Homework 3

Abhijit Prakash Bhatnagar

1. Please include your code! However, you will not be specifically graded on it and your responses to the questions below should be clear enough to understand what you did.

See attached folder

1. (2pt) Define the feature templates you used for implementing the structured perceptron classifier. Propose 3 additional feature templates that (1) will likely be useful for improving NER performance and (2) are not part of your implementation.
2. (2pt) Write out the Viterbi decoding algorithm for the structured perceptron using one recursive equation.
3. (2pts) Ablation Study: In Ablation studies, we remove a subset of features (i.e. a subset of feature templates) to understand how those features affect performance. Design your own ablation study and report how the performance varies with respect to both the dev set and the test set. Try no more than 4 feature vector variations, including your full model with complete feature set.
4. (2pt) While BIO encoding is the most common, there are a number of other encoding schemes (https: //lingpipe-blog.com/2009/10/14/coding-chunkers-as-taggers-io-bio-bmewo-and-bmewo/) also used in practice. In fact, some studies have reported that IO encoding, which does not differentiate ‘B’ from ‘I’, often results in comparable (or even better) performance depending on the data and the task. Why might this be the case? Discuss the potential pros and cons of BIO and IO encoding schemes for NER.
5. (2pt) Error analysis and discussion of results. Including and discussing specific error cases you found will be useful here!