Seminar Topics: Information Extraction

English topics!

Viktor Hangya

hangyav@cis.lmu.de
Structured Sentiment Analysis

▶ Structured analysis:
  ▶ monitor only some target entities
  ▶ monitor only the opinions of some users

▶ Project:
  ▶ Discuss the problem of structured sentiment analysis. What datasets are available? Different variations (targeted or aspect-based).
  ▶ Discuss a few approaches to solve the task.
  ▶ What difficulties do we face?

▶ Resources:
  ▶ Barnes et al., 2022, SemEval 2022 Task 10: Structured Sentiment Analysis Proceedings of the 16th International Workshop on Semantic Evaluation
Multimodal Information Extraction

▶ Extract information from the combination of multiple data modalities:
  ▶ text, picture, sound, video
  ▶ E.g. we often use pictures to make our message clear: sarcastic meme

▶ Project:
  ▶ How can we fuse texts with other modalities for IE?
  ▶ Discuss a few tasks where multimodal IE is important.
  ▶ What model architectures can be used?
  ▶ What are the hard aspects of the task?

▶ Resources:
  ▶ Fersini et al., 2022, *SemEval-2022 Task 5: Multimedia Automatic Misogyny Identification* *Proceedings of the 16th International Workshop on Semantic Evaluation*
Information Extraction from Language Models

- Extract information from trained neural language models instead of texts.
- Language model training:
  - I went to a concert Saturday [?] → night
- Although they aren’t directly trained to perform IE, they can answer questions or recognize entities:
  - Who is the president of the United States? [?] → Joe Biden
  - Who is the president of the [United States]$^{LOC}$?
- Project:
  - What kind of information is learned by pre-trained language models?
  - How can we extract them?
  - Is it superior to traditional techniques?
  - What problems do we face?
- Resources:
  - Liu et al., 2019, Linguistic Knowledge and Transferability of Contextual Representations Proceedings of NAACL-HLT 2019
  - Cao et al., 2021, Knowledgeable or Educated Guess? Revisiting Language Models as Knowledge Bases Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics
Questions?

hangyav@cis.lmu.de