

# CSE490UA Winter 2017

## Assignment 3

Jason You [ID: 1427878]

February 20, 2017

### Contents

<a href="#">1</a>	<a href="#">Bigram Model</a>	<a href="#">2</a>
<a href="#">2</a>	<a href="#">Trigram Model</a>	<a href="#">2</a>

# 1 Bigram Model

## 1. OOVs and smoothing:

The smoothing method I use is to count all the word appear less than the threshold to be  $< UNK >$  (unknown) words. The original words are retained, and the number of of added counts of the  $< UNK >$  is added to the total word count.

To determine the value of the threshold, I use the most basic method of testing. By choosing different thresholds and recording the accuracy on the developement set, I choose 100 to be the best threshold:

Thresholds(count)	Accuracies(%)
2	10.98
3	9.19
5	10.76
10	22.82
15	33.28
25	44.12
50	65.72
90	67.67
<b>100</b>	<b>84.48</b>
110	68.68
150	78.99

The best performance of this model (threshold=100) on test set is: 84.48%. Some of these test is done by adding a nonlinearity (square root) when finding the emission probability.

## 2. Error analysis:

- (a) The tag 'U' and '#' is most predicted when the given word is unknown, and this kind of error is especially obvious when the threshold is low.

**Example:**

[<https://t.co/BQCPKV3aA6>, 'U'] # the true label is 'N'.

- (b) The other most obvious error is to chooes a threshold that is too low, which will lead to very low accuracy. The reason can be that for many labels, the words appeared in the training set are far from enough to cover the majority of the unknown words in the real twitter language.

But when the thresholds are high, this situation is more balanced between different labels, and thus the error rate decrease.

# 2 Trigram Model

1. OOVs and smoothing: Different from the bigram model, the trigram model can have history (pairs of labels) that never be seen in the training data set. So I choose to

handle this situation by assigning zero probability to these unknown label pairs.

For the OOVs, I use the similar method as in the bigram model, which is chooses a proper threshold to bisect the words in to words and `UNK`. The threshold is determined by the accuracies the model run on the dev-set.

Thresholds(count)	Accuracies(%)
10	60.90
50	77.36
75	75.89
<b>100</b>	<b>84.75</b>
125	72.99
150	80.57

The test accuracy is: 84.63%.

## 2. Error analysis:

The error in the trigram model is more general then the bigram model. The most common errors are from the tag "A", "@", "N", which is hard to determined the label when some special symbols are in the words.

## 3. Comparing to the bigram model:

Obviously, trigram model is much more robust than bigram model, but their peak performance of it is similar.