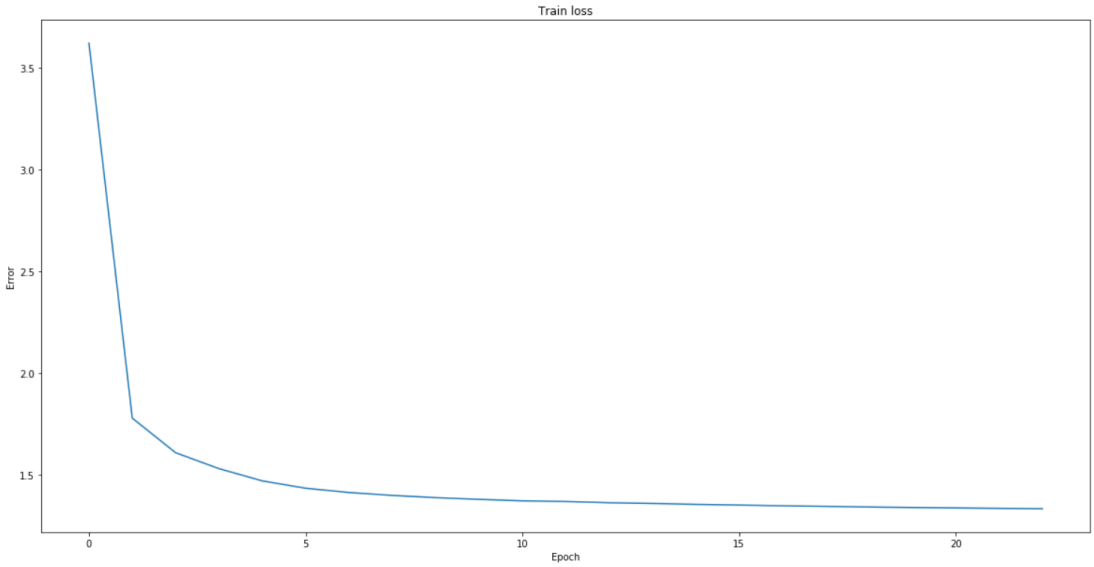
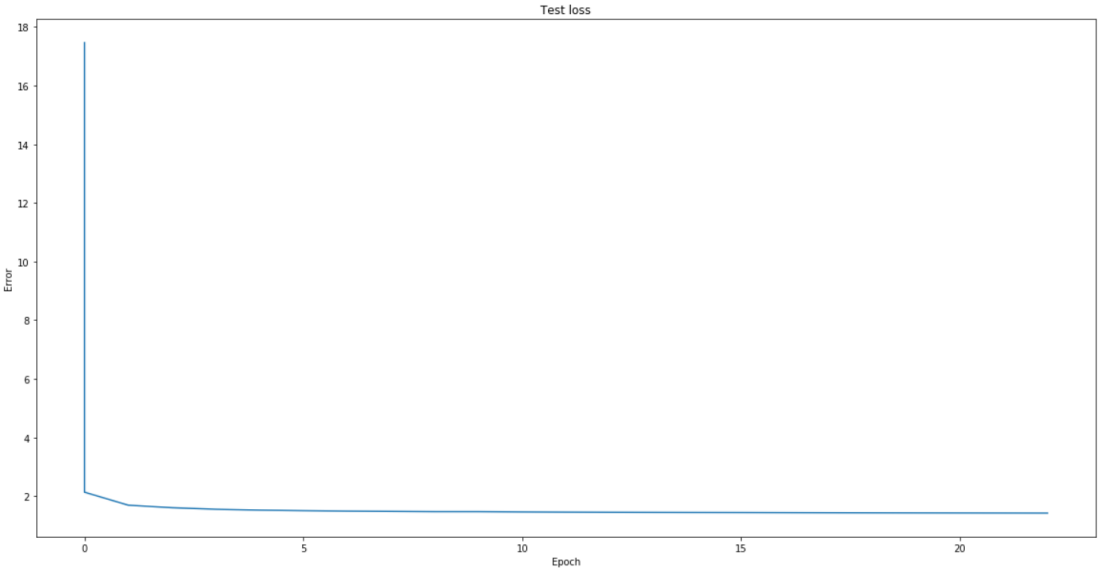
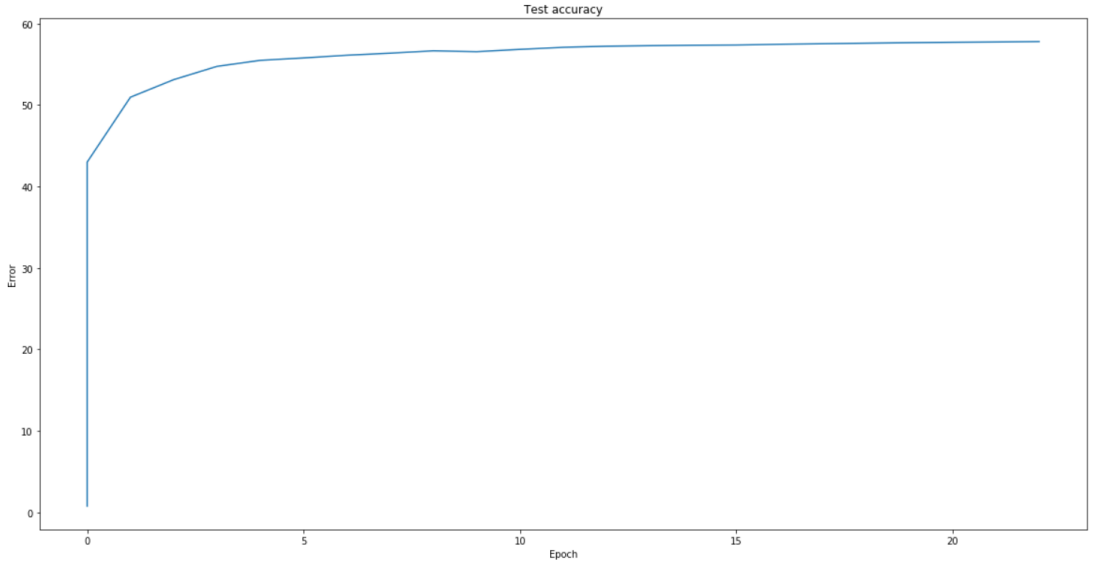
