Language modeling report

Ziyang Li

1. Answers of 3 models

Questions	Unigram	Bigram	Bigram with smoothing
I don't know ____ to go out or not.	whether	whether	whether
We went ____ the door to get inside.	through	through	through
They all had a ____ of the cake.	peace	piece	piece
She had to go to ____ to prove she was innocent.	court	court	court
We were only ____ to visit at certain times.	allowed	allowed	allowed
She went back to ____ she had locked the door.	check	check	check
Can you ____ me?	here	hear	hear
Do you usually eat ____ for breakfast?	serial	both	cereal
She normally ____ with her mouth closed.	choose	both	choose
I'm going to ____ it on the internet.	sell	sell	sell

Table1

2. Probabilities of two different sentences from 3 models

Answers	Unigram	Bigram	Bigram with smoothing
weather/whether	6.3287931e-29 1.6455472e-29	4.0749376e-18 0.0	1.1131109e-28 5.5496871e-32
through/threw	3.1189960e-25 5.3036205e-24	0.0 2.9450239e-17	8.4162054e-31 5.0422618e-27
piece/peace	6.8534830e-22 3.6288740e-22	5.7341405e-20 3.0924635e-18	3.9090147e-28 1.0762597e-26
caught/court	3.5605624e-31 5.4877958e-32	2.2124009e-24 0.0	4.4385175e-36 1.6352960e-40
aloud/allowed	1.1913235e-27 2.9536946e-29	1.7969231e-22 0.0	2.4373735e-33 7.8939640e-37
cheque/check	5.1908078e-31 2.4116585e-32	1.8174693e-24 0.0	3.3150240e-37 6.7472704e-40
hear/here	4.3143609e-13 9.0389086-14	0.0 3.9204532e-12	2.0975554e-19 7.9189672e-18
serial/cereal	5.9955317e-27 1.3190169e-26	0.0 0.0	7.8456993e-34 1.9606786e-34

chews/choose	1.2965203e-25	0.0	2.8150486e-32
	1.1062460e-27	0.0	9.4180885e-33
cell/sell	1.5889823e-24	3.3995225e-16	4.7482399e-25
	7.8181582e-25	0.0	7.4459452e-29

Table2

3. Accuracy

	Unigram	Bigram	Bigram with smoothing
accuracy	60%	80%	90%

Table3

4. Discussion

According to data from table1,2&3, these results are expected. The bigram model with add-1 smoothing works better than unigram and bigram without smoothing and unigram model does not have good performance. The accuracy of bigram model with add-1 smoothing is high but it cannot choose every right answer even if the answer is obvious, I think model based on longer context may fix this problem. Also, the probabilities of some words in bigram model are zero, except for grammatical errors (e.g. here me) the remaining words got zero probabilities probably because training corpus is not enough.