NLP HW1 - Report

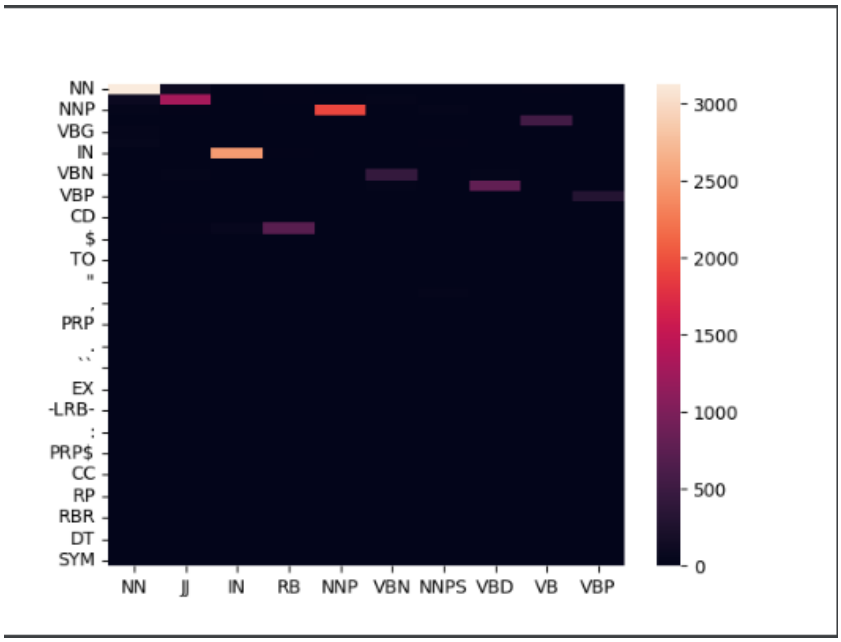
Submitted by: Idan Pogrebinsky 325968565, Ilai Avni 212778229

Model 1:

The accuracy we achieved on “test1” was 95.4 percent.
The features used to train the model were the ones described by [(Ratnaparkhi, 96)], as well as capital letters and numbers in the word and apostrophes in the current or previous word.
The number of features per feature set found seen during training are as follows:

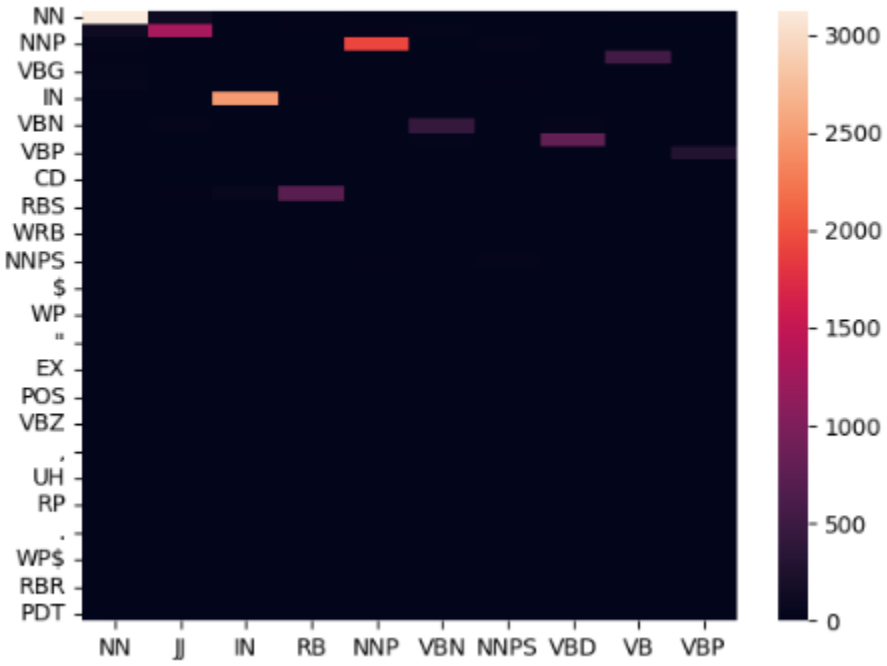
- “f100”: 14382
- “f101”: 4511
- “f102”: 9457
- “f103”: 7665
- “f104”: 1029
- “f105”: 44
- “f106”: 28093
- “f107”: 29452
- Tags of words starting with a capital letter: 32
- Tags of words containing numbers: 2
- Tags of words containing an apostrophe: 12
- Tags of words that come after words containing an apostrophe: 36

Confusion matrix on “test1”:



As we can see, the model predicts ‘NN’, ‘NNP’, ‘VBN’, ‘IN’ and ‘VBP’ fairly well, and misclassifies ‘JJ’ and ‘VB’.

The confusion matrix for threshold = 5:



It is clear that there is no significant difference between the two, therefore we use threshold = 5 for model 2.
We propose overcoming these misclassifications by handcrafting features specifically to distinguish between ‘JJ’ and ‘NN’ and omitting features that are too similar between them and could therefore contribute to confusion between them.
The inference done using this model was done by the Viterbi algorithm we implemented, while adding use of backpointers and Beam Search, which significantly improved the runtime of the algorithm.
The motivation behind backpointers is not needing to calculate the best tagging for the current word given the next two, since this was already done when calculating the values of ‘pi’ during the inference.
The motivation behind using Beam Search was the exponential reduction in the number of branches explored by the algorithm, without any noticeable differences in test accuracy.

Model 2:

The accuracy achieved was 58.4 percent.
The number of features per feature set found seen during training are as follows:

- f100 64
- f101 190
- f102 206
- f103 107
- f104 101
- f105 20
- f106 63
- f107 59
- Tags of words starting with a capital letter: 8
- Tags of words containing numbers: 1
- Tags of words containing an apostrophe: 0
- Tags of words that come after words containing an apostrophe: 0

The features used in this model are the same as in model 1, except the threshold used is 5 instead of 1.
In lieu of a relevant test file for this model, we used the same test file as for model 1, without computing a confusion matrix.

Competition:

We made no specific alterations to the model for the competition.
We expect a ~85 percent accuracy on the competition file.
A possible explanation for the dissonance between the accuracy on the test and competition files is that the train and test sentences come from the same corpus whereas the competition words come from a different one, which results in a more similar distribution between the train and test than the train and competition.

Work Split:

While most of the work was done together, the bulk of the feature implementation and result evaluation was done by Idan, while the Viterbi implementation, as well as the report, were done by Ilai.

Environment Specs:

All code was run on a laptop with an i7 10th generation CPU and 32GB of RAM.