Deep Inference for Automated Proof Tutoring?*

Christoph Benzmüller[1,2], Dominik Dietrich[1], Marvin Schiller[1], and Serge Autexier[1,3]

[1] Dept. of Computer Science, Saarland University, 66041 Saarbrücken, Germany
[2] Computer Laboratory, The University of Cambridge, Cambridge, CB3 0FD, UK
[3] German Research Centre for Artifificial Intelligence (DFKI), Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany

1 Introduction

ΩMEGA [7], a mathematical assistant environment comprising an interactive proof assistant, a proof planner, a structured knowledge base, a graphical user interface, access to external reasoners, etc., is being developed since the early 90's at Saarland University. Similar to HOL4, Isabelle/HOL, Coq, or Mizar, the overall goal of the project is to develop a system platform for formal methods (not only) in mathematics and computer science. In ΩMEGA, user and system interact in order to produce verifiable and trusted proofs. By continously improving (not only) automation and interaction support in the system we want to ease the usually very tedious formalization and proving task for the user.

A very recent application direction of ΩMEGA, studied in the DIALOG project [3], is e-learning in mathematics. The hypothesis is that our system can fruitfully support the tutoring of mathematical proofs.

In 2001 our group opted for a major reimplementation of the ΩMEGA system. A major motivation was to replace the natural deduction (ND) calculus at the logical core of the system by another, ideally more suitable, logical base formalism. What exactly did we have in mind?

1. An earlier study (cf. [5]) of the influence of the ΩMEGA's ND core to its abstract level reasoning mechanism, such as proof planning, revealed a negative influence mainly due to unnatural, structural peculiarities of ND. We conjectured that a less 'low-level' logical core should remedy this problem.
2. Abstract level proofs in ΩMEGA may be unsound. They have to be expanded at logical core level before they can be verified and finally trusted. Expansions to ND level, however, often dramatically increase the proof size and they typically lead to very unnatural and human incomprehensible proofs. Supporting the tutoring of proofs based on these unnatural, low-level proofs thus seemed a suboptimal starting point for our DIALOG project.

For our reimplementation we therefore picked Serge Autexier's CoRe calculus [1] as a promising alternative to ND. CoRe has then been transformed into an 'assertion level' inference mechanism [11] which constitutes the logical kernel of the new ΩMEGA$^{\text{CoRe}}$ system. CoRe and our assertion level inference mechanism are (higher-order) variants of the deep inference approach[1], that is, they support deductions deeply inside a given

* Funded by the DFG SFB378 (Project OMEGA) and by EPSRC under grant EP/D070511/1.

[1] `http://alessio.guglielmi.name/res/cos/index.html`

J. Hertzberg, M. Beetz, and R. Englert (Eds.): KI 2007, LNAI 4667, pp. 435–439, 2007.

s1: Let $(x,y) \in (R \circ S)^{-1}$.	*s6 :* Now let $(x,y) \in S^{-1} \circ R^{-1}$.
t1: Good. correct	*t6:* O.k. correct
s2: It follows that $(y,x) \in (R \circ S)$.	*s7 :* $\Rightarrow \exists z \in M : (x,z) \in S^{-1} \wedge (z,y) \in R^{-1}$.
t2: That's correct. correct	*t7:* Yes. correct
s3: Therefore $\exists z \in M : (y,z) \in R \wedge (z,x) \in S$.	*s8 :* $\Rightarrow \exists z \in M : (z,x) \in S \wedge (y,z) \in R$.
t3: Yes, that's right. correct	*t8:* This deduction is also correct. correct
s4: This implies: $\exists z \in M : (z,y) \in R^{-1} . \wedge (x,z) \in S^{-1}$	*s9:* $\Rightarrow (y,x) \in R \circ S$.
t4: Your deduction is correct. You are on a good way. correct	*t9:* This deduction is again correct. correct
s5: It follows: $(x,y) \in S^{-1} \circ R^{-1}$.	*s10:* $\Rightarrow (x,y) \in (S \circ R)^{-1}$.
t5: Yes. Therefore it is shown that every element from $(R \circ S)^{-1}$ already lies in $S^{-1} \circ R^{-1}$. correct	*t10:* Congratulations! With this you have shown both inclusions. Your solution is now complete. correct

Fig. 1. Example dialog; $s\_$ are student turns and $t\_$ are tutor turns

formula without requiring preceding structural decompositions as needed in ND (or sequent calculus). In ΩMEGA$^{\text{CORE}}$ we thus have a smaller 'distance' between abstract level proofs and their expansions to the verifiable assertion level. Most importantly, we now support reasoning directly at the assertion level, while such a layer did only exist in the old ΩMEGA for *a posteriori* proof presentation purposes.

