# FactorFactory PythonProject源码分析

作者：

审阅者：

修改记录：

|  |  |  |
| --- | --- | --- |
| 版本 | 时间 | 备注 |
| V0.1 | 20190816 |  |

目录：

1. 文件简介
2. 基本概念
3. 数据结构
4. 函数和类
5. 问题说明
6. 附录

## 1. 文件简介

因子工厂包括两部分：因子检测模块，回测模块。

1.1因子检测模块

因子检测模块目前已实现：（1）时间序列IC检测；（2）横截面IC检测。

因子检测模块代码，时间序列相关代码在mainV0.1.py模块下，横截面相关代码在mainV1.1.py模块下，后续会将mainV0.1.py与mainV1.1.py合并精简。

1.2回测模块

回测模块，目前已实现（1）因子数据加工、对接聚宽平台。

## 2. 基本概念

说明整个项目文件涉及的总体思想、命名规范，尤其是在设计数据结构操作或算法时。

### 2.1总体思想（流程）

**数据处理部分**

（1）基础行情数据从TB下载，相同数据频率保存在同一文件夹。如，5分钟线数据保存在min5文件夹下、日线数据保存在day文件夹下。

目前信号加工基于指数数据（即AG000.csv、RB000.csv等等）。

（2）数据加工整理形成Multi\_Index\_Dataframe，后续操作基于这个大的Dataframe变量进行。

Dataframe变量形式如下：index索引有两层，level=0为日期Date、level=1为品种Symbol；colums依次为开盘价Open、最高价High、最低价Low、收盘价Close、成交量Volume（手）、持仓量OpneInterest（手）。

