November 5, 2022 Dillon Harding
CS 410 Tech Review

A Review of GPT Model Advancements Over Generations

Created by the group Open AI, the GPT neural network is a groundbreaking state of the art language model (Shree, 2020). The neural network has seen many iterations over the years, with GPT-4 being released soon per the date of this tech review (Romero, 2022). To date, there have three released iterations of the GPT neural network, named GPT-1, GPT-2, and GPT-3 respectively (Team, 2021). In an effort to explore the advancements in each generation of the GPT language model, it is important to look at a variety of metrics, including but not limited to performance, model architecture, and the model's dataset.

GPT-1 was created as a paper in 2018 as a solution to the issue of natural language models containing high specificity in their training data (Shree, 2020). Prior to the creation of GPT-1, many natural language models were created on small to medium sized data sets, usually with a specific purpose in mind (Shree, 2020). GPT-1 aimed to create a generalized model that could perform a broad scale of natural language tasks (Shree, 2020). To create the model, OpenAi used a large data set at the time known as BookCorpus as their dataset (Shree, 2020). One task that made GPT-1 a state-of-the-art model at the time was its ability to use zero-shot performance, which mimics the human ability to make an educated guess (Shree, 2020). With 117 million hyperparameters, the model was able to beat state of the art language models in most tasks (Shree, 2020).

Only a year after the creation of GPT-1, OpenAI would go on to create their second iteration, GPT-2 (Shree, 2020). Compared to its predecessor, GPT-2 boasted a whopping 1.5 billion hyperparameters, making it nearly 10 times larger in size (Team, 2021). To achieve this, OpenAI utilized web scraping efforts on the popular social media platform Reddit to collect over 40GB of text data (Shree, 2020). The main improvement focus of GPT was that of task conditioning for language models (Shree, 2020). An example of this is the ability to translate English text to another language, such as Japanese or Spanish using the model's relation in understanding (Shree, 2020). The improvements made to the model would allow for GPT-2 to outperform nearly all state-of-the-art models of the time (Shree, 2020).

The most recently released iteration of the GPT neural network is GPT-3, released in 2020 (Team, 2021). GPT-3 implemented a many new sources of data that would drastically increase the number of hyperparameters in the model to 175 billion from GPT-2's 1.5 billion (Team, 2021). This would be mainly due a vast dataset known as CommonCrawl, which collected data from across the internet. With this vast influx of data, GPT-3 gained the ability to perform arithmetic equations, write paragraphs of text, including programming functions, and summarization of text pieces (Team, 2021). GPT-3 is regarded as a state-of-the-art natural language model per date of this tech review (Shree, 2020).

The GPT models, specifically GPT-3, have opened the door for seemingly endless possibilities in terms of automation. Companies may use GPT-3 to create chat-bots able to help customers with general queries, or even to create reports based on sets of text information (Team, 2021).

References:

Shree, P. (2020, November 10). *The journey of open AI GPT models*. Medium. Retrieved November 5, 2022, from https://medium.com/walmartglobaltech/the-journey-of-open-ai-gpt-models-32d95b7b7fb2

Team, 360D. T. M. G. (2021, July 23). *GPT-1, GPT-2 and GPT-3 in artificial intelligence - 360digitmg*. 360digitmg.com. Retrieved November 6, 2022, from https://360digitmg.com/types-of-gpt-in-artificial-intelligence

Romero, A. (2022, April 17). *GPT-4 is coming soon. here's what we know about it*. Towards Data Science. Retrieved November 5, 2022, from https://towardsdatascience.com/gpt-4-is-coming-soon-heres-what-we-know-about-it-64db058cfd45