**第二届易观算法大赛——性别年龄预测运行说明**

**团队：江湖交流**

**队员： Fred flyfoxs**

一 运行过程简介

本程序由队员fred和队员felix分别运行各自程序，后将产出结果加权平均，形成最终结果进行提交。过程如下：

（一）fred程序运行过程

1. 进入code\_fred目录
2. 执行prepare\_data.py产生特征文件
3. 执行first\_layer\_traing.py，产出第一层模型输出，作为第二层模型输入
4. 执行second\_layer\_traing.py，产出第二层模型输出。

（二）felix程序运行过程

1. 进入code\_felix/model/目录
2. 分别运行dnn.py、xgb.py、lgb\_raw.py、sex/\*.py ，作为第一层模型，产出数据，作为第二层模型输入
3. 进入 code\_felix/merge 目录，运行 merge\_dnn\_merge.py，产出第二层模型输出

（三）结果合并过程

进入 code\_felix/merge 目录，运行 file\_merge.py，将 fred和felix产出的csv文件加权合并，产出最终结果进行提交

二 文件和目录结构介绍

input/ 大赛提供的原始数据和经过手工调整的数据

output/ 过程中的结果产出

code\_fred/ 存放 fred的代码

├──prepare\_data.py :特征准备

├──model\_set.py :存放使用的模型

├──first\_layer\_traing.py :第一层模型产出

├──second\_layer\_traing.py :第二层模型产出。结果在output/目录下

code\_felix/ 存放felix的代码

├── merge

│   ├── dnn\_merge.py :使用DNN对各个模型输出的h5文件做融合

│   ├── file\_merge.py :直接对输出结果做简单的加权融合

│   └── utils.py

├── model

│   ├── dnn.py

│   ├── lgb\_raw.py

│   ├── search

│   │   ├── lgb.py : 对LGB模型参数做随机搜索

│   │   └── xgb.py :对XGB模型参数做随机搜索

│   ├── sex

│   │   ├── dnn\_sex.py :使用DNN模型对sex 和age很不预测

│   │   ├── lgb\_age.py :使用LGB 对age做单独预测

│   │   ├── lgb\_sex.py :使用LGB 对sex做单独预测

│   │   ├── xgb\_age.py :使用XGB 对age做单独预测

│   │   └── xgb\_sex.py :使用XGB 对age做单独预测

│   └── xgb.py

├── tiny

│   ├── feature\_filter.py :随机的过滤一些不重要的维度

│   ├── group\_label.py :对APP分组做各种统计

│   ├── knn.py :使用KNN算法补齐APP缺失的分组

│   ├── lda.py :使用LDA算法对APP和分组做降维

│   ├── package.py :对APP的信息做各种汇总

│   ├── tfidf.py :对APP和分组计算TFIDF

│   ├── usage.py :对APP的使用做汇总

│   ├── util.py

│   └── word2vec.py :针对APP的使用信息, 对APP向量化

└── utils\_

├── util\_cache\_file.py

├── util\_date.py

├── util\_log.py

└── util\_pandas.py