Ying Zhang

CS 410 Text Information Systems

Professor: Dr. Zhai

October 23, 2020

Review for PALM: Pre-training an Autoencoding and Autoregressive Language Model for Context-conditioned Generation

PALM is a pre-training an autoencoding and autoregressive language model (Bi 1). It can be used to generate abstracts, question answers, conversational response and other language generation tasks. This model is currently widely used in multiple actual business scenario, such as: CNN/DayilyMail abstractive summarization. At the time of this paper was published, it was the Rank 1 on MS MARCO leaderboard for Question Answering and Natural Language Generation. Now, let's talk about the construction and function of the model, as well as some experimental results of this model.

Unlike other nature learning language models, which either use autoencoding or autoregressive, PALM implement both methods for its model. Specifically, it uses bidirectional autoencoding to understand the given contexts and then uses the autoregressive method to generate a text sequence (answers). To impalement both methods, PALM uses autoencoding-based comprehension as a Transformer encoder and autoregressive generation as Transformer decoder and both Transformers were jointly pre-trained in two different stages (Figure 1).

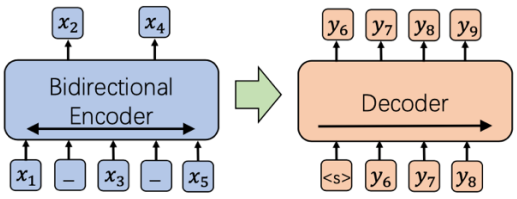

Figure 1: A schema of PALM (Bi 3)

First stage is the Transformer encoder encodes context bidirectionally to predict masked token and second stage is the Transformer decoder generates the text segment after the context that has been analyzed. The goal for the training is to maximize the log-likelihood of the text of the truth from the decoder output.

PALM has very good performance in generative QA, summarization, question generation, and response generation. For generative QA, PALM was leading the answer generation on MS MARCO leaderboard with the RG-L value of 0.498 in Dec. 2019. PALM also has good performance in summarization. From CNN/DailyMail and Gigawor, two of main non-anonymized summarization datasets, the SOTA score are exceed including Google's BERT&T5, Facebook's BART, and Baidu's ERNIE-GEN and other training model. The RG-1, RG-2, and RG-L value were 44.30, 21.12, and 41.41 respectively on CNN/DailyMail and RG-1, RG-2, and RG-L value on Gigaword were 39.45,20.37, and 36.75 respectively (Bin 7). In question generation, PALM leading the SOTA result on SQuAD dataset, with BLEU-4, MTR, and RG-L value of 24.11, 25.85, and 52.38 respectively (Bi 7). Lastly, for response generation, in terms of perplexity on Cornell Movie Dialog corpus, PALM leading the Perlexity(10K Data) and Perplexity(110K data) score value of 45.43 and 21.98 (Bi 8).

In conclusion, PALM is a novel and well perform language model. With the fine-tuning on downstream generation conditioned on context, PALM had the best performance in multiple types of language generation tasks. Yet, according to the paper, with longer training and more steps over a larger corpus can improve the performance even more. (Bi 8) We hope to PALM can be used in wide range of application in real life.

Reference:

Bi, Bin, and Chenliang Li. “PALM: Pre-Training an Autoencoding&Autoregressive Language Model for Context-Conditioned Generation.” Apr. 2020, arxiv.org/pdf/2004.07159.pdf.