HMM project, Tim Ottens & Buck Boon 10549110

March 11, 2018

Abstract

In this paper we describe our experiments on Hidden Markov Models with Laplace Smoothing especially how we optimise the transition and emission α's

1 Introduction

This experiment tries to optimise the parameters that are used in a Laplace-Smoothed Hidden Markov Model. For Laplace smoothing two variables are used, the transition and emission α's. These are used to represent the unknown words that are encountered when using the language model.

We trained multiple models for different combinations of the parameters. transition α now called α and emission α now called β. We used the Penn Treebank Corpus in our experiment which was cut in three parts; training, development and test. The training data was used to train all of our models

2 Model

The model counts all occurrences for each transition and emission pair. It uses the α and β to compensate for combinations that aren't seen in the training set. It uses these counts to create a probability matrix of all $P(c_{prev}|c)$ and $P(x|c)$

3 Parameter Estimation

To estimate the parameters we trained a model for each combination of α and β while calculating for each combination the log perplexity and the tag accuracy (Tabel 1). To calculate these values we used the development set of the PTB.

The accuracty was calculated by trying to predict all the tags for each word using the viterbi pathing algorithm and comparing all the predicted tags to the real tags defined by the test set. The accuracy is then found by dividing the correct predictions by the total number of tags.

4 Experiments

Below you see the data we collected from our testing.

Alpha	Beta	Perplexity	Accuracy
0.01	0.01	4.56627	0.880875
0.01	0.1	4.41098	0.880875
0.01	1	4.68071	0.841057
0.01	10	5.53552	0.751632
0.1	0.01	4.56626	0.880875
0.1	0.1	4.41095	0.880875
0.1	1	4.68069	0.841057
0.1	10	5.53551	0.751632
1	0.01	4.56613	0.880548
1	0.1	4.41072	0.880875
1	1	4.68042	0.841057
1	10	5.53537	0.751632
10	0.01	4.56519	0.878264
10	0.1	4.40871	0.881201
10	1	4.67807	0.842363
10	10	5.53424	0.753264

We find that model 13 has the best results. It uses $\alpha = 10$ and $\beta = 0.1$

When we test our test set on this model we find log perplexity of 4.41 and an accuracy of 0.88

After this revelation we tried some test sentences using the predictive powers of the viterbi algorithm.

Word	Gold	Pred
But	CONJ	ADJ
Rep.	NOUN	NOUN
Marge	NOUN	ADP
Roukema	NOUN	NOUN
-LRB-	.	.
R.	NOUN	NOUN
,	.	.
N.J	NOUN	NOUN
.	.	.
-RRB-	.	.
instead	ADV	ADV
praised	VERB	ADP
the	DET	DET
House	NOUN	NOUN
's	PRT	PRT
acceptance	NOUN	NOUN
of	ADP	ADP
a	DET	DET
new	ADJ	ADJ
youth	NOUN	NOUN
“	.	.
training	NOUN	NOUN
”	.	.
wage	NOUN	NOUN
,	.	.
a	DET	DET
subminimum	NOUN	NOUN
that	ADP	ADP
GOP	NOUN	NOUN
administrations	NOUN	PRT
have	VERB	VERB
sought	VERB	VERB
T-1	X	X
for	ADP	ADP
many	ADJ	ADJ
years	NOUN	NOUN
.	.	.

accuracy: 0.8919

5 Conclusion

We conclude that model 13 is the best model, this model has both the lowest perplexity and the highest accuracy.