Can Artificial Intelligence Serve Computer Algebra Systems?

Matthew England[0000−0001−5729−3420]

Coventry University, Coventry, CV1 5FB, UK
Matthew.England@coventry.ac.uk

Abstract. *We discuss how Machine Learning can optimise Computer Algebra Systems, and how XAI might inspire future CAS development.*
Keywords: *computer algebra, symbolic computation, machine learning*

1. Introduction

A *Computer Algebra Systems* (CAS) performs exact mathematics through the manipulation of symbolic expressions. Examples include the proprietary systems Maple & Mathematica and the open source SymPy & SageMath. *Machine Learning* (ML) refers to statistical techniques that learn rules from data. ML underpins the AI advances of recent years and is used very widely, but rarely in CASs.

There has been attempts for AI to directly perform mathematics, such as [1] which trained a transformer to integrate expressions and find analytical solutions to differential equations. The transformer solved more problems correctly within a time limit than CASs. However, this analysis combined two cases into failure: when a solution could not be found and when the wrong solution is found. While the CASs did not solve problems quickly, they should *never* produce the wrong answer. Hence there is limit appetite amongst CAS developers to use ML in directly producing their output: accuracy less than 100% would lose their USP.

However, CAS algorithms often come with choices that have no effect on the mathematical correctness of the end result, but can have a big impact on the resources required. Further, they can affect how the end result is presented (two very different, but mathematically equivalent, expressions). These choices range from the low level (in what order to perform a search) to the high (which algorithms to use for this problem instance). Too often these choices fall to heuristics, often undocumented and rarely scientifically validated. The author and others have recently considered the use of ML to make such decisions.

2. Prior work with ML to Optimise CASs

The author published the first paper to use ML in making such a choice in for a CAS: in [2] a support vector machine was trained to select the variable ordering for a Cylindrical Algebraic Decomposition (CAD), an important object for performing quantifier elimination in the real numbers. A following EPSRC project experimented with different approaches, cumulating in [3]. We will survey this and similar uses of ML in CAS, explaining why this is a challenging AI domain.

3. Future Promise from ML to Influence CAS Development

The aforementioned work demonstrated that ML can optimise CASs. But can ML influence symbolic computation further? Consider [4] which studied Buchberger's Algorithm to produce a Gröbner Basis: an important tool for studying solutions of polynomial equations. The algorithm must process a list of pairs of polynomials: it can do so in any order but some orders are more efficient than others. Reinforcement learning was used to make this choice in [4]. Of particular interest was a final analysis that identified some simple components of the agent's strategy pointing to new mathematical understanding of the choice. We propose Explainable AI (XAI) techniques to automatically gain such insights. We present ongoing work to make an XAI analysis of the models in [3]. SHAP has identified a number of new metrics for measuring polynomial complexity, including one recently demonstrated as state-of-the-art heuristic for CAD variable selection [5].

Acknowledgement: Supported by EPSRC grant EP/T015748/1: *Pushing Back the Doubly-Exponential Wall of Cylindrical Algebraic Decomposition* (DEWCAD).

References

[1] Lample, G. and Charton, D. Deep learning for symbolic mathematics. In Proc. ICLR 2020. 2020.
[2] Huang, Z., England, M., Wilson, D., Davenport, J. H., Paulson, L., and Bridge, J. Applying machine learning to the problem of choosing a heuristic to select the variable ordering for cylindrical algebraic decomposition. In Proc. CICM 2014 (LNCS 8543), pp. 92–107. Springer, 2014.
[3] Florescu, D. and England, M. A machine learning based software pipeline to pick the variable ordering for algorithms with polynomial inputs. In Proc. ICMS 2020, (LNCS 12097), pp. 302–322. Springer, 2020.
[4] Peifer, D., Stillman, M., and Halpern-Leistner, D. Learning selection strategies in Buchberger's algorithm. In Proc. ICML 2020, (PMLR 119), pp. 7575–7585. 2020.
[5] del Río, T. and England, M. New heuristic to choose a cylindrical algebraic decomposition variable ordering motivated by complexity analysis. In Proc. CASC 2022 (LNCS 13366), pp. 300–317. Springer, 2022.