



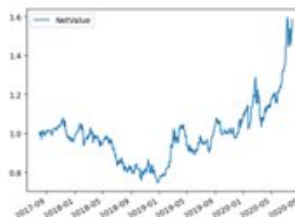
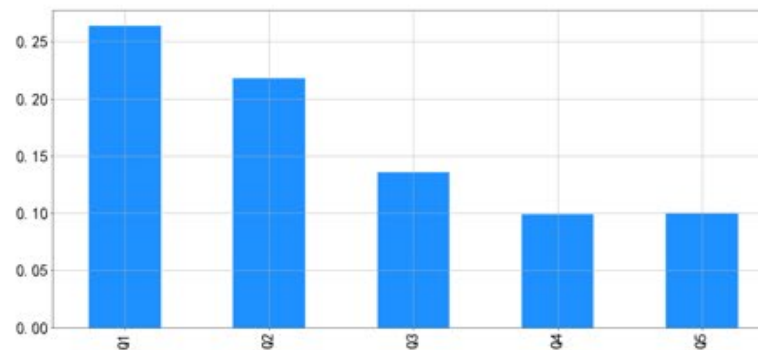
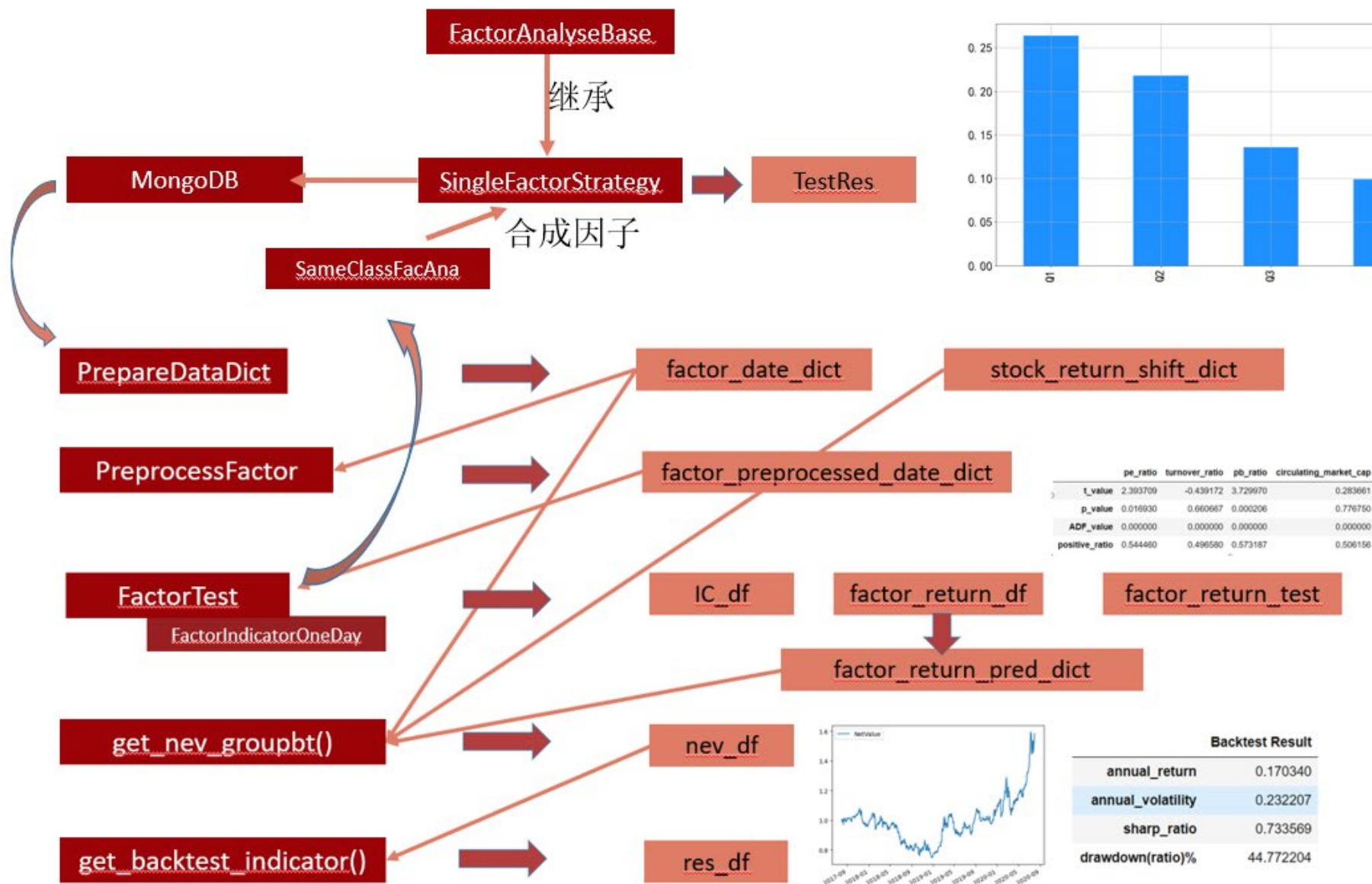
北京大学量化交易协会2020级培训

