Creating a new Commonsense Reasoning Benchmark

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

German Rigau and Javier Alvez

Contact / Kontaktua: german.rigau@ehu.eus

Description / Deskribapena

Large knowledge-bases and complex ontologies are being used in a wide range of knowledge based systems that require practical commonsense reasoning. For evaluating AdimenSUMO[1], we use a benchmark automatically derived from the world knowledge encoded in WordNet and its mapping into SUMO.

Goals / Helburuak

Develop a new commonsense reasoning benchmark derived from existing datasets by applying already available Word Sense Disambiguation algorithms.

Requirements / Betebeharrak

Basic knowledge of Linux command-line interface:

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

Framework / Esparrua

There are several common sense datasets available including BLESS (Baroni and Lenci, 2011), K&H+N (Necsulescu et al., 2015), EVALution (Santus et al., 2015), Subtask 2 of the CogALex-V shared task (Santus et al., 2016a), ROOT09 (Santus et al., 2016b). There are different Word Sense Disambiguation approaches to test and evaluate: knowledge based such as UKB (Agirre et al. 2013) or SyntagRank (Scozzafava et al. 2020) and neural systems using pre-trained language models (Blevins and Zettlemoyer 2020) or sentence transformers (Reimers and Gurevych 2019).

Tasks and plan / Atazak eta plana

- Prepare the existing datasets for the WSD procedure.
- Try and evaluate existing alternative WSD algorithms.
- Create the new Commonsense Reasoning Benchmark.
- Evaluate the new Commonsense Reasoning Benchmark using AdimenSUMO.

References

Eneko Agirre, Oier Lopez de Lacalle and Aitor Soroa. 2013. Random Walks for Knowledge-Based Word Sense Disambiguation. Computational Linguistics. 40:1. ISSN 0891-2017

Alvez J., P- Lucio and Rigau G. Adimen-SUMO: Reengineering an Ontology for First-Order Reasoning. International Journal on Semantic Web and Information Systems (IJSWIS). Volume 8 (4). IGI Global, USA. 2012.

Álvez J., P. Lucio, and G. Rigau, A framework for the evaluation of SUMO-based ontologies using WordNet, IEEE Access, 7, 36075–36093, (2019).

1https://adimen.si.ehu.es/web/adimenSUMO

M Baroni, A Lenci. How we BLESSed distributional semantic evaluation. Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics, EMNLP 2011, pages 1–10, Edinburgh, Scotland, UK, July 31, 2011

Blevins, Terra, and Luke Zettlemoyer. "Moving Down the Long Tail of Word Sense Disambiguation with Gloss-Informed Biencoders". ACL 2020

Necşulescu, Silvia, Sara Mendes, David Jurgens, Núria Bel, and Roberto Navigli. "Reading between the lines: Overcoming data sparsity for accurate classification of lexical relationships." In Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics, pp. 182-192. 2015.

Enrico Santus, Frances Yung, Alessandro Lenci, and Chu-Ren Huang. 2015. Evalution 1.0: an evolving semantic dataset for training and evaluation of distributional semantic models. In LDL@IJCNLP.

Enrico Santus, Alessandro Lenci, Tin-Shing Chiu, Qin Lu, and Chu-Ren Huang. 2016b. Nine features in a random forest to learn taxonomical semantic relations. In LREC 2016.

Santus, Enrico, Anna Gladkova, Stefan Evert, and Alessandro Lenci. "The CogALex-V shared task on the corpus-based identification of semantic relations." In Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V), pp. 69-79. 2016.

Scozzafava, Federico, Marco Maru, Fabrizio Brignone, Giovanni Torrisi, and Roberto Navigli. "Personalized PageRank with syntagmatic information for multilingual word sense disambiguation." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 37-46. 2020.

Reimers, Nils, and Iryna Gurevych. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3973-3983. 2019.