John Blazic
3/28/17
<div style="text-align:center">Proposal: Improving A* Parsing through Matrix Decomposition</div>

Part 1:

The paper that has been provided with this proposal was voted the best paper of EMNLP 2016. I will be reimplementing the parsing algorithm defined in the paper and attempting to improve upon the efficiency. The technique that is defined in this paper contains a number of high-dimensional datasets that are imported from the outside that have been shrunk to smaller dimensions as well as a multi-dimensional matrix of parameters, Θ.

First, I would like to try to implement a separate word embedding model, which takes a sparse matrix formed from a document-index matrix and compresses it down to one with just a handful of dimensions, in this case 50. For reference, Word2Vec uses between 100-1000 dimensions to give a better representation. I would like to try to recreate these matrices and see which of the singular values are useful to find a more appropriate dimensionality, if there is one.

Second, I would like to play with the dimensionality of the Θ matrix that they are using to store the parameters of their model. It is possible that they may be over or under utilizing the parameters here. I believe that improving these 2 matrices could improve the efficiency and accuracy of the model.

Part 2:

This paper is from a conference, EMNLP, Empirical Methods in Natural Language Processing, that is one of the foremost in our field. If I do find that I can improve the results, this could lead to research that I could present at this or another similar conference. Additionally the practice it will give me in modifying neural networks and their parameters will be invaluable going into the industry.

Part 3:

At the time of the report, I intend to have implemented the suggestions above and tested them thoroughly against the results provided (see part 4). These results will give me an idea if the methods I'm intending to use will be good enough to provide an opportunity to submit a paper to a conference. I don't intend for the report to be a full conference paper, but it will include the results that I would submit, some of the conclusions that I would hope to make, as well as some more details on the linear algebra used and the results from that, which may or may not be part of the paper.

Part 4:

There will be a few ways to succeed for this. First, the performance of this and many similar systems are presented in the paper and can be used as a measurement of how well any change I make to the system actually do. Second, in going through and attempting to play with the dimensionality of the matrices involved in the algorithm, I can present the singular values that I find for the matrices that I am attempting to modify. I can then use those values to back up why I made certain changes in an attempt to improve the system and with the aforementioned results from the paper, show any changes in performance.