Assignment 4:HMMs and PoS Tagging Report

Shyam, Varsani
University of Alberta, varsani@ualberta.ca

Kavit, Gami
University of Alberta, gami@ualberta.ca

1. Introduction

The purpose of this report is to indicate the performance of our taggers. In this report we discuss how our two taggers, the HMM and the Brill taggers have performed and the tuning we did to achieve the performance.

2. Tuning Efforts

For the Hmm tagger, initially we achieved an accuracy of 0.242 which is relatively low. However, while tuning, we used the estimator function which helped us achieve a high accuracy of 0.825. This is because while tuning the estimator parameter, we used the LidstoneProbDist in the NLTK package which helped us tune the estimator parameter inorder to achieve a higher accuracy since it finds the maximum likelihood estimate for the resulting frequency distribution and hence finds the best value for the estimator parameter.

For the Brill tagger, initially we chose to use a trigram tagger as the baseline tagger, set max rules at 100 and min accuracy at 0.9 and achieved an accuracy of 0.637 which is relatively low. We tried different combinations of max rules and added a backoff set to a Unigram tagger but the accuracy of the Brill tagger was still relatively low. Therefore, we changed our baseline tagger to a Bigram and tried tuning max rules again. The resultant accuracy of our Brill tagger now increased to 0.768 which is still considered relatively low. Finally we changed the initial tagger to a unigram tagger with the back off set to the RegexpTagger from the NLTK library. We tuned the max rules parameter to find an optimal max number of rules set at 200. The final accuracy of our Brill Tagger is now relatively high at 0.872. Overall, for almost all combinations of the baseline tagger that we tried, we realized that increasing the number of max rules only helped to a certain number, mostly 200. Beyond this number, accuracy either decreased or stayed the same.

3.Tagger Performance

After training our taggers, we were able to achieve an accuracy of 0.825 with the Hmm tagger when being tested using the in-domain test data and an accuracy of 0.773 when tested using the out-of-domain test data.
Our Brill tagger on the other hand was able to achieve an accuracy of 0.872 when being tested using the in-domain test data and an accuracy of 0.859 when being tested using the out-of-domain test data.

4. Tagger Errors

To identify and compare the errors made by our taggers, we created an evaluation function that compares the input and output file for each tagger and prints out a dictionary of the mislabeled items. The Hmm tagger frequently mislabeled 'NN', 'NNP', 'NNS' and 'NNPS' tags. Words that were supposed to be tagged 'NN' would be tagged with either one of the remaining tags mentioned above or sometimes the other way round. This happened frequently because of the ambiguity of these words and the context they are being used in. Our Hmm tagger also mislabeled 'JJ', 'VB', 'VBP', 'VBD' and 'VBG' tags. It misidentified some adjectives as 'NNP' or 'NNS'. This could have happened because of the production rule that was used in tagging, the rules that are generally used oftenly in English language took precedence over those that are used less frequently. The tagger's confusion between the verb tags could have mostly happened because of the context they were being used in, considering the fact that Hmm would only consider the previous word and tag to label the current word with a tag.

In the Brill tagger we noticed that most of the errors were made while predicting JJ and VB. The JJ tags were often predicted to be NN tags by the brill tagger. This is because the tagger misinterpreted the adjectives with Nouns. The VB tags were often misinterpreted as being a NN tag most of the time. This could be because these words can either be a verb or a Noun and hence when the tagger comes across this it would often tag it as a Noun. However, we also noticed that it was being predicted as VBP and VBN in some cases this is because the words are spelled the same way in all the tenses and hence would often predict it wrong. The same was noticed while predicting VBD, VBP and VBN. The NN tags were also being predicted wrongly with different tags. This is because the tagger was not able to differentiate between a noun and an adjective at times and also if the spelling was the same in different tenses. Similar observations were made while labeling NNP tags where they were mislabeled as NN tags most of the time. This could partially be because of the unigram baseline tagger we used. Instead the baseline should have been defined as a trigram which can back off to a bigram, unigram and a default NLTK tagger as needed. This would help rid a lot of the ambiguity around words that remain the same in multiple tenses and have different meanings depending on the context.

4.1 Comparison of Tagger Errors

We noticed that some of the errors in both the Taggers were similar as they both misidentified the tags because the words would have the same spelling in different tenses. This was mostly observed in the NN, JJ and VB tags where these would often be mislabeled as a tag of a different tense or a different part of speech.

All in all, the Hmm tagger had significantly more mismatches compared to the Brill tagger. Since the Hmm tagger only considers the word and tag that appeared before to label the current word hence creating a lot of ambiguity whereas the Brill tagger had a backoff chain it could follow to deal with tagging ambiguous words.