

Supervisor report

for the master thesis
“Analysis of Word Embeddings”

Student: Matthias Thurnbauer

Supervisor: Prof. Dr. Andreas Fischer

Submitted on: January 4th, 2023

Summary

Mr. Thurnbauer investigates in his thesis the performance of conventional word embedding approaches. In recent years, transformer models, such as LaMDA or GPT-3, have obtained huge interest. However, the respective models are incredibly complex and require a huge amount of resources (in terms of both time and data) to train. Mr. Thurnbauer is particularly interested in evaluating, whether a more light-weight, conventional approach can achieve acceptable performance in two narrow aspects, namely construction of a thesaurus and disambiguation of polysemes. Both are evaluated on the German wikipedia corpus. He shows that, by closer investigation of word context, ambiguity can be identified and, in some cases, resolved, all while keeping training times within reasonable bounds.

Evaluation of the content

Mr. Thurnbauer analyses and evaluates the performance of two well-known word embedding models: word2vec and GloVe. He employs the two models to solve the two problems of thesaurus generation and word disambiguation for German language texts. The application of the models with the specific tasks on a large German text corpus is new. The discussion of related work by Mr. Thurnbauer is extensive. He demonstrates good knowledge of the state-of-the-art and his advances on top of it are clearly recognizable.

The methodology of Mr. Thurnbauers approach is convincing and adheres to common scientific standards. His arguments are rigorous and clear. To build on and promote the results by Mr. Thurnbauer, an excerpt from the thesis is currently being prepared for publication in a scientific conference.

Structure and quality of presentation

The structure of the thesis is clear, following common scientific standards. The text is quite readable, especially considering that English is not Mr. Thurnbauers native language. Mr. Thurnbauer presents his results in a clear way, and makes good and extensive use of figures and tables to summarize key results.

Collaboration with the student

Working with Mr. Thurnbauer was quite positive and productive. He delved deep into the subject matter and proved himself highly motivated and self-organized. Mr. Thurnbauer was in all cases quick to react to comments and hints. He was creative in developing new ideas and attentive in accepting comments and constructive critique. Problems were often solved carefully and independently by himself with little to no necessary correction from my side.

Result

Overall, I rate the thesis as: **1.0 (sehr gut, excellent)**

Deggendorf, 20. Januar 2023

Prof. Dr. Andreas Fischer