Proof Granularity as an Empirical Problem?[1]

Marvin Schiller† and Christoph Benzmüller⋆

†Saarland University, Germany
⋆International University Bruchsal, Germany

Internation Conference on Computer Supported Education,
March 23rd-26th, 2009

Studienstiftung des deutschen Volkes   UNIVERSITÄT DES SAARLANDES   International University in Germany

[1] This work has been funded by Studienstiftung des deutschen Volkes

Overview

1. Background: Tutorial DIALOG on Mathematical Proof
2. Step Size - Observations & Modeling
3. Study Environment & Experiments
4. Outlook & Discussion

Background: The DIALOG Project

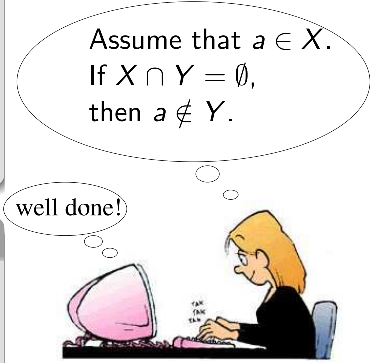
Tutorial Dialog for Mathematics.

Employed Techniques

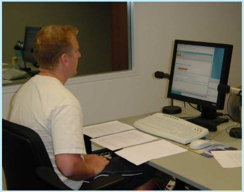
- Dyn. Domain Reasoning for math proofs: math assistant ΩMEGA
- NL processing, dialogue management, teaching model

Research Processes (in spiral model)

- DIALOG system design
- Prototype development
- Empirical studies



Simulation (Wizard-of-Oz)



Let R and S be relations in a set M. It holds that: $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$. Do the proof interactively with the system!

A pair (x, y) is element of $R \circ S$ iff there is a z in M such that $(x, z) \in R$ and $(z, y) \in S$

Correct!

Therefore a pair (x, y) is element of $(R \circ S)^{-1}$ if there is a z in M, such that $(x, z) \in S$ and $(z, y) \in R$

That's not correct!

Simulation (Wizard-of-Oz) [Ctd.]

Tutor: Let R and S be relations in a set M. It holds that: $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$. Do the proof interactively with the system!

Student: Execute $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$!

Tutor: That's what you have to prove. Would you like a hint?

Student: yes

Tutor: At first, try to show $(R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}$.

Student: How do I do that?

Tutor: You have to show that any element in $(R \circ S)^{-1}$ is also in $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$. You could start with: Let $(a, b) \in (R \circ S)^{-1}$. ...Then it holds ...?

Step Size in Mathematics (Granularity)

Mathematical practice: skip intermediate steps when appropriate:

Proof Exercise: $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$

- We show $(R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}$ and $S^{-1} \circ R^{-1} \subseteq (R \circ S)^{-1}$
- We assume $(y, x) \in (R \circ S)^{-1}$ and show $(y, x) \in S^{-1} \circ R^{-1}$
- Hence, $(x, y) \in R \circ S$
- Hence, $\exists z$ s.t. $(x, z) \in R \wedge (z, y) \in S$
- Hence, $\exists z$ s.t. $(z, x) \in R^{-1} \wedge (z, y) \in S$
- Hence, $\exists z$ s.t. $(z, x) \in R^{-1} \wedge (y, z) \in S^{-1}$
- Hence, $\exists z$ s.t. $(y, z) \in S^{-1} \wedge (z, x) \in R^{-1}$
- Hence, $(y, x) \in S^{-1} \circ R^{-1}$

$\vdots$

Step Size in Mathematics (Granularity)

Mathematical practice: skip intermediate steps when appropriate:

Proof Exercise: $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$

We show $(R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}$ and $S^{-1} \circ R^{-1} \subseteq (R \circ S)^{-1}$

- We assume $(y, x) \in (R \circ S)^{-1}$ and show $(y, x) \in S^{-1} \circ R^{-1}$
- Hence, $(x, y) \in R \circ S$
- Hence, $\exists z$ s.t. $(x, z) \in R \wedge (z, y) \in S$

Hence, $\exists z$ s.t. $(z, x) \in R^{-1} \wedge (z, y) \in S$

- Hence, $\exists z$ s.t. $(z, x) \in R^{-1} \wedge (y, z) \in S^{-1}$

Hence, $\exists z$ s.t. $(y, z) \in S^{-1} \wedge (z, x) \in R^{-1}$

- Hence, $(y, x) \in S^{-1} \circ R^{-1}$

$\vdots$

Step Size in the Experiments

Granularity: The question of the appropriate step size/complexity.

