Mathematical Domain Reasoning Tasks in Natural Language Tutorial Dialog on Proofs

Christoph Benzmüller

Joint work with: Quoc Bao Vo (and the SFB378 DIALOG Project)

Saarland University
Saarbrücken, Germany
http://www.ags.uni-sb.de/~chris/dialog/

AAAI-05, July 11th, Pittsburgh, USA
The DIALOG Project in the SFB 378

Computational Linguistics

Ambiguity Resolution and Proof Step Evaluation

Computer Science (Deduction)

Robust Deep and Shallow Linguistic Analysis

Natural Language Tutorial Dialog on Proofs

Non-trivial Mathematical Domain Reasoning
Method: Progressive Refinement

Modeling of the DIALOG System

Empirical Simulation Studies

Partial Implementation of a Demonstrator
WOZ-Experiment → Own Corpus

**Subject (Student) Room:**
- Audio Recording
- Subject GUI

**Cam1**
- Subject
- Audio Control

**Wizard Room:**
- Cam2 Screen
- Overall Control
- Subject Screen
- Audio Control
- Audio Recording

- Wizard GUI
- Wizard (Tutor)

**Overall Control**
- Experiment
- Wizard

© C. Benzmüller 2005
WOZ-Experiment → Own Corpus

Wizard (Tutor) Room:

Cam2 Screen
Overall Control
Experimenter
Wizard GUI
Wizard (Tutor)

Cam1 Screen
Subject Screen
Audio Control
Audio Recording

Subject Room:

Audio Recording
Subject GUI
Cam1
Subject
Audio Control
Cam2
Corpus Example

T1: Bitte zeigen Sie: \(K((A \cup B) \cap (C \cup D)) = (K(A) \cap K(B)) \cup (K(C) \cap K(D))!\)

[Please show: \(K((A \cup B) \cap (C \cup D)) = (K(A) \cap K(B)) \cup (K(C) \cap K(D))!)\]

S1: nach deMorgan-Regel-2 ist \(K((A \cup B) \cap (C \cup D)) = (K(A \cup B) \cup K(C \cup D))\).

[by deMorgan-Rule-2 \(K((A \cup B) \cap (C \cup D)) = (K(A \cup B) \cup K(C \cup D))\) holds.]

T2: Das ist richtig!

[This is correct!]

S2: \(K(A \cup B)\) ist laut deMorgan-1 \(K(A) \cap K(B)\)

[K(A \cup B) is \(K(A) \cap K(B)\) according to deMorgan-1]

T3: Das stimmt auch.

[That is also right.]

S3: und \(K(C \cup D)\) ist ebenfalls laut deMorgan-1 \(K(C) \cap K(D)\)

[and \(K(C \cup D)\) is also \(K(C) \cap K(D)\) according to deMorgan-1]

... 

Get corpus: http://www.ags.uni-sb.de/~chris/dialog/

Total figures 1. exp.: 66 dialogs / av. 12 turns / 1115 sentences
Research Challenges

Perspective of Mathematical Domain Reasoning (MDR):

- Support for resolution of Ambiguities and Underspecification
- Proof Step Evaluation
  - Soundness: proof step verifiable by formal system?
  - Granularity: size/argumentative complexity of proof step?
  - Relevance: proof step needed/useful in achieving the goal?

Perspective of NL Analysis: [. . . not in this talk . . .]

Perspective of Dialog Management: [. . . not in this talk . . .]

Perspective of Tutoring Proofs: [. . . not in this talk . . .]
Research Challenges

Perspective of Mathematical Domain Reasoning (MDR):
- Support for resolution of Ambiguities and Underspecification
- Proof Step Evaluation
  - Soundness: proof verifiable by formal system?
  - Granularity: argumentative complexity of proof step?
  - Relevance: proof step needed/useful in achieving the goal?

Perspective of NL Analysis: [... not in this talk ...]

Perspective of Dialog Management: [... not in this talk ...]

Perspective of Tutoring Proofs: [... not in this talk ...]
Mathematical Domain Reasoning

---

--- declarative abstract level sketches

\[ \text{Communication Gap} \]

--- procedural calculus level proofs ---

---

```
NL Input

 NL Input

 analysis

 Reading-1

 Reading-2

 MDR

 Reading-1

 + PSE-Info

 X

 Reading-2

 + PSE-Info

 tutor module
```
Mathematical Domain Reasoning

\[ A \cup B \text{ contains } B \]

\[ B \in A \cup B \]
\[ B \subseteq A \cup B \]
\[ B \subset A \cup B \]
$B \in A \cup B$

