
量化交易平台开发手册

闫涛
阿尔山金融科技有限公司
北京
{yt7589}@qq.com

Abstract

基于 tushare.org 开放数据集，构建本地量化交易开发平台。

1 概述

1.1 环境搭建

创建开源项目，项目网址：<https://github.com/yt7589/aqp>，本地环境为：
d:/awork/aftdc/incubate/aqp，虚拟环境激活：d:/aadesk/dev/python/quant/Script/activate，
这个是通过 `python -m venv quant` 来创建的虚拟环境。

1.2 整体架构

app_main.py：程序主入口；
app_registry.py：管理程序中所有全局性配置和变量；
controller 目录：所有业务逻辑实现类；
model 目录：所有数据库操作类；

1.3 数据服务商

我们采用的是 tushare.pro 提供的数据服务：<https://tushare.pro>。

1.4 Python 单元测试

我们采用类似测试驱动开发方式，主要功能都要有单元测试，因此需要使用单元测试框架。我们采用 unittest 单元测试框架，下面我们以一个具体的例子为例，来说明怎样使用单元测试框架。我们来看怎样从 t_account_io 表的某个账户中转出现金的 10%。我们首先建立 test 目录，存放所有单元测试用例。

首先建立测试用例目录，其名称为 tdd，注意：不要叫 test 目录，这个目录会有问题。我们建立账户流水控制器类 controller/c_account_io.py，如下所示：

```
1 class CAccountIo(object):
2     def __init__(self):
3         self.name = 'CAccountIo'
4
5     @staticmethod
6     def withdraw(account_id, amount):
7         '''
8         从指定账户取出指定金额，增加资金转出金额记录，并更新用户现金资产
9         '''
10        print('CAccountIo.withdraw')
11
```

```

12 #股票预测:
13 #https://medium.com/m/global-identity?redirectUrl=https%3A%2F%2Ftowardsdatascience.com%2Faifortrading-2edd6fac689d

```

Listing 1: 账户流水类

在测试目录下创建对应的测试类 tdd/controller/t_c_account_io.py, 如下所示:

```

1 import unittest
2 from controller.c_account_io import CAccountIo
3
4 class TCAccountIo(unittest.TestCase):
5     def test_withdraw(self):
6         account_id = 1
7         amount = 100
8         CAccountIo.withdraw(account_id, amount)
9         self.assertEqual('abc', 'abc')
10
11 if '__main__' == __name__:
12     unittest.main()

```

Listing 2: 账户流水测试类

测试入口类 app_test.py, 如下所示:

```

1 from tdd.controller.t_c_account_io import TCAccountIo

```

Listing 3: 单元测试入口类

运行单元测试用例命令为:

```

1 python -m unittest -v app_test.py
2 python -m unittest -v tdd.model.t_m_account_io.TMAccountIo.test_withdraw

```

Listing 4: 测试程序命令行执行

第一种方式是在一个文件中统一进行处理, 第二种方式直接指定预测试的类甚至欲测试的方法。运行结果如下所示:

Figure 1: 单元测试运行结果

```

[(quant) arxanfintechdeMacBook-Pro:app arxanfintech$ python -m unittest -v app_test.py
test_withdraw (tdd.controller.t_c_account_io.TCAccountIo) ... CAccountIo.withdraw
ok

-----
Ran 1 test in 0.000s

OK
(quant) arxanfintechdeMacBook-Pro:app arxanfintech$ █

```

1.5 Flask 入门

Flask 框架目前已经超越 Django, 成为 Python 语言 Web 开发中最流行的框架。

1.5.1 入门程序

使用 flask 进行 Web 开发非常简单, 因为实际应用中 python 只用于提供 API 接口, 因此我们以 API 服务器为例:

```

1 import json
2 from flask import Flask
3
4 app = Flask(__name__)
5

```

```

6 @app.route('/')
7 def hello_world():
8     resp = {'code': 1}
9     resp['result'] = {}
10    resp['result']['userId'] = 101
11    return json.dumps(resp)
12
13 app.run(port=5080, debug=True)