Exercise: z.Z. $(R \circ S)^{-1} = (x, y) \in S^{-1} \circ R^{-1}$

$\vdots$

student] $(x, y) \in (R \circ S)^{-1}$

tutor] Now try to draw conclusions from this!

correct	appropriate	relevant

student] $(x, y) \in S^{-1} \circ R^{-1}$

tutor] This cannot be concluded directly.
You need some intermediate steps!

correct	too coarse-grained	relevant

Step size annotated by tutors as *appropriate*, *too coarse-grained* (too big a step) or *too detailed* (too small a step)

Modeling (Suitable) Granularity

Goal: diagnose student's step size, granularity-adapted proof presentation.

Approach

Modeling/representation of proofs: choice of suitable proof calculus/mechanism (assertion level vs. ND or resolution)
Analysis: granularity-relevant criteria
Classification: classify (multi-inference) proof steps (as *appropriate*, *too big* or *too small*)

Learn classifier from empirical samples.

Approach: Model Student Proofs as Assertion Level Proofs

Student's Proof

Ex: Show
$(R \circ S)^{-1} = S^{-1} \circ R^{-1}$!

Assertion Level Proof

$$\text{Exercise:} \quad \vdash \underbrace{(R \circ S)^{-1}}_{\Gamma} = \underbrace{S^{-1} \circ R^{-1}}_{\Theta}$$

Approach: Model Student Proofs as Assertion Level Proofs

Student's Proof

Ex: Show $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$!

s1: $\boxed{\text{Let } (x, y) \in (R \circ S)^{-1}}$.

Assertion Level Proof

$$\textbf{Def.}\ = \ \dfrac{\dfrac{\overline{\texttt{s1}: (x, y) \in (R \circ S)^{-1} \vdash (x, y) \in \Theta}}{\vdash (R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}}\ \textbf{Def.} \subseteq \qquad \vdash \Theta \subseteq \Gamma}{\texttt{Exercise:} \quad \vdash \underbrace{(R \circ S)^{-1}}_{\Gamma} = \underbrace{S^{-1} \circ R^{-1}}_{\Theta}}$$

Approach: Model Student Proofs as Assertion Level Proofs

Student's Proof

Ex: Show $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$!

s1: Let $(x, y) \in (R \circ S)^{-1}$.

s2: Hence $(y, x) \in (R \circ S)$.

Assertion Level Proof

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\mathtt{s2}: (y,x) \in (R \circ S) \vdash (x,y) \in \Theta}}{\mathtt{s1}: (x,y) \in (R \circ S)^{-1} \vdash (x,y) \in \Theta}\,\text{Def.}^{-1}}{\vdash (R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}}\,\text{Def.} \subseteq \qquad \vdash \Theta \subseteq \Gamma}{\mathtt{Exercise:} \quad \vdash \underbrace{(R \circ S)^{-1}}_{\Gamma} = \underbrace{S^{-1} \circ R^{-1}}_{\Theta}}\,\text{Def.} =$$

Approach: Model Student Proofs as Assertion Level Proofs

Student's Proof

Ex: Show $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$!

s1: Let $(x, y) \in (R \circ S)^{-1}$.

s2: Hence $(y, x) \in (R \circ S)$.

s3: Hence $(y, z) \in R \wedge (z, x) \in S$.

Assertion Level Proof

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\texttt{s3:}\ (y,z) \in R \wedge (z,x) \in S \vdash (x,y) \in \Theta}}{\texttt{s2:}\ (y,x) \in (R \circ S) \vdash (x,y) \in \Theta}\ \text{Def.}\circ}{\texttt{s1:}\ (x,y) \in (R \circ S)^{-1} \vdash (x,y) \in \Theta}\ \text{Def.}^{-1}}{\vdash (R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1} \qquad \vdash \Theta \subseteq \Gamma}\ \text{Def.} \subseteq$$

$$\dfrac{\ldots}{\texttt{Exercise:}\quad \vdash \underbrace{(R \circ S)^{-1}}_{\Gamma} = \underbrace{S^{-1} \circ R^{-1}}_{\Theta}}\ \text{Def.} =$$

Approach: Model Student Proofs as Assertion Level Proofs

Student's Proof

Ex: Show $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$!

s1: Let $(x,y) \in (R \circ S)^{-1}$.

s2: Hence $(y,x) \in (R \circ S)$.

s3: Hence $(y,z) \in R \wedge (z,x) \in S$.

s4: Hence $(z,y) \in R^{-1} \wedge (x,z) \in S^{-1}$.

