Summary

1. BERT is a state-of-the-art language representation model, outperforming in a wide range of competitions (<https://gluebenchmark.com/leaderboard>).

2. The original paper on arxiv (<https://arxiv.org/abs/1810.04805>) is a brief introduction of the ideas and algorithms. Short tutorials with more technical details could be found here <https://mccormickml.com/2019/07/22/BERT-fine-tuning/>.

3. In general, the training process of BERT is very time-consuming. Based on multiple GPUs/TPUs, which though might be hard to afford, the training time could be reduced (<https://arxiv.org/abs/1904.00962>, <https://github.com/google-research/bert>).

Comment

1. There is no doubt that BERT is powerful for natural language inference. However, for our purposes, we only need to answer certain fixed questions, such as ‘risk factors’ and ‘incubation’ (<https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks>), so BERT might be too expensive to be applied.

2. Another way to solve such data-mining problems is by Latent Dirichlet Allocation (<https://www.wikiwand.com/en/Latent_Dirichlet_allocation>). In short, the model could learn the relation between words and topics and the relation between topics and articles, hence it could output the most relevant n articles with respect to certain word like ‘risk factors’. One nice open paper could be found here <https://www.kaggle.com/danielwolffram/topic-modeling-finding-related-articles>.