Adaptive Assertion-Level Proofs (Extended Abstract)*

Christoph Benzmüller
Articulate Software, USA
c.benzmueller@googlemail.com
and Marvin Schiller
Saarland University, Germany
Marvin.Schiller@dfki.de

For nearly two decades human-oriented theorem proving techniques have been in the focus of interest of the ΩMEGA project [1] at Saarland University. And since one decade their application to tutorial natural language dialog on mathematical proofs has been studied in the interdisciplinary project DIALOG [3]. The focus on user-orientedness and mathematical practice in these projects resulted in developments such as assertion level proofs, achievements in the field of tactic proof search and proof planning, and investigations into qualitative aspects of the generated proofs.

The notion of the assertion level proofs was originally devised by Huang (cf. [1]). It characterizes proofs where each inference step is justified by a mathematical fact, such as a definition, theorem or a lemma. Initially, assertion level proofs were generated in a post-processing step from natural deduction proofs. The new ΩMEGA-CoRe system [1] supports reasoning directly at the assertion level.

The ΩMEGA project (among others) has pioneered hierarchical proofs via tactics and proof planning as a human-oriented way of organizing proof search. The notion of island proofs was introduced for proof sketches deliberately omitting proof parts. Users are enabled to formulate proofs by indicating only a subset of relevant intermediary formulas, and automated proof search is used to fill the logical gaps in the argumentation. Traditionally, proof scripts and tactics specify a sequence of actions to be carried out in order to obtain a proof. In contrast to that, the advantage of island proofs (or declarative proof plans/sketches) is that they as such constitute a meaningful and human-readable proof representation, by making the intermediate stages of the proof explicit. This prompted the development of declarative proof scripts (cf. [2]) to enable tactic reasoning at a user-friendly level of representation.

Hierarchical proofs, such as generated by tactics, enable proofs at different levels of abstraction. However, in a user-oriented setting such as e-learning for mathematical proofs, we need to ensure that the step size (granularity) of proofs is appropriate in the given context.

To address this issue we have proposed and investigated metrics for appropriate proof step size, which we align with empirical standards as determined by experiments [5, 6]. Initially, we devise a catalog of criteria of (single- or multi-inference) proof steps that we deem potentially relevant for judging their granularity in context. This process of identifying relevant criteria is supported by a corpus of students' proof steps, collected in the DIALOG project, which are annotated with granularity judgments by human expert judges. The question thus is, what distinguishes steps that were judged as appropriate from those considered inappropriate? We investigate a list of criteria:

Structural properties of proof steps. Does the step introduce new hypotheses or new subgoals? Is a step similar to a sequence of previous steps? Are the manipulations restricted to the same formula part? Can manipulations be applied in parallel? Are the inference steps carried out in forward or backward direction?

*We present joint work with Serge Autexier and Dominik Dietrich.

Aggregation and diversity. How many assertion level inference steps are combined? How many different concepts are employed? What (mathematical) theories do they belong to? Do the employed facts have the status of a definition, theorem or a lemma?

User knowledge. How many of the employed concepts have previously been mastered (or not mastered) by the user?

Explicitness. Does the user name the employed concepts?

We consider the question of appropriate granularity of a proof step as a classification problem. Classifiers – as models for granularity – assign a granularity verdict to any (single- or multi-inference) proof step in view of the granularity criteria. The analysis of the granularity criteria for each proof step is the task of a specialized component of the ΩMEGA system.

We investigate whether such models can be learned from samples of granularity judgments by human experts. For this purpose, we have developed a dedicated experiment environment to present proofs at various step sizes to human judges and to collect their granularity judgments. Based on these samples, we use machine learning techniques to generate models for the experts' judgments in an automated fashion. We compare the performance of decision tree/rule set learning, linear regression and support vector machine learning for this task, and inspect the generated models. The analysis also allows us to determine which of the examined granularity criteria are most useful for modeling the experts' decisions. The learned models enable us to adapt proofs to a user-friendly level of granularity for the purpose of tutorial proof presentation, and to diagnose whether a proof attempt by a student in a tutorial dialog progresses at an adequate pace.

Our experiments provide evidence that granularity judgments allow some room for subjectivity. However, proof representations at the assertion level are confirmed as a useful basis for diagnosing and adapting proof granularity.

References

[1] Serge Autexier, Christoph Benzmüller, Dominik Dietrich, and Jörg Siekmann. ΩMEGA: Resource-adaptive processes in an automated reasoning system. In Crocker and Siekmann [4], pages 389–416.

[2] Serge Autexier and Dominik Dietrich. A tactic language for declarative proofs. In Matt Kaufmann and Lawrence C. Paulson, editors, *Proceedings of the International Conference on Interactive Theorem Proving (ITP 2010), Edinburgh, Scotland*, volume 6172 of *LNCS*, pages 99–114. Springer, 2010.

[3] Christoph Benzmüller, Marvin Schiller, and Jörg Siekmann. Resource-bounded modelling and analysis of human-level interactive proofs. In Crocker and Siekmann [4], pages 263–283.

[4] Matthew W. Crocker and Jörg Siekmann, editors. *Resource-Adaptive Cognitive Processes*. Cognitive Technologies. Springer, 2010.

[5] Marvin Schiller and Christoph Benzmüller. Presenting proofs with adapted granularity. In Bärbel Mertsching, Marcus Hund, and Muhammad Zaheer Aziz, editors, *KI 2009: Advances in Artificial Intelligence, 32nd Annual German Conference on AI, Paderborn, Germany. Proceedings*, volume 5803 of *LNAI*, pages 289–297. Springer, 2009.

[6] Marvin Schiller, Dominik Dietrich, and Christoph Benzmüller. Proof step analysis for proof tutoring – a learning approach to granularity. *Teaching Mathematics and Computer Science*, 6(2):325–343, 2008.