成果展示

Multifactor & ML

2020-11-23

流程说明





■ 数据库

数据库表与字段

- 日行情表(daily_quote): datetime, open, high, low, close, volume, code, name, raw_open, raw_high, raw_low, raw_close, raw_volume, adj_factor, paused, pre_close
- 分钟行情表(minute_quote): datetime, open, high, low, close, volume, code, name
- 股票基本信息表(stock_info): datetime, code, industry_zx1, name, listed_code, is_st, listed_date, expired_date, industry_zx1_code, industry_zx1_name
- 指数成分与权重(index_weight): datetime, index_code, code, weight
- 日因子表(daily_factor): datetime, code, pe_ratio, turnover_ratio, pb_ratio, ps_ratio, pcf_ratio, capitalization, market_cap, circulating_cap, circulating_market_cap, pe_ratio_ly



■ 数据获取

PrepareDataDict

- `get_factor_date_dict(cls, factor_list, start_date, end_date, index_name='000300.SH')`
- `get_stock_return_shift_dict(cls, start_date, end_date, index_name='000300.SH', shift_days=1)`

factor_date_dict - Dictionary (731 elements)

Key	Type	Size	Value
2017-08-01 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-02 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-03 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-04 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-07 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-08 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-09 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-10 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-11 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-14 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-15 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...
2017-08-16 00:00:00	DataFrame	(300, 4)	Column names: pe_ratio, turnover...

2017-08-01 00:00:00 - DataFrame

Index	pe_ratio	turnover_ratio	pb_ratio	listing_market
000001.SZ	8.3408	1.2033	1.0243	1867.75
000002.SZ	12.3792	0.2158	2.4564	2273.89
000008.SZ	41.0197	0.3581	3.4904	152.091
000009.SZ	83.4435	5.2214	4.4032	194.836
000060.SZ	48.6415	2.5508	3.0876	268.868
000063.SZ	-52.249	0.9717	3.2965	782.729
000069.SZ	10.7397	0.419	1.7283	564.654
000100.SZ	26.1424	3.3663	2.0632	344.752
000156.SZ	34.8129	0.2362	2.1026	145.864
000157.SZ	-194.262	0.9276	1.0242	299.443
000166.SZ	22.3868	0.7083	2.2667	436.665
000333.SZ	17.8074	0.542	4.4593	2607.01
000338.SZ	17.7159	2.160	1.7580	202.757



■ 数据获取

PrepareDataDict

- `get_factor_date_dict(cls, factor_list, start_date, end_date, index_name='000300.SH')`
- `get_stock_return_shift_dict(cls, start_date, end_date, index_name='000300.SH', shift_days=1)`

stock_return_shift_dict - Dictionary (731 elements)

Key	Type	Size	Value
2017-08-01 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-02 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-03 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-04 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-07 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-08 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-09 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-10 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-11 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-14 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-15 00:00:00	DataFrame	(300, 1)	Column names: return
2017-08-16 00:00:00	DataFrame	(300, 1)	Column names: return

2017-08-01 00:00:00 - DataFrame

Index	return
000001.SZ	0.00996377
000002.SZ	0.00683177
000008.SZ	-0.00922266
000009.SZ	0.00108814
000060.SZ	0.000822368
000063.SZ	0.00131464
000069.SZ	0.0206074
000100.SZ	-0.0156658
000156.SZ	-0.00953678
000157.SZ	-0.004158
000166.SZ	0.003367
000333.SZ	-0.00946832
000338.SZ	0.00285207



预处理

PreprocessFactor

- **deExtreme()**
- **neutralize()**
- **standardize()**

	pe_ratio	turnover_ratio	pb_ratio	circulating_market_cap
000001.SZ	8.8362	0.6008	0.9306	2588.7273
000002.SZ	7.9952	0.7548	1.6528	2607.5515
000063.SZ	35.6760	1.8267	4.3791	1361.6823
000066.SZ	68.8343	6.8555	6.9036	455.8013
000069.SZ	4.9086	0.8309	0.9017	504.5438
...
603799.SH	167.8822	6.2777	5.2066	461.2399
603833.SH	32.0198	0.2274	6.0645	527.0057
603899.SH	58.7608	0.4088	14.2976	601.4040
603986.SH	148.2462	2.6903	11.3513	922.7435
603993.SH	47.9185	1.6045	2.3831	777.2940

