Duc Minh Le

1. Data

Q1. (3 points) How many word tokens are there in train.en?
395598
Q2. (3 points) How many word types are there in train.en?
24403
Q3. (2 points) How many sentences are there in train.en?
46172
Q4. (2 points) What is the average sentence length (in words) in train.en?
42.1

2. NMT

Q5. (8 points) Show the English translations of the first 5 dev sentences in this format:
<source Vietnamese>
<human reference English translation>
<your system's English translation>
<blank line>
Câu chuyện này chưa kết thúc .
This is not a finished story .
This is not the same thing .

Nó là một trò chơi ghép hình vẫn đang được xếp .
It is a jigsaw puzzle still being put together .
It 's a very very important of that .

Hãy để tôi kể cho các bạn về vài mảnh ghép nhé .
Let me tell you about some of the pieces .
Let me show you a few things .

Ngôn từ , qua bao năm tháng là bạn đồng hành với ông , giờ quay ra chế giễu ông .
Words , for so long his friends , now mocked him .
<unk> , she was in a day , she was <unk> .

Ông rút lui vào yên lặng .
He retreated into silence .
He 's now .

Q6. (3 points) Look at the log file in your output directory. What is the best dev set perplexity your model could get?
dev ppl 31.66
Q7. (3 points) What is the BLEU score on the dev set?
dev bleu 6.6
Q8. (3 points) What is the test set perplexity?
test ppl 35.42
Q9. (3 points) What is the BLEU score on the test set?
test bleu 5.6

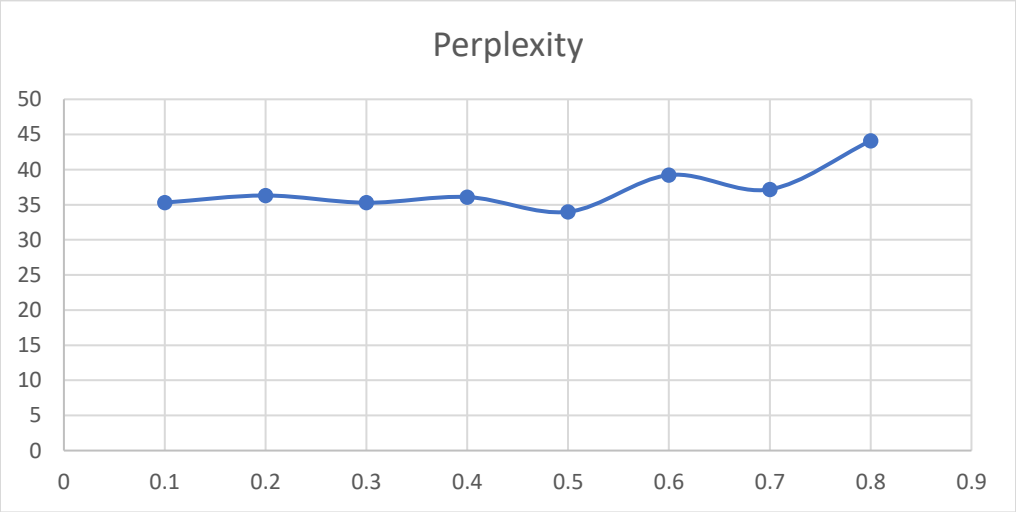
3. Hyper-parameters

Q10. (10 points) Create a results table that includes hyper-parameter values and their effect on test set perplexity and BLEU.

