BERT's recent best practices & applications

Intro

In this tech review, we are going to go over some of the recent advancement in BERT and its variants. We are going to particularly go over its applications on varieties of tasks and the best practices to make it happen in industry.

Content

Deployment

zhihu/cuBERT, Fast implementation of BERT inference directly on NVIDIA (CUDA, CUBLAS) and Intel MKL
xmxoxo/BERT-train2deploy, Bert Model training and deploy

Distillation

To make model smaller and cheaper
qiangsiwei/bert_distill, BERT distillation
kevinmtian/distill-bert, Knowledge Distillation from BERT

Recent Advancement Based on Bert

pytorch/fairseq, Facebook AI Research Sequence-to-Sequence Toolkit written in Python. RoBERTa: A Robustly Optimized BERT Pretraining Approach
facebookresearch/SpanBERT, Code for using and evaluating SpanBERT. , This repository contains code and models for the paper: SpanBERT: Improving Pre-training by Representing and Predicting Spans
thunlp/ERNIE, ERNIE: Enhanced Language Representation with Informative Entities –improve bert with heterogeneous information fusion
zihangdai/xlnet, XLNet: Generalized Autoregressive Pre Training for Language Understanding

Task – Question Answering

deepset/roberta-base-squad2 This is the roberta-base model, fine-tuned using the SQuAD2.0 dataset. It's been trained on question-answer pairs, including unanswerable questions, for the task of Question Answering.
allenai/allennlp-bert-qa-wrapper, This is a simple wrapper on top of pretrained BERT based QA models from pytorch-pretrained-bert to make AllenNLP model archives, so that you can serve demos from AllenNLP.

sogou/SMRCToolkit, This toolkit was designed for the fast and efficient development of modern machine comprehension models, including both published models and original prototypes.

Task – Text Generation

Tiiiger/bert_score, Automatic Evaluation Metric described in the paper BERTScore: Evaluating Text Generation with BERT (ICLR 2020). We now support about 130 models (see this spreadsheet for their correlations with human evaluation). Currently, the best model is microsoft/deberta-xlarge-mnli, please consider using it instead of the default roberta-large in order to have the best correlation with human evaluation
asyml/texar, Toolkit for Text Generation and Beyond https://texar.io, Texar is a general-purpose text generation toolkit, has also implemented BERT here for classification, and text generation applications by combining with Texar's other modules

Task – Text Summarization

santhoshkolloju/Abstractive-Summarization-With-Transfer-Learning, Abstractive summarisation using Bert as encoder and Transformer Decoder
nlpyang/BertSum, Code for paper Fine-tune BERT for Extractive Summarization

Task – Chat bot

yuanxiaosc/BERT-for-Sequence-Labeling-and-Text-Classification, This is the template code to use BERT for sequence lableing and text classification, in order to facilitate BERT for more tasks. Currently, the template code has included conll-2003 named entity identification, Snips Slot Filling and Intent Prediction.

Task – Sentiment Analysis

HSLCY/ABSA-BERT-pair, Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence
Chung-I/Douban-Sentiment-Analysis, Sentiment Analysis on Douban Movie Short Comments Dataset using BERT

Conclusion

The best practices in terms of how to use BERT in terms of deployment and choosing the right variant provide some rule of thumbs.
Also, we could see that BERT is a very versatile in terms of potential applications due to its multi modality.

References

https://arxiv.org/abs/1810.04805
Github.com
Huggingface.co