Assertion Level Proof

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\texttt{s4:}\ (z,y) \in R^{-1} \wedge (x,z) \in S^{-1} \vdash (x,y) \in \Theta}}{(y,z) \in R \wedge (x,z) \in S^{-1} \vdash (x,y) \in \Theta}\ \mathbf{Def}^{-1}}{\texttt{s3:}\ (y,z) \in R \wedge (z,x) \in S \vdash (x,y) \in \Theta}\ \mathbf{Def.}^{-1}}{\texttt{s2:}\ (y,x) \in (R \circ S) \vdash (x,y) \in \Theta}\ \mathbf{Def.} \circ}{\dfrac{\texttt{s1:}\ (x,y) \in (R \circ S)^{-1} \vdash (x,y) \in \Theta}{\vdash (R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}}\ \mathbf{Def.} \subseteq\ \ \vdash \Theta \subseteq \Gamma}\ \mathbf{Def.}^{-1}}{\texttt{Exercise:}\quad \vdash \underbrace{(R \circ S)^{-1}}_{\Gamma} = \underbrace{S^{-1} \circ R^{-1}}_{\Theta}}\ \mathbf{Def.} =$$

Typically: 1 student step $\cong$ 1 or several assertion level steps (experiment: usually 1-3, seldomly more)

Granularity Criteria

Possible criteria for size of a (multi-)inference step ("features")

- How many assertion level inference applications? (*total*)
- What concepts are used? (*concepts*)
- How many concepts are not yet known to the student? (*unmastered*)
- Are the concepts named? (*verb*)
- etc.

Student step	Infs	Features	Verdict
1. We assume $(y,x) \in (R \circ S)^{-1}$ and show $(y,x) \in S^{-1} \circ R^{-1}$	Def.$=$, Def.$\subseteq$	total:2, concepts:2, relations:0, verb:0,...	?
2. Hence, $(x,y) \in R \circ S$	Def^{-1}	total:1, concepts:1, relations:1, verb:0,...	?

...

Example Classifier

Sample ruleset classifier

* total $\in \{0,1,2\} \Rightarrow$ “appropriate”
* unmastered $\in \{2,3,4\} \wedge$ relations $\in \{2,3,4\} \Rightarrow$ “step-too-big”
* total $\in \{3,4\} \wedge$ relations $\in \{0,1\} \Rightarrow$ “step-too-big”
* unmastered $\in \{0,1\} \Rightarrow$ “appropriate”
* _ $\Rightarrow$ ”appropriate”

Student step	Infs	Features	Verdict
1. We assume $(y,x) \in (R \circ S)^{-1}$ and show $(y,x) \in S^{-1} \circ R^{-1}$	Def.$=$, Def.$\subseteq$	total:2, concepts:2, relations:0, verb:0,...	appropriate
2. Hence, $(x,y) \in R \circ S$	Def^{-1}	total:1, concepts:1, relations:1, verb:0,...	appropriate

...

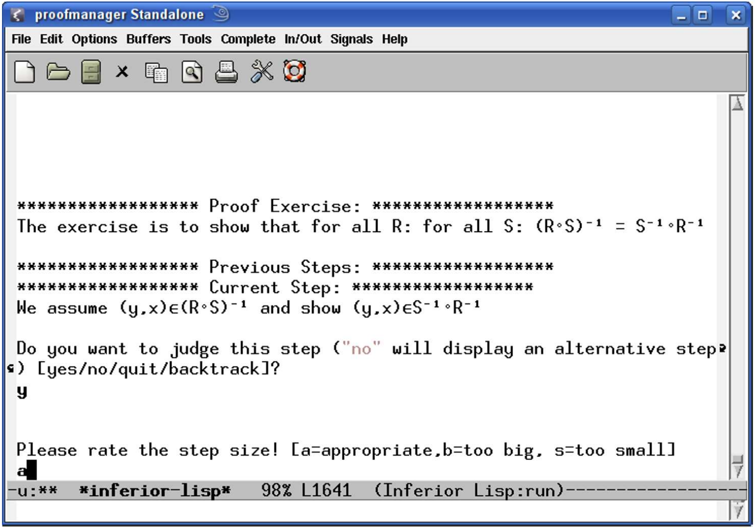
Study Environment - Motivation

- Learn classifiers via annotations from expert tutors using standard machine learning
- WoZ experiments not ideal for focused study on granularity