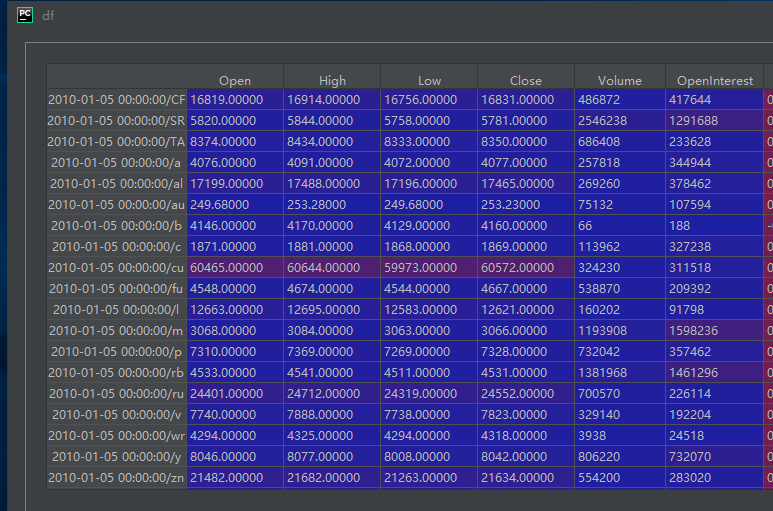


图1

**因子检测部分**

（3）基于大的Dataframe计算因子值，每个新增因子在df中增加一列。如下，新的CR因子、VV因子、DRV因子，在df中增加3列。



图2

注，如后续因子值增加、考虑采用dict存储。

（4）因子的计算，在dataframe中对列进行向量化计算，因子值计算函数为“cal\_因子名称”形式，如下图。

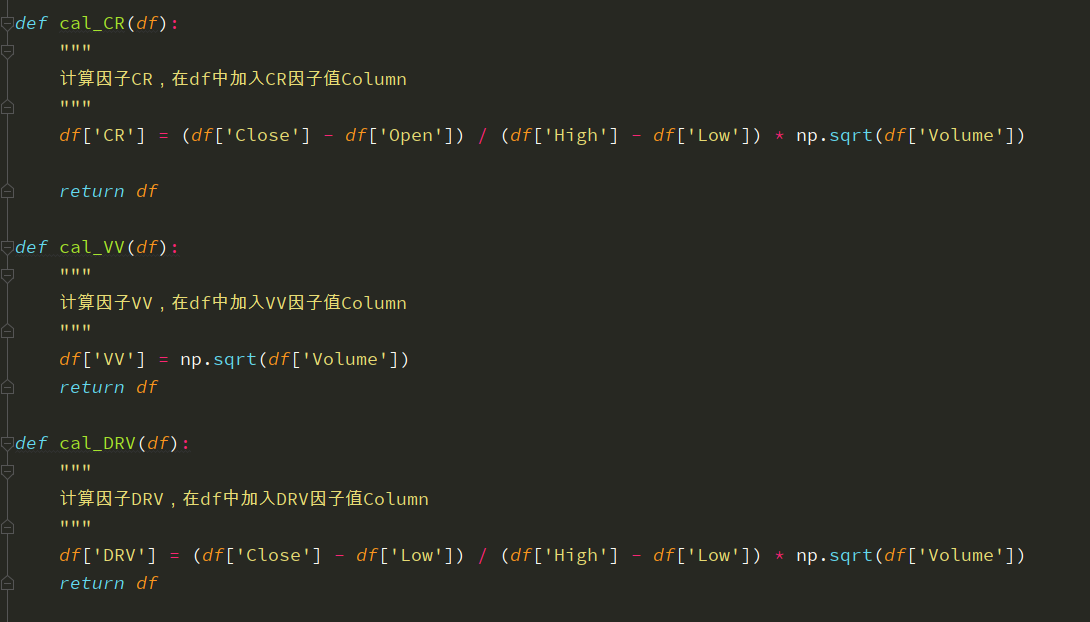


图3

（5）因子加工核心函数handle\_data\_CR\_factor（），主要实现数据载入及因子计算功能。计算新的因子，则在此函数内叠加新的因子计算函数。



图4

（6）基于含有因子值的大的dataframe变量（如图2），进行IC计算。IC计算分为时间序列和横截面两种，时间序列IC计算应用函数indicator\_IC（）；横截面IC计算应用函数indicator\_CS\_IC。

示例，CR因子在横截面（cross\_sectional）上的IC指标，计算结果如下

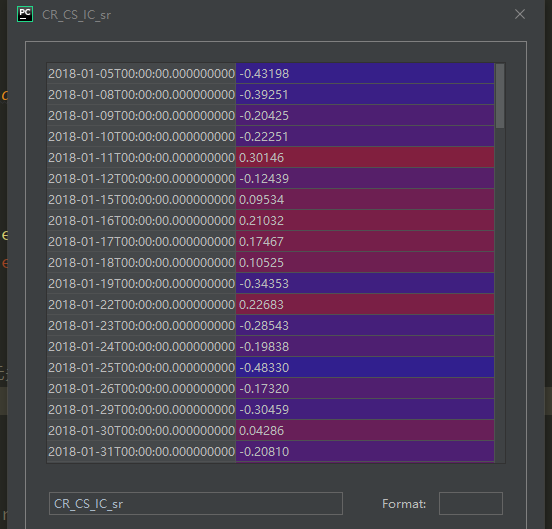


图5

**回测部分**

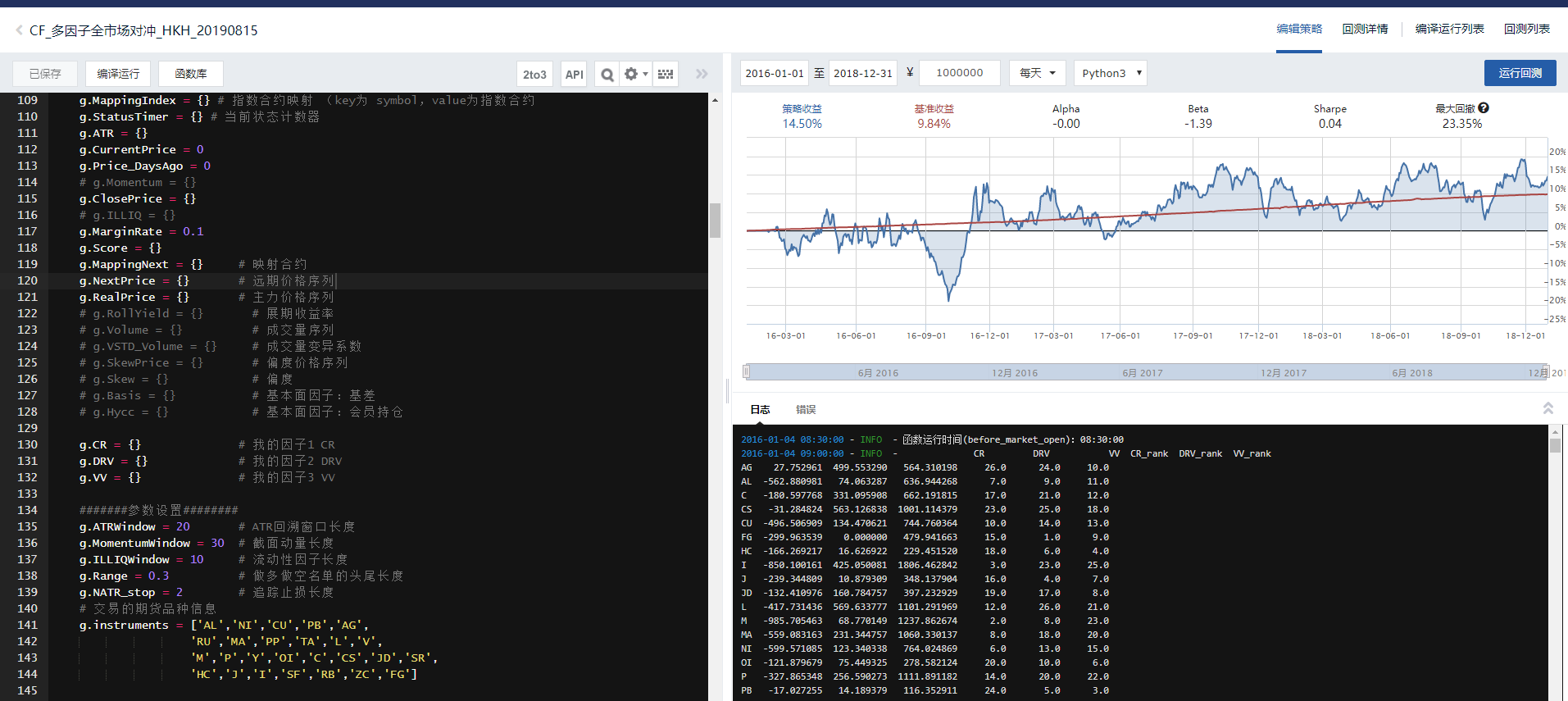
（7）通过mainV1.1.py中handle\_data\_for\_JQ（）函数，导出聚宽回测需要的因子数据。

