Interactive Theorem Provers: Applications in AI, Opportunities, and Challenges

Mohammad Abdulaziz

King's College London, London, United Kingdom

Interactive theorem provers (ITPs) are computer programs in which axioms and a conjecture are stated in a formal language; a user provides the ITP with relatively high-level steps of a formal proof for the conjecture; then, by invoking automated theorem provers, the ITP tries to generate low-level steps that fill the gaps between the steps provided by the user, thus forming a complete formal proof of the conjecture. The ITP also checks the entire formal proof against the axioms, thus confirming the soundness of all derivations in the formal proof.

The first topic I will cover in my talk is *why should the AI community consider ITPs?* The most obvious benefit is that, using ITPs, we can construct correct-by-construction AI software. Indeed, ITPs, more than any other methodology, have been used to construct mathematically proven, correct-by-constuction software in other areas of CS, and they thus offer a promising way to significantly improve the trustworthiness of AI software. Also, using an ITP to reason about mathematically involved AI concepts or algorithms usually leads to new conceptual or abstract insights.

The second topic I will cover is *how can ITPs be (feasibly) applied to reason about and verify concepts and algorithms in AI?* In general, using ITPs to reason about mathematical concepts or software is known to be hard, e.g. it is estimated that the number of proof lines to prove a piece of software correct grows quadratically with the number of lines in that piece of software. This complexity is further compounded if the goal for a verified piece of software is to perform as efficiently as its optimised, unverified counterpart.

Furthermore, in the context of AI, in addition to the general scalability issues faced by ITP-based software verification, there is the extra problem of the relatively mathematically involved theory behind an AI algorithm or a piece of AI software. For instance, to prove that algorithms for planning under uncertainty are 'correct', one needs to build within the ITP an entire background formal theory that formalises notions and theorems about probabilities, Markov Decision Processes, and functional analysis. Such involved mathematical background theory is usually not required for applying ITPs to other domains for which most current ITP-based verification methodologies were developed, like in operating systems and security.

I will discuss those two topics based on examples from my own work on reasoning about and verifying: 1. classical planning algorithms (Abdulaziz, Norrish, and Gretton 2018; Abdulaziz, Gretton, and Norrish 2019; Abdulaziz and Kurz 2023), 2. (temporal) planning semantics and validation software (Abdulaziz and Lammich 2018; Abdulaziz and Koller 2022), 3. algorithms for probabilistic planning (Schäffeler and Abdulaziz 2023), and 4. algorithms for matching, and applications in algorithmic game theory (Abdulaziz, Mehlhorn, and Nipkow 2019; Abdulaziz and Madlener 2023).

I will use examples from my work to 1. show how verifying AI software using ITPs can lead to finding bugs lurking in state-of-the-art, stable, decades old AI software systems, 2. demonstrate how new mathematical concepts and proofs as well as improved algorithms, can be discovered assisted by ITPs, and 3. show the measures and trade-offs my collaborators and I took to improve the feasibility of using ITPs within the context of the mathematically heavy AI software and algorithms.

References

Abdulaziz, M.; Gretton, C.; and Norrish, M. 2019. A Verified Compositional Algorithm for AI Planning. In *ITP*.

Abdulaziz, M.; and Koller, L. 2022. Formal Semantics and Formally Verified Validation for Temporal Planning. In *AAAI*.

Abdulaziz, M.; and Kurz, F. 2023. Formally Verified SAT-Based AI Planning. In *AAAI*.

Abdulaziz, M.; and Lammich, P. 2018. A Formally Verified Validator for Classical Planning Problems and Solutions. In *ICTAI*.

Abdulaziz, M.; and Madlener, C. 2023. A Formal Analysis of RANKING. In *ITP*.

Abdulaziz, M.; Mehlhorn, K.; and Nipkow, T. 2019. Trustworthy Graph Algorithms (Invited Paper). In *MFCS*.

Abdulaziz, M.; Norrish, M.; and Gretton, C. 2018. Formally Verified Algorithms for Upper-Bounding State Space Diameters. *J. Autom. Reason.*

Schäffeler, M.; and Abdulaziz, M. 2023. Formally Verified Solution Methods for Infinite-Horizon Markov Decision Processes. In *AAAI*.