Seminar Topics: Information Extraction

English topics!

Viktor Hangya

hangyav@cis.lmu.de

Named entity recognition

- Topic: **WNUT2017 Shared Task on Novel and Emerging Entity Recognition**

- Overview:
  - Shared task for detecting previously unseen entities.
  - What are the hard aspects of the task?
  - Present training data, annotation process, evaluation metrics.
  - Give a general outline of the approaches.
  - Is NER solved for emerging entities?
- Paper:(link)
  - **Results of the WNUT2017 Shared Task on Novel and Emerging Entity Recognition** Derczynski et al., 2017,
    Proceedings of the 3rd Workshop on Noisy User-generated Text

Named entity recognition

- Topic: **Deep Learning for Emerging Named Entity Recognition**

- Overview:
  - An approach for unseen named entity recognition (WNUT2017).
  - Describe the task briefly.
  - Describe the system's modules and results in more details.
- Paper:
  - **Distributed Representation, LDA Topic Modelling and Deep Learning for Emerging Named Entity Recognition from Social Media** Jansson and Liu, 2017,
    Proceedings of the 3rd Workshop on Noisy User-generated Text

Parallel sentence extraction

- Topic: **BUCC shared task for parallel sentence extraction**

- Overview:
  - Why and for what extrinsic tasks is parallel sentence mining important?
  - How is the training and test data built for the shared task?
  - Give an overview of submitted systems and results.
- Paper:
  - **Overview of the Third BUCC Shared Task: Spotting Parallel Sentences in Comparable Corpora** Zweigenbaum, Sharoff and Rapp, 2018, *Proceedings of the 11th Workshop on Building and Using Comparable Corpora*

Parallel sentence extraction

- Topic: **Sentence representation for parallel sentence extraction**

- Overview:
  - What are sentence embeddings and how to train them?
  - How can they be used for parallel sentence mining?
  - Present the scenarios in which the mined pairs can be used.
- Paper:
  - **Filtering and Mining Parallel Data in a Joint Multilingual Space** Schwenk, 2018,
    Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics
  - **Learning Joint Multilingual Sentence Representations with Neural Machine Translation** Schwenk and Douze, Sharoff and Rapp, 2017,
    Proceedings of the 2nd Workshop on Representation Learning for NLP

Sentiment analysis

- Topic: **Document classification with hierarchical attention**

- Overview:
  - What are the hard aspects of document classification vs. sentence classification?
  - Describe the hierarchical attention classifier.
  - What experiments were performed?
  - What role does contextual information play in classification?
- Paper:
  - **Hierarchical Attention Networks for Document Classification** Yang et al., 2016,
    Proceedings of North American Chapter of the Association for Computational Linguistics: Human Language Technologies

Sentiment analysis

- Topic: **Cross-lingual sentiment classification**

- Overview:
  - How can a classifier trained only on source language annotations classify target language texts?
  - Describe the method for training bilingual document representations.
  - Describe the performed experiments.
  - Can we use this approach to any language pair?
- Paper:
  - **Bridging the Language Gap: Learning Distributed Semantics for Cross-Lingual Sentiment Classification** Zhou, He and Zhao, 2014, *Proceedings of Natural Language Processing and Chinese Computing*

Sentiment analysis

- Topic: **Joint aspect and sentiment classification**

- Overview:
  - Describe the task of aspect-level sentiment analysis.
  - Describe the model for joint aspect and sentiment detection.
  - What is the advantage of joint modeling vs. detecting aspect and sentiment in two consecutive steps?
  - Describe the performed experiments.
- Paper:
  - **Joint Aspect and Polarity Classification for Aspect-based Sentiment Analysis with End-to-End Neural Networks** Schmitt et al., 2018, *Proceedings of Empirical Methods in Natural Language Processing*

Summarization

- Topic: **Abstractive neural text summarization**

- Overview:
  - What is the difference between abstractive and extractive summarization?
  - Describe the model for abstractive summarization.
  - What specific problems are tackled and how?
  - Describe the used data and the performed experiments.
- Paper:
  - **Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond** Nallapati et al., 2016,
    Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning