20890个词语。每个问题最短2个，最长40个，大部分小于15个。

3048个单字。每个问题最短3个，最长58个，大部分小于20个。

Train：pos占比0.519109

初赛 Test：pos loss=17.167525， neg loss=17.371649，

0：baseline。Tf-idf logistic regression 默认参数 基于词语

0.506701

1：Siamese rnn model on words seed8。BiLSTM+BiGRU, trainable=False, gradient clip, SpatialDropout1D, last state+gap+gmp concat，4 dense，epoch6 =0.311498

2: Siamese CNN+LSTM+Timedistributed on words seed 8. Copy the model: <https://www.linkedin.com/pulse/duplicate-quora-question-abhishek-thakur/> change the models to share weights between input 1 and 2. Trainable=False, gradient clip, concat 6 features. 5 dense 200d. Epoch4 =0.275762. epoch8 =0.248068 epoch10=0.253624（early stop）no stopwords

3: chars on 1

4: BIMPM.

5: Siamese cnn model on words seed8.

6：chars on 2 tranable

7：siamese GRU+cnn on words seed 8. maxpool+avgpool 【1,2,3,5】1 dense 300d trainable=True

8：chars on 5

9: chars on 7

10: chars on 4. Trainable=false

11: 7 LSTM seed 11 trainable=True

12: 11 LSTM+GRU seed 12 trainable=True no\_stopwords

13: 12 seed 13

14: siamese LSTM+GRU+cnn+attention on words seed 8. Feed forward attention+maxpool+avgpool 【1,2,3,5】1 dense 300d trainable=False no\_stopwords

15：12 seed 15 trainable=False

16: ESIM+feedforward on words seed 8. Trainable=false no\_stopwords

17: 16 delete feedforward

18: Decomposable attention (16 delete lstm)

19: 11 seed 19 no\_stopwords

20： 16 seed 20 include stopwords

21：20 seed 21

22: 20 seed 22

23:20 seed 23

24：7 seed 24 trainable=false

25: 11 seed 25 trainable = false

26: 7 seed 26 trainable=false no\_stopwords

27: 11 seed 27 trainable=false no\_stopwords

28:16 seed 28

Stack：

1st layer

7,11,12,16， 14（delete gru），17（add 1 lstm can improve to 0.18, but 5min every epoch）

2nd layer

Lr 0.186，xgb 0.170，linear svm （clip） 0.255，rf 0.167，knn 0.181 ，mlp 0.173

3rd layer

rf>xgb>mlp>lr