

Towards United Reasoning for Automatic Induction in Isabelle/HOL

Yutaka Nagashima

yutaka.nagashima@cvut.cz

CIIRC, Czech Technical University in Prague
University of Innsbruck

ABSTRACT

Inductive theorem proving is an important long-standing challenge in computer science. In this extended abstract, we first summarize the recent developments of proof by induction for Isabelle/HOL. Then, we propose *united reasoning*, a novel approach to further automating inductive theorem proving. Upon success, united reasoning takes the best of three schools of reasoning: deductive reasoning, inductive reasoning, and inductive reasoning, to prove difficult inductive problems automatically.

KEYWORDS

deductive reasoning, inductive reasoning, abductive reasoning, proof by induction, Isabelle/HOL

1 INDUCTION IN ISABELLE/HOL

Despite the importance of inductive theorem proving, its automation remains as a long-standing challenge in computer science. To facilitate proof by induction, Isabelle/HOL [10] offers tools to apply induction, such as the *induct* method and the *induction* method.

However, these tools were built for human-interaction rather than for automation. For example, when using the *induct* method proof authors are supposed to specify

- on which terms they apply induction,
- which variables they generalize, or
- which induction rules they use for recursion induction

by passing arguments to the *induct* method.

2 PSL: PROOF STRATEGY LANGUAGE

As our first step towards automatic induction, we developed PSL [7], a programmable, meta-tool framework for Isabelle/HOL. Using PSL, one can specify the following example proof strategy:

```
strategy DInd = Thens [Dynamic(Induct), Auto, IsSolved]
```

When one applies this strategy to a given proof goal, PSL's runtime creates various *induct* methods with arguments using the information within the proof goal and its background proof context. Then, it combines those *induct* methods with *auto* and *is_solved*, where *auto* is the general purpose proof method, and *is_solved* checks if there is a remaining sub-goal or not. Based on such combinations of methods, PSL's runtime system executes an iterative deepening depth-first search, trying to identify a combination of induction arguments, with which Isabelle can discharge the proof goal.

Note that despite the confusing name proof by induction is an example of *deductive reasoning*, as well as many other proof methods of Isabelle/HOL: we can deduce the underlying induction principle from the axioms of Isabelle/HOL.

3 PGT: GOAL-ORIENTED CONJECTURING

For some inductive problems it is not enough for proof authors to apply the *induct* method with arguments, but they have to come up with useful auxiliary lemmas. These auxiliary lemmas have to be strong enough to derive the original goals, but they should be provable at the same time.

We proposed our Proof Goal Transformer (PGT) [9] as an extension to PSL to facilitate such *abductive reasoning*.

The advantage of this goal-oriented approach is that we can identify valuable conjectures out of many conjectures through proof search by combining the conjecturing mechanism with Isabelle's proof automation and counter-example finder.

For example, if PGT produces a conjecture, conjecture, while trying to prove an original goal, goal, PGT inserts conjecture as an assumption of goal, transforming the original goal into the following two sub-goals:

- conjecture → goal
- conjecture

Then, Isabelle attempts to discharge the first sub-goal with a standard proof automation tool, *fastforce*, and attempts to refute the second sub-goal with a counter-example finder, such as *quickcheck*. PGT discards this conjecture, conjecture, if *fastforce* cannot discharge the first sub-goal (conjecture is not strong enough) or *quickcheck* finds a counter-example (conjecture is not provable). This way, when combined with other sub-tools, PGT can focus on valuable conjectures to avoid the explosion of the search space.

4 MELOID: MACHINE LEARNING INDUCTION

It is sometimes necessary to apply nested induction with multiple conjecturing steps. Unlike the aforementioned goal-oriented conjecturing approach, there is no known way to narrow the search space of the *induct* method using counter-example finders: the *induct* method often transforms a given proof goal into a base case and step cases while preserving the provability even when inappropriate arguments are given to the *induct* method.

The sub-goals produced by the *induct* method with inappropriate arguments are often still provable if the original goal is provable, but are only harder to discharge because the inappropriate application of the *induct* method transforms the original goals to forms unsuitable for Isabelle's proof automation.

