

源码文档目录结构说明

data:存放原始数据

----train.txt: 训练集原始数据集 (比赛方提供)
----test.txt: 测试集原始数据集 (比赛方提供)
----stop.txt: 停用词信息 (选手提交文档中包含)
----map.txt: 地域信息编码表 (选手提交文档中包含)

feature:存放各选手自处理后的数据

----lzp: 存放user1处理后的数据
----ljh: 存放user2处理后的数据
----yyt: 存放user3处理后的数据

model: 存放各自选手run.py产生的model

----lzp: 存放user1的model
-----money:罚金模型
-----laws: 法文模型
----ljh: 存放user2的model
-----money:罚金模型
-----laws: 法文模型
----yyt: 存放user3的model
-----money:罚金模型
-----laws: 法文模型

result: 存放各自选手run.py产生的result

----lzp: 存放user1的结果
-----money:罚金
----- fajin.zi.jsonprob.tsv 罚金字级别预测中间结果
----- fajin.ci.jsonprob.tsv 罚金词级别预测中间结果
-----laws: 法文
----- fawen.zi.jsonprob.tsv 罚金字级别预测中间结果
----- fawen.ci.jsonprob.tsv 罚金词级别预测中间结果
----ljh: 存放user2的结果
-----money:罚金
-----jh_penalty_cnn_blending_prob.csv 罚金CNN预测中间结果
-----jh_penalty_dense_cnn_blending_prob.csv 罚金Wide&CNN预测中间结果
-----laws: 法文
-----jh_laws_cnn_blending_prob.csv 法文CNN预测中间结果
-----jh_laws_wide&cnn_blending_prob.csv 法文Wide&CNN预测中间结果
----yyt: 存放user3的结果 -----money:罚金
-----textCNN(SoftMax)_all_prob_blend.csv 罚金1预测中间结果

-----textCNN(SigMoid)_all_prob_blend.csv 罚金2预测中间结果
-----textCNN(SoftMax)_9_prob_blend.csv 罚金3预测中间结果
-----laws: 法文 -----textCNN_laws_all_prob_blend.csv 法文1预测中间结果
----final.fajin.tsv 融合后罚金中间结果
----final.fawen.tsv 融合后法文中间结果
----final.all.json 融合最终结果

code: 存放各自的源码

----lzp: 存放user1的源码
-----run.py: 执行user1源码
----ljh: 存放user2的源码
-----run.py: 执行user2源码
----yyt: 存放user3的源码
-----money:罚金模型源码
-----laws: 法文模型源码
-----preprocess.py: 数据预处理源码
-----run.py执行user3源码
----blend.py执行融合各个用户结果复现源码

复现源码文档执行源码说明

1. 执行各选手下源码目录下的run.py获得各个选手的模型和结果
 - . 执行src目录下的blend.py获得融合结果
 - . 融合结果的最终文件命名为: /result/final.all.json

复现源码执行环境（主要Package）说明

1. Ubuntu14.04+Python2.7
 - . Keras 2.0.9
 - . tensorflow 1.4.0
 - . numpy 1.13.3
 - . pandas 0.20.3
 - . h5py 2.2.1
 - . jieba 0.38
 - . tqdm 4.19.4
 - . genism 0.13.3

