Information Retrieval - Assignment 1 Peer Review

Anonymous

June 12, 2019

1 Summary

In this Assignment usages of word embeddings (word2vec) get shown, first staring of by finding similar words and doing a dimension reduction with SVD to plot a subset of 500 words. Additionally analogy predictions for syntax and semantics are shown. As a second part word embeddings get used for sentiment analysis (classification) and ending in the last part implementing CPTW (Contextually Propogated Term Weights) for classification.

2 Peer-review

I will first give some general remarks about the assignment and then give feedback on each task.

3 General Remarks

I do like the structure of the document and the intro to each Task makes it easy to to catch up what you are talking about and what results you want to present. The only thing that first popped up in my eyes were some typos apart from that your writing switches between *we* or *I*, preferably choose one of them and stick to it. So that the language used in the report is consistent.

4 Report feedback

4.1 Task 1

4.1.1 Warm up

Apparently, a good short explanation to the word embeddings. The similar words are well presented, same for the 2D projection of the subset of 500 words from the word embeddings.

4.1.2 Main Task

Simple and easy explanation about the vector offset method used. But I do question the accuracy of predicting the 4th word which seems too low. Especially how you have calculated the accuracy, isn't clear from the code or the report. And maybe also showing an example of syntax/semantics could be useful to understand the task at hand better.

4.2 Task 2

Also here I like the very short and on point introduction and what representation you have used (by taking the average over all words in a document). The usage of 2 different classifier for the classification at hand is definitely a nice addition to at least have some comparison and showing that you did different experiments.

4.3 Task 3

Since the results of the CPTW are missing, so the last task isn't that meaningful. Especially by only showng the TF-IDF score but as described in the report the result will be delivered as soon as possible and hopefully compared with the TF-IDF to then write down the outcome and the interpretation of those results.

5 Final Remarks & Conclusion

Overall good structured and good presentation of results obtained (either as a table or image). As said before the language of the report can use some brush up and the missing CPTW should also be introduced as soon as possible to finish up the report on assignment 1. Adding some references would also help to give a reader a better understanding of the data and methods used.