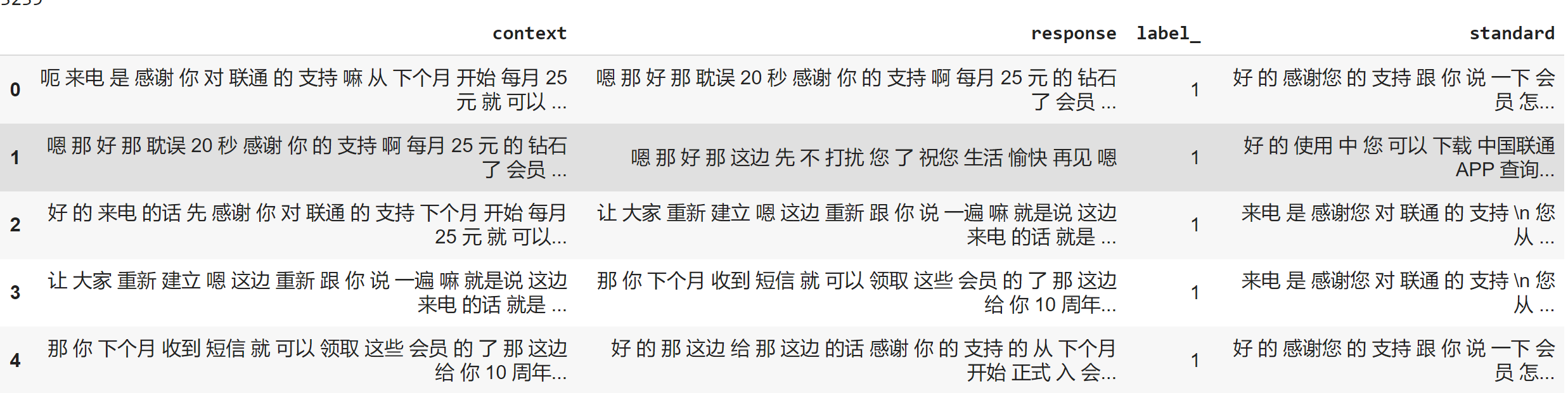
# TFIDF.ipynb

导入dataset和标准话术列表，使用TFIDF模型为dataset中每一行response匹配与其最相似的标准话术。



将标准话术加入dataframe中，并保存至”./data/tfidf.csv”



# Preprocess.ipynb

导入上一步中得到的tfidf.csv，将每条context各自与所有标准话术连接起来，并将错误的话术标注label为0。

对context和candidates进行分词和嵌入，并split成trian, test, validation三部分，用pickle打包成pkl文件，存在“./models-release-1.5/PaddleNLP/dialogue\_model\_toolkit/deep\_attention\_matching/data/data.pkl“

# context\_input.ipynb

生成测试数据：选择一条context，与已有的8条标准话术一一对应做成一个用于测试预训练模型Predict效果的dataframe。