In this short paper we report on our ongoing application and evaluation of the ΩMEGA$^{\text{CORE}}$-system for proof tutoring in the DIALOG project. For this, we apply our novel proof assessment module [6] developed in the DIALOG project to 17 proof dialogs which we have obtained in a previous experiment [4]. We study the 'quality' of the automatically reconstructed proofs and analyse the coverage of our proof assessment module.

2 Evaluation

We have applied our proof assessment module to 17 tutorial dialogs taken from the Wizard-of-Oz experiment reported in [4]. These dialogs consist in alternating utterances by a student and a tutor. The student attempts to solve an exercise from the domain of binary relations, namely to show that $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$ holds for two relations R and S in a set M, where $\circ$ denotes relation composition and $^{-1}$ denotes inversion. The tutor responds to the subsequent proof step utterances of the student. Each step is annotated with a judgment regarding its correctness by the tutor. One example dialog from our evaluation set is shown in Figure 1.

The idea of the proof assessment module is to support automated proof tutoring, that is, to automatically judge about the correctness (and also the granularity and the relevance – not discussed in this paper) of each single student proof step (cf. [2]). For this, the assessment module is initialized with the relevant axioms for this domain, which are automatically transformed into inference rules at the assertion level (cf. [8]) by ΩMEGA$^{\text{CORE}}$ and made available for proving.

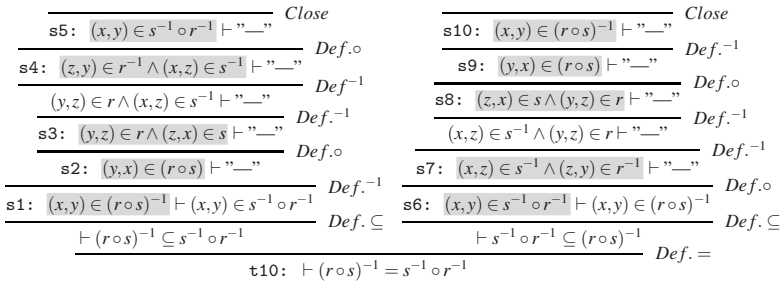

Fig. 2. Annotated ΩMEGA$^{\text{CoRe}}$ assertion level proof for the example dialog

Then, for each of the 17 dialogs, the assessment was performed stepwise by our assessment module. The assessment module maintains an assertion level proof object that represents the current state of the proof under construction, which can include several proof alternatives in the case of underspecified, that is, insufficiently precise, proof step utterances by the student causing ambiguities (cf. [2,6]). For each proof step uttered by the student, the module uses a *depth-limited breadth-first search* (with pruning of superfluous branches) to expand the given proof state to all possible successor states up to that depth. From these, those successor states that match the given utterance wrt. to some filter function (analyzing whether a successor state is a possible reading of the student proof step) are selected. We thus obtain, modulo our filter function, assertion level counterparts to all possible interpretations of correct student proof steps. If for a given utterance, no matching successor state can be reached, the utterance is considered as incorrect.

We compared the results of the automated proof step analysis with the original correctness judgments by the tutors. All steps in the example dialog are correctly classified as valid by our assessment module (used with proof depth four), taking approximately 13.2 seconds on a standard PC.

Figure 2 shows one complete assertion level proof (in sequent notation and annotated by the corresponding student proof steps) that was constructed by the assessment module for the dialog in Figure 1. The number of assertion level steps required (13, excluding the automatic *Close* steps) is still comparable to the number of proof steps as uttered by the student in the original dialog (10), which provides evidence that the ΩMEGA$^{\text{CoRe}}$ assertion level proof is at a suitable level of granularity. Had we used natural deduction as in the old ΩMEGA system, we would have obtained many intermediate steps of rather technical nature making breadth-first proof search for our task infeasible, compare:

$$\frac{(z,y) \in r^{-1} \wedge (x,z) \in s^{-1}}{(y,z) \in r \wedge (x,z) \in s^{-1}}\ Def^{-1}$$

