Proposal for Exist @ IberLEF 2022

march, 2022

At Exist@IberLEF 2022, the idea to evaluate is the comparison of two strategies for prediction assembling. The first one is to combine the prediction made by state of the art and finetuned language models in different languages (i.e., es, en, fr, de, pt, it). The second strategy is to make prediction augmentation for each example by making back-translation and then combine the prediction over these paraphrases.

For both strategies we employ RoBERTa-based pretrained models and introduce the data form MAMI, which is almost aligned with the current task, HAHA 2021, by analyzing which of the tweets related to targets like woman, men and their combination with other elements such as profession, lgbt or age, are or not sexist. Finally we include the data from HAHACKATHON, by just taking positive examples among the offensive ones and contain wildcards such as femeini*, bitch, woman.

The prediction assembling will be evaluated through a simple majority vote and by taking prediction probability from each prediction source (for both strategies individually) to employing a metaclasifier.

If you have an idea of how to formalize the similarity of a language w.r.t. other, would be interesting to evaluate the combination of similar to Spanish and similar to English for the back-translation and prediction augmentation fashion.

Feedback please.