```

Listing 5: Hello World 级 Flask 程序

上述程序中，监听在 5080 端口，返回值为 JSON 格式。

1.5.2 路径中的参数

在很多情况下，我们需要将路径中的部分作为参数，具体处理方式如下所示：

```

1 @app.route('/user/<int:user_id>/<string:user_name>/<email>', methods=['GET'])
2 def process_path_params(user_id, user_name, email):
3     resp = {'code': 1}
4     resp['result'] = {}
5     resp['result']['userId'] = user_id
6     resp['result']['userName'] = user_name
7     resp['result']['email'] = email
8     return json.dumps(resp)

```

Listing 6: 处理 path 中的参数

参数可以带类型说明也可以不带，同时分号两边不能有空格。methods 可以为 GET 或 POST。

1.5.3 处理 QueryString

获取 URL 中的参数可以采用下面的方法：

```

1 @app.route('/get_url_params/t1', methods=['GET'])
2 def get_url_params():
3     resp = {'code': 1}
4     resp['result'] = {}
5     resp['result']['url'] = url_for('get_url_params')
6     resp['result']['userId'] = request.args['userId']
7     # args 类型为 ImmutableMultiDict
8     params = request.args.to_dict()
9     resp['result']['params'] = params
10    return json.dumps(resp)

```

Listing 7: 获取 URL 中的参数

如上所示，url_for 可以打印出完整的 URL 路径，参数为路径中第一个分量。request.args 为不可变字典，也可以将其转变为普通字典。

1.5.4 处理 form

当请求为 POST 时，参数是通过 form 来传递的，这时需要解析 form 中的参数：

```

1 @app.route('/get_post_params/t2', methods=['POST'])
2 def get_post_params():
3     resp = {'code': 1}
4     resp['result'] = {}
5     resp['result']['userName'] = request.form['userName']
6     params = request.form.to_dict()
7     return json.dumps(json)

```

Listing 8: 获取 URL 中的参数

1.5.5 session 处理

略。

1.5.6 cookie 处理

略。

1.5.7 文件上传处理

略。

1.5.8 数据库连接池

首先需要通过 `pip install DBUtils` 模块，该模块提供数据库连接池支持。我们定义数据库操作类 `model/m_mysql.py`，在其中定义数据库增、删、改、查操作，如下所示：

```
1 import pymysql
2 from DBUtils.PooledDB import PooledDB
3
4 RD_DBP = PooledDB(
5     creator=pymysql, # 使用链接数据库的模块
6     maxconnections=6, # 连接池允许的最大连接数，0和None表示不限制连
    接数
7     mincached=2, # 初始化时，链接池中至少创建的空闲的链接，0表示不创
    建
8     maxcached=5, # 链接池中最多闲置的链接，0和None不限制
9     maxshared=3, # 这个参数没多大用，最大可以被大家共享的链接
10    # 链接池中最多共享的链接数量，0和None表示全部共享。PS：无用，因为
    pymysql和MySQLdb等模块的 threadsafety都为1，所有值无论设置为多少，
    _maxcached永远为0，所以永远是所有链接都共享。
11    blocking=True, # 连接池中如果没有可用连接后，是否阻塞等待。True
    ，等待；False，不等待然后报错
12    maxusage=None, # 一个链接最多被重复使用的次数，None表示无限制
13    setsession=[], # 开始会话前执行的命令列表。如：["set datestyle
    to ...", "set time zone ..."]
14    ping=0,
15    # ping MySQL服务端，检查是否服务可用。# 如：0 = None = never, 1 =
    default = whenever it is requested, 2 = when a cursor is created, 4
    = when a query is executed, 7 = always
16    host='127.0.0.1',
17    port=3306,
18    user='quant',
19    password='Quant2019',
20    database='QuantDb', # 链接的数据库的名字
21    charset='utf8'
22 )
23
24 WT_DBP = PooledDB(
25     .....
26 )
27
28 def get_rdb_connection():
29     return RD_DBP.connection()
30
31 def get_wdb_connection():
32     return WT_DBP.connection()
33
34 def close_db_connection(conn):
35     conn.close()
36
37 def query(sql, params):
```

```

38     conn = get_rdb_connection()
39     result = query_t(conn, sql, params)
40     close_db_connection(conn)
41     return result
42
43 def query_t(conn, sql, params):
44     cursor = conn.cursor()
45     cursor.execute(sql, params)
46     rowcount = cursor.rowcount
47     rows = cursor.fetchall()
48     cursor.close()
49     return (rowcount, rows)
50
51 def insert(sql, params):
52     conn = get_wdb_connection()
53     result = insert_t(conn, sql, params)
54     close_db_connection(conn)
55     return result
56
57 def insert_t(conn, sql, params):
58     cursor = conn.cursor()
59     affected_rows = cursor.execute(sql, params)
60     conn.commit()
61     cursor.close()
62     pk = cursor.lastrowid
63     return (pk, affected_rows)
64
65 def delete(sql, params):
66     conn = get_wdb_connection()
67     result = delete_t(conn, sql, params)
68     close_db_connection(conn)
69     return result
70
71 def delete_t(conn, sql, params):
72     cursor = conn.cursor()
73     affected_rows = cursor.execute(sql, params)
74     conn.commit()
75     cursor.close()
76     return (0, affected_rows)
77
78 def update(sql, params):
79     conn = get_wdb_connection()
80     result = update_t(conn, sql, params)
81     close_db_connection(conn)
82     return result
83
84 def update_t(conn, sql, params):
85     cursor = conn.cursor()
86     affected_rows = cursor.execute(sql, params)
87     conn.commit()
88     cursor.close()
89     return (0, affected_rows)