示例代码：

df\_CR = handle\_data\_for\_JQ(df,'CR')

df\_CR.to\_csv('D:/df\_CR.csv')

（8）上传csv因子数据到聚宽研究环境，在聚宽平台上运行策略查看回测情况。回测文件为“聚宽\_CF\_多因子全市场对冲\_HKH\_20190815.txt”，拷贝到策略编辑环境内运行查看。



### 2.2模块、类、函数及变量命名规范

（1）采用“说明\_数据类型”的格式，如保存因子值的dataframe变量、命名为factor\_df。

## 3. 数据结构

列出主要的**Python包**，**模块**，**类**，**函数**和**关键变量**的数据结构。

## 4. 函数和类

### 4.1函数分类

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 分类 | 函数名 | 说明 |
| 1 | 核心函数 | handle\_data\_CR\_factor（） |  |
| 2 | 数据处理函数 | csv\_to\_df（） |  |
| handle\_data\_for\_JQ（） |  |
| 3 | 因子计算函数 | cal\_DRV（） | 计算因子值 |
| cal\_Return（） | 计算收益和下一期收益 |
| 4 | 指标计算函数 | indicator\_CS\_IC（） | 横截面IC计算 |
| indicator\_IC（） | 时间序列IC计算 |
| 5 | 功能函数 | Get（） | 获取可迭代Date、Symbol |
| concat\_func（） | 拼接功能函数 |

### 4.2函数说明

待完善

## 5. 问题说明

### 5.1后续优化

（1）数据清洗；

（2）数据载入优化；

（3）时间序列IC计算，因子参数优化n\rolling period；

## 附录一：

因子工厂设计框架：







基本面因子策略

套利因子策略

动量因子策略

|  |  |  |  |
| --- | --- | --- | --- |
| **动量因子策略** | | | |
|  | 时间序列（通道类） | 横截面（对冲类）排序 | 备注 |
| 因子检测模块  （1）不调用回测：IC  （2）调用回测：IR | 单品种IC | 横截面IC | Part1.因子检测指标（IC\IR）；  Part2.衡量方法（正态分布） |
| 回测模块 | （一般不涉及因子合成）；  0/1指标，驱动买卖；  止盈止损；  开平仓数量； | 因子合成；  因子排序；  资金权重分配 | Part3.组合构建包括：选股、权重分配；  Part4.择时包括：开平仓、止盈止损 |

|  |  |  |  |
| --- | --- | --- | --- |
| **套利因子策略** | | | |
|  | 时间序列（通道类） | ~~横截面（对冲类）排序~~ | 备注 |
| 因子检测模块  （1）不调用回测：IC  （2）调用回测：IR | 单品种IC | ~~横截面IC~~ | Part1.因子检测指标（IC\IR）；  Part2.衡量方法（正态分布） |
| 回测模块 | （一般不涉及因子合成）；  0/1指标，驱动买卖；  止盈止损；  开平仓数量； | ~~因子合成；~~  ~~因子排序；~~  ~~资金权重分配~~ | Part3.组合构建包括：选股、权重分配；  Part4.择时包括：开平仓、止盈止损 |