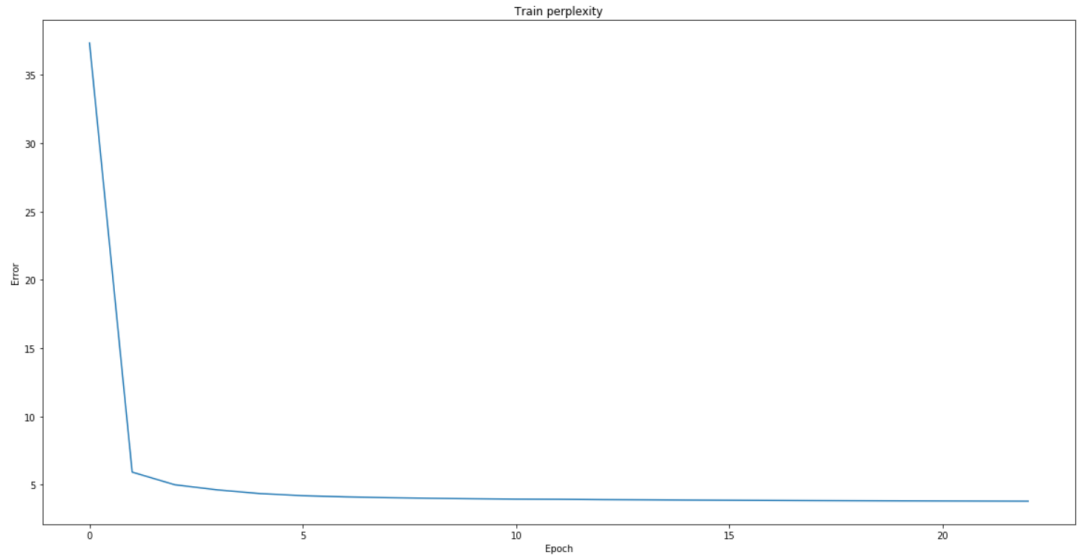
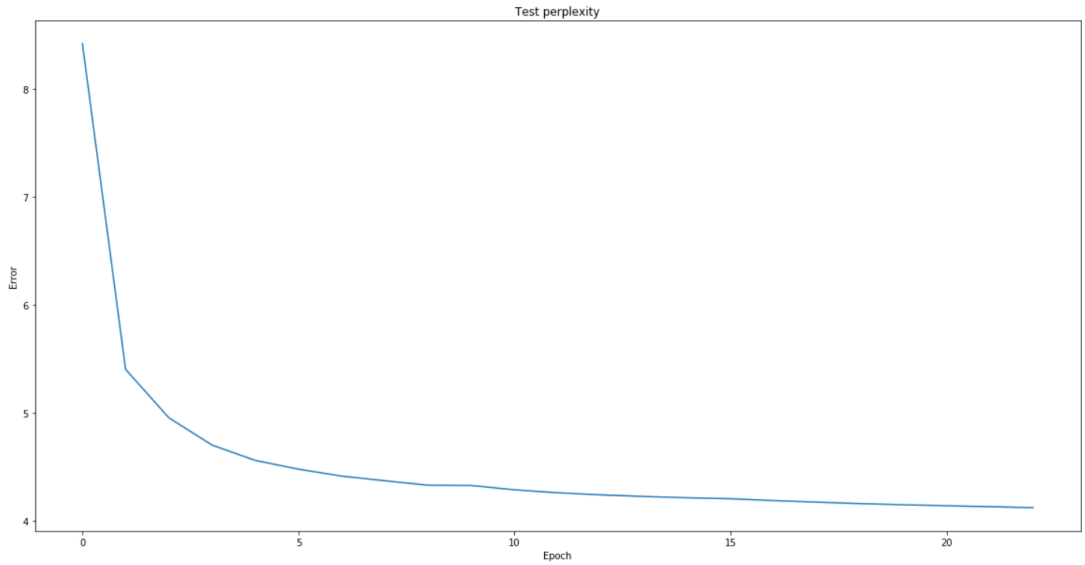
1.


