本次最终的结果包含两个结果的融合，最终为0.745 ：0.6\*gbm + 0.4\*ffm

Gbm参数基本采用默认选项，学习率0.05，ffm采用deepfm，将dnn与fm融合，对于一条样本同一field出现多个特征值得情况进行求和，其中dnn采用三层网络结构，并采用dropout，fm部分隐向量维度为15。

1、gbm采用的有效特征，给出的用户Id特征，和广告id特征进行onehot编码，creativesize是数值特征，对用户特征与广告特征进行两种特征间的交叉，进行计数作为特征，将交叉特征中广告每个特征进行统计，计算比例参数，最终可达到0.738

2、deepfm基本采用原始id特征以及creativesize 特征输入，可达0.736

3、将两者按比例结合，达到0.745

期间尝试：

1、将label==1的进行计数，计算各特征对应比例。验证集计算时将验证集是训练集与验证集一起统计的数据，导致验证集标签泄漏，效果比较好，测试集效果较差。尽管验证集效果较好，但测试数据时提交答案很差，说明label构造特征不太好，有采用一些平滑，主要没有时间信息？

2、gbdt构造特征

没有采用分折数据，而是直接用训练器训练的模型进行叶子节点预测，效果不太行，gbdt输入得是稠密特征，而我输入了稀疏特征，但是纯稠密特征gbdt只能达到0.67左右的auc，

并未采用。

3、一开始先将label==1的同一个广告的所有topic,interst,kw分别进行lda建模，进行降维，建模的topic数进行多个参数进行测试，选择合适的参数。事实证明过拟合。最后将一个广告所有topic,interst,kw进行lda降维，效果仍比较差。

4、将广告中的层级关系进行统计，这些特征并没有提高分数，如advertiserid与aid等

5、采用ffm构造模型，0.73左右。Ffm更喜欢稀疏特征，对稠密特征处理并不太有效。因此将gbdt的特征输出输入到ffm或deepffm，效果不行。

6、采用了dmlc的pslite fm，可以调节batch大小，batch调的适当可以适当提分

7、将三阶交叉计数特征，效果并没有提高

8、正负样本1：10,进行采样，效果变差，改变了测试集分布？最终未采样

总结：

整个比赛过程一个人尝试了很多方法，但是很多时候一味的想当然，如造成一些特征穿越等。构造特征如何更加准确判别特征重要性，目前主要通过树的重要性以及验证集效果来看，需要更好的方法，如训练集与测试集某个特征的方差等一致？对于底层优化算法，感觉对实际比赛提升还是比较少。特征构造经验不够丰富，同时怎样构造有效特征以后要多锻炼。模型调参此次比赛基本没有尝试，这也是深度网络效果不能提高的问题所在，期间只是修改层数与添加dropout，这块需要进一步学习。路漫漫其修远兮，加油小蜗牛。