Tech Review on OpenAI GPT, GPT-2, GPT-3

Qiwei He

Generative Pre-trained Transformer (GPT) models are created by OpenAI, which are the groundbreaking models in the field of natural language processing that deliver valuable language models. They are versatile in performing various NLP tasks including textual entailment and text summarisation (Shree, 2020). And they don't require supervised training for completing these tasks. GPT's language models don't need too many examples for delivering the same level of results compared to supervised training models. The models have evolved from GPT1 to GPT2 to GPT3 by OpenAI in recent years.

Prior to GPT's birth, most other recognized NLP models are more toward supervised training models. The supervised models have crucial shortcomings. Their usages are too narrowly focused on the specific tasks that they are trained for, and they demand a substantial amount of annotated data to train with. The first edition of GPT proposed the idea to make a generative model by utilizing unlabeled data then refine it with selected downstream tests like sentiment analysis. Basically, it is composed of unsupervised pre-training and supervised refinement processes (Shree, 2020). The supervised refinement process would take a linear and a softmax layer applied on the transformer

model aimed to receive the task labels for downstream tasks. The dataset it utilized is from BooksCorpus which contains about 7000 unpublished literature. The performance of GPT1 is quite impressive with its feature having zero-shot performance (Shree, 2020). It basically means the model does not require any example but just the instruction. The GPT1 illustrates the merit of generative pre-training. It also paved the way for other models following this kind of orientation with relatively large datasets and increased parameters.

The GPT2 model evolved from GPT1 as it takes a more substantial dataset and more parameters. One of the changes is the task conditioning taken by GPT2. It means there would be different outputs created for the same input with respect to different tasks. The dataset of GPT2 is tremendous compared to GPT1's. Its dataset is the pulled data from Reddit platform of a size of 40GB from 8 million documents. GPT2 also has 10 times more parameters compared to GPT1's parameters (Shree, 2020).

GPT3, an even more powerful model than GPT2, obtains 175 billion parameters (Shree, 2020). It is 100 times even more than GPT2's. Its strong capability ensures that it is possible to write literature that is similar to human's. The astonishing performance could even apply on the tasks of writing codes given description of the task. There are still some part that can be improved for GPT3. It is still not too competent in performing tasks of natural language inference (Shree, 2020). The change of unidirectionality of GPT models to bidirectional models could address this shortcoming.

Overall, GPT models are groundbreaking in directing the future direction of NLP.

Reference

1. Shree, P. (2020, November 10). *The Journey of Open AI GPT models*. Medium. https://medium.com/walmartglobaltech/the-journey-of-open-ai-gpt-models-32d95b7b7fb2
2. Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I., 2018. Improving language understanding by generative pre-training.
3. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I., 2019. Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), p.9.
4. Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. “Language models are few-shot learners.” *arXiv preprint arXiv:2005.14165* (2020).
5. Rei, M., 2017. Semi-supervised multitask learning for sequence labeling. *arXiv preprint arXiv:1704.07156*.
6. Waswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., 2017. Attention is all you need. In *NIPS*.