Train loss
Error
Epoch
3.5
3.0
2.5
2.0
1.5
0
5
10
15
20


Test loss
Error
Epoch
18
16
14
12
10
8
6
4
2
0
5
10
15
20


Test accuracy
Error
Epoch
60
50
40
30
20
10
0
0
5
10
15
20





2. final accuracy is 58%, final test perplexity is 4.122
3. ***Sample***: Harry stopped and magging up beding bothroad room. "It didn't liev? "Oild Rin," said Scape. "A would one for Conkerfy netrest to could duel, theith 'lood moving herd quickly as he got up to the lesson, station les (Although some words are wrong)
   max: Harry stopped at the stone staircase to the sight of the corner of the corner of the corner of the corner of the corner of the corner of the corner of the corner of the corner of the corner of the corner of the co
   beam search: Harry stopped at the sight of the corner of the corner of the corner of the corner of the corner of the corner of the carriage and the corner of the corner of the wall and should have been to go and the first time to be somethi
   prompt: Harry stopped
4. Sample performs best. Because every time it can choose different word from the distribution. Instead, Max and beam search could produce many repeated words.
5. Sample performs better, beam search is really stupid that it can only repeat words.
   for T=0: Sample did worse, there are many wrong words and the sentence has no meaning. Beam search can produce good words, but repeat them so many times

0<T<1: Sample did better and better. (I think maybe 0.7 sample did best) Beam search still repeats a lot, it has no difference with max ,method.

T>1: Sample has less punctuations, words seems better, but sentence remains bad.

T<0: There are no obvious difference wrt T

New Corpus

1. I choose a famous fiction in China(Wrote in Chinese), character number: 5562048
2. Actually I didn't notice any obvious difference between the two corpus. Sample still looks good but make no sense. Max and beam search have the same output with words repeated.
3. (I am sorry that I use Chinese corpus, but I think it is really interesting!)