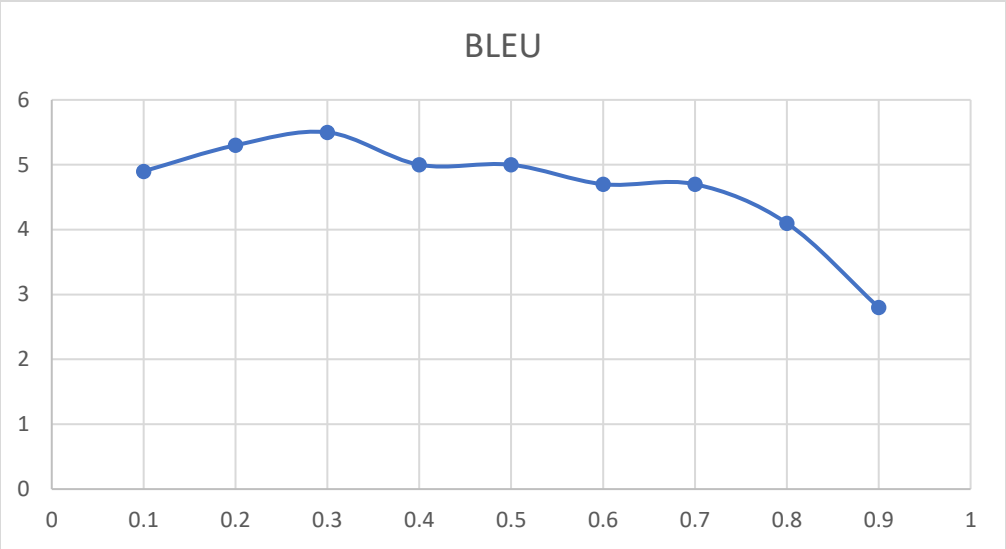
Dropout

Drop out	**Perplexity**	**Bleu**
0.1	35.3	4.9
0.2	36.31	5.3
0.3	35.28	5.5
0.4	36.09	5.0
0.5	33.99	5.0
0.6	39.22	4.7
0.7	37.18	4.7
0.8	44.1	4.1
0.9	54.52	2.8

Q11. (10 points) Plot hyper-parameter values (x-axis) vs. test set perplexity (y-axis).



Q12. (10 points) Plot hyper-parameter values (x-axis) vs. test set BLEU (y-axis).



Q13. (5 points) Did your results match your expectations? Briefly explain your findings.

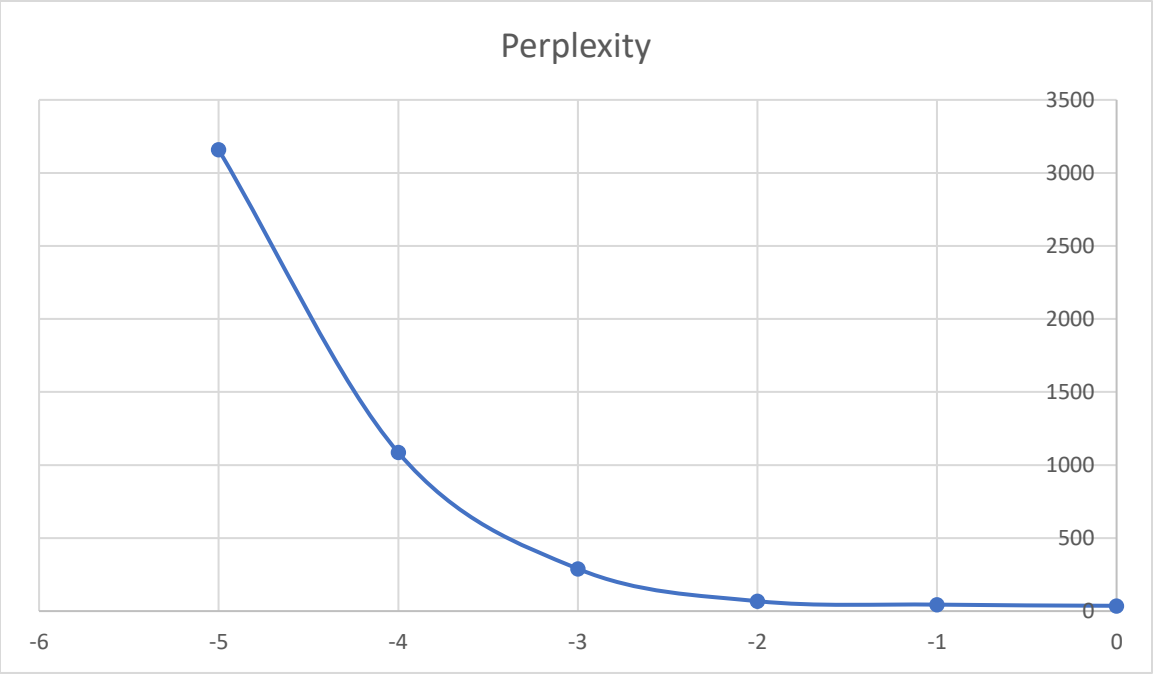
The result matched with my expectation about DropOut. DropOut forces components of NN to remember more information. However, if the DropOut rate is too high, the remaining active hidden units can't archive good performance because of the total number of active hidden units are so small.

