A Discussion of the Formalization of Mathematics and its Implications

CSCI 3702 Final Paper

Allison Liu

May 2 2021

Abstract

The thought processes of machines and humans is fundamentally different. However, in the field of mathematics, projects are being developed to directly formalize the jumps in logic that humans make when writing proofs. Those formalizations can be used to create a database of mathematical knowledge for computers to eventually learn from and build off of. Automating mathematical reasoning has major implications, not only for the field of mathematics, but also on the way humans are able utilize computers as tools.

1 Introduction

Mathematics is a framework that humans have used to understand the world since Pythagoras in the 6th century BC. It is a fundamental lens through which we look to describe the world. Due to its long history, mathematics is often viewed as a comprehensive field of study. In reality, like many STEM fields, there exist many unproven "holes" in the field. Specifically, when it comes to proofs—which are the building blocks for what is true mathematically—these gaps in knowledge are noticeable. Because mathematics is so expansive and has a large, logical foundation, it would be reasonable to consider programming a computer to check and possibly even generate new proofs to fill in the gaps of mathematics that humans have missed. However, while humans and machines are capable of solving the same types of logical problems, the strategies which they use are fundamentally different and highlight the differences between their cognitive processes.

Most of the time, mathematical proofs involve a lot of inductive reasoning, creativity, and intuition, which makes solving these types of problems very difficult for computers. To make the task feasible involves a special subset of math called "formalizing mathematics," or explicating each logical step involved for a proof [1]. Once a computer can interpret each step of logical reasoning for a proof and those steps are stored in a database, we can create a graphical representation of known mathematical facts where each proof step is a node with an edge connecting it to its logical implications. Then, using machine learning techniques, a computer can then verify information already known or even extrapolate new mathematical truths. Currently, people are working on formalizing mathematics to accomplish this overarching goal of automating mathematical reasoning [2]. Specifically, I will focus on an open source mathematical library called LEAN [3].

The task of formalizing mathematics requires an in-depth understanding of the mathematical problem itself, as well as someone proficient in articulating human thought processes. This presents a difficulty; in addition to the technical mathematical knowledge necessary to make the task successful, there is also not a widespread push to formalize mathematics [1].

Machine learning is a computational technique that is presented as a "black box" model, or a model whose steps are difficult to trace and whose conclusions cannot always be explained using human logic. Mathematicians and other researchers in STEM fields are slow to adopt new machine learning techniques, despite the advances that machine learning could bring to their respective fields. For mathematicians, it is something that causes skepticism precisely because it is very difficult to follow the logical processes behind a conclusion. A "gray box" model using traditional pen-and-paper techniques for a theoretical understanding combined with the use of machine learning techniques would be ideal. A "gray box" model should be used in STEM, so that we are able to understand what is happening in the big picture, while also utilizing the capabilities of modern computing power effectively. Furthermore, humans would still be a necessary, as it is important to find a balance between allowing a computer to explore new avenues of the field of mathematics and having a human dictate what is actually important and useful to solve [4].

In this paper, I plan to discuss the differences in cognitive thought processes between computers and humans that makes a task such as formalizing mathematics difficult. I then consider the limitations on what computers can be used for, particularly as it relates to mathematics, and finally, discuss the implications of the success of a project such as LEAN.

2 The thought processes of humans and computers

Since the invention of the computer in the 1940's, we have used computers as tools for performing complex intellectual tasks. Alan Turing once defined the goal of an "intelligent computer" as a machine that could indistinguishably imitate a human [5]. While the field of computer science has moved away from that thought as an overarching goal, we still use human intelligence—including elements of linguistics, psychology, philosophy, neuroscience, and more—to inspire our ideas for “teaching” a computer to process information like a human and output results we can understand [6].

The reason that artificially intelligent machines are often spoken about with anthropomorphic language, like “learn”, “teach”, “think”, and “intelligent” is because the ways in which computers and humans process information is often analogous. Computers and computer programs nowadays have the capability to display characteristics reminiscent of human intelligence, so it becomes natural to anthropomorphize them and talk about them as if they were human. In theory, the field of mathematics would be a great place to start, because it is something that can be represented symbolically, in a way that a computer could interpret easily.