temperature: 0.7

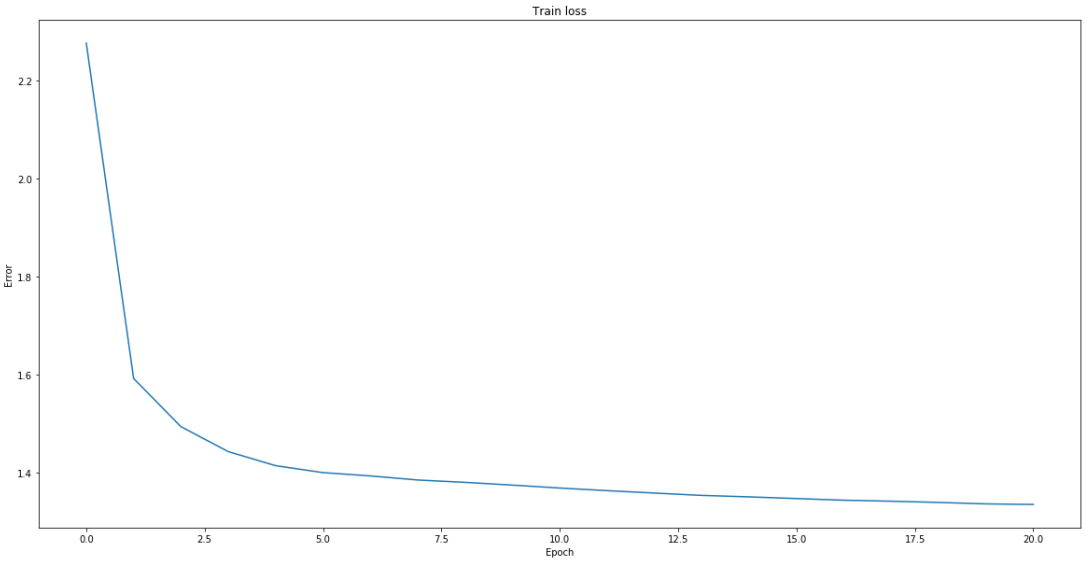
generated with max　　他们，这些人，都是在这一刻，一些人影，却是在这一刻，一股异常的劲气，却是犹如一股无形火焰般的暴涌而出，旋即一股磅礴斗气，便是犹如一股无形火焰般的暴涌而出，旋即一股磅礴斗气，便是犹如一股无形火焰般的暴涌而出，旋即一股磅礴斗气，便是犹如一股无形火焰般的暴涌而出，旋即一股磅礴斗气，便是犹如一股无形火焰般的暴涌而出，旋即一股磅礴斗气，便是犹如一股无形火焰般的暴涌而出，旋即一股磅礴斗气，便是犹如一股无形火焰般的暴涌而出，旋即一股磅礴斗气，便是犹如一股无形火焰般

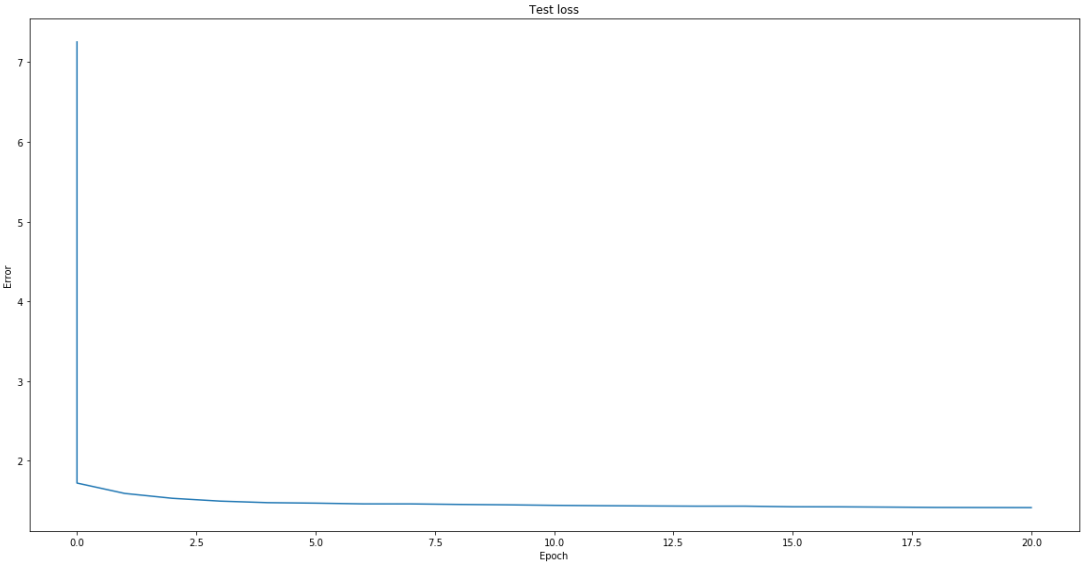
generated with sample　他这里召着不清楚地盯着这几个曹单“呵呵契，身么越毒立，再没有哥怀，然而他才手指嘭的双翼轻雷动。枫临死，所以“天先误阳，都是紫色不少……”第二侵骨骸进在某种区渐澈扰，萧玉深吸收一会，一股然后略微紧握了压起，“席调要来的，那我便你会让你泽去，传承极为的三人。”加上面币的名斗者也是听了他这二袖，男子，目光投黛处，原立在山脉之中生出一般悻事，一个两人缠赛，被这以其他的确实刚兜……啰会镀规的三名斗材也前不壬

generated with beam　他。“呵呵，我们，我们的确是有些不少。”“呵呵，小医仙，我们就不会再度出现，我们，我们就不会让你这些家伙的话，我们也不会再加玛帝国。”“呵呵，我们，我们也不知道。”“呵呵，小医仙，我们也不知道。”“呵呵，小医仙，我们也不知为什么？”“呵呵，我们就算是我们不知为何，我们的确是有些不少。”“呵呵，我们，我们也不知道，我们就算是我们，我们就算是我们的人，都是我们不知为何，我们这些家伙的话，我

