BERT (language model)

Bidirectional Encoder Representations from Transformers (BERT) is a technique for NLP (Natural Language Processing) pre-training developed by Google. BERT was created and published in 2018 by Jacob Devlin and his colleagues from Google.[1][2] Google is leveraging BERT to better understand user searches.[3]

The original English-language BERT model used two corpora in pre-training: BookCorpus and English Wikipedia.[1]

Contents

Performance

When BERT was published, it achieved state-of-the-art performance on a number of natural language understanding tasks:[1]

- GLUE (General Language Understanding Evaluation) task set (consisting of 9 tasks)
- SQuAD (Stanford Question Answering Dataset) v1.1 and v2.0.
- SWAG (Situations With Adversarial Generations)

Analysis

The reasons for BERT's state-of-the-art performance on these natural language understanding tasks are not yet well understood.[4][5] Current research has focused on investigating the relationship behind BERT's output as a result of carefully chosen input sequences,[6][7] analysis of internal vector representations through probing classifiers,[8][9] and the relationships represented by attention weights.[4][5]

History

BERT has its origins from pre-training contextual representations including Semi-supervised Sequence Learning,[10] Generative Pre-Training, ELMo,[11] and ULMFit.[12] Unlike previous models, BERT is a deeply bidirectional, unsupervised language representation, pre-trained using only a plain text corpus. Context-free models such as word2vec or GloVe generate a single word embedding representation for each

word in the vocabulary, where BERT take into account the context for each occurrence of a given word. For instance, whereas the vector for "running" will have the same word2vec vector representation for both of its occurrences in the sentences "He is running a company" and "He is running a marathon", BERT will provide a contextualized embedding that will be different according to the sentence.

On October 25, 2019, Google Search announced that they had started applying BERT models for English language search queries within the US.[13] On December 9, 2019, it was reported that BERT had been adopted by Google Search for over 70 languages.[14]

Recognition

BERT won the Best Long Paper Award at the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).[15]

See also

- Transformer (machine learning model)
- Word2vec
- Autoencoder
- Document-term matrix
- Feature extraction
- Feature learning
- Neural network language models
- Vector space model
- Thought vector
- fastText
- GloVe
- TensorFlow

References

1. Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (11 October 2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". arXiv:1810.04805v2 (https://arxiv.org/abs/1810.04805v2) [cs.CL (https://arxiv.org/archive/cs.CL)].
2. "Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing" (http://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html). *Google AI Blog*. Retrieved 2019-11-27.
3. "Understanding searches better than ever before" (https://blog.google/products/search/search-language-understanding-bert/). *Google*. 2019-10-25. Retrieved 2019-11-27.
4. Kovaleva, Olga; Romanov, Alexey; Rogers, Anna; Rumshisky, Anna (November 2019). "Revealing the Dark Secrets of BERT" (https://www.aclweb.org/anthology/D19-1445). *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. pp. 4364–4373. doi:10.18653/v1/D19-1445 (https://doi.org/10.18653%2Fv1%2FD19-1445).
5. Clark, Kevin; Khandelwal, Urvashi; Levy, Omer; Manning, Christopher D. (2019). "What Does BERT Look at? An Analysis of BERT's Attention" (https://doi.org/10.18653/v1/w19-4828). *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Stroudsburg, PA, USA: Association for Computational Linguistics: 276–286. doi:10.18653/v1/w19-4828 (https://doi.org/10.18653%2Fv1%2Fw19-4828).
6. Khandelwal, Urvashi; He, He; Qi, Peng; Jurafsky, Dan (2018). "Sharp Nearby, Fuzzy Far Away: How Neural Language Models Use Context". *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics: 284–294. arXiv:1805.04623 (https://arxiv.org/abs/1805.04623). Bibcode:2018arXiv180504623K (https://ui.adsabs.harvard.edu/abs/2018arXiv180504623K). doi:10.18653/v1/p18-1027 (https://doi.org/10.18653%2Fv1%2Fp18-1027).

7. Gulordava, Kristina; Bojanowski, Piotr; Grave, Edouard; Linzen, Tal; Baroni, Marco (2018). "Colorless Green Recurrent Networks Dream Hierarchically". *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics: 1195–1205. arXiv:1803.11138 (https://arxiv.org/abs/1803.11138). Bibcode:2018arXiv180311138G (https://ui.adsabs.harvard.edu/abs/2018arXiv180311138G). doi:10.18653/v1/n18-1108 (https://doi.org/10.18653%2Fv1%2Fn18-1108).
8. Giulianelli, Mario; Harding, Jack; Mohnert, Florian; Hupkes, Dieuwke; Zuidema, Willem (2018). "Under the Hood: Using Diagnostic Classifiers to Investigate and Improve how Language Models Track Agreement Information". *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Stroudsburg, PA, USA: Association for Computational Linguistics: 240–248. arXiv:1808.08079 (https://arxiv.org/abs/1808.08079). Bibcode:2018arXiv180808079G (https://ui.adsabs.harvard.edu/abs/2018arXiv180808079G). doi:10.18653/v1/w18-5426 (https://doi.org/10.18653%2Fv1%2Fw18-5426).
9. Zhang, Kelly; Bowman, Samuel (2018). "Language Modeling Teaches You More than Translation Does: Lessons Learned Through Auxiliary Syntactic Task Analysis" (https://doi.org/10.18653/v1/w18-5448). *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Stroudsburg, PA, USA: Association for Computational Linguistics: 359–361. doi:10.18653/v1/w18-5448 (https://doi.org/10.18653%2Fv1%2Fw18-5448).
10. Dai, Andrew; Le, Quoc (4 November 2015). "Semi-supervised Sequence Learning". arXiv:1511.01432 (https://arxiv.org/abs/1511.01432) [cs.LG (https://arxiv.org/archive/cs.LG)].
11. Peters, Matthew; Neumann, Mark; Iyyer, Mohit; Gardner, Matt; Clark, Christopher; Lee, Kenton; Luke, Zettlemoyer (15 February 2018). "Deep contextualized word representations". arXiv:1802.05365v2 (https://arxiv.org/abs/1802.05365v2) [cs.CL (https://arxiv.org/archive/cs.CL)].
12. Howard, Jeremy; Ruder, Sebastian (18 January 2018). "Universal Language Model Fine-tuning for Text Classification". arXiv:1801.06146v5 (https://arxiv.org/abs/1801.06146v5) [cs.CL (https://arxiv.org/archive/cs.CL)].
13. Nayak, Pandu (25 October 2019). "Understanding searches better than ever before" (https://www.blog.google/products/search/search-language-understanding-bert/). *Google Blog*. Retrieved 10 December 2019.
14. Montti, Roger (10 December 2019). "Google's BERT Rolls Out Worldwide" (https://www.searchenginejournal.com/google-bert-rolls-out-worldwide/339359/). *Search Engine Journal*. Search Engine Journal. Retrieved 10 December 2019.
15. "Best Paper Awards" (https://naacl2019.org/blog/best-papers/). *NAACL*. 2019. Retrieved Mar 28, 2020.

External links

- Official GitHub repository (https://github.com/google-research/bert)