

BERT vs. XLNet in Multilabel Text Classification

Gaoussou Youssouf Kebe, Cynthia Matuszek, Francis Ferraro

Department of Computer Science and Electrical Engineering

University of Maryland, Baltimore County

{gaoussoul, cmat, ferraro}@umbc.edu

Abstract

The complex nature of emotions makes it also one of the hardest text classification tasks. We present a comparative study of state of the art language representation models XLNet and Bert in sentiment analysis, specifically multi-label classification of tweets among 6 basic emotions (anger, disgust, fear, joy, sadness and surprise).

1 Background & Related Work

Giving computers the ability to recognize human emotion is a long-standing goal of Artificial Intelligence. Despite the multi-modal nature of emotion recognition, natural language processing has established itself as one of the most popular ways to tackle the problem. The complex nature of emotions makes it also one of the hardest text classification tasks. In this work, we compare two state of the art text classification models XLNet and Bert in a multilabel twitter emotion analysis task.

Transformers (Vaswani et al., 2017) are a novel type of sequence transduction model that forgoes recurrence and relies entirely on attention, thereby considering more global relationships in longer input and output sequences. Consequently, they have been also been used in recent attempts to tackle the emotion analysis problem. BERT (Devlin et al., 2018) introduces a masked bidirectional language modeling objective where context learned from both directions is used to predict randomly masked tokens. XLNet (Yang et al., 2019) takes this concept farther by considering different permutations of the tokens in the input sequence.

Balazs et al. (2018) use a pre-trained ELMo (Peters et al., 2018) model (among other components) to classify tweets Luo and Wang (2019) and use a pre-trained BERT model to classify dialogues from the TV show Friends and Facebook chat logs. In contrast to our work, both assign a document to

a single label. Our work is closest to (Ying et al., 2019) where a pre-trained BERT model is used to classify tweets from the same SemEval-2018 Task 1 dataset used in our work.

2 Methodology

The labels in our work are based on the Ekman model (Ekman, 1999) of six basic emotions: anger, disgust, fear, happiness, sadness, and surprise. We use the SemEval-2018 Task 1: Affect in Tweets (Mohammad et al., 2018) dataset. The data is initially pre-processed by adding the prefix NOT_ to every words in a sentence that comes after negation words and deleting all links and mentions. Stop words like pronouns and articles are also filtered out. Both models (XLNet and Bert) are fine-tuned and trained on the training set. The output of the fine-tuned models are turned into probabilities using a sigmoid function. A fixed threshold is used for multi-label classification, and a label is attributed to a tweet only if the probability exceeds that threshold. After 10-fold cross validation is done on the training set to optimally set both models up, the following results are obtained:

	XLNet		Bert	
	Micro	Macro	Micro	Macro
Precision	0.790	0.771	0.756	0.740
Recall	0.714	0.655	0.701	0.617
F1 score	0.750	0.691	0.727	0.644

Table 1: Initial results (XLNet vs Bert)

3 Discussion & Future Work

In this work, we demonstrate that XLNet outperforms Bert on the problem of multilabel emotion analysis. We expect that this difference would apply to other similar multilabel text classification problems, and will explore this in future work.

References

- Jorge A. Balazs, Edison Marrese-Taylor, and Yutaka Matsuo. 2018. [IIDYT at IEST 2018: Implicit emotion classification with deep contextualized word representations](#). *CoRR*, abs/1808.08672.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. [BERT: pre-training of deep bidirectional transformers for language understanding](#). *CoRR*, abs/1810.04805.
- Paul Ekman. 1999. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16.
- Linkai Luo and Yue Wang. 2019. [Emotionx-hsu: Adopting pre-trained BERT for emotion classification](#). *CoRR*, abs/1907.09669.
- Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.
- 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.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). *CoRR*, abs/1706.03762.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*.
- Wenhao Ying, Rong Xiang, and Qin Lu. 2019. [Improving multi-label emotion classification by integrating both general and domain-specific knowledge](#). In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 316–321, Hong Kong, China. Association for Computational Linguistics.