Abstract

This project focused on understanding the side effects of specializing a large language model such as GPT-2 XL on the task of Knowledge Graph Generation. It was theorized that this kind of specialization may lead to a decrease in the performance of the models in other language modelling and understanding tasks such as multiple choice selection, text entailment, word sense disambiguation and other similar tasks. In order to compare the performance of the models, a specialized model (COMET) was compared with a baseline model (GPT-2 XL) on several general language tasks. The specialized model was "fully trained" on the task of Knowledge Graph generation while the baseline model was not. Both models had their base layers frozen and only the language modelling head was trained on the respective tasks. The results of the experiments indicate that the specialized model does not perform as well as the baseline model on some tasks.

Keywords: Knowledge graphs, Language Models, GPT-2 XL, COMET-DISTIL, Glue Benchmark