Application of Machine Learning and Artificial Intelligence Techniques to Automated Theorem Provers (ATPs)

By Ryan Jaeger

The Pennsylvania State University, University Park, PA 16801

Introduction

A. Background

Automated theorem proving, one of the central topics of automated reasoning, centers around the use of computers to formally prove a given conjecture. Automated theorem provers, or ATPs, take a conjecture and set of initial conditions as inputs and then follow a logical calculus to rigorously prove the conjecture. While some mathematicians may be skeptical of the ability of a computer to complete a rigorous proof, in 1997, William McCune of the University of New Mexico developed a custom ATP to successfully prove the previously unsolved Robbin's conjecture, a topic in Boolean algebra [5]. This achievement indicated to the mathematical community that the idea of automating mathematical proofs may not be so outlandish.

However, despite the success with the Robbin's conjecture, automated theorem provers have failed to gain widespread traction from mathematicians [1]; Bundy asserts that the limitations of the power and usability of ATPs are the reason behind this. He claims that ATPs are not powerful enough to prove "novel conjectures of interest to mathematicians." Blanchette and Kühlwein concur, arguing that the reasoning power behind ATPs is "far behind what is considered standard for a human mathematician" [5]. Because computers struggle with mathematical concepts such as symmetry and analogy, it can take significantly longer for an ATP to prove a theorem than a mathematician. Beyond the lack of power, the difficulty of use has also prevented theorem provers from becoming particularly popular. Human input is often required to translate between mathematical notation and prover syntax and to guide the prover if errors occur [1]. Furthermore, the output from an ATP is complex and "notoriously unreadable" as described by Melis and Siekmann [7]. Because so much expertise is required to gain value from an automated theorem prover, this acts as a large barrier to entry. Together, these shortcomings of ATPs have stagnated their usage.

B. Topic of Inquiry

Over the past few decades, the fields of artificial intelligence (AI) and machine learning (ML) have significantly developed and techniques from them are being applied across fields. In this literature review, I intend to examine whether there exists compelling evidence to suggest that techniques from machine learning and artificial intelligence can be applied to automated theorem provers to improve power or usability. To evaluate the weight of the evidence, I will first assess the utility of AI/ML methods in improving the processes of existing theorem provers. Then, I will consider new, novel automated theorem provers that make use of AI/ML concepts in their design and function.

C. Scope

In order to put reasonable constraints on the scope of this literature review, I focus exclusively on first-order theorem provers; higher-order provers are not considered. Furthermore, I exclude interactive theorem provers, which work with humans step-by-step to develop a proof, in favor of automated theorem provers, which can solve problems from end to end without continued human guidance. Finally, the contents of this review will focus primarily around the results of utilizing artificial intelligence and machine learning for automated reasoning rather than the mechanics of any of these methods.

Applications of Artificial Intelligence/Machine Learning to Existing ATPs

The general approach for automated theorem provers involves a search of existing axioms and heuristics to find the correct combination that leads to a rigorous proof for a given conjecture and conditions. In this process, there are multiple steps in which AI/ML techniques have been applied in hopes of streamlining the proving process.

A. Axiom Selection

When given a conjecture to prove, the automated theorem prover will search through its libraries for any mathematical axioms that may be related to the problem at hand. This library of axioms can often be extensive, and the search process can become time-consuming. An axiom selection algorithm works to identify the most relevant axioms to provide a starting point for the proving process. For example, when given a conjecture that involves pi and the sine function, the axiom selection algorithm will rank axioms containing pi and the sine function as the most relevant [5]. In their experiment, Blanchette and Kühlwein explore whether or not dependency trees, a machine learning algorithm, can be applied to axiom selection algorithms to eliminate any extraneous axioms from being attempted during the proving process. While the concept presented by Blanchette and Kühlwein appears promising, their paper is only an introduction to the technique and offers no substantive evidence of it success. Specifically, they mention that the validity of this idea cannot be found without attempting on multiple ATPs, during which the axiom selection improvements may be confounded by hardware, time limit, and ATP version.

