GPT-3

The Good, The Bad, and The Ugly

Brayden Turner

CS 410

Text Information Systems

Getting computers to understand natural language to the level that humans do can be applied to a near limitless amount of applications, but it isn't an easy feat. A lot of different tools have tried to tackle this application to varying degrees of success. One of the latest technologies created, GPT-3, has taken the understanding of natural language to a whole different level. Created by OpenAI in 2020 and currently only seeded to researchers and developers, GPT-3 was built to, given a string of words, predict the next word. GPT-3 has over 175 billion parameters and trained on a dataset containing hundreds of billions of words.

OpenAI has released a few version of GPT since 2018, each version building on the learnings from the previous one. The key innovations that helped take GPT-3 to new heights was changing how sentences are represented and training the model using unsupervised learning. Traditionally in NLP models, sentences are represented as fixed length vectors. This approach kind of gets in the way of all the various applications and free form texts that humans produce. The next big thing that GPT-3 improved upon was using more unsupervised learning methods. A lot of other models used supervised learning with labeled data from humans. This approach can be very time consuming, and at times even impossible with the amount of text data we produce. These two things combined make GPT-3 much more powerful.

Only a few developers have been lucky enough to be invited to use GPT-3 to apply to NL problems. These researchers have created a variety of applications ranging from impersonating celebrities to creating computer code. While GPT-3 is trained on a lot of different texts, you can focus it on certain areas to more directly apply it to your application. Here are a few things developers have created in recent months: question-answer based search engine, chat with historical figures, computer code generator based on text descriptions, text-based RPG games, creative fiction writer, compose guitar tabs, etc. The thing about GPT-3, though, is

it wasn't really designed to be good at any of these applications. It was generally trained by OpenAI and then was fed a few expected input/output examples to get it to conform to the application. It takes what it's learned from the corpus of human speech and applies it to the style that's asked of it, similar to how we may apply knowledge.

GPT-3 isn't without it's shortcomings, however. Although the things built on top of it are impressive, they are cherry-picked to build hype around it. Feed it some text to write a science fiction novel and you can quickly see it gets in the ballpark, but it's not Frank Herbert. Sometimes it can go in circles or not really tell a compelling or grounded story. It also has a lot of the same problems that have plagued other generative natural language models for years. It can produce extremely hateful and toxic output depending on the subject matter passed in. It can continue to perpetuate gender stereotypes based on adjectives it uses around things like "male" or "female". A machine learned model is only as good as the data it's trained on, and there is a lot of hate speech and radicalized language on the internet. This falls on OpenAI to address and mitigate, and will be a difficult task given the complexity of the model.

The problem a lot of these models lack is common sense. GPT-3 may generate a news article where Barack Obama and Harry Potter are discussing the colony on mars, which we quickly know isn't reality. This isn't something that can necessarily be fixed with more data or more compute time, however. For starters, it's extremely impractical. And it most likely isn't solely a data problem. Some have suggested that to improve GPT-3 is to more broadly utilize human feedback. We are fairly good at identifying when something is obviously wrong, and could help steer the model in to more common sense territory. Another way may be to combine it with non-text data as well, to get closer to how we learn to associate knowledge and connections. Despite the leap that language models have made in recent decades,

we don't have to worry about the next breakout author being a Python program just yet.

References

[1]OpenAI, "GPT-3 Powers the Next Generation of Apps," OpenAI, Mar. 25, 2021. https://openai.com/blog/gpt-3-apps/ (accessed Oct. 28, 2021).

[2]K. Piper, "GPT-3, explained: OpenAI's new language AI is uncanny, funny-and a big deal - Vox," Vox, Aug. 13, 2020. https://www.vox.com/future-perfect/21355768/gpt-3-ai-openai-turing-test-language (accessed Oct. 28, 2021).

[3]"What is GPT-3 and Why Does it Matter? | IT Business Edge," IT Business Edge, Oct. 19, 2021. https://www.itbusinessedge.com/development/what-is-gpt-3/ (accessed Oct. 28, 2021).

[4]"Is GPT-3 the 'Reborn Doll' of Artificial Intelligence? | Mind Matters," Mind Matters, Oct. 12, 2021. https://mindmatters.ai/2021/10/is-gpt-3-the-reborn-doll-of-artificial-intelligence/ (accessed Oct. 28, 2021).

[5]W. D. Heaven, "OpenAI's new language generator GPT-3 is shockingly good—and completely mindless | MIT Technology Review," MIT Technology Review, Jul. 20, 2020. https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/ (accessed Oct. 28, 2021).

[6]C. Metz, "Meet GPT-3. It Has Learned to Code (and Blog and Argue). - The New York Times," The New York Times - Breaking News, US News, World News and Videos, Nov. 24, 2020. https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html (accessed Oct. 28, 2021).

[7]J. Vincent, "OpenAI's latest breakthrough is astonishingly powerful, but still fighting its flaws - The Verge," The Verge, Jul. 30, 2020. https://www.theverge.com/21346343/gpt-3-explainer-openai-examples-errors-agi-potential (accessed Oct. 28, 2021).

[8]L. O'Sullivan and J. P. Dickerson, "Here are a few ways GPT-3 can go wrong – TechCrunch," TechCrunch, Aug. 07, 2020. https://techcrunch.com/2020/08/07/

here-are-a-few-ways-gpt-3-can-go-wrong/?
guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS88&guce_referr
er_sig=AQAAAJiDWNFKSewbAwlyjq-11doaue49eDxNQaP58ykGc8O9bkDwTWMBdc
TKF9KcZ7F7q08Q1KVmTa1JVEXXp_k9XuhQBORaRrBEgjG2trsm8c5Bgx-J57U-
NcgxywZqbP0alSLz_QIiH1awsHt2zfRDVLC_jSIgmbpipr1CZdT9IIVB (accessed Oct.
28, 2021).