



清华大学
Tsinghua University

FBDQA期末大作业 暨 AI量化模型竞赛规则说明

金融大数据与量化分析

Financial Big Data and Quantitative Analytics



1.1 竞赛规则概要

■ **预测目标：** 利用过往及当前数据预测未来中间价的移动方向

■ **输入数据：**

■ **行情频率：** 3秒一个数据点（也称为1个tick的snapshot）

■ **每个数据点包括当前最新成交价/五档量价/过去3秒内的成交金额等数据**

■ **训练集中每个数据点包含5个预测标签的标注**

■ **允许利用过去不超过100tick（包含当前tick）的数据，预测未来N个tick后的中间价移动方向**



1.2 竞赛规则概要： mid-price

■ **预测目标：** 利用过往及当前数据预测未来中间价的移动方向

■ 中间价定义：

$$\text{midprice} = \begin{cases} \frac{\text{ask1} + \text{bid1}}{2}, & \text{if ask1} \neq 0 \text{ and bid1} \neq 0 \\ \text{ask1}, & \text{if bid1} = 0 \text{ (例如跌停)} \\ \text{bid1}, & \text{if ask1} = 0 \text{ (例如涨停)} \end{cases}$$

■ 预测时间跨度：5、10、20、40、60个tick, 5个预测任务

■ 即在t时刻，分别预测t+5tick, t+10tick, t+20tick, t+40tick, t+60tick以后：

■ 最新中间价相较t时刻的中间价：下跌/不变/上涨



1.3 竞赛规则概要：移动方向

■预测目标：最新中间价的移动方向

■如何定义midprice的移动方向？

■ x = 待预测时刻的midprice - 当前时刻的midprice

$$\phi(x) = \begin{cases} 0 & x < -\alpha \\ 1 & -\alpha \leq x \leq \alpha \\ 2 & \alpha < x \end{cases}$$

■ 0表示下跌，1表示不变，2表示上涨

■ 对于预测窗口为5tick, 10tick, alpha取0.05%

■ 对于预测窗口20tick, 40tick, 60tick, alpha取0.1%





1.4 竞赛规则概要：评分标准

■ 模型评测指标：

■ 客观指标1（30分）：F0.5

- 对每个预测任务，将所有参赛队伍结果按排名分5档，按F0.5得分从高到低排列
- 通过正态分布将得分分布至15-30分之间
- 取5个预测任务中的最高得分

■ 客观指标2（30分）：P&L

- 对每个预测任务，将所有参赛队伍结果按排名分5档，按最终Profit高低从高到低排列
- 通过正态分布将得分分布至15-30分之间
- 取5个预测任务中的最高得分

■ 主观评价（40分）：

- 形式完整性（10分）、方案合理性（10分）、结果有效性（10分）、新颖性（10分）
- 经过路演后由评委打分得出



02 数据说明



2.1 文件位置

- 数据集发布位置: <https://cloud.tsinghua.edu.cn/published/fbdqa-2021a-mmpchallenge/>
- 请使用FBDQA2021A_MMP_Challenge.zip

