IIIDYT at IEST 2018: Implicit Emotion Classification With Deep Contextualized Word Representations

Anonymous EMNLP submission

Abstract

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est. Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

1 Introduction

- High level description of the task (Klinger et al., 2018) and its *raison d'être* (Francés para Edison)
- High level description of the model used and results obtained

2 Related Work

- briefly describe previous Wassa papers (Baziotis et al., 2018; Duppada et al., 2018; Abdou et al., 2018) (Should we do this despite not having based our work in any of them?)
- Talk about elmo (Peters et al., 2018)

3 Proposed Model

3.1 Preprocessing

- Replaced `[#TRIGGERWORD#]`, `@USERNAME`, `[NEWLINE]`, and `http://url.removed`, by `__TRIGGERWORD__`, `__USERNAME__`, `__NEWLINE__`, and `__URL__` respectively.
- Tokenized by using an emoji-aware modified version of the `twokenize.py`[1] script.

To incorporate emoji knowledge we used an emoji database[2] which we also modified for avoiding conflict with emoji sharing unicode codes with common glyphs used in twitter such as the hashtag symbol. (Link to appendix for specifics; mention `keycaps`)

3.2 Model Architecture

- Elmo (Peters et al., 2018) as word representations
- BiLSTM (Graves and Schmidhuber, 2005; Graves et al., 2013) as context fine-tuner (?)
- max-pooling as word aggregation method
- slanted triangular learning rate schedule with mostly default params (just set the η_{max} parameter to 0.001.) as lr schedule (Howard and Ruder, 2018)
- 0.5 dropout after word layer, 0.1 after sentence layer (Srivastava et al., 2014)
- Only external feature is the elmo pre-trained language model

3.3 Ensembling

We trained several models with the previously-mentioned architecture, but different initial parameters (by changing the random seed). We fed them the test data, and averaged the predicted output probabilities. We tried all the possible combinations of averaging 2 to 10 models, and found out that a specific combination of 6 models yielded the best results (show bar graph). This provides evidence for the fact that having the same number of independent classifiers as class labels provides the best results (Bonab and Can, 2016).

[1] https://github.com/myleott/ark-twokenize-py

[2] https://github.com/carpedm20/emoji/blob/e7bff32/emoji/unicode_codes.py

4 (Failed) experiments

- scalar gates, vector gates and elmo
- dropouts?
- POS tags
- combined pooling
- different learning rate schedules
- emoji vs no emoji
- sentence encoding lstm hidden size
- transformer architecture
- show different accuracies when using different training sizes?

5 Conclusions and Future Work

References

Mostafa Abdou, Artur Kulmizev, and Joan Ginés i Ametllé. 2018. Affecthor at semeval-2018 task 1: A cross-linguistic approach to sentiment intensity quantification in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 210–217, New Orleans, Louisiana. Association for Computational Linguistics.

Christos Baziotis, Athanasiou Nikolaos, Alexandra Chronopoulou, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, Shrikanth Narayanan, and Alexandros Potamianos. 2018. Ntua-slp at semeval-2018 task 1: Predicting affective content in tweets with deep attentive rnns and transfer learning. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 245–255, New Orleans, Louisiana. Association for Computational Linguistics.

Hamed R. Bonab and Fazli Can. 2016. A theoretical framework on the ideal number of classifiers for online ensembles in data streams. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, pages 2053–2056, New York, NY, USA. ACM.

Venkatesh Duppada, Royal Jain, and Sushant Hiray. 2018. Seernet at semeval-2018 task 1: Domain adaptation for affect in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 18–23, New Orleans, Louisiana. Association for Computational Linguistics.

Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech Recognition with Deep Recurrent Neural Networks. In *Proceedings of the 2013 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6645–6649, Vancouver, Canada. IEEE.

Alex Graves and Jürgen Schmidhuber. 2005. Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. *Neural Networks*, 18(5-6):602–610.

J. Howard and S. Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. *ArXiv e-prints*.

Roman Klinger, Orphée de Clercq, Saif M. Mohammad, and Alexandra Balahur. 2018. Iest: Wassa-2018 implicit emotions shared task. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Brussels, Belgium. Association for Computational Linguistics.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958.