$A \cup B$ contains $B$

$B \subseteq A \cup B$

$B \subset A \cup B$

type checking
Mathematical Domain Reasoning

A \cup B \text{ contains } B

B \in A \cup B

B \subseteq A \cup B

B \subset A \cup B

\text{ theorem proving}
\[ \mathcal{P}((A \cup C) \cap (B \cup C)) = \mathcal{P}(C \cup (A \cap B)) \]

\[ \mathcal{P}((A \cup C) \cap (B \cup C)) = \mathcal{P}(C \cap (A \cup B)) \]

\[ \mathcal{P}((A \cup C) \cap (B \cup C)) = \mathcal{P}(C \cup (A \cap B)) \]

Type checking
\( K((A \cup C) \cap (B \cup C)) = K(C) \cup (A \cap B) \)

\( K((A \cup C) \cap (B \cup C)) = K(C \cup (A \cap B)) \)

\( K((A \cup C) \cap (B \cup C)) = K(C) \cup (A \cap B) \)

\( K((A \cup C) \cap (B \cup C)) = K(C \cup (A \cap B)) \)

Theorem proving
Given: (DM-1) $X \cup Y = \overline{X} \cap \overline{Y}$
(DM-2) $X \cap Y = \overline{X} \cup \overline{Y}$

Task: Please show $(A \cup B) \cap (C \cup D) = (A \cap B) \cup (C \cap D)$

New: By deMorgan $(A \cup B) \cap (C \cup D) = (A \cup B) \cup (C \cup D)$. 
Proof Step Evaluation

Given: (DM-1) \( X \cup Y = \overline{X \cap Y} \)
(DM-2) \( X \cap Y = \overline{X \cup Y} \)

Task: Please show \( (A \cup B) \cap (C \cup D) = (\overline{A} \cap \overline{B}) \cup (\overline{C} \cap \overline{D}) \)

New: By deMorgan \( (A \cup B) \cap (C \cup D) = (\overline{A} \cup \overline{B}) \cap (\overline{C} \cup \overline{D}) \)

Soundness: yes
Granularity: 1x(DM-2)
Relevance: yes
Proof Step Evaluation: How?

Discourse:

(1) A ∧ B
(2) A → C
(3) C → D
(4) F → B

New:

We show E.

(1) …
(2) …
(3) …
(4) …

(G) D ∨ E

PSE:

Soundness

Granularity

Relevance

(G') E

(G) …
Proof Step Evaluation: How?

**Discourse:**

1. $A \land B$
2. $A \Rightarrow C$
3. $C \Rightarrow D$
4. $F \Rightarrow B$

**New:**

We show $E$.

$\vdash (G') E$ (G)

**PSE:**

**Soundness**
- $(G') \vdash ? (G)$
- any proof

**Granularity**

**Relevance**
Proof Step Evaluation: How?

Discourse:

(1) \( A \land B \)
(2) \( A \Rightarrow C \)
(3) \( C \Rightarrow D \)
(4) \( F \Rightarrow B \)
?
(G) \( D \lor E \)

New:

We show \( E \).

PSE:

Soundness

- \((G') \vdash ? (G)\)
- any proof

Granularity

- size-of(\((G') \vdash ? (G)\))
- cognitively adequate proofs

Relevance
Proof Step Evaluation: How?

Discourse:

(1) \( A \land B \)
(2) \( A \to C \)
(3) \( C \to D \)
(4) \( F \to B \)

\(?\)

\((G) \ D \lor E\)

New:

We show \(E\).

PSE:

Soundness

\(-\) \( (G') \vdash? (G) \)

any proof

Granularity

\(-\) \( \text{size-of}((G') \vdash? (G)) \)

cognitively adequate proofs

Relevance

\(-\) \( (1), (2), (3), (4) \vdash? (G') \)

detours?, shorter proofs?
Granularity and Relevance call for
cognitively adequate abstract level proofs

+ enumeration of (some) proof alternatives

- One candidate: knowledge based proof planning [Bundy88]
- Original motivation: widen range of automatable maths
- New motivation: support for proof step evaluation
Related Work

- **Motivation:** [Moore93] Flexible tutorial NL dialog supports active learning
- **Closest related:** [Zinn04] analyzes well structured text-book proofs for soundness
- **NL analysis:** shallow techniques and keyword spotting probably not suitable
- **MDR:** Comparison against ‘golden standard solutions’ [GreaserEtAl00] no suitable
- **Dialog modeling:** Autotutor [PersonEtAl00], Geometry Tutor [MatsudaVanLehn03], Trindi and Siridus [TraumLarsson03], Beetle [Zinn03]
Conclusion

Finding ‘a proof’ automatically

Essential criterion: soundness

Traditional TP challenges: completeness, efficiency

Reasoning about human-constructed proof(step)s

Novel criteria: granularity & relevance

Novel TP challenges: qualitative aspects of proofs, enumeration of proofs

■ Lots of ongoing work in all corners of the DIALOG Project