BERT VISION

Improving span annotation and classification task performance using parameter-efficient model architectures trained on BERT's hidden state activations

Siduo Jiang, William Casey King, Cristopher Benge

July 21, 2020

Submitted in partial fulfillment of the requirements
for the final project of MIDS W266
to the
Faculty of Graduate Studies
of the School of Information
at the University of California, Berkeley

Project GitHub Repository: BERTVision

Publication submitted to the graduate studies faculty of the University of California, Berkeley for evaluation and consideration: *Daniel Cer PhD, Joachim Rahmfeld, Mark Butler, Sid J Reddy PhD, Mike Tamir PhD, Drew Plant, Gurdit Chahal*

This document was typeset using LaTeX, with a mixture of `classicthesis` developed by André Miede. The bibliography was processed by Biblatex. Robert Slimbach's Minion Pro acts as both the text and display typeface. Sans-serif text is typeset in Slimbach and Carol Twombly's Myriad Pro; monospaced text uses Jim Lyle's `Bitstream Vera Mono`.

CONTENTS

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce maximus nisi ligula. Morbi laoreet ex ligula, vitae lobortis purus mattis vel. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Donec ac metus ut turpis mollis placerat et nec enim. Duis tristique nibh maximus faucibus facilisis. Praesent in consequat leo. Maecenas condimentum ex rhoncus, elementum diam vel, malesuada ante.

1.1 Introduction

This sentence requires citation [1]. This sentence requires multiple citations to imply that it is better supported [2], [3]. Finally, when conducting an appeal to authority, it can be useful to cite a reference in-text, much like [4] do quite a bit. Oh, and make sure to check out the bear in Figure 1.1.

$$A = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{21} & A_{22} & A_{32} \end{bmatrix} \tag{1.1}$$

Figure 1.1: *Figure Example*

1.1.1 NLP Tasks

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce maximus nisi ligula. Morbi laoreet ex ligula, vitae lobortis purus mattis vel. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Donec ac metus ut turpis mollis placerat et nec enim. Duis tristique nibh maximus faucibus facilisis. Praesent in consequat leo. Maecenas condimentum ex rhoncus, elementum diam vel, malesuada ante. Fusce pulvinar, mauris pretium placerat venenatis, lectus ex tempus lacus, id suscipit libero lorem eu augue. Interdum et malesuada fames ac ante ipsum primis in faucibus.

Donec nec nibh sagittis, finibus mauris quis, laoreet augue. Maecenas aliquam sem nunc, vel semper urna hendrerit nec. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Maecenas pellentesque dolor lacus, sit amet pretium felis vestibulum finibus. Duis tincidunt sapien faucibus nisi vehicula tincidunt. Donec euismod suscipit ligula a tempor. Aenean a nulla sit amet magna ullamcorper condimentum. Fusce eu velit vitae libero varius condimentum at sed dui.

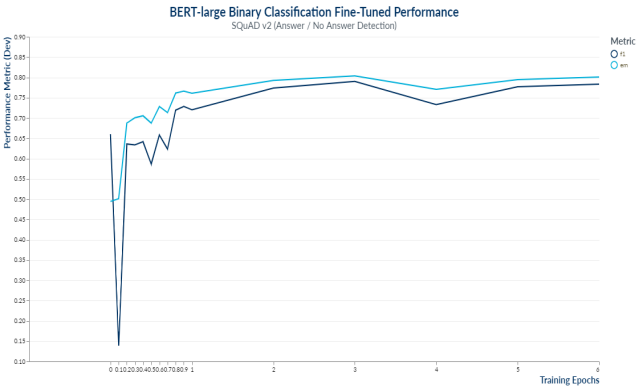

Figure 1.2: *Image Figure Example*

BACKGROUND

Space: the final frontier. These are the voyages of the starship Enterprise. Its continuing mission: to explore strange new worlds. To seek out new life and new civilizations. To boldly go where no one has gone before!

2.1 Going Deeper

This sentence requires citation [1]. This sentence requires multiple citations to imply that it is better supported [2], [3]. Finally, here is an example of a table:

Table 2.1: *Model : RESET test and Breusch-Pagan test p-values (Naive)*

RESET (power=2)	Breusch-pagan
0.636	0.017

Here, now, I've referenced the table 2.1! You see, not that bad really.

Shall we try a list? We shall!

- PRBARR : 'probability' of arrest.
- PRBCONV : 'probability' of conviction.
- PRBPRIS : 'probability' of prison sentence.
- AVGSEN : average sentence in days.
- POLPC : police per capita.

METHODS

We're no strangers to love. You know the rules and so do I. A full commitment's what I'm thinking of. You wouldn't get this from any other guy. I just wanna tell you how I'm feeling. Gotta make you understand.

Never gonna give you up, Never gonna let you down, Never gonna run around and desert you. Never gonna make you cry, Never gonna say goodbye, Never gonna tell a lie and hurt you.

