

dataPredict项目总结

###项目回顾

这次这个项目我们是采取的一开始小组合作完成任务，之后再独立编写代码完成任务。小组阶段时我们分为了如下几个步骤：小组讨论确定预测目标目标label -> 确定features -> 确定loss function -> 小组分工 之后肖总发布了自己写的一个初步的框架之后，我发现我们搭建的框架有很多不足。我就在肖总的框架下，又重新写了一份代码。

- 原本框架

按照流水线的思想来设计框架，这样灵活度较低。与肖总的框架相比最明显的区别就是在feature的灵活度方面有明显的差距。

```
processor
dataGenerator.py
learningModel.py
picDrawer.py
preProcessor.py
run.py
util.py
```

- 新框架

这个框架OOP的设计比我搭建的框架更加明显，职责分配很明确。我之前设计框架的时候就是没有专门特征处理IndicatorGallexy的模块。导致preProcessor模块与dataGenerator模块的职责耦合性有点大。

```
fastResearchData.py
indicatorGallexy.py
main.py
model.py
modelEngine.py
util.py
```

- 本次新尝试的东西

Argparser: 让你写的python项目可变的参数由外部传入，并且他让可以通过阅读你的--help使用你的软件

Unit test: 老实说以前每个模块下面写个if __name__ == "__main__"进行单元测试是没有用过的，发现特别好用

sphinx : 用于生成文档的一个工具，虽然没用成功。

小组讨论 : 吸取了上次项目失败的教训。本次项目开始时花了较多时间在讨论上。并且带着她们搭建了框架，所以她们对框架的理解也比较清晰。之后开展工作也更加方便。

- 待改进

- 特征选取模块 — 选取方法，速度问题
- 特征处理优化 — 正则化，归一化等函数的设计