An Exploration of Contextualized Word Vectors for Sentiment Analysis

Adam Hare, Lisa Schut, Hans Hanley, Aditya Agarwal

Department of Computer Science, University of Oxford, Oxford, UK

DEPARTMENT OF COMPUTER SCIENCE

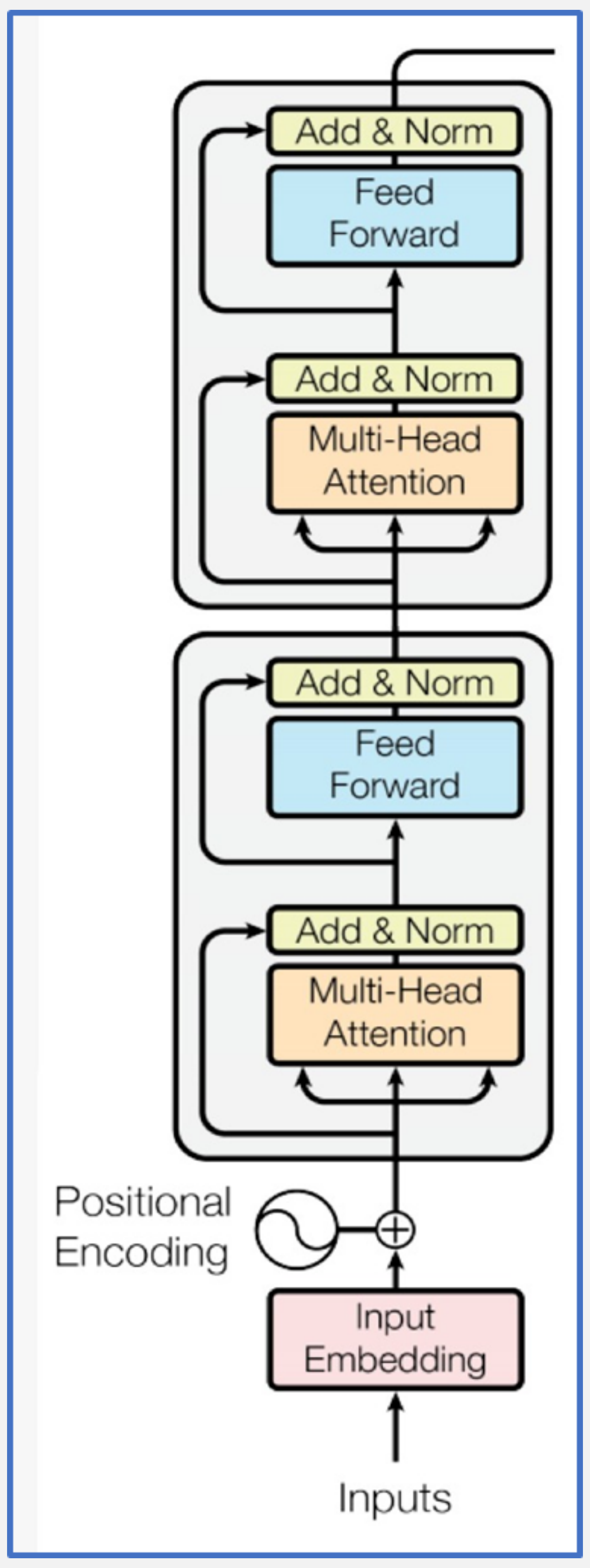
Abstract

Word embeddings that capture context-specific information have shown promise in downstream Natural Language Processing tasks. High quality sequence-to-sequence machine translation learns contextual information from encoder inputs, making it an ideal candidate for transfer learning. This work focuses on utilizing Transformer and CNN machine translation models to learn contextualized word representations for use in sentiment classification. By using positional embeddings and bi-attentive classification networking with the CNN and the Transformer, we see better results on the Stanford Sentiment Treebank (SST) compared to non-contextualized word embeddings. We achieve up to 4.71% increase in accuracy for CNN on SST-2 and up to 4.25% increase in accuracy for Transformer on SST-5. Compared to contextualized word representations trained using recurrent neural networks (CoVe), our approach is more computationally efficient.

Background

- Since 2014, the sequence-to-sequence models have become the state of the art in machine translation.
- McCann et al. (2017) use a LSTM encoder-decoder model with Luong attention to perform neural machine translation.
- Using the encoder hidden states, McCann et al. (2017) generated contextualized word embeddings to improve performance in downstream tasks like sentiment analysis.