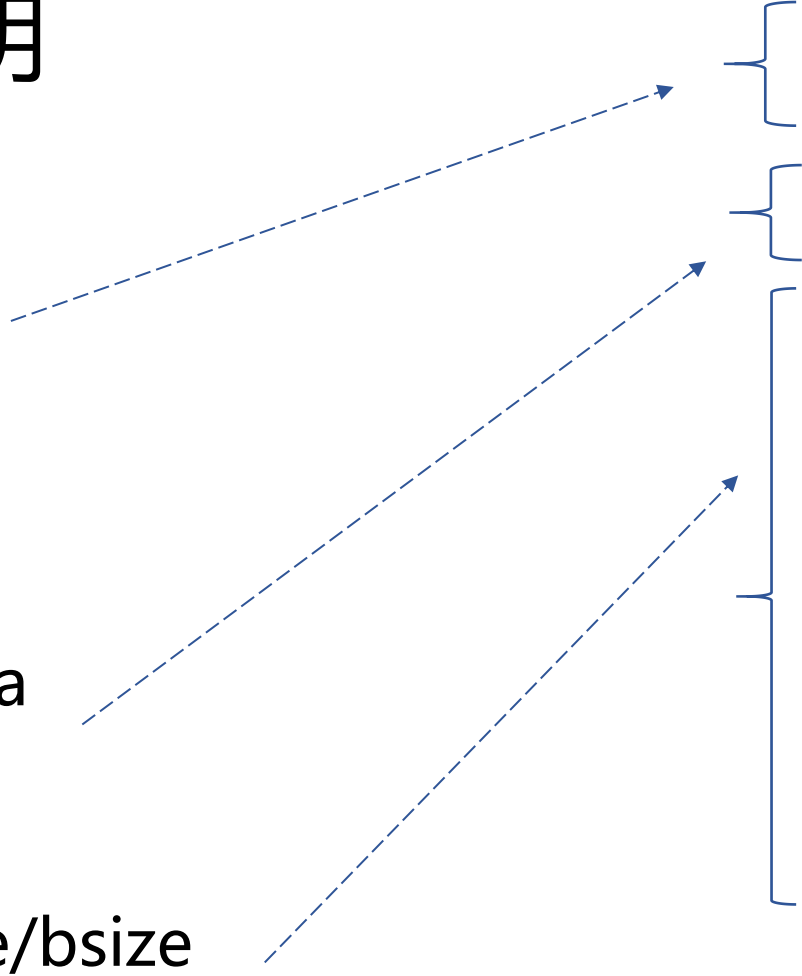
2.2 数据说明

- 共79个交易日，10只标的的股票
- 所有行情数据存在train_data.csv

| date | time | sym | close | amount_delta | n_midprice | n_bid1 | n_bsize1 | ... | n_ask5 | n_asize5 | label5 | label10 | label20 | label40 | label60 |
|------|----------|-----|-----------|--------------|------------|-----------|--------------|-----|-----------|--------------|--------|---------|---------|---------|---------|
| 0 | 09:40:03 | 0 | 0.015691 | 2594895.0 | 0.015270 | 0.015130 | 1.961524e-06 | ... | 0.016531 | 1.645685e-05 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0 | 09:40:06 | 0 | 0.015410 | 837314.0 | 0.015410 | 0.015130 | 3.424355e-06 | ... | 0.016811 | 1.828539e-05 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0 | 09:40:09 | 0 | 0.015130 | 184807.0 | 0.015270 | 0.015130 | 3.158386e-06 | ... | 0.016531 | 1.657321e-05 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0 | 09:40:12 | 0 | 0.015691 | 500046.0 | 0.015551 | 0.015410 | 4.155771e-07 | ... | 0.016811 | 1.828539e-05 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0 | 09:40:15 | 0 | 0.014290 | 1785635.0 | 0.014710 | 0.014290 | 9.308927e-06 | ... | 0.016251 | 4.920433e-06 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 78 | 14:49:45 | 9 | -0.001504 | 2305964.0 | -0.001562 | -0.001620 | 1.455123e-06 | ... | -0.000926 | 2.297563e-06 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 78 | 14:49:48 | 9 | -0.001620 | 3132273.0 | -0.001678 | -0.001736 | 8.971435e-07 | ... | -0.001157 | 5.525091e-07 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 78 | 14:49:51 | 9 | -0.001736 | 2364061.0 | -0.001793 | -0.001851 | 7.601650e-07 | ... | -0.001273 | 1.641116e-07 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 78 | 14:49:54 | 9 | -0.001851 | 20587972.0 | -0.002256 | -0.002661 | 3.260351e-08 | ... | -0.001388 | 3.282232e-08 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 78 | 14:49:57 | 9 | -0.001967 | 1526620.0 | -0.002372 | -0.002777 | 3.225340e-07 | ... | -0.001504 | 4.247045e-06 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

