Tips for connLM Training

Wantee Wang

2015-08-15T22:59:54+08:00

Contents

1 Overfitting

When training corpus is large, the model is easy to be overfitted. I found a simple way to verify this. Getting two equal size sample texts from the head and tail of training corpus, then run `connlm-test` to calculate their PPLs respectively. If the two PPL value is far from each other, then the model is obviously overfitting by the later part of data.

Online version at http://wantee.github.io/2015/08/15/tips-for-connlm-training/