Multilingual Emotion Detection

Proposers: Rodrigo Agerri

Contact: rodrigo.agerri@ehu.eus

Description

Detecting and classifying emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust), in text has a large number of applications in computational social science and social media (opinion mining, gender bias detection, fake news, etc.).

Labeled training data is a crucial resource required for building supervised machine learning systems. However, compiling human annotated datasets for emotion is expensive in terms of time and money. This is especially true for languages other than English, for which very few datasets exist annotated for emotions (Mohammad et al. 2018).

In this project we propose to use deep learning techniques (Conneau et al. 2020) to investigate emotion detection from a multilingual perspective, including both supervised and unsupervised approaches.

Furthermore, we propose to frame Emotion Detection in Political Discourse or in Twitter, as an auxiliary task to fake news detection.

Objectives

The candidate may choose between the following objectives:

1. Unsupervised emotion detection systems (lexicon-based) (Mohammad and Turney 2010).
2. Experiment with new transfer learning approaches for Emotion Detection in low resource languages (Basque, or any other language of interest for the candidate).
3. Semi-automatic generation of training data for domain (Qadir and Riloff 2013, 2014) and low-resourced languages via bootstrapping and data augmentation.

The master thesis can be written in Basque or English.

Tasks and Plan

- 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

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In ACL 2020.

Mohammad, S. M. and Turney, P. D. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, CAAGET '10, pages 26–34, Stroudsburg, PA, USA. Association for Computational Linguistics.

Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA, June 2018.

Qadir, A. and Riloff, E. (2013). Bootstrapped learning of emotion hashtags #hashtags4you. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 2–11, Atlanta, Georgia. Association for Computational Linguistics.

Qadir, A. and Riloff, E. (2014). Learning emotion indicators from tweets: Hashtags, hashtag patterns, and phrases. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1203–1209, Doha, Qatar. Association for Computational Linguistics.