

LLM

王铭泽 ZY2203115

wmz20000729@buaa.edu.cn

1 Introduction

Large-scale language models (LLMs) typically refer to language models that consist of trillions (or more) parameters trained on vast amounts of text data, such as GPT-3, PaLM, Megatron, and LLaMA. Existing LLMs primarily employ model architectures similar to smaller language models, such as Transformers, and utilize pretraining objectives, such as language modeling. The key distinction lies in the significant scaling of model size, pretraining data, and computational resources (multiplied by a large factor) in LLMs. They exhibit a better understanding of natural language and are capable of generating high-quality text based on given context, such as prompts. This capacity improvement can be partially described by the power law scaling, wherein performance roughly follows a substantial increase with model size.

2 Experiment

2.1 Q-A

Question: 'Which year did I start playing basketball?'

Context: 'I was born in 2000. I started playing basketball after I turned ten years old.'

2.1.1 Deepset/Roberta-base-squad2

Res: '2000'

Therefore, it is evident that the model's understanding of the content is shallow and lacks cognition. It can only recognize the concept of "time" but fails to grasp the specific context of my question regarding when I started playing basketball. It merely identifies that my question pertains to a time-related query.

2.1.2 Distilbert-base-cased-distilled-squad

Res: '2000'

Similarly, both the aforementioned model and this model have a similar size of around 250MB, which cannot be considered as a large-scale model. As a result, it naturally leads to a moderate level of understanding. From the outcomes, it is apparent that the model can only recognize the concepts of years and time but does not possess the ability to comprehend the addition of time and specific time points.

2.1.3 Bert-large-uncased-whole-word-masking-finetuned-squad

Res: '2010'

However, this time a large-scale model, specifically BERT-Large, was utilized. It was fine-tuned on the SQuAD dataset, and the resulting model was used for inference. It was observed that the model successfully recognized the concepts of time and specific time points and accurately performed addition operations on them. The output results were correct.

2.2 summarization

Article: " New York (CNN)When Liana Barrientos was 23 years old, she got married in Westchester County, New York. A year later, she got married again in Westchester County, but to a different man and without divorcing her first husband. Only 18 days after that marriage, she got hitched yet again. Then, Barrientos declared "I do" five more times, sometimes only within two weeks of each other. In 2010, she married once more, this time in the Bronx. In an application for a marriage license, she stated it was her "first and only" marriage. Barrientos, now 39, is facing two criminal counts of "offering a false instrument for filing in the first degree," referring to her false statements on the 2010 marriage license application, according to court documents. Prosecutors said the marriages were part of an immigration scam. On Friday, she pleaded not guilty at State Supreme Court in the Bronx, according to her attorney, Christopher Wright, who declined to comment further. After leaving court, Barrientos was arrested and charged with theft of service and criminal trespass for allegedly sneaking into the New York subway through an emergency exit, said Detective Annette Markowski, a police spokeswoman. In total, Barrientos has been married 10 times, with nine of her marriages occurring between 1999 and 2002. All occurred either in Westchester County, Long Island, New Jersey or the Bronx. She is believed to still be married to four men, and at one time, she was married to eight men at once, prosecutors say. Prosecutors said the immigration scam involved some of her husbands, who filed for permanent residence status shortly after the marriages. Any divorces happened only after such filings were approved. It was unclear whether any of the men will be prosecuted. The case was referred to the Bronx District Attorney's Office by Immigration and Customs Enforcement and the Department of Homeland Security's Investigation Division. Seven of the men are from so-called "red-flagged" countries, including Egypt, Turkey, Georgia, Pakistan and Mali. Her eighth husband, Rashid Rajput, was deported in 2006 to his native Pakistan after an investigation by the Joint Terrorism Task Force. If convicted, Barrientos faces up to four years in prison. Her next court appearance is scheduled for May 18. "

2.2.1 Bart-large-cnn

Res: 'Liana Barrientos, 39, is charged with two counts of "offering a false instrument for filing in the first degree" In total, she has been married 10 times, with nine of her marriages occurring between 1999 and 2002. She is believed to still be married to four men.'

This summary covers most of the content, condensing it while maintaining consistency with the main ideas of the original text.

2.2.2 Distilbart-cnn-12-6

Res: 'Liana Barrientos, aged 39, is facing charges of committing two counts of "offering a false instrument for filing in the first degree." Throughout her history, she has entered into matrimony a total of 10 times, with the majority of her marriages taking place between 1999 and 2002. Currently, it is suspected that she remains legally married to four individuals.'

The abstract should have a slightly higher complexity, but at the same time, it should cover most of the content in a concise manner, while remaining faithful to the main ideas and the original text.

2.2.3 PG-cnn-12-6

Res: 'Liana Barrientos, who is 39 years old, is facing charges of "offering a false instrument for filing in the first degree." Surprisingly, she has entered into matrimony a remarkable 10 times, predominantly between 1999 and 2002. Interestingly, it is believed that she is presently legally wedded to four distinct individuals.'

This summary effectively captures the essence of the content, concisely condensing it while staying true to the primary concepts presented in the original text.

2.3 text-generation

Context: "Hello, I'm a language model,"

2.3.1 GPT2

Res: 'Hello, I'm a language model, a system model. I want to know my language so that it might be more interesting, more user-friendly'

Clearly, this sentence is well-generated, with a clear logical structure and easy readability, demonstrating its power and effectiveness.

2.3.2 Bloom-560m

Res: 'I'm a language model, not a language model

Undoubtedly, this sentence is excellently crafted, exhibiting a coherent logical structure and effortless readability, showcasing its impressive potency and efficacy.

2.3.3 GPT4

Res: 'I'm a language model, I do not possess personal experiences or emotions. I am here to assist you with generating text, providing information, or answering questions to the best of my abilities. If there's anything specific you would like assistance with, please let me know and I'll be glad to help!'

It is evident that the model is powerful, and the text generated by the model possesses more emotions and vividness.

3 Conclusion

This article selected a total of 9 models and 3 downstream tasks to evaluate their performance on each task. It is not difficult to observe that as the model's parameter size increases, its level of "intelligence" also tends to be higher, which holds true for most tasks. Additionally, the effectiveness of the model also relies on the similarity between the fine-tuning dataset and the text during inference.