```

Listing 9: 获取 URL 中的参数

如??所示，定义了读、写数据库连接池，后缀为 _t 的函数用于数据库事务中使用。

2 数据处理

数据处理包括从 tushare.org 网站获取数据，将其转化为量化平台所需的数据格式。

2.1 获取沪深市场所有挂牌股票

获取在沪深两市挂牌的所有股票的基本信息。

2.1.1 接口定义

获取股票基本信息接口为 stock_basic，其参数为：

Table 1: stock_basic 接口输入参数说明

名称	类型	必选	描述
is_hs	str	N	是否沪深港通标的，N 否 H 沪股通 S 深股通
list_status	str	N	上市状态：L 上市 D 退市 P 暂停上市
exchange	str	N	交易所：SSE 上交所 SZSE 深交所 HKEX 港交所

返回值为：

Table 2: stock_basic 接口返回结果说明

名称	类型	描述
ts_code	str	TS 代码
symbol	str	股票代码
name	str	股票名称
area	str	所在地域
industry	str	所属行业
fullname	str	股票全称
enname	str	英文全称
market	str	市场类型（主板/中小板/创业板）
exchange	str	交易所代码
curr_type	str	交易货币
list_status	str	上市状态：L 上市 D 退市 P 暂停上市
list_date	str	上市日期
delist_date	str	退市日期
is_hs	str	是否沪深港通标的，N 否 H 沪股通 S 深股通

调用格式为：

```

1 import tushare as ts
2 pro = ts.pro_api()
3 data = pro.stock_basic(exchange='', list_status='L',
4                         fields='ts_code,symbol,name,area,industry,list_date')
5 data = pro.query('stock_basic', exchange='', list_status='L',
6                 fields='ts_code,symbol,name,area,industry,list_date')
```

Listing 10: 获取股票基本信息

见 10 所示结果示例：

	ts_code	symbol	name	area	industry	list_date
0	000001.SZ	000001	平安银行	深圳	银行	19910403
1	000002.SZ	000002	万科A	深圳	全国地产	19910129
2	000004.SZ	000004	国农科技	深圳	生物制药	19910114
3	000005.SZ	000005	世纪星源	深圳	房产服务	19901210
4	000006.SZ	000006	深振业A	深圳	区域地产	19920427
5	000007.SZ	000007	全新好	深圳	酒店餐饮	19920413

