

Homework 5 - DA2210

November 3, 2021

Philip Salqvist

phisal@kth.se

1992-04-17

Contents

1	Reading An Article	1
1.a	1
1.b	1
1.c	1
1.d	4
1.e	4
1.f	4
1.g	4

1 Reading An Article

1.a

I accessed the article through ACL Anthology's website. The full reference can be found in the collection of references [1].

1.b

The article was published in a conference proceeding from a conference of the North American Chapter of the Association for Computational Linguistics. The scientific field at the venue was mainly Human Language Technologies and the proceedings was later published in Association for Computational Linguistics. Since the article isn't published in a peer reviewed journal, but rather a conference proceeding, I assume it isn't peer reviewed.

1.c

Complete citation. Authors, Date of publication, Title (book or article), Journal, Volume, Issue, pages: J. Delvin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, 2019, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Association for Computational Linguistics, Volume 1, couldn't find issue, 4171–4186.

If web access: <https://aclanthology.org/N19-1423/>; 2021-11-02.

Key Words: Google, AI, Language Understanding, Language Representation.

General subject: Machine Learning

Specific subject: Language Understanding.

Hypothesis: There exists two main approaches when it comes to applying

pre-trained language representations to down-stream tasks. These are fine tuning and feature based. Both of these use unidirectional language models to learn general language representations. The authors hypothesize that the use of unidirectional models are sub optimal, and that using bidirectional models would be an improvement, especially for the fine-tuning approach.

Methodology: The bidirectional model BERT is compared to other existing top performing models. A number of different benchmarks/datasets are used, including The General Language Understanding Evaluation (GLUE) benchmark, The Stanford Question Answering Dataset (SQuAD v1.1), an extension of SQuAD v1.1 called SQuAD v2.0 and finally The Situation With Adversarial Generations (SWAG) dataset. A base version of BERT, with the same model size as OpenAI, is used for comparison. In addition to this, a larger version of BERT is used as well. These are referred to as *BERT_{BASE}* and *BERT_{LARGE}*.

Result(s): Both *BERT_{BASE}* and *BERT_{LARGE}* outperforms all systems with substantial margin on the GLUE benchmark. In the SQuAD v1.1 benchmark the best BERT system outperforms the previous top performing system. In SQuAD v2.0 an improvement is observed compared to previous best performing system. Finally, when observing the results from the SWAG benchmark test we can see that *BERT_{LARGE}* is outperforming both ESIM+ELMo and OpenAI GPG.

Summary of key points: A broad set of Natural Language Processing (NLP) tasks have previously been solved by using language model pre-training. Although, previously the models that have been used are constructed in a unidirectional manner. In this article, the authors show that their new bidirectional model, called BERT, outperforms most existing models on a number of NLP tasks.

Context (how this article relates to other work in the field; how it ties in with key issues and findings by others, including yourself): To solve tasks related

to NLP, a number of approaches have been developed. The two main ones are Unsupervised Feature-Based and Unsupervised Fine-tuning approach. When it comes to the former, a great example of advancement is the ELMo system which advanced state-of-the-art in NLP. The latter approach have been used by OpenAI GPT, and achieved state-of-the-art results on many tasks based on the GLUE benchmark. However, these two approaches have mainly been used with unidirectional models. This article's main contribution is to take this further, and prove that similar or even better results can be achieved using a unidirectional model to solve the same tasks.

Significance (to the field; in relation to your own work): The article is highly significant and influential in the field of Natural Language Processing. Not only at Google where the BERT model was developed, but clearly in the scientific field as well since the article has over 24 000 citations according to semantic scholar.

Important Figures and/or Tables (brief description; page number): Table 1-4 on pages 4176-4177. Presents the BERT models performance on different benchmarks where it is compared to other top performing models.

Cited References to follow up on (cite those obviously related to your topic AND any papers frequently cited by others because those works may well prove to be essential as you develop your own work):

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding with unsupervised learning. Technical report, OpenAI.

Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning, pages 160–167. ACM.

1.d

I would like to further explore an article written by Alec Radford et al. at OpenAI. [2] The creators of BERT have used an OpenAI model in their comparisons because of the resemblance between the two models, which makes the article especially relevant to further my understanding in this topic. Furthermore, I would like to read this article written by Lample et al., since it deals with NLP but on multilingual data which is interesting research. The article is also cited quite frequently which probably makes it a valuable source on the subject and thus worth reading. [3].

1.e

By examining the structure of the article by Killourhy et al. we can see that the IMRD structure is followed quite strictly, but not perfectly. [4] On the other hand, the article written by Obermeyer et al. have a less rigid structure, and doesn't follow IMRD as strictly. I would consider the subfields of computer science that are more closely related to the social sciences less bounded to the format of IMRD. Subfields that are more related to the hard sciences would probably benefit more of a rigid structure such as IMRD. [5]

1.f

The article I have chosen has been cited 24 379 times, according to semantic scholar.

1.g

The authors have contributed with state-of-the art results in NLP research by implementing a bidirectional model BERT, that in large degree outperforms previous research that have been using unidirectional models.

References

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), June, 2019, Minneapolis, Minnesota. Association for Computational Linguistics. <https://aclanthology.org/N19-1423/>.
- [2] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2012. *Improving language understanding with unsupervised learning*. Technical Report, OpenAI.
- [3] Guillaume Lample, A. Conneau. 2019. *Cross-lingual Language Model Pretraining*. NeurIPS. <https://www.semanticscholar.org/paper/Cross-lingual-Language-Model-Pretraining-Lample-Conneau/ec4eba83f6b3266d9ae7cabb2b2cb1518f727edc>
- [4] Kevin S. Killourhy, Roy A. Maxion. 2009. *Comparing anomaly-detection algorithms for keystroke dynamics*. IFIP International Conference on Dependable Systems & Networks, 29 June-2 July, Lisbon, Portugal. IEEE. <https://ieeexplore.ieee.org/document/5270346>.
- [5] Ziad Obermeyer, Brian Powers, Christine Vogeli, Sendhil Mullainathan. 2019. *Dissecting racial bias in an algorithm used to manage the health of populations*. Science, 366, 6464 (25 Oct 2019), pp. 447-453. DOI: 10.1126/science.aax2342