This blows up the search space of PSL when attacking hard inductive problems that require multiple applications of induction: each invocation of *Dynamic (Induct)* produces many variants of *induct* methods, each of which produces inappropriate sub-goals, and the nested applications of *Dynamic (Induct)* produces again

many inappropriate sub-goals for each sub-goal produced by the previous application of the induct method.

To keep the size of search space tractable, we are currently developing MeLoId [3]. MeLoId is a supervised learning framework to suggest promising arguments to the induct method without completing a proof.

The overall architecture of MeLoId is similar to that of PaMpeR [6], which suggests promising proof methods for a given proof goal based on supervised learning on human-written proof corpora. Upon success, MeLoId converts each invocation of the induct method in proof corpora and the corresponding proof goal into a list of boolean values. Then, it applies a machine learning algorithms to this simplified data and learns how human proof authors use the induct method.

In 2018 we conducted our first small scale preliminary experiment using around 40 feature extractors written in Isabelle/ML. The result was not convincing: even though the feature extractors did manage to distill the essence of some undesirable combinations of arguments for the induct method, they turned out to be unable to extract the essence of promising combinations of arguments for most cases.

The problem was that for a feature extractor to distill induction heuristics, such extractor has to be able to conduct a complex abstract reasoning on different inductive problems across various problem domains. Therefore, it is harder to develop useful extractors for MeLoId directly in Isabelle/ML than to develop the feature extractor used in PaMpeR.

For example, a good feature extractor for MeLoId would reason the syntactic structures of both the proof goals and the definitions of constants appearing in the goals with respect to the induction terms passed as arguments to the induct method, whereas many feature extractors of PaMpeR simply check the existence of atomic terms of certain names within proof goals.

Due to this technical challenge, we decided that it is infeasible to develop useful feature extractors for MeLoId in Isabelle/ML. And we developed a domain-specific language, LiFtEr [4], designed to write feature extractors for MeLoId. LiFtEr allows experienced Isabelle users to encode their induction heuristics as assertions in a style independent of any problem domain.

We plan to write many feature extractors in LiFtEr, extract a database from the Archive of Formal Proofs (AFPs) [1] using those extractors, and learn how Isabelle experts choose arguments for the induct method, so that MeLoId can recommend a few promising combinations of arguments for the induct method for a given goal.

Note that MeLoId's numerical evaluation on induction heuristics is an instance of *inductive reasoning*. For example, even if all invocations of the induct method in the AFPs are compatible with a certain induction heuristic, such information does not guarantee that the induction heuristic is always correct. We can only state that the induction heuristic is *probably* reliable.

5 UNITED REASONING

We envision *united reasoning*, a fully automatic inductive theorem prover embedded in Isabelle/HOL. Upon success, united reasoning combines the forces of PSL's deductive reasoning, PGT's abductive reasoning, and MeLoId's inductive reasoning, transforming

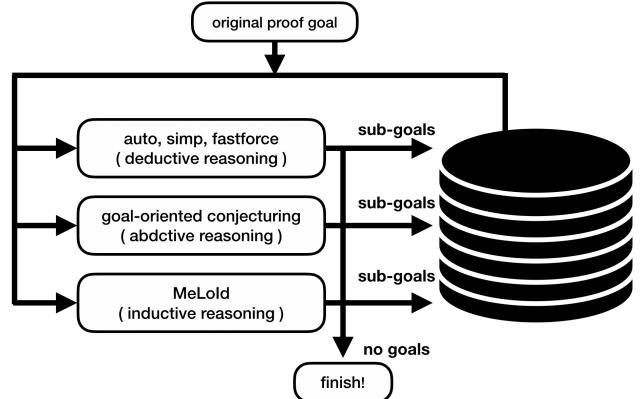


Figure 1: United Reasoning

PSL's depth-first search into a best-first search, so that we can automatically prove difficult inductive problems that involve nested inductions and many conjecturing steps while keeping the search space at a manageable size.

Figure 1 illustrates the overall architecture of united reasoning: when the system receives a proof goal, the system passes the goal to three reasoning mechanisms: deductive reasoning by Isabelle's standard proof automation tools, abductive reasoning by PGT's goal-oriented conjecturing, and inductive reasoning with MeLoId. If deductive reasoning discharges the proof goal, the system stops working, printing the proof method, with which it discharges the proof goal. If there are remaining sub-goals, united reasoning stores such sub-goals in a priority queue. Then, it keeps applying three reasoning mechanisms to the most promising set of sub-goals, which is stored at the top of the priority queue, until the system discharges the original proof goal completely or the queue becomes empty.