B. Heuristic Selection

With a similar motivation as Blanchette and Kühlwein to increase the power of automated theorem provers by controlling the parameters of proof search, Bridge et al. explore whether machine learning can be applied to the heuristic selection. In automated reasoning, the term heuristics refers to a collection of standard settings for parameters and general set of conditions for approaching a given problem; the optimal heuristic is different for each problem, so it can be time-intensive for a computer to find the appropriate one. Through an evaluation of two machine learning algorithms, support vector machines and Gaussian processes, against 6118 problems from the TPTP library, Bridge et al. were able to develop a machine learner that selects an appropriate heuristic better than random chance would [4]. While this learner did not outperform typical human-inputted heuristic choices, it achieved the same level of success over a large sample of problems. This success indicates that it is possible for human input to become obsolete for heuristic selection without any loss of power.

C. Search Control

Often, ATPs incorporate the methodologies from previously proved theorems to inform their decision making when given a new conjecture; however, they "perform poorly in the presence of large fact libraries" [3]. Loos et al. seek to establish the ability of deep neural networks to be effectively leveraged to provide guidance during the proof search process. Because deep neural networks can be slow, incorporating them into ATPs would require a high quality of output to balance the tradeoff of efficiency. Over an analysis of a deep-neural-network-guided proof search of 91,877 problems from the Mizar Mathematical Library, Loos et al. found major improvements to proving power; this boosted ATP found proofs for 7.36% of problems previously unsolved by ATPs. While this method finds proofs to difficult conjectures, when placed under a time constraint, it actually proves fewer than its contemporaries. The success of this method over such a large quantity of problems suggests this method should be reliable, yet, the speed tradeoff must be seriously considered.

D. Analysis

The current developments of machine learning applications to improving the processes of existing ATPs seems to either increase power or usability but not both. The deep neural network used by Loos et al. provided the ability to solve more difficult problems with an ATP, yet it takes significantly longer to do so, thus decreasing its usability for mathematicians. To the contrary, the heuristic selection employed in the research from Bridge et al. did not indicate an improvement in power, yet did make obsolete human input in heuristic selection. While these two applications have shown results, further research would be needed for Blanchette and Kühlwein's work to establish its credibility.

Application of Artificial Intelligence/Machine Learning to New ATPs

Beyond simply improving the existing processes of automated theorem provers, principles and techniques from machine learning and artificial intelligence have also been applied to make entirely original automated reasoning tools.

A. Bootstrapping ATP

In their 1997 paper from the International Joint Conference on Artificial Intelligence, Denzinger et al. contend that multiple ideas of artificial intelligence can be applied to ATPs and present the design for a new ATP that incorporates these concepts. Specifically, they suggest "the combination of case-based reasoning, several similarity concepts, a cooperation concept of distributed AI and reactive planning enables a system to learn from previous successful proof attempts" [2]. Denzinger et al.'s design incorporates a bootstrapping framework, where a prover will learn how to solve harder proofs by first learning how to solve easy proofs. The integration of these multiple techniques results in a prover that outperformed 'current conventional theorem provers" given that "enough exercise in the domain" had been provided [2]. While this explanation of results sounds promising, it requires previous knowledge to be successful, which may not always be the case in practice. Also, while it may have been more successful than provers in 1997, there is no guarantee it is competitive today.

B. Proof Planning

Akin to Denzinger et al., Melis and Siekmann look to present a "new paradigm" within automated theorem proving which employs alternative aspects of artificial intelligence. In this case, human-oriented aspects of AI, like hierarchal planning and meta-level reasoning, are emphasized over the computer-oriented techniques of machine learning [7]. Rather than searching for a working proof in a sea of logical calculus, knowledge-based proof planning, as proposed by Melis and Siekmann, seeks to plan the proof at the abstract level, then recursively expand the proof from there by leveraging mathematical domain knowledge and guidance of control-rules. The Omega system presented in their findings successfully "solved all the well-known challenge theorems including those that cannot be solved by any of the existing traditional systems" [7]. In fact, Melis and Siekmann compare their own work to atomic fission, suggesting "metaphorically speaking, we have shown the atom can be split and indeed it gives off energy." While I have my reservations in concluding that this advancement is analogous to atomic fission, it is clear that the Omega proof-planning system has value. Additionally, because this paper was published in 1999, I have similar concerns as [2] about its applicability today.

