

Opponent Review

for the master thesis

"Analysis of Word Embeddings"

Reviewer: Prof. Dr. Phillipp Torkler
Student: Matthias Thurnbauer
Matriculation No: 00629555

Summary

Mr. Thurnbauer's analyses in his master thesis the performance and applicability of static word embedding techniques for improving classic information retrieval systems. Static word embeddings are limited in their performance compared to more recent transformer based methods, but can be used in situations with limited training data and computational resources. For these scenarios, the applicability of the static word embedding methods *Word2vec* and *GloVe* are tested for the German language to address the problem of thesaurus generation to solve the vocabulary mismatch problem, to solve the disambiguation problem of static word embeddings by using a k-means clustering, and to test the influence of the window size hyperparameter as well as the impact of different normalization techniques on word embeddings.

Review

Overall, the thesis is clearly structured and written in a concise and precise language. Research questions and goals of the thesis are clearly stated and motivated. The thesis' goals are clearly separated from related work and described with respect to recent developments. Technical background information is described, but could have been a bit more verbose.

The description, presentation and interpretation of the results address all research questions in detail and show a clear understanding of both, the used word embedding

techniques and clustering algorithms. Failed experiments and limitations are mentioned and discussed as well.

In conclusion, the quality of the analysis, the structure and quality of presentation completely fulfill the scientific standard of a master thesis and I rate the thesis as **1.0**.

Deggendorf, 03.02.2023

Prof. Dr. Phillipp Torkler