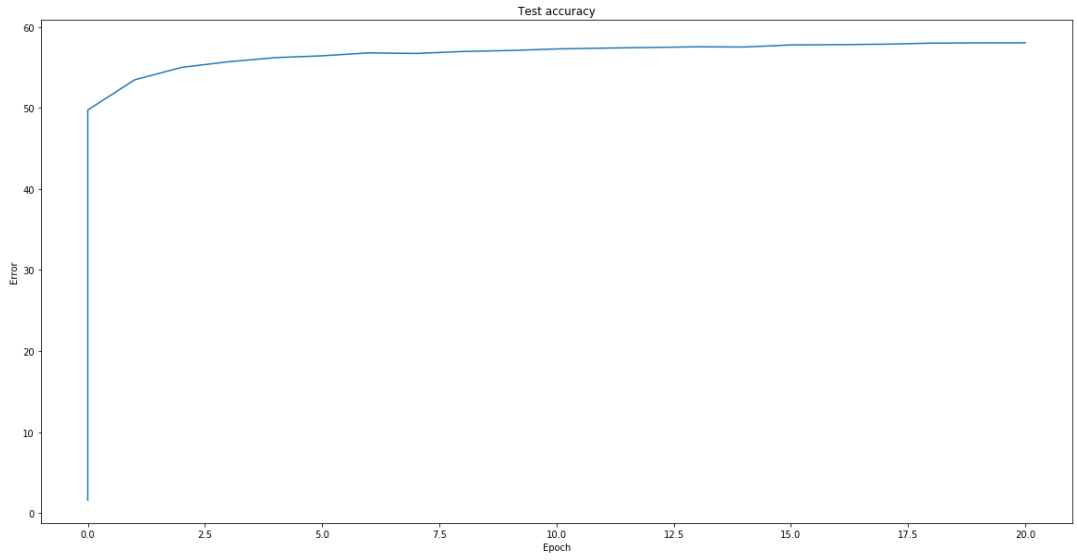
LSTM

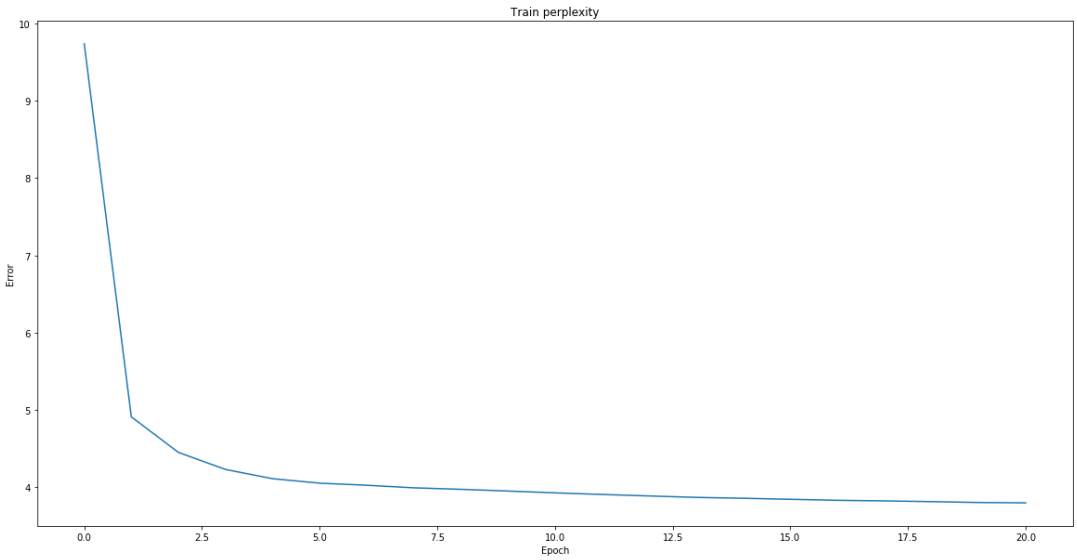
1. I didn't run into any difficulty, since I just changed the initial function and forward function, replace the gru with lstm.
2. it is a little better than GRU, 1 percent higher.

