

Advanced Text Mining Methods

Carsten Jentsch, Kai-Robin Lange

TU Dortmund, Department of Statistics

13.07.2022

Schedule and Requirements

Requirements

- Presentation in English
- Bachelor and Master students are welcome
- Basic knowledge in Text Mining is being expected (personal experience or lecture „Text as data“)

Schedule

- Registration: Until 30th of September using the following link:
<https://umfragen.tu-dortmund.de/index.php/334556?lang=en>
- Distribution of topics: 1st of December
- Presentations: February 2023
- Reports due: March 2023

Motivation

Research Questions

- What do people talk about? – Topic Models
- How do words relate to each other? – Word embeddings
- Development of a word's meaning over time? – Diachronic embeddings
- Positive or negative attitude towards a subject? – Sentiment analysis
- Changing texts over time? – Change point detection
- Predict other quantities using texts? – Forecasting (e.g. GDP)

Projects

Goal: Acquiring deeper knowledge on a particular use case of text data

Theory-oriented works

- Comparing topic/embedding models (theoretically or exemplary)
- The Theory behind Transformer models
- ...

Practice-oriented pipelines

- Semantic Role Labeling
- Coreference Resolution
- Diachronic Embeddings
- Change point detection
- Machine Translation
- Sentiment Analysis
- ...

Questions?

kalange@statistik.tu-dortmund.de