Project Milestone

Yucen Li (Lily)

April 5, 2018

Sorry, I can’t make it to today’s meeting! I forgot about it until earlier today. I also sent you an email, but I wasn’t sure if you’d see it.

Major Changes

There have been no major changes to my project.

Accomplishments

Since the last meeting, I have changed the experimental setup so each sentence of training data has the same number of sentences per language. However, I’m still unsure if this is the setup that leads to the fewest confounding variables, as this still means that there is more training data for languages. I’m thinking about setting a fixed amount per training data and instead changing the number of sentences per language that is input.

Meeting Milestone / Revisions

Because the results I’ve gotten seem to be fairly unpredictable, I’m spending more time analyzing the results rather than conducting new experiments. The experimental analysis has been taking longer than I originally intended, but I do think it’s necessary to understand what’s happening before moving on with the tests.

Surprises

I’ve done analysis on the UD treebanks to see properties such as average sentence length, avg max depth of tree of a sentence, number of children, etc. Most of the numbers between the languages have been pretty similar, but there were a few properties that seemed to be different. For example,
the average depth of a tree in Spanish was about 1 higher than the average for the other languages. Because of this, I’ve decided to look more into where the sentences came from. For example, if one treebanks consists of mostly tweets, it is fairly unlikely to be able to correctly parse something from news.

**Looking Ahead**

When I did pairwise training, it seemed that training on Spanish really hurt the Italian score, which is surprising since the two languages are similar. I plan to manually analyze some of the resulting parses that the model develops in after training on Spanish and Italian to see if there is a pattern to how the sentences are being parsed that’s related to a difference between the grammar, or to see if the parse is random and not indicative. Furthermore, there are actually two different Spanish treebanks in Universal Dependency dataset, so I also plan on training between the two treebanks to see if there is a difference between the quality of the datasets.