Human-Oriented Proof Techniques are Relevant for Proof Tutoring
(Extended Abstract)

Marvin Schiller and Christoph Benzmüller
Saarland University, Germany and Articulate Software, USA

Automated theorem proving techniques are increasingly used within interactive environments for the teaching of proof techniques. Examples are the Advanced Geometry Tutor (Matsuda and VanLehn, 2005), the geometry learning environment ANGLE (Koedinger and Anderson, 1993), the EPGY theorem proving environment (McMath *et al.*, 2001), the Baghera platform (Trgalova and Chaachoua, 2009) for the assessment of proofs in geometry, the Carnegie Proof Lab (Sieg, 2007), the educational variant of the TPS system ETPS (Andrews *et al.*, 2004), the ProofWeb (Kaliszyk *et al.*, 2008) system for practicing natural deduction proofs, the P-Logic tutor (Lukins *et al.*, 2002), the EASy (Gruttmann *et al.*, 2008) system for the e-assessment for mathematical proofs, the tutorial proof checker Tutch (Abel *et al.*, 2001) and the RISC proof navigator (Schreiner, 2006) for education in proof verification.

Most of these systems use automated proof mechanisms based on natural deduction or resolution calculi. Often, the users are required to stick to the employed calculi step by step. Evaluations of several among the above mentioned systems have revealed this as a severe limitation for teaching common mathematical practice. Gruttmann *et al.* (2008, p. 613) report that "EASy requires small steps to be executed one by one, which normally could be combined in a paper-based proof. Some students mentioned that a computer-supported proof could lead to a reduction of mathematical skills." Lukins *et al.* (2002) consider the use of proof search to enable more flexibility as relevant further work. Abel *et al.* (2001) allow for aggregations of steps, but the reported evaluations of Tutch have been carried out with an earlier single-step version of the system. McMath *et al.* (2001) use the resolution-based theorem prover Otter for proof search, but restrict its use to a few seconds to disallow "large leaps of logic". However, the authors report that the verification may fail for various reasons, including incorrectness, big step size, but also ineffectiveness of the employed proof strategies. According to McMath *et al.* (2001), examining Otter's output is insufficient to determine the cause for rejection, and does not help to provide adequate feedback to the learner.

The mismatch between the machine-oriented proofs employed e.g. by Otter and common mathematical practice limits their use for modeling and diagnosing students' proofs. Therefore, we argue for the use of human-oriented, mathematically intelligent proof search techniques to support the computer-assisted teaching of mathematical proofs: A good match between common practice in mathematics and the employed automated reasoning techniques enables more accurate modeling of the user's proof attempt. An accurate model and diagnosis of the user's input is the basis for effective feedback from the teaching environment to the student. Our position is supported by results obtained in the DIALOG project at Saarland University (Benzmüller *et al.*, 2010).

Within the DIALOG project, automated reasoning techniques were investigated for the use within an interactive teaching environment for proofs. The design of the system is guided by the investigation of human one-on-one tutoring, which is generally considered an effective form of teaching. Using simulation studies, we collected corpora of tutorial dialogs on proofs to study the essential requirements for a dialog based proof tutoring system.

In order to model and analyze the proof steps entered by a learner, we require a dynamic domain reasoner supporting human-oriented proof techniques. We consider the mathematical assistant system ΩMEGA (Autexier *et al.*, 2010) as the candidate of choice for this role.

The proof attempts entered by the student are reconstructed stepwise using human-oriented assertion level proof search provided by ΩMEGA. The assertion level proof mechanism enables proofs where each step is justified by a mathematical fact, such as a definition, theorem or lemma. The formal proof reconstructions by ΩMEGA of the student's proof steps are then used as the basis for the stepwise analysis of the proof attempt.

Simulation studies revealed that human expert tutors analyze a learner's proof steps not only with respect to their correctness, they also assess whether the learner progresses at an acceptable pace (an aspect we call granularity) and whether they are relevant (Benzmüller and Vo, 2005). For an adequate assessment of these aspects, it is necessary to model the student's input as closely as possible. We have analyzed the proof steps entered by the learners in the simulation studies, and determined that these proofs are more suitably represented at the assertion level rather than via natural deduction, not to speak of resolution (Benzmüller *et al.*, 2007).

Furthermore, we have conducted a study on the modeling of appropriate granularity for proof tutoring. Our goal is to diagnose the granularity of steps input by the student, but also to meter out the granularity of proofs steps generated by the proof assistant when presenting them to the user. In recent experiments, four expert tutors provided granularity judgments for variations of proofs in different mathematical domains. The difficulty of the proofs roughly corresponds to the beginner level at university. The employed proof steps were modeled formally in ΩMEGA, where each step corresponded to one inference application at the assertion level or to the aggregation of several inference steps. We also added a few single natural deduction steps for comparison.

Our analysis of the judgments by the tutors showed that combinations of one or two assertion level inference applications were generally considered of appropriate granularity in the presented context of proof tutoring. Of those steps generated from one single assertion level inference, a proportion of 92%, 70%, 83% and 96% were considered appropriate by the four different judges, respectively. The natural deduction steps we mixed in for comparison were considered to small by three of the four experts. In (Schiller and Benzmüller, 2009) we present a number of other aspects that are relevant for judging whether a step is of appropriate granularity, which are related to the employed concepts, the structure of proofs, and explicitness. However, the large percentage of steps considered of appropriate size by our tutors provides evidence that the assertion level is rather close to the target granularity for proof tutoring. Furthermore, proofs at the assertion level provide a suitable basis for determining other aspects of proof steps relevant for the diagnosis, since information on the employed facts and the structure of the proof are represented more explicitly here rather than, for example, in the form of a resolution proof.