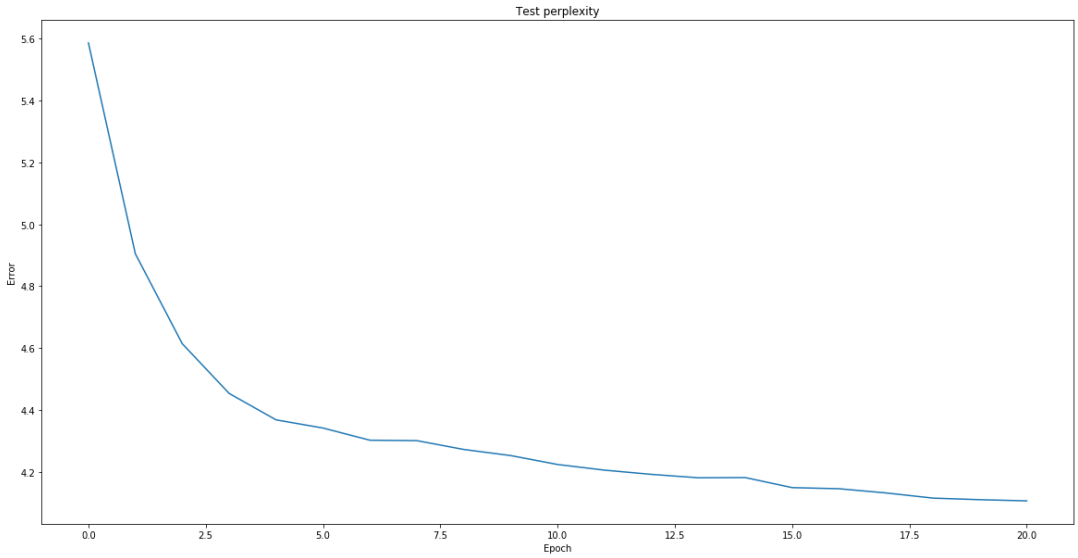

Train loss
Error
Epoch
2.2
2.0
1.8
1.6
1.4
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0


Test loss
Error
Epoch
7
6
5
4
3
2
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

.


Test accuracy
Error
Epoch
60
50
40
30
20
10
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0





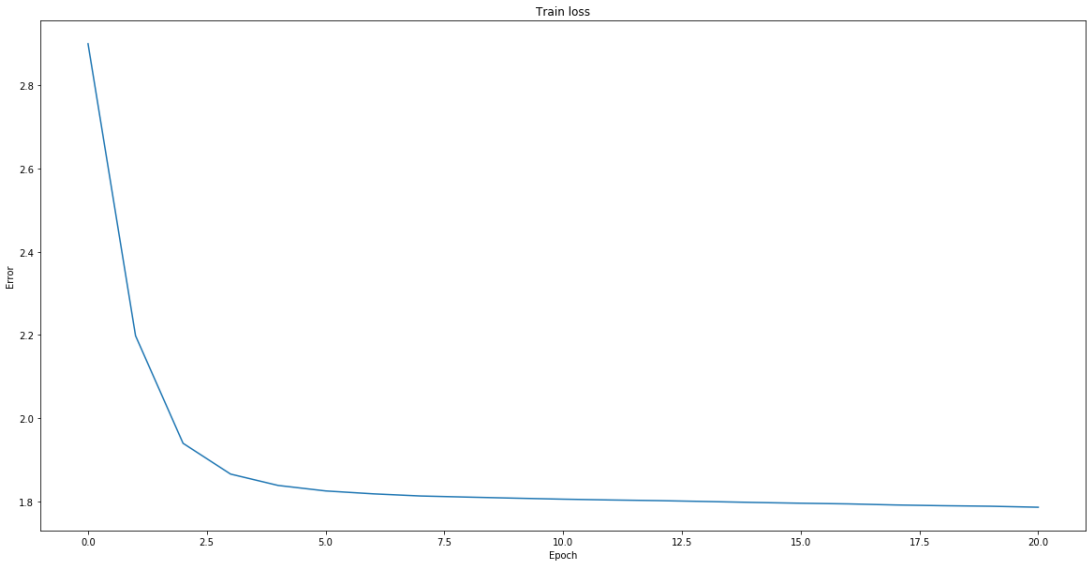
generated with max: His blue eyes were light, bright, and sparkling behind the corridor and the stand and started and the stand and started and the stand and started and the stand and started and the stand and started and the stand and started and the stand and started and the stand and started and