Listing 11: 获取股票基本信息结果示例

2.1.2 区域信息

如代码 11 所示，地区是以字符串形式返回的。我们可能需要按地区来统计股票表现，因此需要将地区统计出来，放到单独的一个表中进行管理。

数据库设计 数据库结构表结构如下所示：

```
1 create table t_area(  
2     area_id int primary key auto_increment,  
3     area_name varchar(200)  
4 );
```

Listing 12: 地区表数据结构

信息处理 当我们读到返回结果的一行时，我们取出地区信息，然后查询 `t_area` 表中是否包含该地区，如果包含则返回对应的 `area_id`，否则将该地区添加到 `t_area` 表中，并返回其 `area_id`。

2.1.3 行业信息

数据库设计

信息处理

2.1.4 股票信息

接口定义

获取并处理数据

2.2 获取日线行情数据

3 量化模型

3.1 时间序列分析

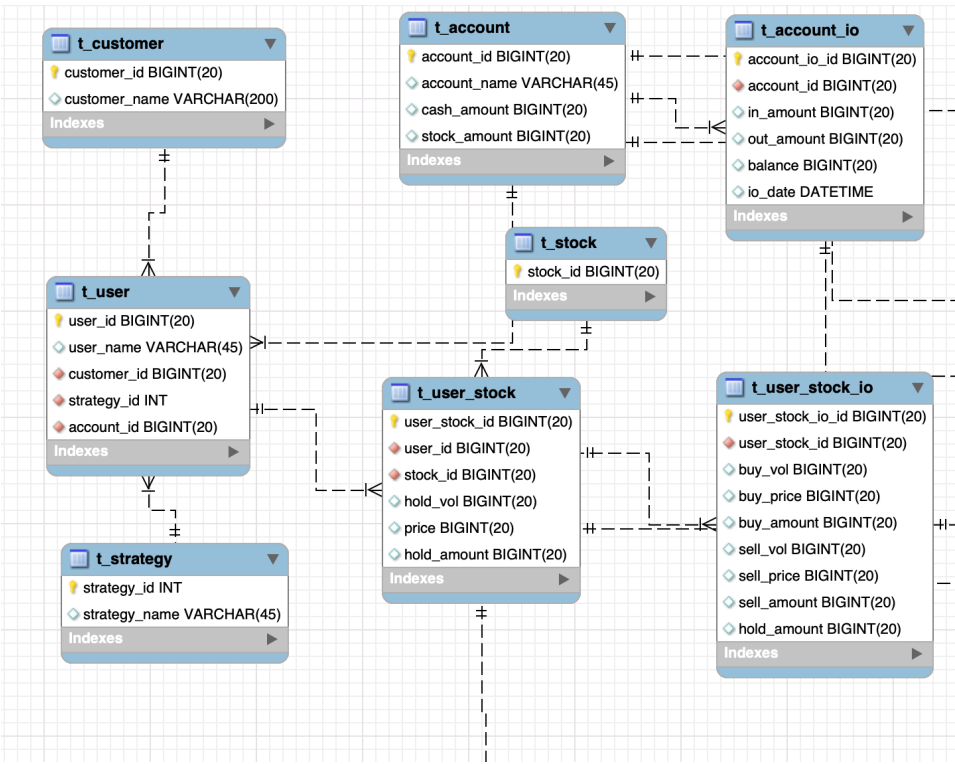
4 回测系统

4.1 数据库系统设计

客户是自然人，用 `t_customer` 表示。客户加上账户再加上量化策略，形成我们系统的用户，用 `t_user` 来表示。账户中具有现金资产和股票资产。账户具有资金的入和出，有股票的买入和卖出。用户持有一些股票，可以对股票进行买卖。用户可以买入和卖出指定数量股票，与账户资金变动相关联股票参数表：买入费率、印花税；卖出费率、印花税等，计入交易成本中。

4.1.1 ER 图

Figure 2: 数据库表 ER 图



4.1.2 客户表

表示自然人客户，结构如下所示：

Table 3: 客户表 (t_custome)

字段	名称	类型	描述
customer_id	客户编号	bigint	主键且自增长
customer_name	客户姓名	varchar	真实姓名

4.1.3 策略表

主要包括股票交易策略：包括 SVM、XGBoost、LSTM、ARIMA、GARCH 等，客户可以订购某个策略而成为我们的用户。

Table 4: 策略表 (t_strategy)