However, there still remain fundamental differences in the way humans and computers process information. For one thing, machine learning is very data-inefficient compared to a human. A human child is able to generalize information from a single instance and then can re-shape their schema, or framework for interpreting the world, when they discover something that is contradictory to their current one. For example, if a child sees a dog and a parent tells them that it is a dog, the image of a furry four-legged object as a dog is incorporated into the child’s world-view or schema. The next day, the child sees a fox and says “I see a dog”, the parent says “No, that’s a fox” and the child’s schema changes. The child can learn from a single instance. A computer, however, must be taught with hundreds, or thousands, of examples in order to interpret an image of a dog. This makes deriving useful information from small data sets very difficult or even impossible for computers.

Furthermore, humans and computers can be tricked in different ways, and we must be cautious and aware of the implications of those situations. Human thinking is inherently biased and our judgement is not objective. It has been demonstrated that people are extremely susceptible to being swayed by stereotypes, rather than considering probability or base-rate frequencies of the occurrence of certain events [7]. We also have noticeable misconceptions of chance (ex. small numbers fallacy) and the framing of problems affects us greatly [7]. Computers can aid in helping humans be more objective in this way, however, we must still be conscious of the fact that computers are not completely objective because we can impart human biases into the computer programs we write. This is especially the case when it comes to machine learning techniques where we train computers using human generated information.

There are many examples in the literature of machine learning algorithms being tricked very easily. By using a filter that makes an indistinguishable difference to the human eye, it is easy to trick a neural network into classifying a panda as a gibbon [8] or a banana as a toaster [9]. It also becomes important to understand the process behind the decisions and outputs a computer gives us.

When it comes to mathematical proofs specifically, humans use high-level reasoning and problem-specific insights to prove theorems. It is incredibly difficult for a computer to follow the logical thought process behind a proof without specific instructions. We skip a lot of steps when we think, without even realizing it. Consider the following story: "It's Kevin's birthday this weekend. Kate thinks he would like a kite. She runs upstairs and shakes her piggy bank. It makes no sound." Even young children can make a very solid conclusion about what happened in the story. But it takes a lot of context to understand it. There are many jumps in logic made in every step of the story that would be very difficult for a computer to understand without lots of prior training. The same is true of mathematics. For a program such as LEAN to be successful, humans will have to figure out how to explicate many of the formal mathematical truths we have discovered (and the logical jumps taken) and re-frame them so they can be understood by a computer.

3 The potential power of LEAN and its impacts

In 2017, researchers developed an artificially intelligent computer program called *AlphaGO* which successfully defeated the one of the world's best GO players [10]. This task does not sound that impressive in 2021...computers have been competitive with humans at games for decades. However, it was actually an incredible accomplishment for the field of machine learning research. GO is an incredibly intricate game that has been played for centuries. There are people who go to special schools and even dedicate their entire lives to learning the game. When the computer defeated the human, people were at first surprised at its moves and believed the computer had made a mistake. Truly though, the computer had learned a new strategy for winning the game–one that humans hadn't discovered in centuries of playing the game. What if we could use this strategy for math, a field that has major implications for human lives?

If we can teach a computer to make jumps in logic with math, it would show potential for doing it with other topics, like interpreting text or speech, or solving research problems, or computer vision. LEAN and other formalizing mathematics tools created based off of machine learning techniques are the beginning of the future of mathematics. Building these tools for mathematics would progress our understanding of how computers interpret human thought and would aid in the development of more advanced machine learning tools, in particular for natural language processing and tasks for which we want computers to imitate human thought.