4. Hyper-parameters

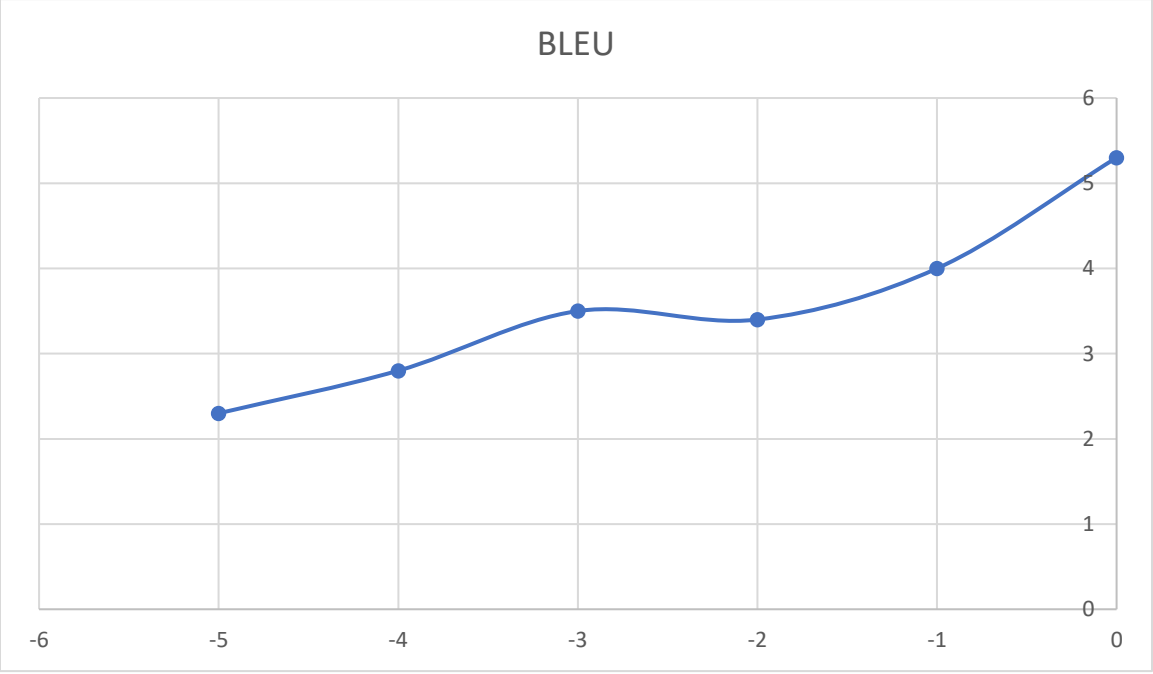
Q14. (10 points) Create a results table that includes different training data sizes and their effect on test set perplexity and BLEU.

Data Size	**Perplexity**	**Bleu**	**Log scale**
1	36.31	5.3	0
½	44.48	4.0	-1
¼	68.12	3.4	-2
1/8	289.95	3.5	-3
1/16	1086.13	2.8	-4
1/32	3158.47	2.3	-5

Q15. (10 points) Plot training data size (x-axis, log scale) vs. test set perplexity (y-axis).



Q16. (10 points) Plot training data size (x-axis, log scale) vs. test set BLEU (y-axis).



Q17. (5 points) How does training data size affect NMT performance? Briefly explain your results.

Training data size has a big impact on the NMT performance. The bigger training data size is, the better NMT model is (smaller perplexity and larger bleu values). Although the performance of NMT is not linearly increased with the size of the training data size, however the data size should be large enough to ensure the quality of NMT.