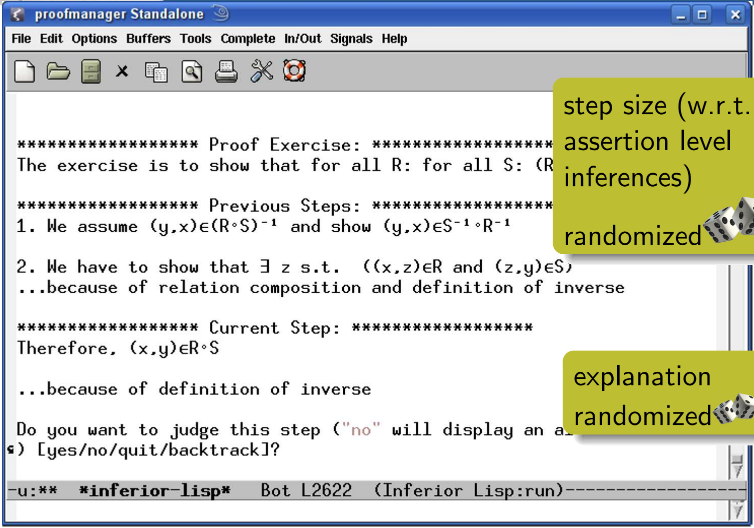
Idea

- Automate student's role using ΩMEGA
- More control over "student"
- More granularity annotations in less time (compared to WoZ)

Study Environment



Study Environment



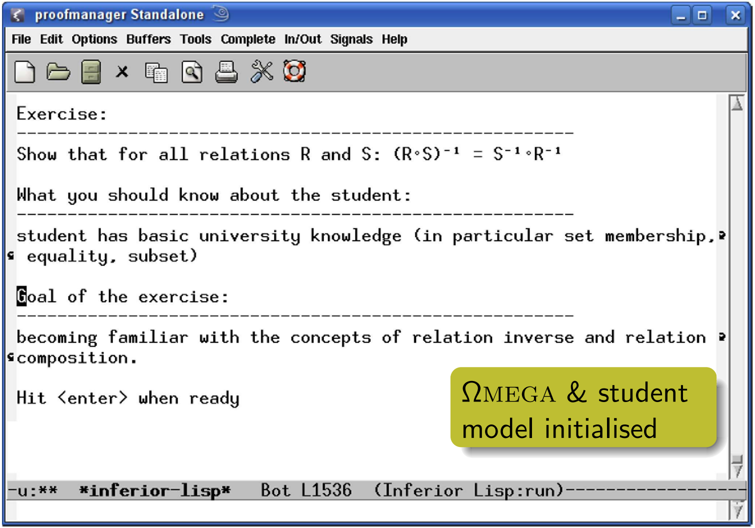
step size (w.r.t. assertion level inferences)

randomized

explanation

randomized

Study Environment



ΩMEGA & student model initialised

Ongoing Work

Ongoing Evaluation

- 2 experiments with 2 expert tutors (using different exercises in naive set theory, relations, topology)

	Tutor 1	Tutor 2
Steps annotated:	135	207
Perf. learnt classifier[1] -mean correct - κ	86.7% κ=0.68	68.9% κ=0.47
Interrater reliability[2]	κ=0.37	

[1] best rule-based classifier, evaluated on full dataset using 10-fold cross validation

[2] on common subset of 108 steps

Discussion & Outlook

Discussion

- Empirical modeling of granularity
- ...independent of introspection/justification of experts' judgments
- Thus, we imitate the behavior of expert tutors
- Is it desirable/possible to establish a best practice for judging proof step granularity?

Outlook

- Further experiment sessions planned with different experts
- Measure performance of learnt classifiers, agreement between tutors
- What are the most useful granularity criteria for the classification task?

Thank you!

Questions ?

Diversity in Wizard-of-Oz Corpora

Proof exercise: $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$

Student X		Student Y	
st[0]:	$(R \circ S)^{-1} = \{(y,x) \mid (x,y) \in (R \circ S)\}$	st[0]:	One needs to show equality between two sets.
tu[0]:	This statement is correct.	tu[0]:	That's right! How do you proceed?
st[1]:	$(R \circ S)^{-1} = \{(y,x) \mid \exists z (z \in M \land (x,z) \in R \land (z,y) \in S\}$	st[1]:	I use the extensionality principle.
tu[1]:	This formula is also correct.	tu[1]:	That's right.
st[2]:	$(R \circ S)^{-1} = \{(y,x) \mid \exists z (z \in M \land (z,x) \in R^{-1} \land (y,z) \in S^{-1}\}$	st[2]:	Let $(s,r) \in (R \circ S)^{-1}$. According to the definition of the inverse relation it then holds that $(r,s) \in (R \circ S)$.
tu[2]:	This is correct. You are on a good way.	tu[2]:	That's right!