2.2 数据说明

- 信息戳
 - 日期
 - 时间
 - 股票代码
- 成交数据
 - amount_delta
 - n_close
- 量价数据
 - ask/bid/asksize/bsize



| 字段 | 含义 | 说明 |
|--------------|--------------|---|
| date | 日期 | sequential标号：既保留跨标的的可比性，也隐去实际时间 |
| time | 时间戳 | 保留实际时间戳，3s一档行情 |
| sym | 标的(仅序号) | |
| close | 最新价/收盘价 | 以涨跌幅表示 |
| amount_delta | 成交量变化 | 从上个tick到当前tick发生的成交金额 |
| n_midprice | 中间价 | 标准化后的中间价，以涨跌幅表示 |
| n_bid1 | 买一价 | 标准化后的买一价，以下类似 |
| n_bsize1 | 买一量 | |
| n_bid2 | 买二价 | |
| n_bsize2 | 买二量 | |
| n_bid3 | 买三价 | |
| n_bsize3 | 买三量 | |
| n_bid4 | 买四价 | |
| n_bsize4 | 买四量 | |
| n_bid5 | 买五价 | |
| n_bsize5 | 买五量 | |
| n_ask1 | 卖一价 | |
| n_ask1 | 卖一量 | |
| n_ask2 | 卖二价 | |
| n_ask2 | 卖二量 | |
| n_ask3 | 卖三价 | |
| n_ask3 | 卖三量 | |
| n_ask4 | 卖四价 | |
| n_ask4 | 卖四量 | |
| n_ask5 | 卖五价 | |
| n_ask5 | 卖五量 | |
| label5 | 5tick价格移动方向 | 当前tick中间价相对于5tick之前的移动方向，0为下跌，1为不变，2为上涨 |
| label10 | 10tick价格移动方向 | |
| label20 | 20tick价格移动方向 | |
| label40 | 40tick价格移动方向 | |
| label60 | 60tick价格移动方向 | |
| | | |

2.3 字段说明

| 字段 | 含义 | 说明 |
|-------------------|---------|--|
| date | 日期 | 去掉实际日期，但是保留序号，取值范围0-79，保留跨标的的可比性 |
| time | 时间戳 | 保留实际时间戳，3s一档行情 |
| sym | 股票标的 | 仅序号，10只股票分别为0-9 |
| n_close | 最新成交价 | 无量纲，以涨跌幅表示，-1 ~ 1，即跌100%到涨100%（实际取值范围更小） |
| amount_delta | 成交量变化 | 从上个tick到当前tick发生的成交金额，单位元 |
| n_midprice | 中间价 | 无量纲，以涨跌幅表示 |
| n_bid1-n_bid5 | 买一价-买五价 | 无量纲，以涨跌幅表示 |
| n_ask1-n_ask5 | 卖一价-卖五价 | 无量纲，以涨跌幅表示 |
| n_bsize1-n_bsize5 | 买一量-买五量 | 无量纲，以换手率表示 |
| n_asize1-n_asize5 | 卖一量-卖五量 | 无量纲，以换手率表示 |



03 测试方法



3.1 测试方法

- 测试方式为交互式调用
 - 参赛者需要提交源代码及模型，模型文件不超过2GB，FP32精度
- 代码需提供提供一个Predictor类，对一组数据点进行一次预测
 - 类应至少具有__init__方法和predict方法：
 - `def predict(x: dataframe) -> List[int]`
 - 评测程序会多次调用predict，得到单个预测并汇总计算P&L

3.2 predict方法

- 考虑到同学可能会需要多个时间步（不超过100tick）的数据用于单次预测，单次会输入形如[100,26]的数据：
 - 100行为过去100个tick的数据
 - 每行长度为26，与发布数据相同(去除标注项)
- 函数应返回5个int类型预测标签（标签为0/1/2）

3.3 调用方式

- 为了公平性与数据保密性，评测程序会打乱测试点输入顺序
 - 单次输入的100个tick为连续数据点
 - 不同次输入的数据无先后关系
 - date：日期会置为0，无意义
 - sym：可能的范围0-10（可能有不来自于10只训练集股票的数据，仅作泛化性的额外参考，不列入最终积分）
 - time：time会保留实际交易时间，以备参赛模型会根据时段不同做推理



