MATHEMATICS

COMPUTERS ARE LEARNING TO DO MATH, BUT CAN THEY LEARN TO THINK ALONG THE WAY: A LOOK AT COMPUTER-ASSISTED PROOFS

Written by Jackson Kunde
Edited by Walid Lakdari

Since its nascent days in the minds of 19th-century mathematicians and its earliest implementations breaking codes during World War II, computer science has been intertwined with the study of mathematics. Mathematical logic formed the basis of computing, while computers allowed for fast arithmetic calculations. In short, mathematics advanced computing, and computing advanced mathematics. While these sciences are inextricably linked, it wasn't until the 1970s that computer science found applications in the abstract branches of mathematics known as "pure" math. Now, artificial intelligence is developing new algorithms and assisting some of the most brilliant pure mathematicians of our lifetime, like fields medalist Terrance Tao (Tao, 2023).

The First Computer-assisted Proof

The first theorem that relied heavily on computer assistance is the four-color theorem. The problem is as follows: take any map, for example, the map of the United States and color each state such that no two states with the same colors are touching. For example, we may color California and Wisconsin the same color, but not California and Oregon as they border one another. This idea can be described as a problem in graph theory where each state is a vertex with an edge between each border state. Mathematicians posed the question: Can we color the map using only four colors?

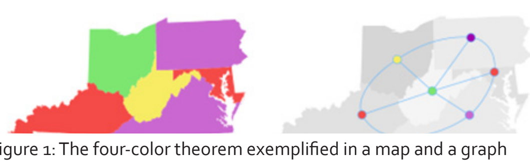

Figure 1: The four-color theorem exemplified in a map and a graph (Source: Quanta Magazine)

The problem originated in the 1850s when South African mathematician Francis Guthrie noticed that the counties of England could be colored with only four colors. He speculated that any map could be colored using only four colors (Richeson, 2023). For over a hundred years, no one was able to produce sufficient mathematical proof of Gunthrie's conjecture. When mathematicians Kenneth Appel and Wolfgang Haken at the University of Illinois took on the problem in 1976, they had a new trick up their sleeve: computers. They devised a proof that would give computers the ability to test many different cases to find a solution to the theorem. Over six months and thousands of hours of computing time, Appel and Haken were able to check thousands of configurations exhaustively (Richeson, 2023). The reward for their work: a long sought-after proof of the four-color problem. With this breakthrough, the relationship between pure mathematics and computing was established.

Computers Start Learning

That proof was established nearly 50 years ago. Since then, computer development has exploded —the IBM computer that Illinois mathematicians relied on is now less powerful than the modern cell phone (Love, 2014). With this improved hardware, scientists have developed methods to teach computers how to learn new tasks, such as games and algorithms. One of Google's artificial intelligence labs, DeepMind, is a pioneer behind an approach that teaches computers known as "reinforcement learning." Training a computer using a reinforcement learning algorithm is very similar to training your dog to do tricks. Unfortunately, we cannot simply explain to dogs how to roll over; instead, we give dogs treats if they complete a trick correctly or get close to completing the trick. This is how researchers can teach computers complicated tasks; rather than trying to explain and encode a good approach to a game, which we often cannot even verbalize or define, we can give the computer a list of possible "moves" that it could choose from and reward it if it performs well in the game.

Through this approach, DeepMind built artificial intelligence which achieved superhuman performance in a variety of different games. In the Chinese strategy game Go, they developed artificial intelligence that Lee Sedol, one of the best Go players of all time (DeepMind). In Chess, their artificial intelligence, AlphaZero, learned to play chess better than many grandmasters (Silver et. al, 2017). Furthermore, this approach is not only limited to existing games, DeepMind discovered that they could teach AI new tasks by formulating them as games like Chess and Go. It was this approach that allowed DeepMind to teach computers to develop new mathematical algorithms.

DeepMind taught artificial intelligence to generate new, faster algorithms to do matrix multiplication. Matrix multiplication is a mathematical operation that underlies many of the calculations that computers perform regularly. It is used in computer graphics, physics simulations, and even in the algorithms used to train artificial intelligence. Due to matrix multiplication's ubiquity, developing efficient algorithms for it could improve computing speeds across numerous computer programs.

To accomplish this, DeepMind created a single-player game where each move that the AI could take would correspond to a step in the matrix multiplication algorithm. The reward for the player would be finding fewer steps to successfully multiply the matrix.

The "game" of matrix multiplication is extremely difficult — the number of possible algorithms is far greater than the number of possible games of Chess or Go. In spite of the difficulty, though, the computer successfully discovered a variety of new algorithms to multiply matrices without any prior knowledge of previous algorithms. It rediscovered current state-of-the-art algorithms and developed new ones. If simple algorithms would take 100 steps and state-of-the-art approaches could solve the problems in 80 steps, a computer could learn to do it in just 76 (Fawzi et. al, 2022). This research demonstrated that computers could advance modern mathematics through learning.

