A Comprehensive Technology Review of BERT

Hannah Benig, CS410 Fall 2023

I. Introduction

BERT (Bidirectional Encoder Representations from Transformers) is a family of language models introduced in 2018 by Google. It can be fine-tuned with only one additional output layer to create state-of-the-art models for various tasks, including question answering and language inference, without substantial task-specific architecture modifications(Devlin et al.). BERT considers the context for each occurrence of a given word, unlike word2vec or GloVe, which generate a single-word embedding representation for each word in the vocabulary (Devlin, Chang). The purpose of this review is to outline BERT's architecture and training process, highlight its capabilities and applications, and compare BERT with ChatGPT.

II. BERT Architecture and Training Process

BERT is trained on a dataset of 3.3 billion words from passages in Wikipedia and Google's BooksCorpus. BERT's architecture consists of 12 transformer layers, 768 hidden layers, 12 attention heads, 110 million parameters, and 4 TPUs for processing, and takes 4 days to train. BERT is the first deeply bidirectional, unsupervised language representation pre-trained using only a plain text corpus. The bidirectional nature of the model allows for information to be passed in both directions (left-to-right and right-to-left), but this process alone would allow the model to "see" the next/previous word rather than having to predict the word. To avoid "giving the model the answer," BERT utilizes a masked language model (MLM) - this is the first step in BERT's training process. MLMs mask/hide a word in a sentence, which forces BERT to bidirectionally use words on either side of the hidden word to predict it (Muller). While this may seem like a simple concept or implementation, BERT was the first model to successfully integrate this architecture. The second step in BERT's training process is Next Sentence Prediction (NSP). NSP achieves a similar goal to MLM, but instead of predicting a word in a sentence, we now want to predict the following sentence and ensure it correctly relates to our current sentence. Finally, BERT is fine-tuned to unify the outputs of the first two steps.

III. BERT's Capabilities and Applications

Since BERT has been trained on thousands, if not millions, of articles, it only makes sense that it would have many applications within natural language processing (NLP). For example, sentiment analysis, determining how positive or negative book or movie reviews are, question answering, when integrated in a chatbot, and text prediction, helping you write emails or making suggestions about what the next word or phrase should be (Muller).

Recently, researchers in India compared BERT and GloVe to detect sarcasm in tweets. The team used a labeled dataset of tweets containing 5,000 entries with an equal number of sarcastic and non-sarcastic entries. They pre-processed the data by removing null values,

performing tokenization and removing punctuation, converting each entry to lowercase, removing stopwords, removing/replacing slang into its normal, plain English form, removing noise from the text, and stemming the entries. For each encoding, they tested multiple Scikit-learn classifiers. For both BERT and GloVe, the Logistic Regression classifier performed the best, with BERT having an F-measure of 0.630 and GloVe having an F-measure of 0.690. They concluded that even though BERT gives contextual word representations, it was unable to perform at the same level as GloVe on this dataset (Khatri et al.)

IV. Comparison of BERT and ChatGPT

While there is no question that BERT has already achieved amazing results for countless problems within the NLP world. When ChatGPT-4 was released in November 2022, the landscape of natural language processing was forever changed. The question then becomes, can BERT ever achieve the same level of popularity as ChatGPT? On the surface level, we see that while BERT and ChatGPT are both language processing models that use transformer architectures, however, they are designed for separate tasks (Odor). BERT uses bidirectional transformers to complete tasks such as sentiment analysis, question answering, and entity recognition, as mentioned in the previous section. ChatGPT uses generative transformers to provide human-like responses to questions asked in natural language.

ChatGPT and BERT are both trained on over 110 million parameters. From this data, they can understand the underlying structures and complex patterns of language. BERT is currently used at Google to optimize the results returned to users related to their queries. With some fine-tuning, it may not be impossible to say that the search results that BERT returns could be themselves queried and the response summarized to return a straightforward answer in natural language, thus making it conversational like ChatGPT. There is still more work to be done, but it's not impossible to think we may see a conversational chatbot that utilizes BERT and competes with ChatGPT.

V. Conclusion

It's crazy to think of all the ways that natural language processing occurs all around us, everything from a seemingly simple Google search to analyzing the sentiment of tweets to predict or monitor threats. In this article, we outlined BERT’s architecture and training process, highlighted its capabilities and applications, and compared BERT with ChatGPT.

VI. References

Devlin, Jacob and Chang, Ming-Wei. “Open Sourcing BERT: State-of-the-Art Pre-Training for Natural Language Processing.” *Google Research Blog*, 2 Nov. 2018, blog.research.google/2018/11/open-sourcing-bert-state-of-art-pre.html.

Devlin, Jacob, et al. “Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” *arXiv.Org*, 24 May 2019, arxiv.org/abs/1810.04805.

Khatri, Akshay, et al. “Sarcasm Detection in Tweets with BERT and GloVe Embeddings.” *arXiv.Org*, 20 June 2020, arxiv.org/abs/2006.11512.

Muller, Britney. *Bert 101 - State of the Art NLP Model Explained*, Hugging Face, 2 Mar. 2022, huggingface.co/blog/bert-101.

Odor, Emmanuel. *Bert Google vs ChatGPT-4: A Deep Dive into Natural Language Processing Technologies*, Linkedin, 23 Mar. 2023, www.linkedin.com/pulse/bert-google-vs-chatgpt-4-deep-dive-natural-language-processing-odor.