04 提交包准备



4.1 必要文件

为正确评测模型，需按引导文字完成以下准备：

1. 模型权重文件（这里以.pth文件为例）。

2. 提交说明：columns.json

- 明确你要测哪一版数据

FBDQA2021A_MMP_Challenge

- 明确你用到了哪些特征列

[https://cloud.tsinghua.edu.cn/published/fbdqa-2021a-](https://cloud.tsinghua.edu.cn/published/fbdqa-2021a-mmpchallenge/)

mmpchallenge/

3. 环境依赖说明：requirements.txt

4. 根据模板写自己的预测类，Predictor.py

金融大数据机器学习策略仿真平台

首页 × 比赛提交 ×

前期准备 公榜评测 私榜评测

1. 准备好训练完成的权重文件 (*.pth)
2. 准备训练所用特征列文件columns.json内容格式样例如下

```
1 {"data": "t5", "feature": ["ask1", "ask2", "ask3", "asize1", "asize2", "asize3", "bid1", "bid2", "bid3", "bsize1", "bsize2", "bsize3"]}
```

3. 在程序根目录下执行下面命令导出 requirements.txt

```
1 pipreqs ./ --encoding=utf-8 --debug
```

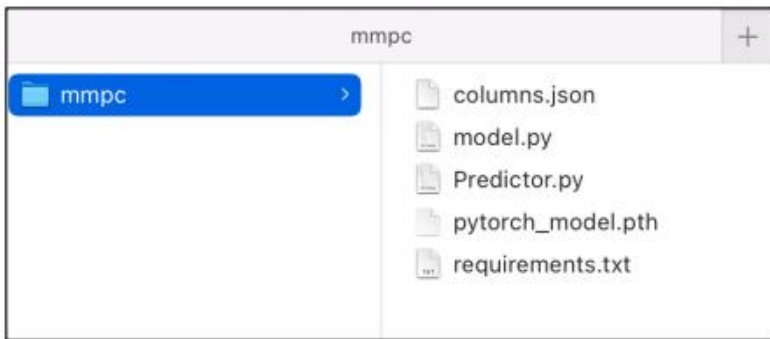
4. Predictor 类模板(Predictor.py)

```
1 import os
2 class Predictor():
3     def __init__(self, pkl_path):
4         #进行你需要的初始化，如加载模型等
5         self.model = self.load_model(pkl_path)
6         self.device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
7         self.model.to(self.device)
8
9     def predict(self, x):
10        #进行你需要的预处理（例如选择合适的输入列，生成所需特征，规范化等等）
11        x_hat = preprocess(x)
12        x_hat.to(self.device)
13        #进行预测
14        y = model(x_hat)
15        #转化成list(如果需要)
16        y = y.numpy().tolist()
17        返回预测结果
```


4.2 示例

➤ 典型文件夹结构：

- columns.json -> 存储了提交关键信息的字典，包括数据集、特征、label
- Predictor.py -> 必须提供Predictor类，以及predict方法。评测器通过调用该方法进行评测。
- model.py -> 如果你采用了深度模型，记得在Predictor.py中import。
- model.pth -> 你模型的权重文件。如果是tensorflow可以命名为model.h5。请在Predictor类的__init__函数中加载模型权重。



文件夹结构示例



4.3 准备

- 将训练的模型拷贝到 “zip包验证程序/mmpc” 目录下（不要在mmpc下建子文件夹）
- 将写好的columns.json 文件 拷贝到 “zip包验证程序/mmpc” 目录下
- 将 requirements.txt 文件 拷贝到 “zip包验证程序/mmpc” 目录下

4.4 修改验证代码

根据你的模型修改model.py和 Predictor.py 文件

a、model.py 为模型定义

b、Predictor.py 为推理类 （模型加载的路径 一定要是 './mmpc/...'）**环境构建需要，平台会把 './mmpc/' 替换成程序运行路径**

修改zip包验证程序中main.py 需要你指定你测试数据位置



4.5 验证&打包

构建运行环境

```
1 conda create -n test python=3.8 #平台默认python环境是3.8
2 conda activate test
3 pip install -r zip包验证程序/mmpc/requirements.txt
```

