Embedding for scientific texts

Validation & Benefits

Arjan Meijer

University of Amsterdam & Elsevier

Supervisors

- University of Amsterdam: Dr. R. Cushing
- Elsevier: R. Karimi & J.A. Schijvenaars

Word Embedding

- High-dimensional, scalable, dense vector representation of a word
  - Capture semantic and syntactic information (Lai et al.)

- Enables calculation with text
  - Apple – Apples ≈ Car – Cars
  - Men – King ≈ Women - Queen

Research questions

- Have word embeddings a higher accuracy for academic texts than TF-IDF?

- Which metric(s) can be used to measure the accuracy of word embeddings fro scientific articles?

- Which parameter values create embeddings with the highes accuracy for academic texts?

Progress

- Research environment

- Validation of word embeddings on categorization tasks
  - Usage of already existing word embeddings

Overview

- Supervisors: R. Cushing, R. Karimi & J. A. Schijvenaars

- Research questions:
  - Have word embeddings a higher accuracy for academic texts than TF-IDF?
  - Which metric(s) can be used to measure the accuracy of word embeddings fro scientific articles?
  - Which parameter values create embeddings with the highest accuracy for academic texts?

- Progress:
  - Research environment
  - Validation of word embeddings on categorization tasks.