Ling/CSE 472: Introduction to Computational Linguistics

5/21: Word Embeddings

Word Embeddings

- => J&M slides

Reading questions

- The concept of information retrieval is really interesting to me! Is the measurement of the best matching document d from our query q done using the cosine function shown in section 6.4? If so, how does comparing one word, q to an entire piece of text d work?

Reading questions

- Are word embeddings only used in very specific areas of NLP or are they fairly standard nowadays?

- Are there any examples of practical uses of word embeddings that might be surprising?

Reading questions

- Is there a way to measure the dissimilarity of two target words? The reading mentions seeing how close the dot product of the target words is 0 but is there a function for the dissimilarity?

- Would word2vec also be able to capture similarities or 'neighboring' words to the target word even if they were far separated from the word itself? How long can the distance from the target word to its neighbors in the sentence be?

Reading questions

- Fasttext representations of unknown words: Why was the method of representing words with their constituent n-grams chosen, and does it actually produce useful embeddings (for example ones that will represent the unknown word's similarity to known words with some accuracy)?

- based on teh reading, it seems like there are some ways to do word embedding involving typos and other incorrect or incomplete words. Is this done in practice on a large scale as many text online are not grammatical but meaningful for people to understand.

Reading questions

- In Figure 6.16, adjectives are shown with their comparatives and superlatives to have similar patterns in their vectors between the three forms. Would this still apply for adjectives which don't use the -er -est morphology? i.e. would words like good->better->best or fun->more fun->most fun still show the same pattern on the vector space? Semantically, these groupings of comparatives and superlatives mean the same as regulars like slow/slower/slowest, but could their irregular morphology change their vectors?

Reading questions

- Using the vector space to visualize analogies and historical semantic trends was very cool! Would it be possible to make estimations about how a given word's semantics evolved over time using historical trend data of words with similar embeddings? Could we even potentially make predictions of how a word's semantics will evolve in the future given its past?

Reading questions

- Biases seem like a really tough thing to address, as they are inherently part of the sense of a certain word and seemingly cannot simply be 'erased' from the word through de-biasing. In this way, are biases always something doomed to be part of technology like word2vec, or are there any other ways that they can be addressed? Would a word being in a non-biased context in training data possibly help this?

Reading questions

- It is mentioned in the reading that word embeddings don't just reflect bias, but in fact amplifies them. I wonder what exactly in the models contributes to this increase in bias (for example a heavier correlation between some gender and stereotypes)? Is it something intrinsic to word embeddings that just could not be avoided or might it be improved by maybe better designed vector representations?

Remaining schedule

- 5/23: Linguistic semantics + NLP, Octopus paper, Stochastic Parrots
- 5/25: Grammar Matrix + AGGREGATION projects
- 5/30: Meaning Making with Artificial Agents, Wrap-up exercises
- 6/1: Term project presentations (8 groups x 10 minutes)

NLP/Compling in the news

- https://www.npr.org/2023/05/18/1176806824/striking-movie-and-tv-writers-worry-that-they-will-be-replaced-by-ai

- https://www.washingtonpost.com/technology/2023/05/17/tiktok-ban-montana/