字段	名称	类型	描述
strategy_id	策略编号	bigint	主键且自增长
strategy_name	策略名称	varchar	真实姓名

4.1.4 账户表

客户拥有账户，每个客户对应的用户具有唯一的账户。账户中有现金和股票资产。

Table 5: 账户表 (t_account)

字段	名称	类型	描述
account_id	账户编号	bigint	主键且自增长
account_name	账户名称	varchar	易于记忆的名称
cash_amount	现金资产	bigint	以分为单位
stock_amount	股票资产	bigint	以分为单位

其有对应的历史表 t_account_hist，除上述字段外，还加上 hist_date 字段，用于记录每一天的资产。

4.1.5 账户流水表

显示用户现金账户资金进出情况，与股票流水表主键相同，用于表示股票买卖过程中资金的变化情况。

Table 6: 账户流水表 (t_account_io)

字段	名称	类型	描述
account_io_id	账户流水编号	bigint	主键且自增长
account_id	账户编号	bigint	t_account 表外键
in_amount	转入金额	bigint	以分为单位
out_amount	转出金额	bigint	以分为单位
balance	余额	bigint	以分为单位
io_date	发生日期	datetime	

4.1.6 用户股票表

用于表示用户当前拥有的股票。

Table 7: 用户股票表 (t_user_stock)

字段	名称	类型	描述
user_stock_id	账户流水编号	bigint	主键且自增长
user_id	用户编号	bigint	t_account 表外键
stock_id	股票编号	bigint	以分为单位
hold_vol	持有量	bigint	
price	价格	bigint	
hold_amount	转出金额	bigint	以分为单位
balance	余额	bigint	以分为单位
io_date	发生日期	datetime	

有对应的历史表，在上述字段基础上添加 hist_date 字段，用于记录每一天的资产。

4.1.7 用户股票流水表

记录用户股票买卖情况，主键与 t_account_id 相同，股票买卖是账户流水的一个子类。

Table 8: 用户股票表 (t_user_stock_io)

字段	名称	类型	描述
user_stock_io_id	账户流水编号	bigint	主键且自增长
user_stock_id	用户编号	bigint	t_account 表外键
buy_vol	买入量	bigint	手数
buy_price	买入价格	bigint	
buy_cost	买入成本	bigint	
buy_amount	买入金额	bigint	包括印花税、手续费、所得税等成本
sell_vol	卖出量	bigint	以分为单位
sell_price	卖出价格	bigint	以分为单位
sell_cost	卖出成本	bigint	包括印花税、手续费、所得税等成本
sell_amount	卖出金额	bigint	
hold_vol	持仓量	bigint	
hold_price	收盘价	bigint	
hold_amount	市值	bigint	

对应有历史表，除上述字段外，还有 hist_date 字段。

4.2 准备实验数据

创建一个新客户：

```
1 INSERT INTO 'QuantDb'. 't_customer' ('customer_id', 'customer_name')
VALUES (1, '王俊锋');
```

Listing 13: 创建新客户

添加一个新策略：

```
1 INSERT INTO 'QuantDb'. 't_customer' ('customer_id', 'customer_name')
VALUES (1, '王俊锋');
```

Listing 14: 创建新策略

创建账户并添加初始现金：

```
1 INSERT INTO 'QuantDb'. 't_account_io' ('account_io_id', 'account_id',
'io_date', 'in_amount', 'out_amount', 'balance') VALUES (1,
100000000, '2019-02-26', 100000000, 0, 1);
2 INSERT INTO 'QuantDb'. 't_account' ('account_id', 'account_name',
'cash_amount', 'stock_amount') VALUES (1, '股票投资账户', 100000000,
0);
```

