

Information Extraction Seminar Topics

Dario Stojanovski

Ludwig Maximilian University of Munich

stojanovski@cis.lmu.de

October 21, 2019

- ▶ Reminder: these topics should all be in English

Self-training for Temporal Relation Extraction

- ▶ Matching events to specific times. Important in the medical domain for tracking progression of illnesses.
- ▶ Challenging to obtain large labeled datasets. How to handle scenarios with limited labeled instances?
- ▶ Explain how self-training works for this task. Contrast to other approaches to semi-supervised learning.
- ▶ Analysis of improvements and errors.

Sources and possible papers

- ▶ Lin, Chen, et al. "Self-training improves Recurrent Neural Networks performance for Temporal Relation Extraction." Proceedings of the 9th International Workshop on Health Text Mining and Information Analysis (LOUHI 2018)

Cross-lingual Named Entity Recognition

- ▶ Motivation for cross-lingual NER.
- ▶ How is cross-lingual NER different than monolingual NER?
- ▶ What are the potential benefits?
- ▶ Explain what kind of data is needed and how to use it?
- ▶ Explain the method and features used in the paper.

Sources and possible papers

- ▶ Tsai, Chen-Tse, et al. "Cross-Lingual Named Entity Recognition via Wikification." Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL)

Automatically Labeled Data for Event Extraction

- ▶ Overview of hand-labeled data for event extraction. What problems do we face when using this data?
- ▶ Can automatically labeled data solve them?
- ▶ Explain the process of generating automatically labeled data.
- ▶ Present the method for event extraction.

Sources and possible papers

- ▶ Chen, Yubo, et al. "Automatically Labeled Data Generation for Large Scale Event Extraction." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics

Aspect-based Sentiment Analysis

- ▶ Motivation for aspect-based sentiment analysis.
- ▶ How is it different than traditional sentiment analysis?
- ▶ How does the data look like? Is it challenging to create?
- ▶ Neural networks - what architecture is best suited for the task?

Sources and possible papers

- ▶ Wang et al. "Attention-based LSTM for Aspect-level Sentiment Classification" Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing

Hyperpartisan News Detection

- ▶ Fake vs Hyperpartisan news - differences and similarities.
- ▶ Present the shared task and explain how is the task defined.
- ▶ What is the data? How was it created, annotated etc. How were the teams evaluated?
- ▶ What are the shared task conclusions?
- ▶ Present the basic approach of the best team.

Sources and possible papers

- ▶ Kiesel et al. "SemEval-2019 Task 4: Hyperpartisan News Detection." Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)
- ▶ Jiang et al. "Team Bertha von Suttner at SemEval-2019 Task 4: Hyperpartisan News Detection using ELMo Sentence Representation Convolutional Network" (SemEval-2019)