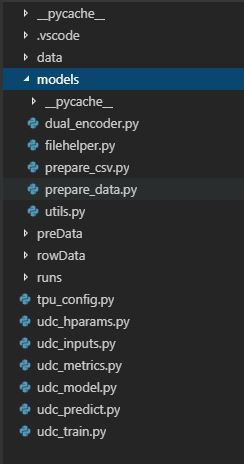
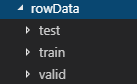
## 代码结构以及流程说明

1. 整体结构



1. 文件说明
   1. Tpu\_config.py , 返回是否使用tpu的配置
   2. Udc\_hparams.py，程序的主要配置信息
   3. Udc\_input.py， 构造CNN算法模型的输入方法
   4. Udc\_metrics.py, 旧版本遗留代码，程序改造为tpu模式后，已经无用
   5. Udc\_model.py, 构造算法模型
   6. Udc\_predict.py, 预测方法
   7. Udc\_train.py, 训练方法
   8. Models
      1. dual\_encoder.py， 定义模型中具体使用的算法
      2. filehelper.py, 读写文件的帮助方法
      3. prepare\_csv.py, 从原始聊天记录制作csv文件的入口方法
      4. prepare\_data.py, 从csv文件制作tfrecords文件的方法
      5. utils.py, 从原始聊天记录制作csv文件的具体实现方法
2. 数据文件夹
   1. rowData文件夹下有三个子文件夹，分别为test,train,valid。用来存放原始聊天记录，分别用于测试，训练，验证。



* 1. data文件夹，将csv文件制作为tfrecords文件时，tfrecords文件以及相关vacab文件的存放文件夹
  2. preData文件夹，预测数据存放的文件夹。预测数据分两个文件per.csv和canned.csv；

pre.csv是预测数据，canned.csv是ResponseSet

* 1. runs文件夹保存训练出来的模型

## 流程说明

1. 使用prepare\_csv.py,将原始的聊天记录生成csv文件
2. 使用prepare\_data.py,再将csv文件制作为tfrecords文件，同时生成字典文件vocab\_processor.bin和vocabulary.txt
3. 使用udc\_train.py，训练模型，输入数据为tfrecords文件，模型输出到runs文件夹
4. 使用udc\_predict.py进行预测