Since the goal-oriented conjecturing removes most irrelevant conjectures and MeLoId is expected to give us a few promising applications of the induct method, we hope that the system can have a few number of the most promising sets of sub-goals near the top of the priority queue.

So far, it is unclear how we should prioritize the three reasoning mechanisms: since existing proof corpora present only one proof for each theorem, a naive application of supervised learning does not seem to be a promising approach to learning when to apply which school of reasoning. We expect that the approach based on evolutionary computation to theorem proving [5] may help us to attack this problem; however, this still remains as our future work.

The most famous approach to automating proof by induction is called the waterfall model. Compared to the original waterfall model, which “*uses no search*” and “*is designed to make the right guess the first time, and then pursue one goal with power and perseverance*” [2], we are designing unified reasoning with search in mind. Unlike Moore, the creator of the waterfall model, we plan to make united reasoning a search-based software and place less importance on the speed of proof search because we trace the proof search using the writer monad transformer in Isabelle/ML [8] and produce efficient proof scripts upon successful (potentially slow) proof search as we did so with PSL.

ACKNOWLEDGMENTS

This work was supported by the European Regional Development Fund under the project AI & Reasoning (reg. no.CZ.02.1.01/0.0/0.0/15_003/0000466).

REFERENCES

- [1] Gerwin Klein, Tobias Nipkow, Larry Paulson, and Rene Thiemann. 2004. *The Archive of Formal Proofs*. <https://www.isa-afp.org/>
- [2] J. Strother Moore. 1973. *Computational logic : structure sharing and proof of program properties*. Ph.D. Dissertation. University of Edinburgh, UK. <http://hdl.handle.net/1842/2245>
- [3] Yutaka Nagashima. 2018. Towards Machine Learning Mathematical Induction. *CoRR* abs/1812.04088 (2018). arXiv:1812.04088 <http://arxiv.org/abs/1812.04088>
- [4] Yutaka Nagashima. 2019. LiFET: Language to Encode Induction Heuristics for Isabelle/HOL. (2019). arXiv:1906.08084
- [5] Yutaka Nagashima. 2019. Towards evolutionary theorem proving for isabelle/HOL. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion, GECCO 2019, Prague, Czech Republic, July 13-17, 2019*. 419–420.
- [6] Yutaka Nagashima and Yilun He. 2018. PaMpeR: proof method recommendation system for Isabelle/HOL. In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, ASE 2018, Montpellier, France, September 3-7, 2018*. 362–372. <https://doi.org/10.1145/3238147.3238210>
- [7] Yutaka Nagashima and Ramana Kumar. 2017. A Proof Strategy Language and Proof Script Generation for Isabelle/HOL. In *Automated Deduction - CADE 26 - 26th International Conference on Automated Deduction, Gothenburg, Sweden, August 6-11, 2017, Proceedings (Lecture Notes in Computer Science)*, Leonardo de Moura (Ed.), Vol. 10395. Springer, 528–545. https://doi.org/10.1007/978-3-319-63046-5_32
- [8] Yutaka Nagashima and Liam O'Connor. 2016. Close Encounters of the Higher Kind Emulating Constructor Classes in Standard ML. *CoRR* abs/1608.03350 (2016). arXiv:1608.03350 <http://arxiv.org/abs/1608.03350>
- [9] Yutaka Nagashima and Julian Parsert. 2018. Goal-Oriented Conjecturing for Isabelle/HOL. In *Intelligent Computer Mathematics - 11th International Conference, CICM 2018, Hagenberg, Austria, August 13-17, 2018, Proceedings*. 225–231. https://doi.org/10.1007/978-3-319-96812-4_19
- [10] Tobias Nipkow, Lawrence C. Paulson, and Markus Wenzel. 2002. *Isabelle/HOL - a proof assistant for higher-order logic*. Lecture Notes in Computer Science, Vol. 2283. Springer.