4 Difficulties and limitations of formalizing techniques

Mathematics is an incredibly expansive and diverse field and even giving a human no constraints for exploring the field of math would be too much. We would need to give a computer, which can process orders of magnitudes of information more than we can as humans, some direction for it to be useful. Even if a mathematical problem solver eventually becomes successful and we are able to formalize much of what we know, mathematicians will still be necessary to continue to formalize or verify new discoveries made by humans. In any case, it is necessary to determine the kinds of knowledge that are important for humanity and that must be a task done by humans. There is a concept in machine learning (specifically reinforcement learning) called exploitation and exploration. Exploitation is the algorithm's tendency to make similar decisions to ones that have benefited it in the past, and exploration is the algorithm's tendency to explore new areas of the space that it has not explored yet. The goal of the programmer is to set parameters to strike a balance between exploration and exploitation to optimize the effectiveness and efficiency of the algorithm. In the case of mathematical problem solvers, mathematicians and computer programmers both serve as the people adjusting the parameters of where in the field of math should be explored. Only searching for mathematical truths in areas that we have already explored a lot and know are useful may be good, but the field of math is so expansive that we could also be missing out on new discoveries in areas that we may not have even known existed and that are incredibly interesting. Balancing the two areas must be done by a person, because giving a computer free reign to explore in all directions or to just go down a "rabbit hole" of the same types of math would not be a good use of computing power or time. The balance must be tuned and updated every once in a while.

Due to the "black box" model of machine learning, it is often impossible, or at the very least incredibly difficult, to trace the logical process a computer took to arrive at a particular conclusion [2]. As we use math to help us form an understanding of the world, it becomes necessary to also understand the process of arriving at a solution. The black box model does not exactly lend itself to doing so easily. One of the reasons machine learning has been slow to catch on in many areas of STEM research is because of the lack of understanding we have of its results. Many researchers enter the field of research because they are inherently curious and desire genuine understanding, so getting a result from a machine that is unexplainable is highly unsatisfying and not all that helpful. If LEAN or similar mathematical problem solvers are to be successful, we must also figure out a way to program them such that it is possible to follow the computer's logical thought process. This is difficult, because the computer is able to parse much more information efficiently than we are as humans, and there are often many "hidden layers" buried in the data that we are unable to distinguish. Ideally, we can come up with a sort of "gray box" model where we are able to combine a physical or symbolic understanding of the problem with the aid of computational tools. A good start would be to use the problem solvers simply as tools to verify old discoveries/theorems or even to check conjectures.

Our lack of understanding of the computer's process presents a ethical problem for using machine learning techniques in general.

5 Ethical Considerations

As we consider the implications of an artificially intelligent mathematical computer program that is in the beginning stages of development, a discussion on ethics is warranted. Whenever we create computer programs that we allow to make judgements and decisions that affect the lives of real people, it is necessary to have a discussion on the ethics of what we are creating [11]. The ethics of machine learning is constantly being debated as we enter an age where computers become an important part of every day life. Unfortunately, machines are not as unbiased when making decisions as we sometimes assume they are, even though they are binary machines [12]. Because computers output results based on information produced by humans, they are also subject to the biases of human thought. Computational results become even more dangerous when we cannot explain how or why a computer produced results it did.

Additionally, the anthropomorphic rhetoric we use to discuss machine learning can be very useful when creating or understanding a model, but it can become dangerous if taken too literally [13]. Human thought and computational thought are analogous and we can use concepts from both to help understand the other; however, we cannot view computers as substitutes for human decision-makers. We cannot allow computers to make decisions that affect humans without regulation, because the biases and judgements imparted by humans cannot be removed or explained [11]. It is not possible to ask a computer to justify its decisions like we can with a person. These days, we allow black box machine learning algorithms dictate whether a person will be able to get a credit card, or take a loan out on a house, or how much they pay for insurance, but there is nothing regulating the use of these computer programs. Companies used artificially intelligent resume parsers for years until it was discovered that many of then were gender-biased [11]. Joy Buolamwini discovered many facial recognition algorithms could not recognize Black people with as much accuracy [14]. With the development of any artificially intelligent machine, it is imperative to ensure ethical considerations are kept in mind and to constantly re-evaluate those ideas. With a mathematical proof-solver, this would be important because mathematics is often the basis of these algorithms.

The future of machine learning will be figuring out a method to track the steps taken by an algorithm to arrive at a solution. Perhaps mathematics would be a good place to start, because it can be represented symbolically using logical steps. While time-consuming and tedious to do so, it is possible to track the solution using some proof-assistant programs [2].