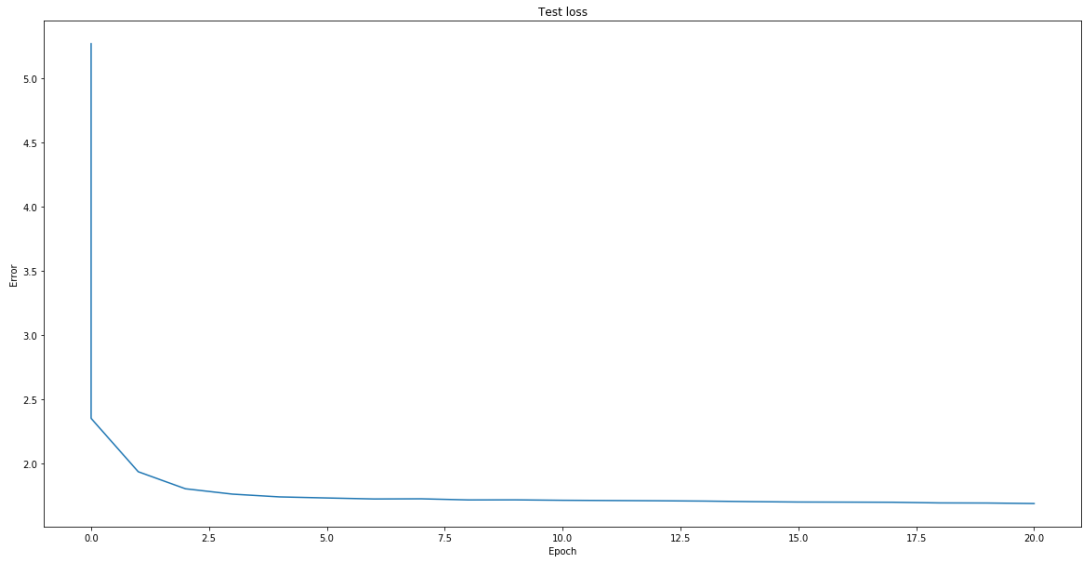
generated with sample: His blue eyes were light, bright, and sparkling behind its plint round away, from the enough Goody wand at to ~profele, they missed appoints. The finby our of SNED!I)EDET SYEUB?K !"S is fortstand. The kill storuz3be or that Blaston: He saw on the feed to

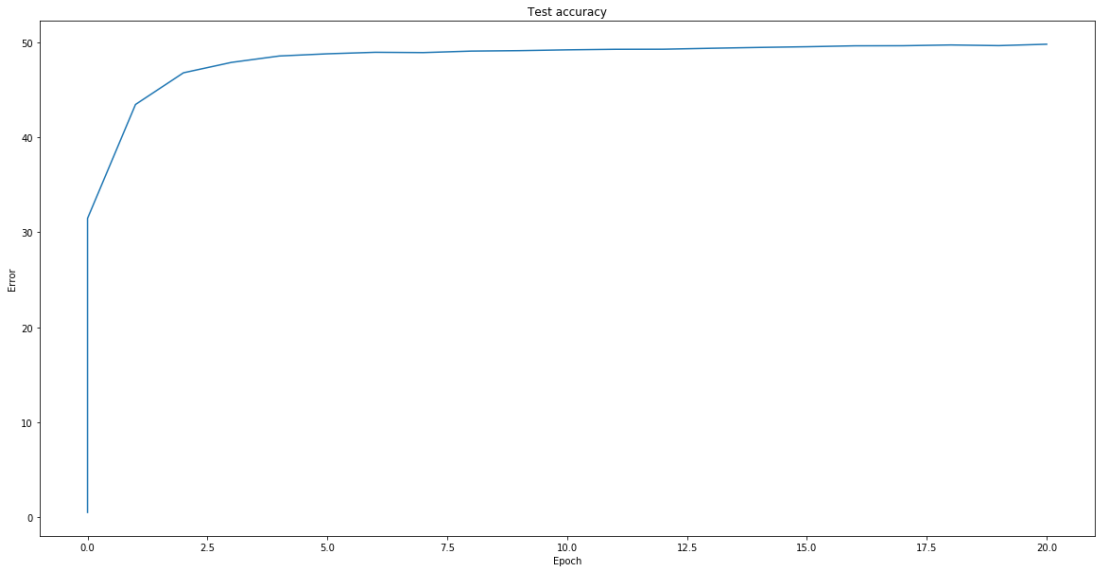
generated with beam: His blue eyes were light, bright, and sparkling behind the corridor and the stand and started and the stand and was a long of the corridor and the could be an everyone was a stand and was a long of the corridor and the could be an everyone was a long of

