Information Extraction
Seminar topics

Dario Stojanovski

stojanovski@cis.lmu.de

- Reminder: these topics should all be in English

Multilingual Named Entity Recognition

- Why do we need multilingual NER?
- How is multilingual NER different than monolingual NER?
- Quick overview of the two main approaches
- Explain what kind of data is needed and how to use it
- Present a method for NER

Sources and possible papers
- Kim et al. "Multilingual named entity recognition using parallel data and metadata from wikipedia." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 2012.

Distant supervision for Relation Extraction

- Pros and cons of supervised data for RE
- Explain distant supervision for RE
- Appropriate sources of data for distant supervision
- Explain a model suitable to make use of this data
  - Features
  - Training algorithm
- How to evaluate

Sources and possible papers
  - Mintz et al. "Distant supervision for relation extraction without labeled data." Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2. Association for Computational Linguistics, 2009.

CNN for RE

- Pros and cons of pipeline approaches for RE

- Explain CNN for RE

- Explain the different features

- Discuss how CNN compares against other methods

Sources and possible papers

- Zeng et al. "Relation classification via convolutional deep neural network." *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 2014.

Twitter Sentiment Analysis

- Challenges when doing Twitter Sentiment Analysis

- Preprocessing steps

- Explain pipeline steps

- Pros and cons of the lexicon-based and learning-based methods

Sources and possible papers:

- Zhang et al. "Combining lexicon-based and learning-based methods for twitter sentiment analysis." *HP Laboratories, Technical Report HPL-2011* 89 (2011).

Aspect-based Sentiment Analysis

- Motivation for aspect-based sentiment analysis

- Overview of the SemEval-2016 Task 5: Aspect-Based Sentiment Analysis

- Present training data, subtasks, evaluation metrics …

- Short system overview of one of the participants (preferably a participant with high scores across all metrics)

Sources and possible papers:
- Pontiki et al. "SemEval-2016 task 5: Aspect based sentiment analysis." *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*. 2016.

Event Extraction and Classification

- Event extraction in the biomedical domain: motivation and challenges

- Explain the pipeline for EE

- Explain the novelty of the method

- Discuss error analysis

Sources and possible papers:

- Miwa et al. "Event extraction with complex event classification using rich features." *Journal of bioinformatics and computational biology* 8.01 (2010): 131-146.

Document Event Extraction

- Why do we need the whole document for EE?

- Explain the document-level method

- Discuss sentence-level and document-level approaches

Sources and possible papers:
- Liao et al. "Using document level cross-event inference to improve event extraction." *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010.

Joint Event Extraction

- Compare pipeline with joint methods for EE

- CNNs for EE: explain the architecture

- Overview of additional features

- How different aspects of the model affect performance?

Sources and possible papers:

- Chen et al. "Event extraction via dynamic multi-pooling convolutional neural networks." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Vol. 1. 2015.

Coreference Resolution

- Two possible papers (choose one of them)

- A pipeline approach vs. joint model

- Advantages of one approach against the other

- Advanced topic

Sources and possible papers:

- Raghunathan et al. "A multi-pass sieve for coreference resolution." *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2010.
- Lee et al. "End-to-end neural coreference resolution." *arXiv preprint arXiv:1707.07045* (2017).