Methods

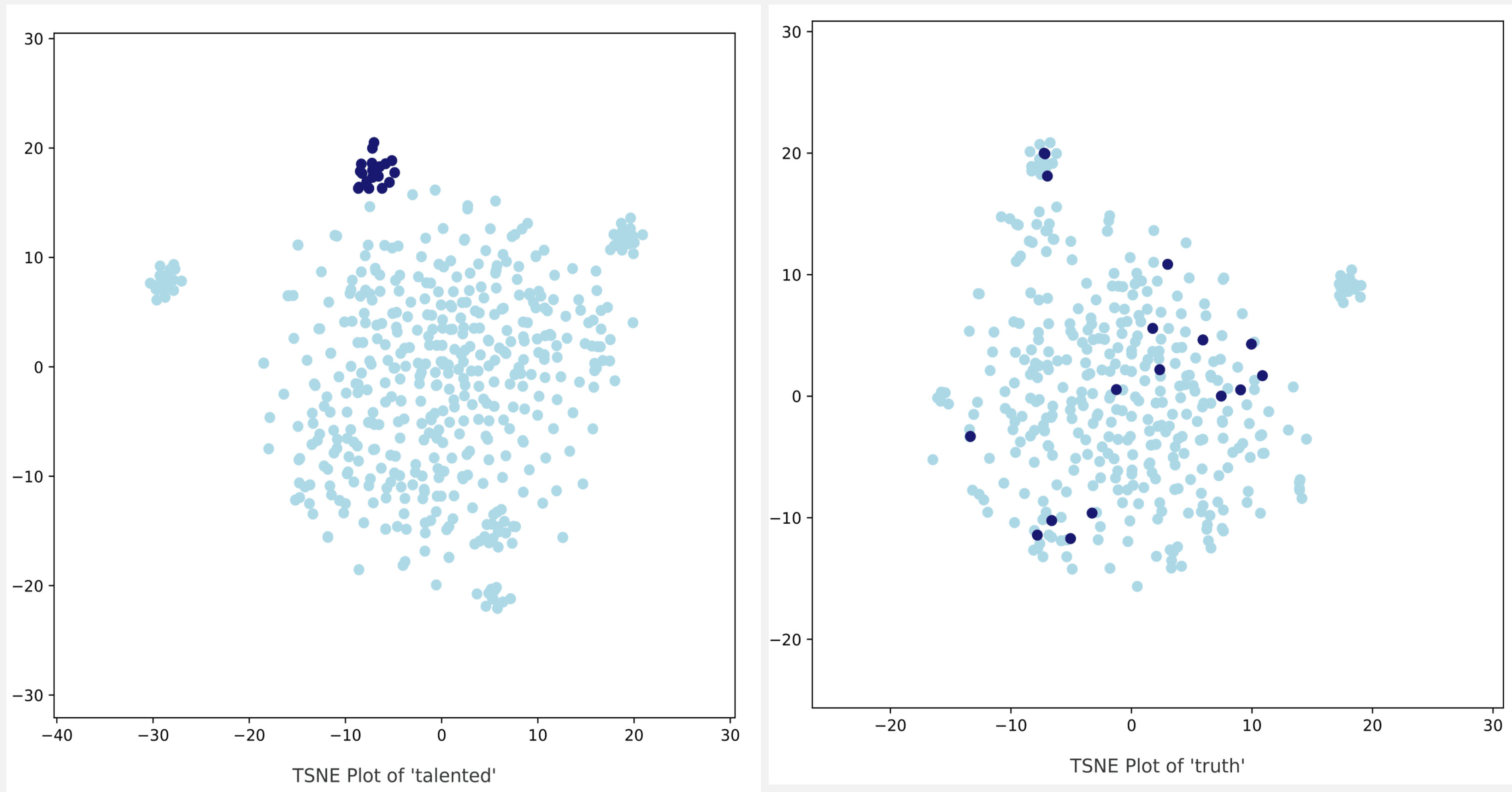


- We train Transformer and CNN based encoder architectures for machine translation to obtain contextualized word embeddings.
- Both models are parallelizable unlike the LSTM based encoders used by McCann et al. (2017).
- We use positional embeddings in both the models to incorporate temporal information from the sentence.
- Transformer, in particular, makes use of multi-headed attention [Vaswani et al. 2017].
- The decoder for the machine translation uses a RNN based architecture.
- We train the translation models on WMT-2016 (English-German).
- Finally, we use a bi-attentive classification network [McCann et al.] to perform sentiment analysis.

Our implementation is available at
https://github.com/adi2103/AML-CoVe

Visualizing Contextualized Embeddings

- We generated TSNE plots for CoVe using scikit-learn.
- CoVe embeddings are context dependent and rarely the exact same.
- Phrases in a sentence sometimes obtain similar embeddings.
- Words whose semantic meaning depends more on the context show more variation in their embeddings (e.g. "truth").
- Words whose semantic meaning is context independent show a smaller variation in their embeddings (e.g. "talented").



Results - Stanford Sentiment Treebank

Embeddings	SST-2		SST-5	
	Validation	Test	Validation	Test
Random	77.46	76.30	34.43	34.04
Glove	81.50	78.30	46.81	45.92
CoVe	81.07	80.29	45.16	47.07
Transformer	80.35	67.98	48.54	45.48
CNN	81.50	80.67	43.09	42.15
Transformer + CoVe	81.65	74.35	48.80	46.10
CNN + CoVe	82.08	82.54	44.19	42.82

Conclusions

- The results suggest that including contextual embeddings may improve performance on sentiment classification.
- The Transformer and CoVe performed the best on the SST-5. The Transformer and CNN performances improved when combined with the CoVe. This suggests that they learn additional underlying information from the machine translation datasets.
- The Transformer and CNN benefit from being nonsequential, allowing for more efficient training.

Works Cited

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.
- Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.
- McCann, B., Bradbury, J., Xiong, C., & Socher, R. (2017). Learned in translation: Contextualized word vectors. In Advances in Neural Information Processing Systems (pp. 6294-6305).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).