Title

Elliot Bobrow

November 11, 2024

1   Introduction

Theorem proving has become a relatively mature field, and there are many systems that can judge the correctness of a proof—see interactive theorem provers such as Coq, Lean, and Agda; or automated theorem provers such as Z3 and Vampire. However, there is less research into judging a proof readable; that is, are its steps of an appropriate size? For example, the following is a "correct" proof that $((A \setminus B) \cup C) \setminus D \subseteq (A \cup C) \setminus D$ for any sets A, B, C, and D:

Proof. Let $x \in ((A \setminus B) \cup C) \setminus D$. Then $x \in (A \cup C) \setminus D$. □

However, all human readers would likely agree that this proof is insufficient because its step is too large. A reader may not be able to follow from the first step to the second, and furthermore, if this proof were submitted by a student as part of an assignment, the grader would not be able to tell if the student had genuine mastery of the material. On the flip side, in a complicated proof in abstract algebra, it would be unnecessary and likely even distracting to spell out every detail about manipulating the underlying sets of each algebraic object. There must be some way to determine whether a given proof step is an appropriate size.

A system that could judge the legibility of a proof would be particularly applicable in a teaching setting, where students are just learning how to write proofs and could benefit from immediate feedback on their steps.

This paper contributes... ?

2   Literature Review

Much of the existing work on the subject of proof granularity comes from Marvin Schiller [6, 7]. In his experiments, he generated proof steps using the Ωmega system [2], and then asked four experts to rate the proof steps as either too small, too big, or appropriate. Using that data, he generated a decision tree which classifies a step as appropriate if it contains 1 or 2 assertion level rules, and too big otherwise.

This brings our attention to "assertion level" steps. The concept of the assertion level was initially introduced by Huang [5]. He proposes that there are three possible levels of justifications in proofs, ranging from least to most abstract. The "logic level" is the level of natural deduction rules, which is the level that computers reason at. The "assertion level" encapsulates the application of axioms, definitions, or theorems (collectively called assertions). Finally, the "proof level" involves meta-arguments within proofs. He finds that the assertion level is the most common, and proposes an algorithm for reconstructing natural deduction proofs at the assertion level.

Following Huang, others have proposed new logical systems to skip the middleman entirely, baking assertions directly into the logic. The *supernatural deduction* extends natural deduction with new inference rules generated from assertions of the form $\forall \overline{x}(P \Leftrightarrow \varphi)$ where P is atomic [3]. This was later extended to the sequent calculus with the *superdeduction calculus* [4]. Finally, Autexier and Dietrich presented *atomic metadeduction,* an extension of *superdeduction* that allows for the generation of rules from arbitrary assertions [1].

3 Possible topics

3.1 Applying Atomic Metadeduction to relevance systems

3.2 My system of axioms

3.3 Could write a whole section on everything wrong with Schiller

References

[1] Serge Autexier and Dominik Dietrich. "Atomic Metadeduction". In: *KI 2009: Advances in Artificial Intelligence*. Ed. by Bärbel Mertsching, Marcus Hund, and Zaheer Aziz. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 444–451. ISBN: 978-3-642-04617-9.

[2] Serge Autexier et al. "ΩMEGA: Resource-Adaptive Processes in an Automated Reasoning System". In: *Resource-Adaptive Cognitive Processes*. Ed. by Matthew W. Crocker and Jörg Siekmann. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 389–423. ISBN: 978-3-540-89408-7. DOI: `10.1007/978-3-540-89408-7_17`. URL: `https://doi.org/10.1007/978-3-540-89408-7_17`.

[3] Paul Brauner, Gilles Dowek, and Benjamin Wack. "Normalization in Supernatural deduction and in Deduction modulo". working paper or preprint. 2007. URL: `https://inria.hal.science/inria-00141720`.

[4] Paul Brauner, Clement Houtmann, and Claude Kirchner. "Principles of Superdeduction". In: *Twenty-Second Annual IEEE Symposium on Logic in Computer Science - LiCS 2007*. Wroclaw, Poland: IEEE Computer Society, July 2007. DOI: `10.1109/LICS.2007.37`. URL: `https://inria.hal.science/inria-00133557`.

[5]   Xiaorong Huang. "Reconstructing proofs at the assertion level". In: *Automated Deduction — CADE-12*. Ed. by Alan Bundy. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 738–752. ISBN: 978-3-540-48467-7.

[6]   Marvin R. G. Schiller. "Granularity Analysis for Tutoring Mathematical Proofs". In: *DISKI*. 2011. DOI: `https://doi.org/10.1111/tops.12012`.

[7]   Marvin R. G. Schiller. "Mechanizing Proof Step Evaluation for Mathematics Tutoring - The Case of Granularity". Universität des Saarlandes, Nov. 2005. URL: `https://page.mi.fu-berlin.de/cbenzmueller/papers/2005-schiller-thesis.pdf`.