Casting NLP tasks as Question Answering for Multilingual and Cross-lingual approaches

Proposer(s) / Proposatzailea(k): Rodrigo agerri

Contact / Kontaktua: rodrigo.agerri@ehu.eus

Description / Deskribapena

McCann et al. (2018) introduced the decaNLP challenge, which converts a set of 10 core tasks in NLP to Question Answering, as shown in the following example:

More recently, Coreference Resolution was also reformulated as a Question Answering problem with impressive results (Wu et al. 2020).

Figure 1: An illustration of the paradigm shift from coreference resolution to query-based span prediction.

Previous work has been done for English and we would like to experimentally analyze the benefits of casting tasks as NLP from a crosslingual and multilingual point of view, helping to generate datasets and systems for other languages.

Goals / Helburuak

1. Experiment with existing datasets using deep learning approaches.
2. Re-formulate another popular task as QA (Opinion Mining, etc.)
3. Generate multilingual and cross-lingual systems via automatic translation of existing English datasets.
4. Compare with zero-shot approaches (Conneau et al. 2020) and few-shot approaches.

Tasks and plan / Atazak eta plana

- Month 1: Start of the project, defining the objectives and tasks.
- Month 2: Start experiments. Optionally, it is recommended for the candidates to attend the "Seminar on language technologies. Deep Learning (LAP 18). https://ixa.si.ehu.es/master/programa_html
- Months 3-5: Experiments and final development.
- Final month: Writing up.

References

Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, Richard Socher (2018). The Natural Language Decathlon: Multitask Learning as Question Answering. In arxiv.

Michael, Julian and Luke Zettlemoyer. "Inducing Semantic Roles Without Syntax." *In Findings ACL* (2021).

Schick, Timo and Hinrich Schütze. "True Few-Shot Learning with Prompts -- A Real-World Perspective." (2021). In arxiv.

Wei Wu, Fei Wang, Arianna Yuan, Fei Wu, Jiwei Li (2020). CorefQA: Coreference Resolution as Query-based Span Prediction. In ACL 2020.