Listing 15: 创建账户并入资

有了上述数据之后，我们就可以从 2018-01-01 开始，利用 SVM 模型，以收盘价为成交价，进行回溯研究了。

具体股票买卖逻辑为：选定某一支股票，利用 50% 现金购买该股票；每日利用 SVM 进行预测，如果涨则拿剩余现金的 10% 买入，如果跌卖出持有股份的 10% 买入：先将钱从账户现金中转出，减少账户现金资产；在股票流水中增加进项，然后增加股票持有量，将钱加入的账户的股票资产下卖出：先减少股票持有量，股票流水中转出资金，账户股票资金减少，账户增加资金进入，将现金增加到现金资产上交易的结算价格按每日收盘价计算绘制每日用户总资产（现金资产加股票资产）

5 深度学习入门

5.1 Tensorflow 底层技术

5.1.1 依赖控制

由于 tensorflow 相当于一门语言，我们在定义计算图时，相当于进行编程，这些程序只有在 session 中才会执行，而 session 中是根据依赖关系来决定执行哪些代码，所以必须通过依赖关系来告诉 tensorflow 执行哪些代码，如下所示：

5.2 线性回归

线性回归的定义：假设有一个问题，观察到的样本为 $\mathbf{x} \in R^n$ 且共有 m 个训练样本，同时每个样本 \mathbf{x}_i 对应一个数值 y_i ，并且我们假设其对应关系为： $y = \mathbf{w} \cdot \mathbf{x} + b$ ，整个问题可以表示为：

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{w} + \mathbf{b} \quad (1)$$

其中矩阵 $\mathbf{X} \in R^{m \times n}$ ，每一行代表一个样本。

我们的代价函数定义为最小平方误差函数：

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^m (y_i - \mathbf{w} \mathbf{x}_i + b_i) \quad (2)$$

这时我们的任务就变为：

$$\arg \min_{\mathbf{w}, \mathbf{b}} \frac{1}{n} \sum_{i=1}^m (y_i - \mathbf{w} \mathbf{x}_i + b_i) \quad (3)$$

由于这个问题比较简单，解这个问题有两种方法：解析法和迭代法。其中解析法的解为：

$$\begin{aligned} \bar{\mathbf{x}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i \end{aligned} \quad (4)$$

其解析解为：

$$\begin{aligned} \hat{\mathbf{w}} &= \frac{\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(y_i - \bar{y})}{\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2} \\ \hat{\mathbf{b}} &= \bar{y} - \mathbf{w} \bar{\mathbf{x}} \end{aligned} \quad (5)$$

如果是迭代法则为梯度下降算法，参数更新公式为：

$$\begin{aligned} \mathbf{w}_{t+1} &= \mathbf{w}_t - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{w}_t} \\ \mathbf{b}_{t+1} &= \mathbf{b}_t - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{b}_t} \end{aligned} \quad (6)$$

在讲解具体的代码之前，我们先来讲解一下 Python 中的 `__call__` 函数。我们知道，在 Python 中，类实例也是可调对象，只需要在类定义中添加 `__call__` 函数定义，即使用类实例加括号的形式进行调用了。我们利用这一特性定义线性回归类，如下所示：

```
1 class LinearRegression(object):
2     def __init__(self):
3         self.w = tf.get_variable('w',
4                                   dtype=tf.float32, shape=[],
5                                   initializer=tf.zeros_initializer())
6
7         self.b = tf.get_variable(
8             'b', dtype=tf.float32, shape=[],
9             initializer=tf.zeros_initializer())
10
```

```
11  
12     def __call__(self, x):  
13         return self.w * x + self.b
```

Listing 16: 线性回归类定义

参考文献: [I.MLearning \[1999\]](#)—[A.NikolaosAI \[1999\]](#)—[Bakry et al. \[2015\]](#)

References

Andreas Nikolaos A.NikolaosAI. *A Book He Wrote*. His Publisher, Erewhon, NC, 1999.

Amr Bakry, Mohamed Elhoseiny, Tarek El-Gaaly, and Ahmed M. Elgammal. Digging deep into the layers of cnns: In search of how cnns achieve view invariance. *CoRR*, abs/1508.01983, 2015.
URL <http://arxiv.org/abs/1508.01983>.

Ivan Marc I.MLearning. Some related article I wrote. *Some Fine Journal*, 99(7):1–100, January 1999.