

Scientific Paper Recomm. via Topic-Chronological Content

Background

Our capacity to produce and store data has been continuously increasing in the last few years. In academia, this is no exception, and many academic repositories store valuable information for researchers. However, the process of organizing and recommending documents, from scientific repositories, is not a trivial task. We want to explore the new models and architectures in NLP to organize prospective suggestions based on their chronological order and their topic similarity producing a semantic timeline.

Goal

Create a recommendation tool for scientific literature capable of organizing its suggestions considering its topic-chronological relevance. Space for literature review on visualization.

Tasks

- Review the literature on scientific paper recommendation;
- Prepare data from one or more repositories;
- Integrate NLP architectures to extract features from spec. dataset;
- Develop/Evaluate a system that organizes its recommendations in a topic-chronological manner;



Corinna Breitinger
breitinger@uni-wuppertal.de



Terry L. Ruas
ruas@uni-wuppertal.de



Semantic Feature Extraction for NLP

Background

The relationship between words in a sentence often tell us more about their semantic content than its actual words individually. Semantic analysis is arguably one of the oldest challenges in Natural Language Processing (NLP) and still present in almost all its downstream applications. However, the extraction of features that describe the semantic aspect of documents is not an easy exercise. We devised a group of approaches that are able to capture these underlying semantic features and use them in NLP tasks.

Goal

Improve document similarity using semantic features and recent advances in NLP/Word Embeddings/Transformers/Reformers

Tasks

- Review the literature on text classification using semantic features;
- Extend devised approaches to recent state-of-the-art word embeddings/transformers techniques;
- Evaluate your approach in specific datasets.



Terry L. Ruas
ruas@uni-wuppertal.de



(In)stability and Usability of Word Embeddings

Background

With the increasing advances in word embedding models, their use is almost required in NLP downstream tasks or applications. However, how robust and stable pre-trained word embeddings models are? Even though the popularity of word embedding models makes them attractive for most NLP challenges there is little discussion about their semantic aspects concerning the semantic features they represent.

Goal

Explore the (In)stability of recent word embeddings algorithms and pre-trained models towards NLP tasks and applications.

Tasks

- Review literature on (In)stability of word embeddings models;
- Select NLP tasks to evaluate similar/equivalent models;
- Present and discuss (possible) main aspects for the instability of word embeddings models;
- Propose and implement improvements to mitigate explored problems.

Anastasia Zhukova
zhukova@uni-wuppertal.de



Terry L. Ruas
ruas@uni-wuppertal.de

