Combining BERT with Static Word Embeddings for Categorizing Social Media

Israa Alghanmi, Luis Espinosa-Anke, Steven Schockaert

Cardiff University, UK

{alghanmiia,espinosa-ankel,schockaerts1}@cardiff.ac.uk

Abstract

- Pre-trained neural language models (LMs) have achieved impressive results in various natural language processing tasks, across different languages.
- A striking example is the performance of AraBERT, an LM for the Arabic language, which is successful in categorizing social media posts in Arabic dialects, despite only having been trained on Modern Standard Arabic.
- Our hypothesis is that the performance of LMs for social media can nonetheless be improved by incorporating static word vectors that have been specifically trained on social media.
- We show that a simple method for incorporating such word vectors is indeed successful in several Arabic and English benchmarks.
- We also find that similar improvements are possible with word vectors that have been trained on traditional text sources (e.g. Wikipedia)

Motivation

- Obtaining complementary strength by combining BERT that was trained on Wikipedia with static vectors trained on social media
- Training static embeddings is not as expensive as pretraining LMs
- Dealing with emerging terms and new vocabulary

Methodology

Challenge

- Directly combining BERT with static word vectors is not possible due to different tokenization strategies

Approach

- Combining representations obtained from BERT and static word vectors at sentence level.

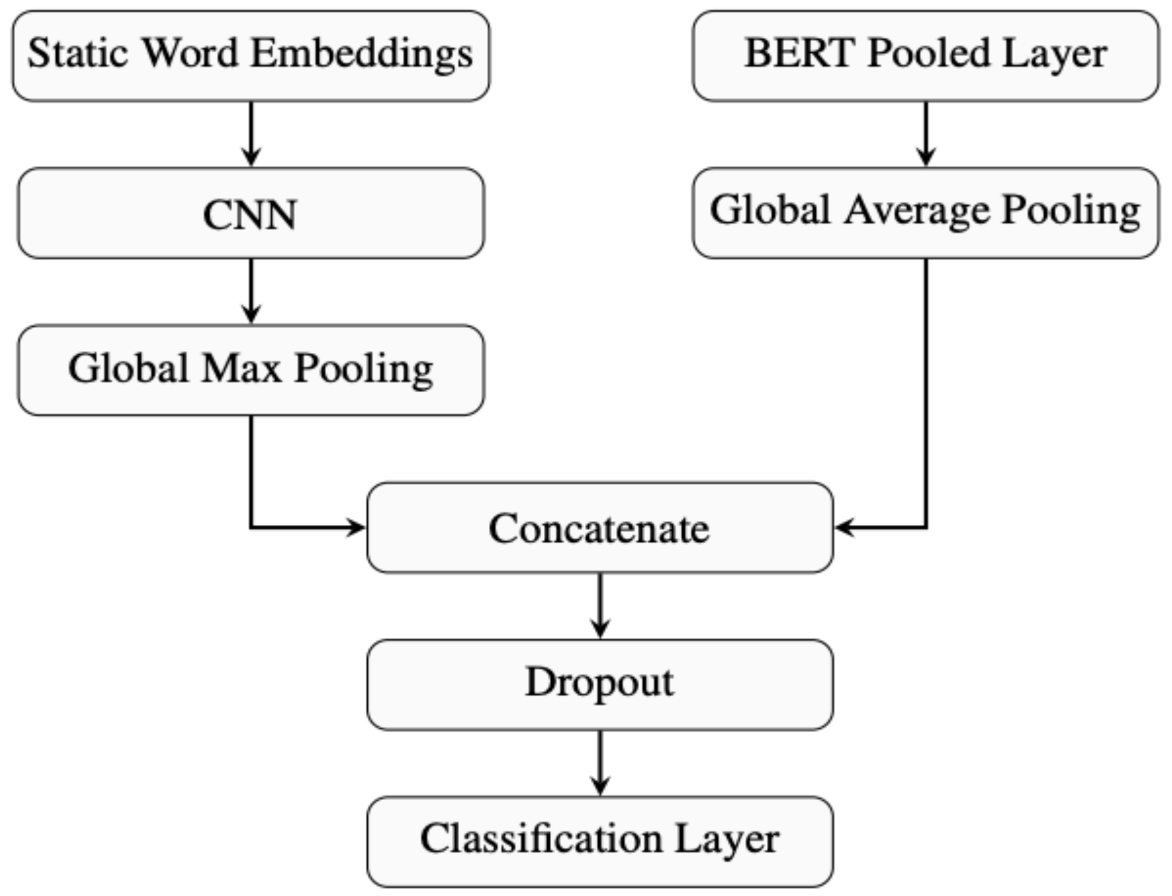

Figure 1: The CNN based proposed architecture

1. Fine-tune BERT model on its own
2. Extract the output from the last BERT layer and freeze BERT model
3. Train the CNN and combined classification layer

Results

Model	Embeddings	AJGT	SemEval-2017	L-HSAB	ArsenTD-Lev
CNN	AraVec-twi	89.9	55.6	60.9	47.5
AraBERT	-	93.3	63.1	71.0	51.8
CNN + AraBERT	AraVec-twi	**93.4** (93.0, 93.9)	64.1 (63.5, 64.4)	**72.1** (71.7, 72.3)	49.2 (47.1, 51.5)