In summary, our studies provide evidence that human-oriented proof search techniques and human-oriented proof representation techniques are highly relevant for successful proof tutoring.

References

Abel, A., Chang, B.-Y. E., and Pfenning, F. (2001). Human-readable machine-verifiable proofs for teaching constructive logic. In *IJCAR Workshop on Proof Transformations, Proof Presentations and Complexity of Proofs (PTP-01)*, pages 33–48, Università degli Studi di Siena, Italy.

Andrews, P. B., Brown, C. E., Pfenning, F., Bishop, M., Issar, S., and Xi, H. (2004). ETPS: A system to help students write formal proofs. *J. Autom. Reasoning*, **32**(1), 75–92.

Autexier, S., Benzmüller, C., Dietrich, D., and Siekmann, J. (2010). ΩMEGA: Resource-adaptive processes in an automated reasoning system. In M. W. Crocker and J. Siekmann, editors, *Resource-Adaptive Cognitive Processes*, Cognitive Technologies, pages 389–416. Springer.

Benzmüller, C., Dietrich, D., Schiller, M., and Autexier, S. (2007). Deep inference for automated proof tutoring. In J. Hertzberg, M. Beetz, and R. Englert, editors, *KI 2007: Advances in Artificial Intelligence. 30th Annual German Conference on AI*, number 4667 in LNAI, pages 435–439. Springer.

Benzmüller, C., Schiller, M., and Siekmann, J. (2010). Resource-bounded modelling and analysis of human-level interactive proofs. In M. W. Crocker and J. Siekmann, editors, *Resource-Adaptive Cognitive Processes*, Cognitive Technologies, pages 263–283. Springer.

Benzmüller, C. E. and Vo, B. Q. (2005). Mathematical domain reasoning tasks in natural language tutorial dialog on proofs. In M. Veloso and S. Kambhampati, editors, *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*, pages 516–522, Pittsburgh, Pennsylvania, USA. AAAI Press / The MIT Press.

Gruttmann, S., Böhm, D., and Kuchen, H. (2008). E-assessment of mathematical proofs: Chances and challenges for students and tutors. In *Proceedings of the 2008 International Conference on Computer Science and Software Engineering, Wuhan, China, December 12th-14th*, volume 5, pages 612–615. IEEE Computer Society.

Kaliszyk, C., Raamsdonk, F. v., Wiedijk, F., Wupper, H., Hendriks, M., and Vrijer, R. d. (2008). Deduction using the ProofWeb system. Technical Report ICIS–R08016, Radboud University Nijmegen. `www4.in.tum.de/~kaliszyk/docs/cek_r.pdf`.

Koedinger, K. and Anderson, J. R. (1993). *Computers as cognitive tools*, chapter Reifying implicit planning in geometry: Guidelines for model-based intelligent tutoring system design, pages 15–46. Erlbaum, Hillsdale, NJ.

Lukins, S., Levicki, A., and Burg, J. (2002). A tutorial program for propositional logic with human/computer interactive learning. In J. L. Gersting, H. M. Walker, and S. Grissom, editors, *SIGCSE*, pages 381–385. ACM.

Matsuda, N. and VanLehn, K. (2005). Advanced geometry tutor: An intelligent tutor that teaches proof-writing with construction. In C.-K. Looi, G. I. McCalla, B. Bredeweg, and J. Breuker, editors, *AIED*, volume 125 of *Frontiers in Artificial Intelligence and Applications*, pages 443–450. IOS Press.

McMath, D., Rozenfeld, M., and Sommer, R. (2001). A computer environment for writing ordinary mathematical proofs. In R. Nieuwenhuis and A. Voronkov, editors, *LPAR*, volume 2250 of *LNCS*, pages 507–516. Springer.

Schiller, M. and Benzmüller, C. (2009). Presenting proofs with adapted granularity. In B. Mertsching, M. Hund, and Z. Aziz, editors, *KI 2009: Advances in Artificial Intelligence*, volume 5803 of *LNCS*, pages 289–297. Springer.

Schreiner, W. (2006). Program verification with the risc proofnavigator. In *Teaching Formal Methods: Practice and Experience, BCS-FACS Christmas Meeting, London, UK, December 15, 2006*. Electronic Workshops in Computing (eWiC), British Computer Society. `http://www.bcs.org/server.php?show=conMediaFile.4033`.

Sieg, W. (2007). The apros project: Strategic thinking & computational logic. *Logic Journal of IGPL*, pages 359–368.

Trgalova, J. and Chaachoua, H. (2009). Automatic analysis of proof in a computer-based environment. In F.-L. Lin, F.-J. Hsieh, G. Hanna, and M. de Villiers, editors, *Proof and Proving in Mathematics Education ICMI'19 : 19th International Conference on Mathematics Instruction, Taipei, Taiwan*, pages 226–231.