Embeddings for Word Sense Disambiguation

Proposer(s) / Proposatzailea(k): names / izenak

German Rigau

Contact / Kontaktua: german.rigau@ehu.eus

Description / Deskribapena

Word Sense Disambiguation (WSD) is a long-standing but open problem in Natural Language Processing (NLP). WSD corpora are typically small in size, owing to an expensive annotation process. Current unsupervised vector representations of words and senses offer new challenges and possibilities for WSD ().

Goals / Helburuak

Test different knowledge-based, supervised and unsupervised alternative WSD software packages.

Requirements / Betebeharrak

Basic knowledge of Linux command-line interface:

- execution of programs through the command line
- handling of text files

Framework / Esparrua

There are different systems to try and evaluate on different languages. We will need to decide the systems, corpora, gold-standards and metrics for its empirical evaluation (Vial et al. 2018).

Tasks and plan / Atazak eta plana

- Try IMS+emb (Iacobacci et al., 2016) with different types of embeddings and meta-embeddings, etc.
- Try alternative supervised WSD systems such as EWISE (Kumar et al. 2019), etc.
- Try alternative unsupervised WSD systems based on BERT (Devlin et al. 2019), etc.

References

Kumar, Sawan, Sharmistha Jat, Karan Saxena, and Partha Talukdar. Zero-shot Word Sense Disambiguation using Sense Definition Embeddings. ACL 2019. Florence, Italy.

Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2016. Embeddings for word sense disambiguation: An evaluation study. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 897–907, Berlin, Germany. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics

Vial, Loïc, Benjamin Lecouteux, and Didier Schwab. Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural Word Sense Disambiguation. In Proceedings of 10th Global WordNet Conference (2019). Wroclaw, Poland.

Loïc Vial, Benjamin Lecouteux, and Didier Schwab. 2018. UFSAC: Unification of Sense Annotated Corpora and Tools. In Language Resources and Evaluation Conference (LREC), Miyazaki, Japan, May.