Research Proposal

Nirbhay P. Tandon
968675
Email: N.P.Tandon@2021.ljmu.ac.uk
Project Supervisor: Mr. Ankit Jha

Abstract

Attention based Transformer architectures have become the norm of modern day Natural Language Processing. Google began this trend back in 2017 with their paper *Attention Is All You Need*[1], by introducing the Transformer architecture that works solely on attention mechanisms. The purpose of our work will be to explore a new kind of Transformer architecture. Compare & contrast its performance against the SQuAD 2.0 Dataset[2] based other architectures such as BERT[3], RoBERTa[4], etc. **Add additional details around how neural nets & lstms compare to transformers. Fill in the final objective of the research proposal**

Contents

3 Research Questions

Highlighted below are some of the questions that we will answer through our research.

4 Aims & Objectives

Through our research we aim to establish the efficiency of our new Transformer architecture. We shall implement the existing models that are available via libraries such as HuggingFace[5], PyTorch & Tensorflow, run the SQuAD 2.0[2], to obtain benchmark scores & then compare the results with our proposed architecture. We hope to establish our proposed transformer architecture as a competent enough contender to be used within both industry & academia.

5 Research Methodology

To implement this research we shall break the project down into 5 phases. These are outlined below.

5.1 Literature Review

In this phase we will review the research that has been published already around the different kinds of architectures, shortlist some of the most widely used ones, compare their results using the SQuAD2.0 [2] & outline the pros and cons of each of these architectures. The rationale is to review & understand as much of the research as possible so that we can avoid potential pitfalls, not duplicate our efforts by reinventing the wheel & organize a better approach to perform our research. ***to add some pointers about how existing research has been done***

5.2 Research Benchmarks

Here we shall focus on obtaining benchmark scores for the shortlisted architectures above using the dataset[2]. The parameters used will be validation & test set scores of the models. The training shall be carried out on each of the models for a 100(***TBD, 100 epochs per model for Squad will take over 100 hours of training, not sure if its worth it***) epochs. We shall also look at the specificity/recall of these results to better understand if our work was done correctly or not.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, Aidan N, Kaiser, L. and Polosukhin, I. (2017). Attention Is All You Need. [online] arXiv.org. Available at: https://arxiv.org/abs/1706.03762.

[2] Rajpurkar, P., Jia, R. and Liang, P., 2018. Know what you don't know: Unanswerable questions for SQuAD. arXiv preprint arXiv:1806.03822.

[3] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[4] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[5] Huggingface.co. 2021. Transformers — transformers 4.4.2 documentation. [online] Available at: ¡https://huggingface.co/transformers/¿ [Accessed 4 April 2021].

[6] Hochreiter, S., Bengio, Y., Frasconi, P. and Schmidhuber, J., 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

[7] Akbik, A., Blythe, D. and Vollgraf, R., 2018, August. Contextual string embeddings for sequence labeling. In Proceedings of the 27th international conference on computational linguistics (pp. 1638-1649).

[8] Al-Rfou, R., Choe, D., Constant, N., Guo, M. and Jones, L., 2019, July. Character-level language modeling with deeper self-attention. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 3159-3166).

[9] Zhou, J., Cao, Y., Wang, X., Li, P. and Xu, W., 2016. Deep recurrent models with fast-forward connections for neural machine translation. Transactions of the Association for Computational Linguistics, 4, pp.371-383.

[10] Luong, M.T., Le, Q.V., Sutskever, I., Vinyals, O. and Kaiser, L., 2015. Multi-task sequence to sequence learning. arXiv preprint arXiv:1511.06114.

[11] Bahdanau, D., Cho, K. and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[12] Zhang, Y., Zhong, V., Chen, D., Angeli, G. and Manning, C.D., 2017, September. Position-aware attention and supervised data improve slot filling. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 35-45).

[13] Levy, O., Seo, M., Choi, E. and Zettlemoyer, L., 2017. Zero-shot relation extraction via reading comprehension. arXiv preprint arXiv:1706.04115.

[14] Yih, S.W.T., Chang, M.W., Meek, C. and Pastusiak, A., 2013. Question answering using enhanced lexical semantic models.

[15] Weissenborn, D., Wiese, G. and Seiffe, L., 2017. Making neural qa as simple as possible but not simpler. arXiv preprint arXiv:1703.04816.

[16] Graves, A., 2013. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.

[17] Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural networks, 61, pp.85-117.

[18] Chung, M.I., Kushner, W. and Damoulakis, J., 1985, April. Word boundary detection and speech recognition of noisy speech by means of iterative noise cancellation techniques. In ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 10, pp. 1838-1838). IEEE.

[19] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L., 2018. Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

[20] Zhang, Z., Yang, J. and Zhao, H., 2020. Retrospective reader for machine reading comprehension. arXiv preprint arXiv:2001.09694.

[21] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. and Soricut, R., 2019. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

[22] Baevski, A. and Auli, M., 2018. Adaptive input representations for neural language modeling. arXiv preprint arXiv:1809.10853.

[23] Zhang, Z., Wu, Y., Zhou, J., Duan, S., Zhao, H. and Wang, R., 2020, April. Sg-net: Syntax-guided machine reading comprehension. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 05, pp. 9636-9643).