

Advanced Data Vis Project Progress Report

Maks Cegielski-Johnson
u0836524

April 1, 2018

1 Project Summary

1. What is an overview of your project?

Visualizing word vectors with high dimensionality techniques.

2. Why is the project worth pursuing?

Word vectors are commonly used in NLP research, but since the dimensionality of vectors is quite large, it would be useful to have a tool that allows for easy exploration of this space to collect insights.

3. What are your project objectives?

Create a visualization that allows user interaction to explore the space of high dimensional word vectors.

4. What are the questions you would like to answer?

How can visualization be used to successfully determine whether two words are similar, or what the degree of similarity is between two words.

5. What data will you plan to use?

Word2Vec and GloVe vectors, given enough time, perhaps WordNet similarities.

6. How can we evaluate how successful your project is once it is completed?

Since this tool is rather exploratory, if correct insights can be made with my tool. If my tool shows two words being similar that aren't similar, is there an underlying similarity that isn't clear between these words, or is my tool incorrect.

2 Timelines

- ~~Week 1 (March 5) - Project proposal~~
- ~~Week 2 (March 12) - get the data, prepare it for visualizing, set up Python frameworks~~
- ~~Week 3 (March 19) - try out different HD visualizations~~
- Week 4 (March 26) - try out different HD visualizations
- Week 5 (April 2) - try out different HD visualizations / try out TOPO visualizations
- Week 6 (April 9) - make website / ~~d3 visualizations~~
- Week 7 (April 16) - make website / d3 visualizations
- Week 8 (April 23) - finish website / prepare presentation, report