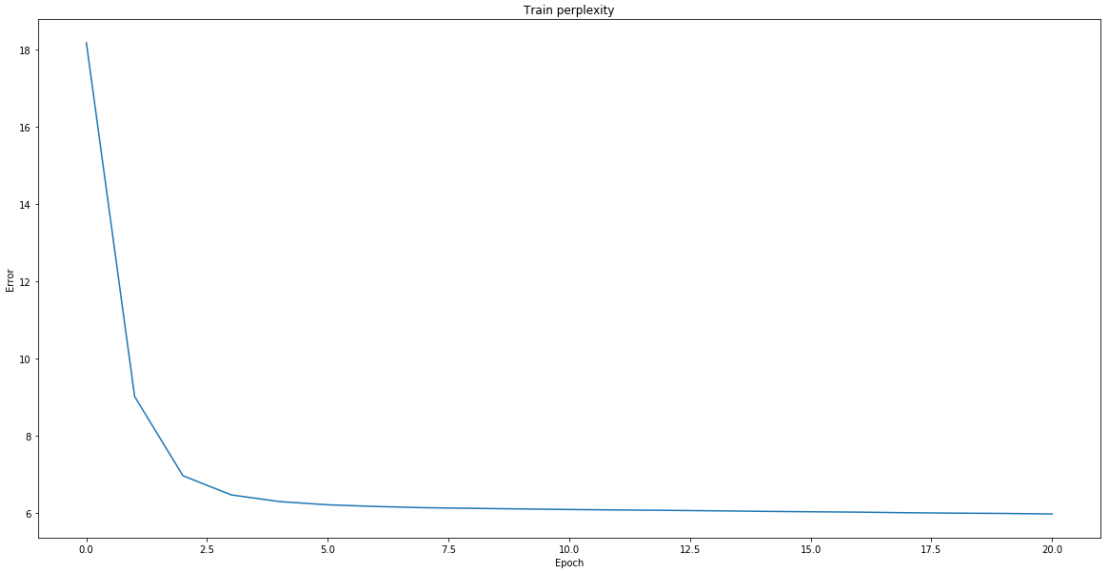
Transformer

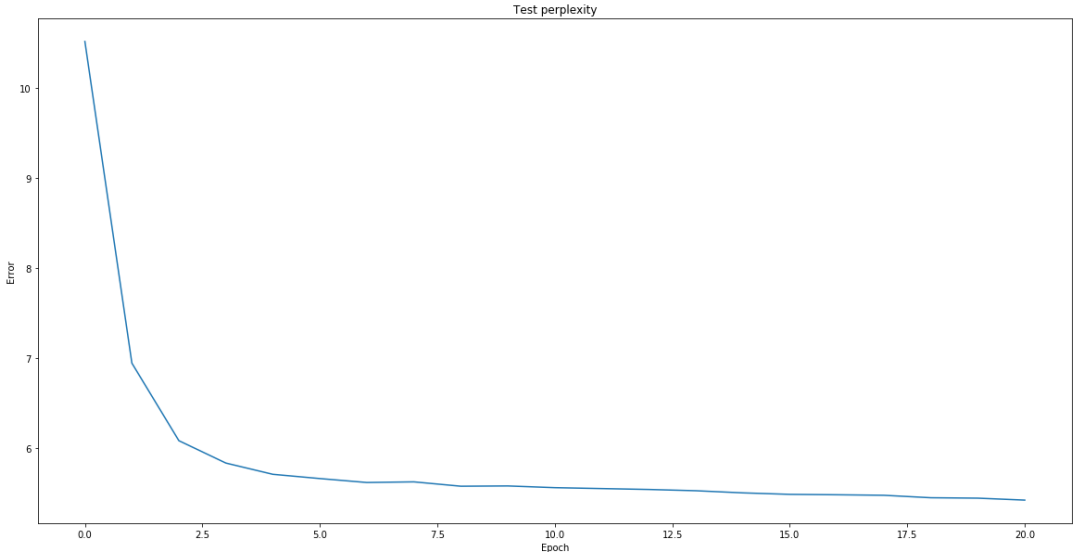
1. I am confused about lots of parameters in transformer, and it took me long time to figure out good parameters for training.
2. The accuracy is worse than GRU











3. generated with max: Harry Potter and the t t t

t t
generated with sample : Harry Potter and thedy ale feve,rorl!"OUSu ond iton, be us alsarer ilyos), s tos ulyinthe f timaly cely s u-: hit bel s baving nve, tilod he f ak ms uza l s ft nd fe onde now telet inihal the me Vo oporou f. tisit l f t
generated with beam: Harry Potter and thed hed the he he the he he hed hed thed he he ted hed thed he he he the ted he he he t hed hed he he he t he ted thed t hed he he he hed ted hed he t hed t hed he he te the hed he he the he te he hed t