LSTM + AraBERT	AraVec-twi	93.1 (92.8, 93.6)	63.7 (63.3, 64.1)	72.0 (71.5, 72.5)	**52.2** (51.7, 52.7)
CNN + AraBERT	AraVec-wiki	**93.4** (92.8, 93.6)	**64.3** (63.9, 64.5)	71.8 (71.6, 72.1)	50.9 (48.0, 53.1)
LSTM + AraBERT	AraVec-wiki	93.1 (92.8, 93.6)	63.7 (63.4, 63.9)	71.9 (71.5, 72.5)	51.9(51.5, 52.5)

F1 scores (%) for Arabic datasets

Model	Embeddings	Irony	OffensEval	Hate	Stance
CNN	GloVe-twi	57.2	75.1	47.0	29.0
BERT	-	67.3	78.5	**50.6**	52.9
CNN + BERT	GloVe-twi	**68.4** (67.4, 69.7)	79.4 (79.2, 79.7)	48.1 (47.6, 48.5)	54.3 (52.9, 56.0)
LSTM+ BERT	GloVe-twi	68.3 (67.6, 68.9)	79.5 (79.2, 79.8)	47.7 (47.5, 47.8)	54.1 (53.1, 55.5)
CNN + BERT	GloVe-wiki	67.7 (66.5, 68.7)	79.4 (79.0, 79.6)	48.1 (47.8, 48.3)	**54.6** (53.6, 55.5)
LSTM+ BERT	GloVe-wiki	68.3 (67.6, 68.9)	**79.6** (79.2, 79.8)	47.7 (47.5, 47.8)	54.1 (53.1, 55.5)

F1 scores (%) for English datasets

Conclusion

Adding static vectors has benefits in many cases, even when these vectors have been trained on Wikipedia

This simple approach could be an alternative to continuously updating language models when dealing with emerging terms, such as trending hashtags and a promising strategy for improving language models for low-resource languages.

GitHub Repository

https://github.com/israa-alghanmi/BERT-GloVe

References

- *Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.*
- *Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543.*
- *Abu Bakr Soliman, Kareem Eissa, and Samhaa R ElBeltagy. 2017. Aravec: A set of arabic word embedding models for use in arabic nlp. Procedia Computer Science, 117:256–265*
- *Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. Arabert: Transformer-based model for arabic language understanding.*
- *Zhuosheng Zhang, Yuwei Wu, Hai Zhao, Zuchao Li, Shuailiang Zhang, Xi Zhou, and Xiang Zhou. 2020. Semantics-aware BERT for language understanding. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, pages 9628–9635.*