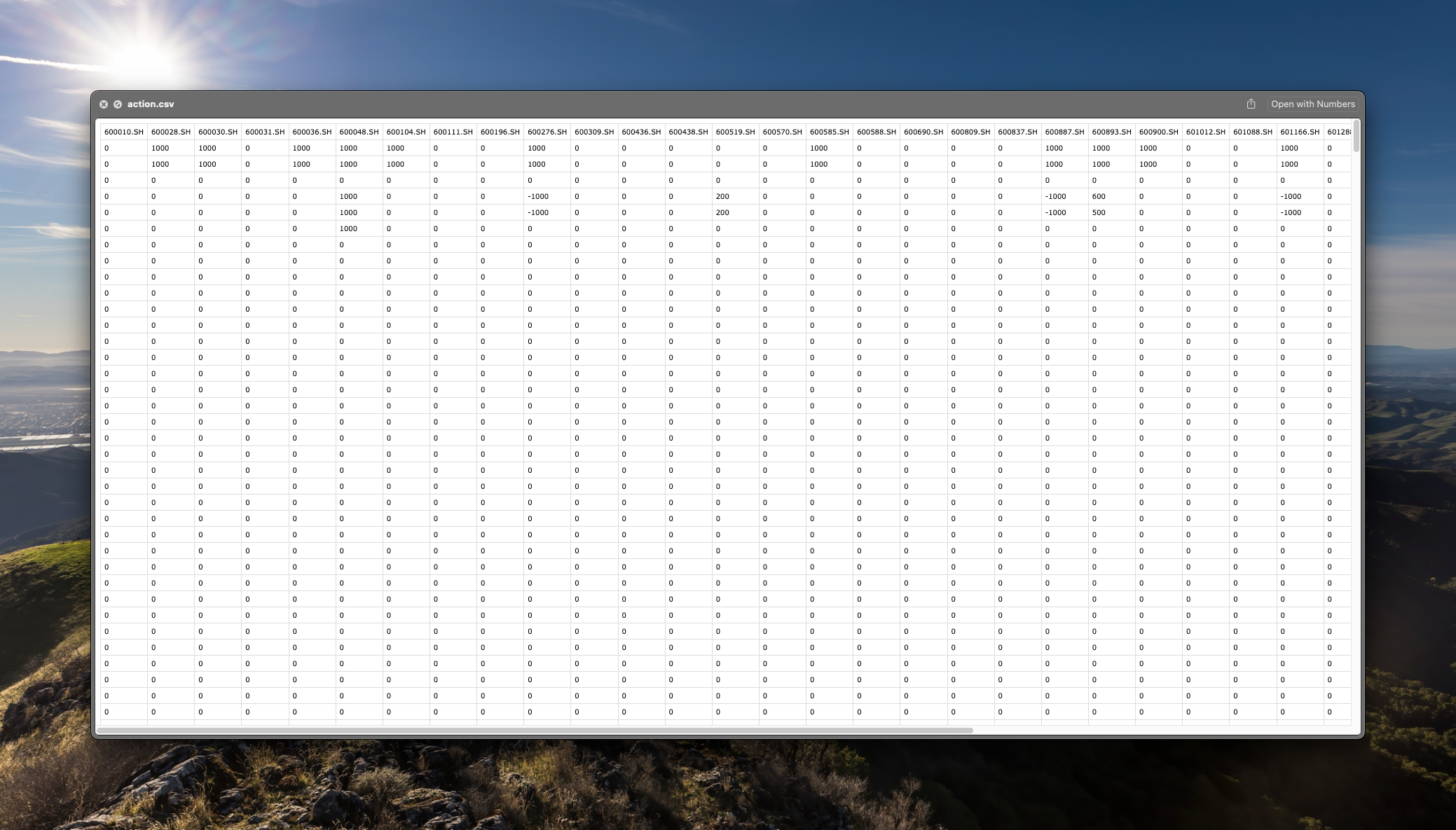
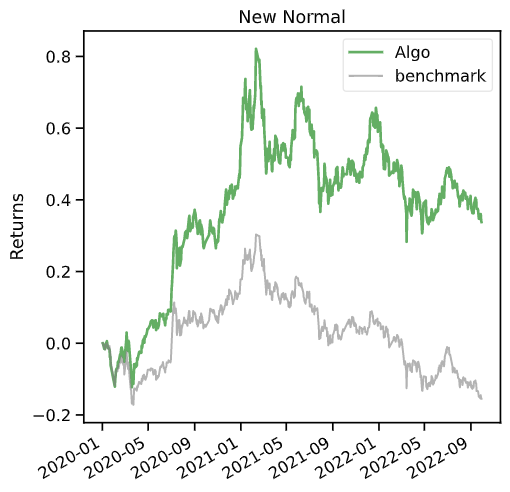
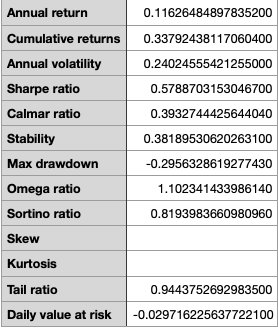
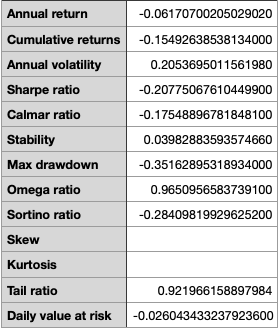


Figure 9 SAC actions



**Figure 10 SAC performance: 从左到右为baseline，SAC和累计收益对比**

表现优于baseline和DDPG。

* 1. A2C

A2C的表现

* 1. PPO

## 小组分工

* 1. 条目

## 总结与展望

* 1. 条目
  2. 条目

## 参考文献

[1]

[2]