These findings, although ground-breaking, have various drawbacks. To continue to make strides on any problem in mathematics we will have to formulate a new game with all the necessary rewards and possible moves similar to what researchers at DeepMind did for matrix multiplication. This will mean that each problem would come with its own set of "learned rules" that would have to be input into the computer each time – a process that is difficult, time-consuming, and simply not possible for every single problem out there. Furthermore, the new approaches that AI models might produce are not guaranteed to be understandable to humans. Though we could make use of the findings provided by artificial intelligence, we will not necessarily understand why they work. This is a problem. Mathematical proofs are formed from logical steps and arguments that explicate the validity of a new theorem. Without that support, it would be difficult for future mathematicians to build from the work of AI.

An AI That Can Explain Why

In recent years, a new type of artificial intelligence known as large language models (LLMs) has skyrocketed in popularity. These models have tantalized researchers with the idea that a computer system could not only develop new ideas but also explain the underlying logic that led them to their ideas. These models learn from vast amounts of textual data, encompassing sources such as books, articles, websites, and a substantial portion of the internet. OpenAI's GPT4, one of the most well-known LLMs, learned from approximately thirteen trillion words (OpenAI, 2023). During training, the model is presented with input sentences and must predict the next word. To put that enormous amount of data into perspective, if someone were to read 300 words per minute, which is around average, it would take them about 82,000 years to read all the words that GPT-4 was trained on.

Figure 2: A visual of how GPT predicts the next word from a sequence

One of the compelling features of this approach is how much simpler this training process is than the last approach. Instead of creating an entirely new game with actions and rewards, these models can use the same information, textbooks, articles, and blog posts that humans use to learn about different subjects. If the last approach was like training a dog, LLMs are more like training humans.

This approach has yielded some very impressive results, OpenAI's GPT-4 scored a 5 on the AP Biology exam, scored over 1500 on the SAT, and passed the bar exam (OpenAI, 2023). Beyond this AI's competitive college application and its status as a licensed attorney, scientists across domains are interested in how intelligent these systems can become. While some scientists are hopeful that these models can be the first artificial general intelligence and surpass human intelligence, many say that LLMs are no more than "stochastic parrots" that simply predict the next word (Bubeck et al, 2023). The ability to do mathematics, a purely logical, yet exceptionally creative science, may be a proxy for the model's intelligence.

"The ability to do mathematics, a purely logical, yet exceptionally creative science, may be a proxy for the model's intelligence."

Unlike the standardized tests mentioned earlier, mathematics has proven very difficult for these AI models. Early versions released in 2022 would often fail at middle and high-school-level mathematics (Frieder et al, 2023). More recent models, like GPT-4, have improved, and early tests have shown that GPT-4 is proficient at basic math while also being able to do some university-level mathematics. However, while GPT-4 can do upper-level problems, this math is far from research-level mathematics, which not only requires significant domain knowledge but also requires clever, original approaches. When GPT-4 has been asked to complete math problems that require a clever solution, such as questions from the International Mathematical Olympiad, it "fails spectacularly" (Frieder et al, 2023). Similarly, GPT-4 fails when asked to answer questions that require advanced approaches and knowledge such as graduate-level mathematics (Frieder et al, 2023).

One method that has improved LLMs' mathematical reasoning is "process supervision." Essentially, if the LLM is the student, process supervision is the teacher asking the student to "show their work." Researchers train another AI model to evaluate LLM solutions based on the mathematical steps that it took to arrive at their final answers. This process has been shown to lead to more correct answers from models and could have the added benefit of forcing LLMs to explain the rationale behind their solutions, rather than just outputting an answer (Lightman et al, 2023).

Conclusion

In conclusion, as we imagine the development of artificial general intelligence, the symbiotic relationship between mathematics and computer science continues. A computer system that can understand mathematics and meaningfully contribute would not only revolutionize the field but would prove there is a model that can follow logical steps and be creative — in other words, a computer that can think.

Works Cited

Tao, T. (2023, October 24). Embracing change and resetting expectations. Microsoft Unlocked. https://unlocked.microsoft.com/ai-anthology/terence-tao/

OpenAI. (2023, March 27). GPT-4 technical report. arXiv.org. https://arxiv.org/abs/2303.08774

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023, April 13). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv.org. https://arxiv.org/abs/2303.12712

Richeson, D. S., & Quanta Magazine moderates comments to facilitate an informed, substantive. (2023, August 8). Only computers can solve this map-coloring problem from the 1800s. Quanta Magazine.
https://www.quantamagazine.org/only-computers-can-solve-this-map-coloring-problem-from-the
-1800s-20230329/

Toal, S. J. (2017, October 18). Celebrating the four color theorem. College of Liberal Arts & Sciences at Illinois. https://las.illinois.edu/news/2017-10-18/celebrating-four-color-theorem

Overbye, D. (2013, April 28). Kenneth I. Appel, mathematician who harnessed computer power, dies at 80. The New York Times.

Frieder, S.; Pinchetti, L.; Griffiths, R.-R.; Salvatori, T.; Lukasiewicz, T.; Petersen, P.C.; Chevalier, A.; Berner, J. Mathematical capabilities of chatgpt. (2023). Cornell University

Love, D. (2014, May 19). The specs on this 1970 IBM mainframe will remind you just how far technology has come. Business Insider.
https://www.businessinsider.com/ibm-1970-mainframe-specs-are-ridiculous-today-2014-5 DeepMind. (n.d.). AlphaGo. Google DeepMind. https://deepmind.google/technologies/alphago/

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2017, December 5).

Image: DeepMind