We've known each other for so long. Your heart's been aching but you're too shy to say it. Inside we both know what's been going on. We know the game and we're gonna play it. And if you ask me how I'm feeling, don't tell me you're too blind to see.

Never gonna give you up, Never gonna let you down, Never gonna run around and desert you. Never gonna make you cry, Never gonna say goodbye, Never gonna tell a lie and hurt you. Never gonna give you up, Never gonna let you down, Never gonna run around and desert you.

Never gonna make you cry, Never gonna say goodbye, Never gonna tell a lie and hurt you. Never gonna give, never gonna give (Give you up). (Ooh) Never gonna give, never gonna give (Give you up).

We've known each other for so long. Your heart's been aching but you're too shy to say it. Inside we both know what's been going on. We know the game and we're gonna play it. I just wanna tell you how I'm feeling. Gotta make you understand.

Never gonna give you up, Never gonna let you down, Never gonna run around and desert you. Never gonna make you cry, Never gonna say goodbye, Never gonna tell a lie and hurt you. Never gonna give you up, Never gonna let you down, Never gonna run around and desert you. Never gonna make you cry, Never gonna say goodbye, Never gonna tell a lie and hurt you. Never gonna give you up, Never gonna let you down, Never gonna run around and desert you, Never gonna make you cry.

3.1 Tools of the Trade

This sentence requires citation [1]. This sentence requires multiple citations to imply that it is better supported [2], [3]. I'm just adding some random text here to make you feel good... I certainly that it's working.

RESULTS

We're talking away, I don't know what I'm to say - I'll say it anyway; Today's another day to find you. Shying away; I'll be coming for your love, okay?
Take on me (take on me), Take me on (take on me), I'll be gone In a day or two.
So needless to say, I'm odds and ends But I'll be stumbling away. Slowly learning that life is okay. Say after me, "It's no better to be safe than sorry"
Take on me (take on me), Take me on (take on me), I'll be gone In a day or two

4.1 Conclusion

This sentence requires citation [1]. This sentence requires multiple citations to imply that it is better supported [2], [3]. I'm just adding some random text here to make you feel good... I certainly that it's working.

4.2 Future Work

This sentence requires citation [1]. This sentence requires multiple citations to imply that it is better supported [2], [3]. I'm just adding some random text here to make you feel good... I certainly that it's working.

BIBLIOGRAPHY

[1] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, ISBN: 9781950737130. arXiv: 1810.04805.

[2] Y. Kim, Y. Jernite, D. Sontag, and A. M. Rush, "Character-Aware neural language models," in *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, 2016, ISBN: 9781577357605. arXiv: 1508.06615.

[3] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuad: 100,000+ questions for machine comprehension of text," in *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2016, ISBN: 9781945626258.

[4] I. Tenney, D. Das, and E. Pavlick, "BERT rediscovers the classical NLP pipeline," in *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 2020, ISBN: 9781950737482. DOI: 10.18653/v1/p19-1452. arXiv: 1905.05950.

[5] M. Lin, Q. Chen, and S. Yan, "Network in network," in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017.

[7] A. Conneau, H. Schwenk, Y. L. Cun, and L. Barrault, "Very deep convolutional networks for text classification," in *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, 2017, ISBN: 9781510838604. DOI: 10.18653/v1/e17-1104.

[8] A. R. Kuefler. (2016). "Merging Recurrence and Inception-Like Convolution for Sentiment Analysis," [Online]. Available: https://cs224d.stanford.edu/reports/akuefler.pdf (visited on 06/17/2020).

[9] G. Limaye, M. Pandit, and S. Vinay. (2019). "BertNet: Combining BERT language representation with Attention and CNN for Reading Comprehension," [Online]. Available: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/default/15783457.pdf (visited on 06/17/2020).

[10] D. Takeuchi and K. Tran. (2019). "Improving SQUAD 2.0 Performance using BERT + X," [Online]. Available: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/default/15737384.pdf (visited on 06/17/2020).

[11] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, *Stand-alone self-attention in vision models*, 2019. arXiv: 1906.05909 [cs.CV].

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

[13] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *CoRR*, vol. abs/1610.02357, 2016. arXiv: 1610.02357. [Online]. Available: http://arxiv.org/abs/1610.02357.

[14] X. Ma, Z. Wang, P. Ng, R. Nallapati, and B. Xiang, *Universal text representation from bert: An empirical study*, 2019. arXiv: 1910.07973 [cs.CL].

[15] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. de Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for NLP," *CoRR*, vol. abs/1902.00751, 2019. arXiv: 1902 . 00751. [Online]. Available: http://arxiv.org/abs/1902.00751.

[16] X. Liu, W. Li, Y. Fang, A. Kim, K. Duh, and J. Gao, "Stochastic answer networks for squad 2.0," *CoRR*, vol. abs/1809.09194, 2018. arXiv: 1809 . 09194. [Online]. Available: http://arxiv.org/abs/1809.09194.

APPENDIX

LIST OF FIGURES

LIST OF TABLES