	pe_ratio	turnover_ratio	pb_ratio	circulating_market_cap
000001.SZ	-0.425896	-0.607188	-0.675929	1.188617
000002.SZ	-0.435147	-0.508527	-0.555971	1.202834
000063.SZ	-0.130658	0.178191	-0.103130	0.261888
000066.SZ	0.234084	3.399917	0.316191	-0.422280
000069.SZ	-0.469099	-0.459773	-0.680730	-0.385468
...
603799.SH	1.323611	3.029746	0.034318	-0.418173
603833.SH	-0.170876	-0.846409	0.176816	-0.368503
603899.SH	0.123275	-0.730194	1.544342	-0.312314
603986.SH	1.107615	0.731460	1.054958	-0.069621
603993.SH	0.004010	0.035837	-0.434668	-0.179472



■ 计算IC, factor return并检验

FactorIndicatorOneDay

- `get_IC1d()`
- `get_factor_return1d()`

FactorTest

- `get_IC_factorReturn_df()`
- `get_ICIR()`
- `factor_return_test()`
- `get_factor_return_pred_dict()`



■ 计算IC, factor return并检验

FactorIndicatorOneDay

- `get_IC1d()`
- `get_factor_return1d()`

FactorTest

- `get_IC_factorReturn_df()`
- `get_ICIR()`
- `factor_return_test()`
- `get_factor_return_pred_dict()`

IC_df

	pe_ratio	turnover_ratio	pb_ratio	circulating_market_cap
2017-08-01	-0.0316	0.262246	-0.0267056	0.0814291
2017-08-02	0.0687036	0.0477312	-0.0199637	-0.230348
2017-08-03	0.127288	-0.21538	-0.137755	-0.043083
2017-08-04	0.0940068	0.197804	0.144142	-0.0376303
2017-08-07	0.0342588	0.191294	0.0839037	-0.0799568
...
2020-07-27	-0.0201215	0.0482884	0.17743	0.0302028
2020-07-28	0.197471	0.261535	0.274566	-0.0633793
2020-07-29	-0.0375924	0.0945677	0.0369779	-0.0245057
2020-07-30	0.205474	0.319881	0.221404	-0.0265277
2020-07-31	0.0911606	0.223944	-0.0503561	-0.190525

factor_return_df

	pe_ratio	turnover_ratio	pb_ratio	circulating_market_cap
2017-08-01	-1.581429	13.124153	-1.336486	4.075139
2017-08-02	3.945660	2.741216	-1.146518	-13.228951
2017-08-03	6.133111	-10.377674	-6.637439	-2.075867
2017-08-04	5.122604	10.778704	7.854548	-2.050547
2017-08-07	1.818693	10.155222	4.454189	-4.244660
...
2020-07-27	-0.906093	2.174482	7.989890	1.360066
2020-07-28	3.938807	5.216642	5.476565	-1.264179
2020-07-29	-1.685743	4.240670	1.658187	-1.098903
2020-07-30	8.183833	12.740587	8.818310	-1.056574
2020-07-31	2.008211	4.933356	-1.109313	-4.197138



■ 计算IC, factor return并检验

FactorIndicatorOneDay

- get_IC1d()
- get_factor_return1d()

FactorTest

- get_IC_factorReturn_df()
- get_ICIR()
- **factor_return_test()**
- get_factor_return_pred_dict()

factor_return_test

	pe_ratio	turnover_ratio	pb_ratio	circulating_market_cap
t_value	2.393709	-0.439172	3.729970	0.283661
p_value	0.016930	0.660667	0.000206	0.776750
ADF_value	0.000000	0.000000	0.000000	0.000000
positive_ratio	0.544460	0.496580	0.573187	0.506156



■ 计算IC, factor return并检验

FactorTest

- get_IC_factorReturn_df()
- get_ICIR()
- factor_return_test()
- **get_factor_return_pred_dict()**

factor_return_pred_dict - Dictionary (731 elements)

Key	Type	Size	Value
2017-08-01 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-02 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-03 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-04 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-07 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-08 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-09 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-10 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-11 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-14 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-15 00:00:00	Series	(4,)	Series object of pandas.core.series mod...
2017-08-16 00:00:00	Series	(4,)	Series object of pandas.core.series mod...

