How Much Knowledge Can You Pack Into The Parameters of a Language Model?

Shubham Gupta

February 11, 2020

1 Introduction

- This is a new paper which explores the limits of using their new T5 model in a context-free QA domain.

- As with the T5 model itself, it is very interesting to see these one-model-to-rule-them-all architectures as they exhibit some form of generalization.

- I found this paper from Adam's Roberts twitter thread which is available here

- **Core Idea**: Bigger models(such as T5) can perform much better on QA tasks with finetuing than the models that explicitely look up knowledge.