Seminar Topics: Information Extraction

English topics!

Alexandra Chronopoulou

achron@cis.lmu.de

Topic: Sentiment Analysis using Distant Supervision

Learning Distributional Word Embeddings from emojis

Overview:

- Why are word2vec embeddings not suitable for social media (e.g. Twitter, microblogging etc)?
- Why are noisy labels (hashtags,emojis) useful in social media sentiment analysis?
- Present training data, model.
- Present experiments conducted to evaluate the approach.

Paper:
emoji2vec: Learning Emoji Representations from their Description
Eisner et al., 2016, *Proceedings of the 4th International Workshop on Natural Language Processing for Social Media at EMNLP 2016*

Topic: Sentiment Analysis using Distant Supervision

Learning Representations from Emoji Prediction in Tweets

Overview:

- Preprocessing steps
- What is transfer learning?
- Present model (Bi-LSTM, attention,..).
- Present experiments. Why is pretraining important? Why does the model perform well on different domains?

Paper:

Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm

Felbo et al., 2017, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*

Topic: Fake News Detection

The Fact Extraction and VERification (FEVER) Shared Task

Overview:

- Shared task for correctly classifying whether factoid claims could be `SUPPORTED`, `REFUTED` or labeled as `NOTENOUGHINFO` using evidence from Wikipedia.
- What is the purpose of the task?
- Describe dataset, scoring metric introduced to evaluate submission.
- Give a general outline of the approaches.

Paper:

The Fact Extraction and VERification (FEVER) Shared Task

Thorne et al., 2018, *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER) at EMNLP 2018*

Topic: Fake News Detection

Stance Detection

Overview:

- Why is stance detection useful for fake news detection?
- Classify the attitude expressed in a text **without** labels about the target.
  e.g. *"@RealDonaldTrump is the only honest voice of the @GOP"*

  positive stance towards Donald Trump but
  negative stance when annotated with respect to Hillary Clinton
- Describe the methods/models used (independent encoding, conditional encoding, bi-directional conditional encoding)
- Briefly describe experiments and results.

Paper:

Stance Detection with Bidirectional Conditional Encoding
Augenstein et al., 2016, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*

Relation Extraction from Clinical Texts

Overview:

- Motivation for information extraction from biomedical/clinical texts.
- Explain the limitations of traditional (non-neural) approaches (cooccurrence, rule-based, feature-based, kernel methods)
- Each word is represented with 6 discrete features (that encode positional, semantic, syntactic info etc). Describe their contribution and present the model used.
- Describe experiments. Which features are most helpful for the task?

Paper: **Relation extraction from clinical texts using domain invariant convolutional neural network**, Sahu et al., 2016, *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*

Named Entity Recognition using Joint Word- and Character-level Embeddings

Overview:

- Why is transfer learning important for NER?
- Describe the model. How is each word, sentence represented?
- Explain the difference of *source* and *target* datasets.
- Explain the impact of transfer learning for various train set sizes of the target dataset based on the provided experiments.
- Explain the impact transferring the parameters up to each layer (token embeddings, + character embeddings, + character LSTM, etc)

Paper: **Transfer Learning for Named-Entity Recognition with Neural Networks**, Lee et al., 2017, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*

Event Detection using Recurrent Neural Networks

Overview:

- Motivation for automatic event extraction.
- What is an event nugget?
- Describe the proposed model (RNN, branch embedding, word embedding).
- Describe the experiments and results.

Paper: **Event Nugget Detection with Forward-Backward Recurrent Neural Networks**, Ghaeini et al., 2018, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2018)*