执行验证程序

```
1 cd zip包验证程序/
2 python main.py
```

如果程序运行没有问题，就可以打包了

打包

将 zip包验证程序 文件下 mmpc 文件打包成zip文件，就可以上传到平台。

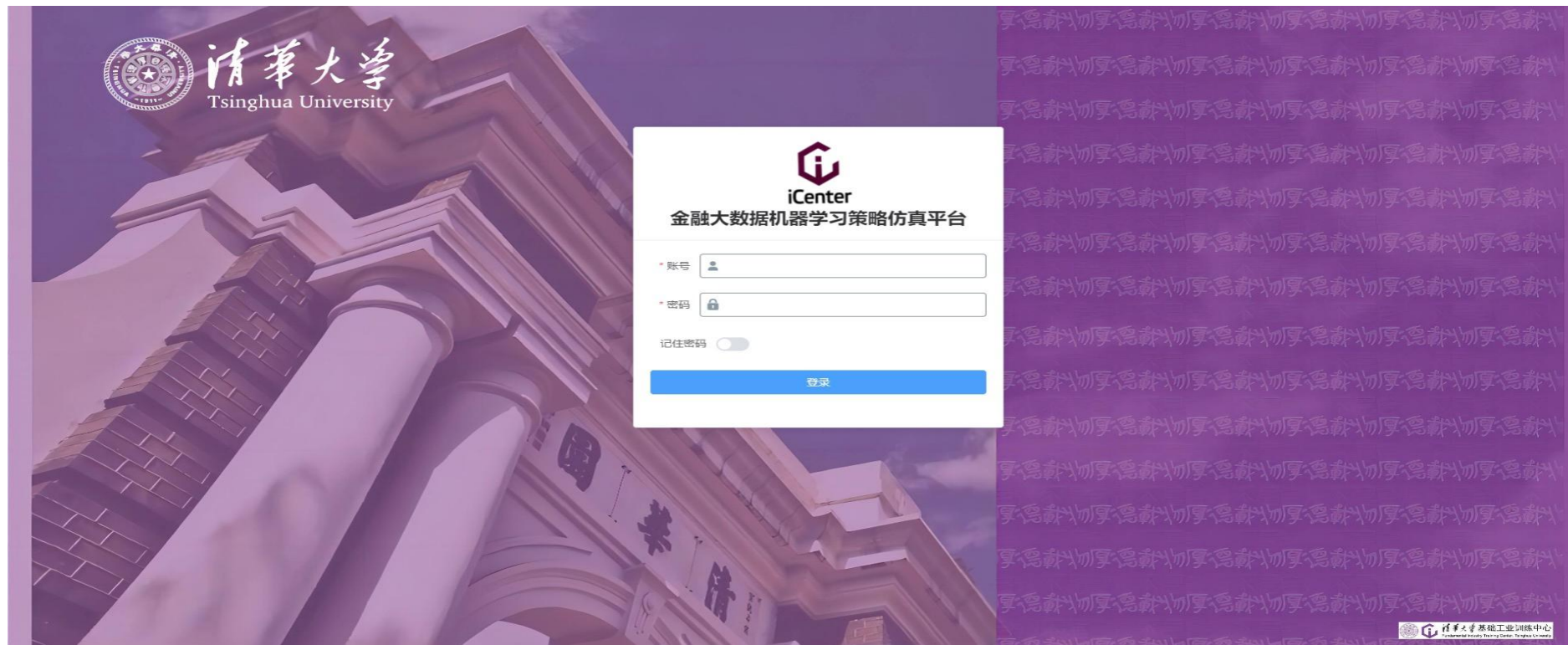


04 平台提交



登录界面

链接: <http://101.6.160.66:18080/>



提交&评测界面

- ◆ 点击“创建评测任务”即可新建评测任务、上传训练好的模型
- ◆ 公榜评测**每隔8小时**可以提交一次
- ◆ 最终评测以最后一次提交为准

金融大数据机器学习策略仿真平台

首页

教学

数据处理

机器学习展示

比赛提交

用户管理

前期准备

公榜评测

私榜评测

用户名

创建评测任务

收益率曲线

| ID | 用户 | 创建时间 | 状态 | 操作 |
|------|----|------|----|----|
| 暂无数据 | | | | |

| ID | 用户 | precision | recall | f_score |
|------|----|-----------|--------|---------|
| 暂无数据 | | | | |



