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

Sentiment Analysis, also known as Opinion Mining, is a discipline whose objective is to automatically identify sentiment in written text (Balazs and Velásquez, 2016).

The WASSA 2018 Implicit Emotion Shared Task (IEST) (Klinger et al., 2018), consists in predicting the emotion of a word that has been excluded from a tweet. Its aim is to find ways to automatically learn the link between situations and the emotion they trigger.

This problem can be reduced to sentence classification, in which we want to classify a sentence, or more specifically a tweet, into one of several categories.

In this paper we present the system that obtained second place, proposed by the team IIIDYT at IEST. Our system was composed by a single pre-trained ELMo layer for encoding words (Peters et al., 2018), a Bidirectional Long-Short Memory Network (BiLSTM) (Graves and Schmidhuber, 2005; Graves et al., 2013), for enriching word representations with context, a max-pooling operation for creating sentence representations from said word vectors, and finally a Dense Layer for projecting the sentence representations into label space (See Figure 3).

2 Related Work

- briefly describe previous Semeval 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?)
- mention previous Wassa papers
- talk about some sentence classification tasks? Works that fall into the broad category of mapping sentences to labels.
- 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

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

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

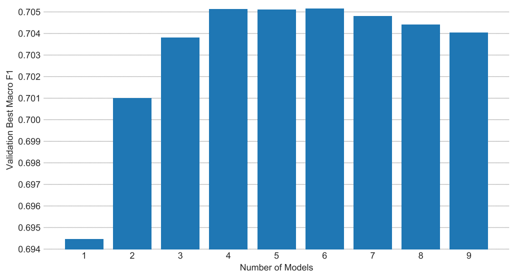

Figure 1: Best Macro F1 vs. Number of Ensembled Models

- 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).

4 (Failed) experiments

- scalar gates, vector gates and elmo
- dropouts?
- POS tags
- combined pooling
- different learning rate schedules
- emoji vs no emoji

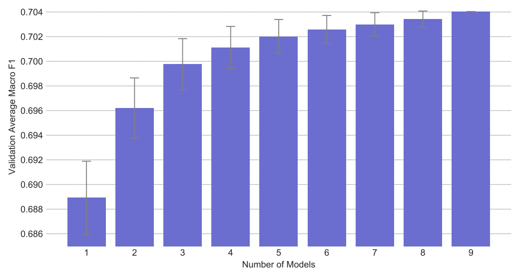

Figure 2: Best Macro F1 vs. Number of Ensembled Models

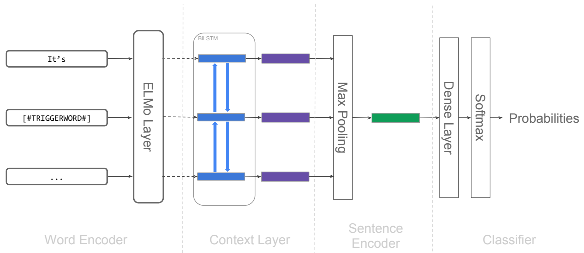

Figure 3: Proposed architecture

- sentence encoding lstm hidden size
- transformer architecture
- show different accuracies when using different training sizes?

Variation	**Accuracy (%)**	**Δ%**
Submitted	**69.23**	-
No emoji	68.36	- 0.87
No ELMo	65.52	- 3.71
Concat Pooling	68.47	- 0.76
LSTM hidden=4096	69.10	- 0.13
LSTM hidden=1024	68.93	- 0.30
LSTM hidden=512	68.43	- 0.80
POS emb dim=100	68.99	- 0.24
POS emb dim=75	68.61	- 0.62
POS emb dim=50	69.33	+ 0.10
POS emb dim=25	69.21	- 0.02

Table 1: Ablation study results

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.

Jorge A. Balazs and Juan D. Velásquez. 2016. Opinion mining and information fusion: A survey. *Information Fusion*, 27:95 – 110.

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.

Jeremy Howard and Sebastian 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.