How the size of n used in n-gram model affects the results?

We train bigram, trigram, 4-gram and 5-gram model based on the English training data, to study the effect of different models on the result. We use maximum likelihood estimation with add-one smoothing.

The experiment results are shown in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2 | 3 | 4 | 5 |
| Perplexity of test data based on the model | 11.29 | 8.87 | 8.05 | 9.79 |
| Training time | 0.10s | 0.16s | 2.17s | 68.47s |
| Size of model file | 16KB | 462KB | 14.7MB | 419.4MB |
| Line of model file | 900 | 26100 | 757770 | 21975330 |

The results indicates that 4-gram model has the lowest perplexity. Generally, the larger n is, the more information the model can record.