Core

$$\dfrac{\dfrac{A := (z,y) \in r^{-1} \wedge (x,z) \in s^{-1}}{(y,z) \in r^{-1}}\ \wedge E}{(y,z) \in r}\ Def.^{-1} \qquad \dfrac{A}{(x,z) \in s^{-1}}\ \wedge E \Bigg}{(y,z) \in r \wedge (x,z) \in s^{-1}}\ \wedge I$$

Natural Deduction

The 17 dialogs in the evaluation contain a total of 147 proof steps. All the steps within a dialog are passed to the assessment module sequentially until a step that is labeled as correct cannot be verified, in which case we move on to the next dialog. This way, we correctly classify 141 out of the 147 steps (95.9%) as correct or wrong. Among the remaining six steps are three where the verification fails, and further three remain untouched.

3 Concluding Remarks

Our initial question whether moving from ΩMEGA's previous ND based logical core to assertion level reasoning in ΩMEGA$^{\text{CORE}}$ was a reasonable decision, can (preliminarily) be answered with 'Yes':

In ΩMEGA$^{\text{CORE}}$ we obtain more adequate formal counterparts of the human proofs as was possible before. Most importantly, we directly search for these proofs at the assertion level which enables us to employ a simple depth-limited breadth-first search algorithm in our proof step assessment module. Interestingly, already a depth limit of just four assertion level steps enables our approach to correctly classify 95.9% of the proof steps in our corpus.

Related to our work are the EPGY Theorem Proving Environment [9], using Otter to justify or reject proof steps proposed to the environment, and the computational framework by Claus Zinn [10] for the analysis of textbook proofs. Our approach differs in the following ways: (i) We address the problem of underspecification and multiple interpretations of student proof step utterances, (ii) we construct one, or several, global, coherent proof object(s) for each dialog instead of just looking from step to step, (iii) we are not just interested in the correctness of proof steps but also in their granularity and relevance; for this adequate formal proofs are even more important.

References

1. Autexier, S.: The core calculus. In: Nieuwenhuis, R. (ed.) Automated Deduction – CADE-20. LNCS (LNAI), vol. 3632, Springer, Heidelberg (2005)
2. Benzmüller, C., Vo, Q.B.: Mathematical domain reasoning tasks in natural language tutorial dialog on proofs. In: Proc. AAAI-05, AAAI Press/The MIT Press (2005)
3. Benzmüller, C., et al.: Natural language dialog with a tutor system for mathematical proofs. In: Ullrich, C., Siekmann, J.H., Lu, R. (eds.) Cognitive Systems. LNCS (LNAI), vol. 4429, Springer, Heidelberg (2007)
4. Benzmüller, C., et al.: Diawoz-II - a tool for wizard-of-oz experiments in mathematics. In: Freksa, C., Kohlhase, M., Schill, K. (eds.) KI 2006. LNCS (LNAI), vol. 4314, Springer, Heidelberg (2007)
5. Benzmüller, C., et al.: Proof planning: A fresh start? In: Proc. of IJCAR 2001 Workshop: Future Directions in Automated Reasoning, Siena, Italy (2001)
6. Dietrich, D., Buckley, M.: Verification of Proof Steps for Tutoring Mathematical Proofs. In: Proc. AIED 2007 (to appear)
7. Siekmann, J., et al.: Computer supported mathematics with omega. J. Applied Logic 4(4), 533–559 (2006)

8. Autexier, S., Dietrich, D.: Synthesizing Proof Planning Methods and Oants Agents from Mathematical Knowledge. In: Borwein, J.M., Farmer, W.M. (eds.) MKM 2006. LNCS (LNAI), vol. 4108, Springer, Heidelberg (2006)
9. McMath, D., Rozenfeld, M., Sommer, R.: A computer environment for writing ordinary mathematical proofs. In: Nieuwenhuis, R., Voronkov, A. (eds.) LPAR 2001. LNCS (LNAI), vol. 2250, pp. 507–516. Springer, Heidelberg (2001)
10. Zinn, C.: A computational framework for understanding mathematical discourse. Logic J. of the IGPL 11, 457–484 (2003)
11. Dietrich, D.: The tasklayer of the omega system. Master's thesis, Universität des Saarlandes, Saarbrücken, Germany (2006)