6 Conclusion

Ultimately, formalizing mathematics would allow us to automate mathematics. The development of a computer program such as LEAN or a similar mathematically intelligent solver could help to verify current knowledge, make the field of mathematics more complete, and potentially make the discovery of new mathematical truths more efficient as it could lead us down branches of math never explored before as it creates new proofs. Furthermore, the task of algorithmically automating math could give insights into the connection between human and computational thought processes which could carry over into other fields of information processing. This would allow for the improvement of computational learning tasks in other areas, such as natural language processing, computer vision, and image processing. Mathematicians would still be necessary to determine what areas of math are actually useful and interesting to study though, as the potential for the computer to get lost in the depths of a particular area of mathematics is much too high [4]. LEAN and similar software will be the beginning of the future of math. As with all artificially intelligent machines, we must implement measures to ensure the ethical use of anything we create, which includes being able to explain the process that machines often take to arrive at conclusions.

Word Count: 2909

References

[1] K. Buzzard, *Formalising mathematics: An introduction.* [Online]. Available: https://xenaproject.wordpress.com/2021/01/21/formalising-mathematics-an-introduction, (accessed: 01.23.2021).

[2] K. Yang and J. Deng, "Learning to prove theorems via interacting with proof assistants," *CoRR*, vol. abs/1905.09381, 2019. arXiv: 1905.09381. [Online]. Available: http://arxiv.org/abs/1905.09381.

[3] L. de Moura, S. Kong, J. Avigad, F. van Doorn, and J. von Raumer, "The lean theorem prover (system description)," in *Automated Deduction - CADE-25*, A. P. Felty and A. Middeldorp, Eds., Cham: Springer International Publishing, 2015, pp. 378–388, ISBN: 978-3-319-21401-6.

[4] J. E. Pretz, A. J. Naples, and R. J. Sternberg, "Recognizing, defining, and representing problems," in *The Psychology of Problem Solving*, J. E. Davidson and R. J. Sternberg, Eds. Cambridge University Press, 2003, pp. 3–30. DOI: 10.1017/CBO9780511615771.002.

[5] A. M. TURING, "I.—COMPUTING MACHINERY AND INTELLIGENCE," *Mind*, vol. LIX, no. 236, pp. 433–460, Oct. 1950, ISSN: 0026-4423. DOI: 10.1093/mind/LIX.236.433.

[6] R. French, "Moving beyond the turing test," *Communications of the ACM*, vol. 55, pp. 74–77, Dec. 2012. DOI: 10.1145/2380656.2380674.

[7] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974, ISSN: 00368075, 10959203. [Online]. Available: http://www.jstor.org/stable/1738360.

[8] I. J. Goodfellow, J. Shlens, and C. Szegedy, *Explaining and harnessing adversarial examples*, 2015. arXiv: 1412.6572 [stat.ML].

[9] T. B. Brown, D. Mané, A. Roy, M. Abadi, and J. Gilmer, "Adversarial patch," *CoRR*, vol. abs/1712.09665, 2017. arXiv: 1712.09665. [Online]. Available: http://arxiv.org/abs/1712.09665.

[10] D. Silver, J. Schrittwieser, and K. Simonyan et. al., "Mastering the game of go without human knowledge," *Nature*, vol. 550, pp. 354–359, 2017. DOI: 10.1038/nature24270.

[11] B. Christian, *The Alignment Problem: Machine Learning and Human Values.* W. W. Norton Company, 2020, ISBN: 9780393635829.

[12] K. Kertysova, "Artificial intelligence and disinformation," *Security and Human Rights*, vol. 29, pp. 55–81, Dec. 2018. DOI: 10.1163/18750230-02901005.

[13] D. Watson, "The rhetoric and reality of anthropomorphism in artificial intelligence," *Minds and Machines*, vol. 29, pp. 417–440, 2019. DOI: 10.1007/s11023-019-09506-6.

[14] S. Kantayya, *Coded Bias (Documentary).* 2020.