C. Machine Learning Feedback Loop

Whereas the two previous authors emphasize artificial intelligence over machine learning for creating new automated theorem provers, Josef Urban introduces a metasystem that utilizes iterative steps of automated reasoning and machine learning. The system, known as MaLARea (Machine Learner for Automated Reasoning), proves theorems during the reasoning phase and then learns from successful proofs during the machine learning phase to better inform reasoning processes in the future [6]. Using the metric of challenge problems within the Mizar library, the first iteration of MaLARea solved 142 out of the 252, more than any comparable system. This concept of directly incorporating machine learning and automated reasoning into a single tool seems very promising. However, at the time of publication in 2007, MaLARea was still in its earliest stages with many potential extensions and improvements, so its applicability today also remains untested.

D. Analysis

Despite some concerns about the relevancy of specific results from Denzinger et al. and Melis and Siekmann, I argue that they establish the foundational framework that artificial intelligence can be effectively applied to the field of automated reasoning. Further research would need to occur to determine whether or not the provers developed by Denzinger et al. and Melis and Siekman hold any pertinent value today. I contend that the development of MaLARea is very promising within the field of automated reasoning; the creation of system that simultaneously solves conjectures while improving its ability to solve future conjectures is an indispensable symbiosis that will only increase power and usability in the long run.

Conclusion

A. Summary of Findings

In this literature review, I explored the state of machine learning techniques and artificial intelligence concepts in modern automated theorem provers to determine if there is compelling evidence to suggest that these applications of ML/AI are successful. First, I evaluated how machine learning affected axiom selection, heuristic selection, and search control with varying levels of improvement along the criteria of power and usability. Then, I determined the value behind machine learning and AI inspired automated theorem provers, which tend to be conceptually strong, but the supporting evidence may no longer be relevant. Fundamentally, I have found that there does exist compelling evidence to suggest that machine learning and artificial intelligence can be applied to automated reasoning. Specifically, the development of the MaLARea prover and the application of deep neural networks to proof searches sharply increased the proving power of ATPs, one of the major shortcoming of contemporary provers.

B. Future Work

In terms of future research on this topic, I suggest exploring further synergistic systems between machine learning and automated reasoning, in the vein of MaLARea. Furthermore, more research must be conducted on axiom selection to determine whether there lies any credibility in using dependency trees for this purpose. Likewise, research ought to be conducted to compare the provers developed in the 1990s by Denziger et al. and Melis and Siekmann to determine their relevancy now.

References

[1] A. Bundy, "Automated theorem provers: a practical tool for the working mathematician?" *Annals of Mathematics and Artificial Intelligence*, vol. 61, no. 1, pp. 3-14, Jun 2011.

[2] J. Denziger, M. Fuchs, and M.Fuchs, "High Performance ATP Systems by Combining Several AI Methods," in *1997 International Joint Conference on Artificial Intelligence*, Nagoya, Japan, pp. 102-107.

[3] S. Loos, G. Irving, C. Szegedy, and C. Kaliszyk, "Deep Network Guided Proof Search," arXiv: 1701.06972 [cs.AI], Jan 2017.

[4] J. P. Bridge, S. B. Holden, and L. C. Paulson, "Machine Learning for First-Order Theorem Proving," *Journal of Automated Reasoning*, vol 53, no. 2, pp. 141-172, Aug 2014.

[5] J. C. Blanchette and D. Kühlwein, "A Survey of Axiom Selection as a Machine Learning Problem," in *Infinity, Computability, and Metamathematics*, College Publications, pp. 1-15.

[6] J. Urban, "MaLARea: a Metasystem for Automated Reasoning in Large Theories," in *Proceedings of the CADE-21 Workshop on Empirically Successful Automated Reasoning in Large Theories,* Bremen, Germany, 2007, pp. 45-58.

[7] E. Melis and J. Siekmann, "Knowledge-based proof planning," *Artificial Intelligence*, vol. 115, no. 1, pp. 65-105, 1999.