Seminar Topics: Information Extraction

English topics!

Alexandra Chronopoulou

achron@cis.lmu.de

Debiasing models used for toxic language detection

- **Toxic language detection**: task of automatically identifying text that is offensive/hateful
- Toxic language primarily targets members of **minority groups**
- Dataset biases, that can be caused by a **problematic data creation process**, create a challenge to detoxifying NLP models
- Target: enable toxic language detection **without suppressing marginalized voices**
- Recent interest in developing **debiasing methods for standard natural language understanding (NLU) tasks**

Debiasing models used for toxic language detection

1. Adversarially remove racial information from text
   - Elazar and Goldberg, 2018, **Adversarial Removal of Demographic Attributes from Text Data**, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*

2. Detection of biases in toxic language
   - Sap et al., 2019, **The Risk of Racial Bias in Hate Speech Detection**, In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*
   - Clark et al., 2019, **Don't Take the Easy Way Out: Ensemble Based Methods for Avoiding Known Dataset Biases**, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*

3. Automated debiasing for toxic language detection
   - Zhou et al., 2021, Challenges in Automated Debiasing for Toxic Language Detection, In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*

Language models become domain experts

- **Language models** (like BERT) are trained on large-scale **open-domain corpora** $\rightarrow$ general language representations
- To perform well in specific, more narrow domains (legal, medical, etc) they need **domain-specific knowledge**
- Does a language model know that "Paracetamol can treat cold"? Yes, if multiple occurrences of the phrase in the pretraining corpus
- What if there are not? One solution: **fine-tuning** – but computationally expensive
- How can we make a language model a **domain expert**?
- **Knowledge graphs** (KGs) serve as a good solution and can be integrated in the LM

Language models become domain experts

1. Learning words and entities using attentive distant supervision
   - Cao et al., 2018, **Joint representation learning of cross-lingual words and entities via attentive distant supervision**, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*

2. Incorporating entities into language models
   - Zhang et al., 2019, **ERNIE: Enhanced Language Representation with Informative Entities**, In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*

3. Knowledge-Enabled Bidirectional Encoder Representation from. Transformers (K-BERT)
   - Liu et al., 2019, **K-BERT: Enabling Language Representation with Knowledge Graph**, In *Proceedings of the AAAI Conference on Artificial Intelligence*

Document-level Relation Extraction

- **Relation Extraction** (RE) is the task of identifying **relational facts** between entities from plain text
- It is important for large-scale **knowledge graph construction**
- RE requires **reading** and **reasoning** over multiple sentences in a document
- Most work focuses on **sentence-level** RE, although at least 40.7% facts sampled from Wikipedia can be extracted only using **multiple** sentences

Document-level Relation Extraction

1. A large document-level relation extraction dataset
   - Yao et al., 2019, **DocRED: A Large-Scale Document-Level Relation Extraction Dataset**, In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*

2. Using hierarchy to extract document-level relations
   - Tang et al., 2020, **HIN: Hierarchical Inference Network for Document-Level Relation Extraction**, In *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*

3. Cross-document mention-level and entity-level graphs to infer relations
   - Zeng et al., 2020, **Double Graph Based Reasoning for Document-level Relation Extraction**, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*

Nested Named Entity Recognition

- **Named entity recognition**: identifying text spans associated with proper names and classifying them according to their semantic class such as *person, organization*, etc
- **Mention detection**: text spans *referring to named, nominal or prominal entities* are identified and classified according to their semantic class
- In the Fig. below, a PERSON named entity is nested in an entity mention of type LOCATION

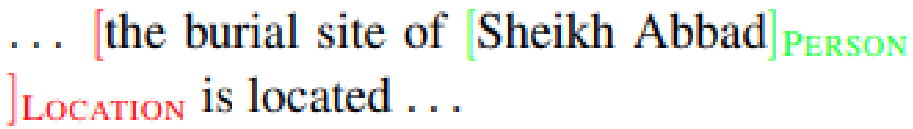

Fig. from Katiyar and Cardie, 2018.

- Most existing methods would **miss the nested entity** - and nested entities are fairly **common**

Nested Named Entity Recognition

1. Mention hypergraph model for nested entity detection
   - Lu and Roth, 2015, **Joint Mention Extraction and Classification with Mention Hypergraphs**, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*
2. Neural network-based methods for *simple* NER
   - Chiu and Nichols, 2016, **Named Entity Recognition with Bidirectional LSTM-CNNs**, In *Transactions of the Association for Computational Linguistics*
   - Lample et al., 2016, **Neural Architectures for Named Entity Recognition**, In *Proceedings of the North American Chapter of the Association for Computational Linguistics*
3. Neural-network based approach for *nested* NER
   - Katiyar and Cardie, 2018, **Nested Named Entity Recognition Revisited**, In *Proceedings of the North American Chapter of the Association for Computational Linguistics*