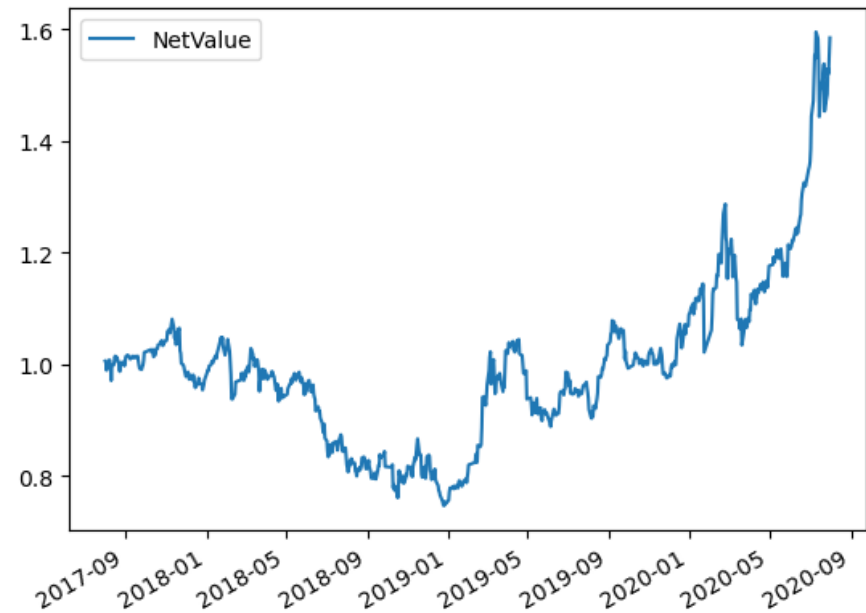
2017-08-01 00:00:00 - Series

Index	7-08-01 00:00
pe_ratio	-1.58143
turnover_ratio	13.1242
pb_ratio	-1.33649
circulating_market_cap	4.07514



分组回测

get_nev_groupbt()



get_backtest_indicator()

```
res_df
```

Backtest Result	
annual_return	0.170340
annual_volatility	0.232207
sharp_ratio	0.733569
drawdown(ratio)%	44.772204



困难与解决方案

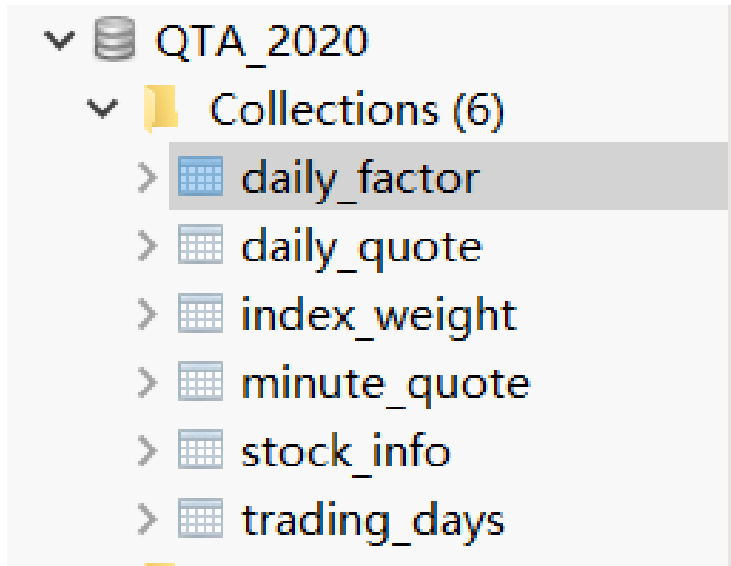
数据库被黑

```
View Document
81.68.114.243:27017 HOW_TO_RECOVERY_BASE README
{
  "_id" : ObjectId("5f9c5cd96203fbb5a43e7ca9"),
  "content" : "All your data is a backed up. You must pay 0.04 BTC to 1757buFtrykguUzQNQgUSrdoQuyEJEF6CW 48 hours for recover it. After 48 hours expiration we will leaked and exposed all your data. In case of refusal to pay, we will contact the General Data Protection Regulation, GDPR and notify them that you store user data in an open form and is not safe. Under the rules of the law, you face a heavy fine or arrest and your base dump will be dropped from our server! You can buy bitcoin here, does not take much time to buy https://localbitcoins.com with this guide https://localbitcoins.com/guides/how-to-buy-bitcoins After paying write to me in the mail with your DB IP: recoverybase@cock.li\n"
}
```

原因：云服务器数据库服务端口暴露在公网。

解决方法：禁止数据库端口接入，通过SSH端口映射到本地来连接数据库。

数据来源



解决方法：神秘的暑期实习力量、卑微爬取JQ

■ 下一阶段任务

- 补充分钟行情数据
- 补充财务报表数据
- 丰富因子库
 - 行情
 - 财务分析
- 研究因子合成
 - 传统方法
 - **ML**方法
- **ML**模型预测收益



北京大学量化交易协会