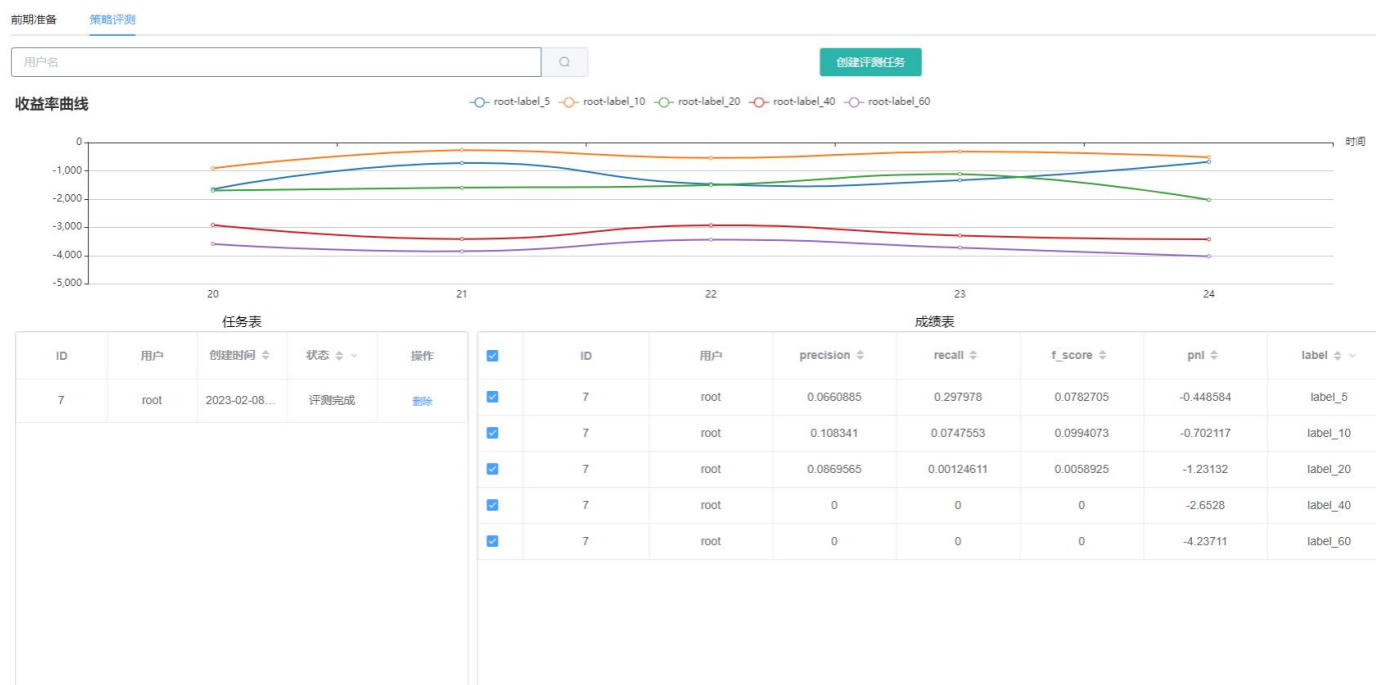
创建评测任务与模型上传

- 点击上传zip文件按钮，选择准备好的zip文件
- 点击创建，评测任务就创建好了。
- 可以在任务表中看到创建的任务。



成绩查看与对比

- 模型上传后自动回测并计算
 - 预测准确率 (precision)
 - 召回率 (recall)
 - F_Score
 - 收益率 (P&L)
 - 夏普率 (Sharpe)
 - 最大回撤
- 可任意勾选策略进行净值曲线比对。





清华大学
Tsinghua University

THANKS