Advise-Me

Domain reasoners for analysing tasks

Bastiaan.Heeren@ou.nl

Multiplier Event

October 19, 2018

Heerlen

Open Universiteit
www.ou.nl

Introduction

Using ICT for assessing mathematics achievement:

- Targeted at 12-15 years old pupils
- Setting up algebraic expressions and equations, and simplifying and solving them

Detailed assessment of free-form answers to math story problems:

- Analysis of intermediate steps
- Determining the high-level solution approach
- Identification of misconceptions

Advise-Me at ECTEL-2018

Fine-grained Cognitive Assessment based on Free-form Input for Math Story Problems

Bastiaan Heeren[1], Johan Jeuring[1,2], Sergey Sosnovsky[2], Paul Drijvers[3], Peter Boon[3], Sietske Tacoma[3], Jesse Koops[4], Armin Weinberger[5], Brigitte Grugeon-Allys[6], Françoise Chenevotot-Quentin[6], Jorn van Wijk[1], and Ferdinand van Walree[1]

[1] Faculty of Management, Science & Technology, Open University of the Netherlands, P.O.Box 2960, 6401 DL Heerlen, The Netherlands, bastiaan.heeren@ou.nl
[2] Dept. of Inf. and Comp. Sciences, Universiteit Utrecht, Utrecht, The Netherlands
[3] Freudenthal Institute, Utrecht University, Utrecht, The Netherlands
[4] Cito, Arnhem, The Netherlands
[5] Dept. of Educational Technology, Saarland University, Saarbrücken, Germany
[6] Laboratoire de Didactique André Revuz, Université Paris Est Créteil, Paris, France

Abstract. We describe an approach to using ICT for assessing mathematics achievement of pupils using learning environments for mathematics. In particular, we look at fine-grained cognitive assessment of free-form answers to math story problems, which requires determining the steps a pupil takes towards a solution, together with the high-level solution approach used by the pupil. We recognise steps and solution approaches in free-form answers and use this information to update a user model of mathematical competencies. We use the user model to find out for which student competencies we need more evidence of mastery, and determine which next problem to offer to a pupil. We describe the results of our fine-grained cognitive assessment on a large dataset for one problem, and report the results of two pilot studies in different European countries.

Keywords: Math story problems · Step-based assessment · Free-form input · Solution strategies · User modelling

Open Universiteit
www.ou.nl

Free-form input

numworx Setting up algebraic expressions > Bastiaan Heeren

LESSON Setting up algebraic expressions

Task

05 **Magical trick?**

A student says to her peer:
"Choose a number, add 8, multiply the result by 3, subtract 4, add the initial number, divide by 4, add 2, and subtract the initial number.
You will end up with 7."

Is this true for any starting number? Explain your answer.

Your work

f(x)

(x+8)*3-4+x/4+2-x = 7

(3(x+8)-4+x)/4+2-x = 7

(3x+24-4+x) /4+2 = 7

(4x+20) /4+2-x = 7

x+5+2-x = 7

7=7

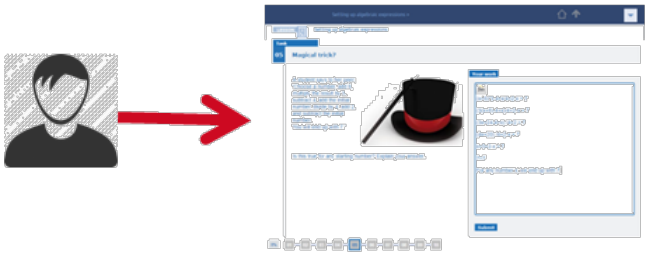
For any number x, we end up with 7.

Submit

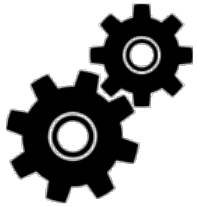
0% 01 02 03 04 05 06 07 08 09 10

High-level architecture

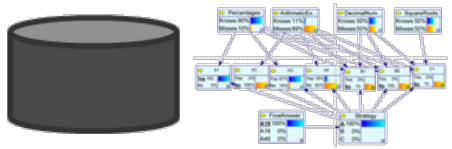
Assessment Environment: user interface, offers tasks



Domain Reasoner (expert model): recognizing steps, (partial) answers, mistakes, etc., collecting evidence



Reporting Module: presents skill levels of individuals and groups, for teachers and pupils.



Task Sequencer: selecting which (sub)task to offer next (or how); at its simplest, task sequence could be static

User and Task Model: storing evidence per user, interpreting evidence as (un)certainties for knowledge components

Open Universiteit
www.ou.nl

Examples of solution approaches

((5+8)*3-4+5)/4+2-5 = 7

- global, arithmetic

5+8 = 13; 13*3 = 39;
39-4 = 35; 35+5 = 40;
40/4 = 10; 10+2 = 12;
12-5 = 7

- step-by-step, arithmetic

5+8*3-4+5/4+2-5 = 7

- global, arithmetic, priority mistakes

5+8 = 13*3 = 39-4 = 35+5
= 40/4 = 10+2 = 12-5 = 7

- step-by-step, arithmetic, chaining

(x+8)*3 = 3x+24
3x+24-4 = 3x+20
3x+20+x = 4x+20
(4x+20)/4 = x+5
x+5+2 = x+7
x+7-x = 7

- step-by-step, algebraic

x+8*3-4+x/4+2-x
= 2x+20/4+2-x
= 2x+5+2-x
= x+7

- simplification mistakes

Datasets for testing

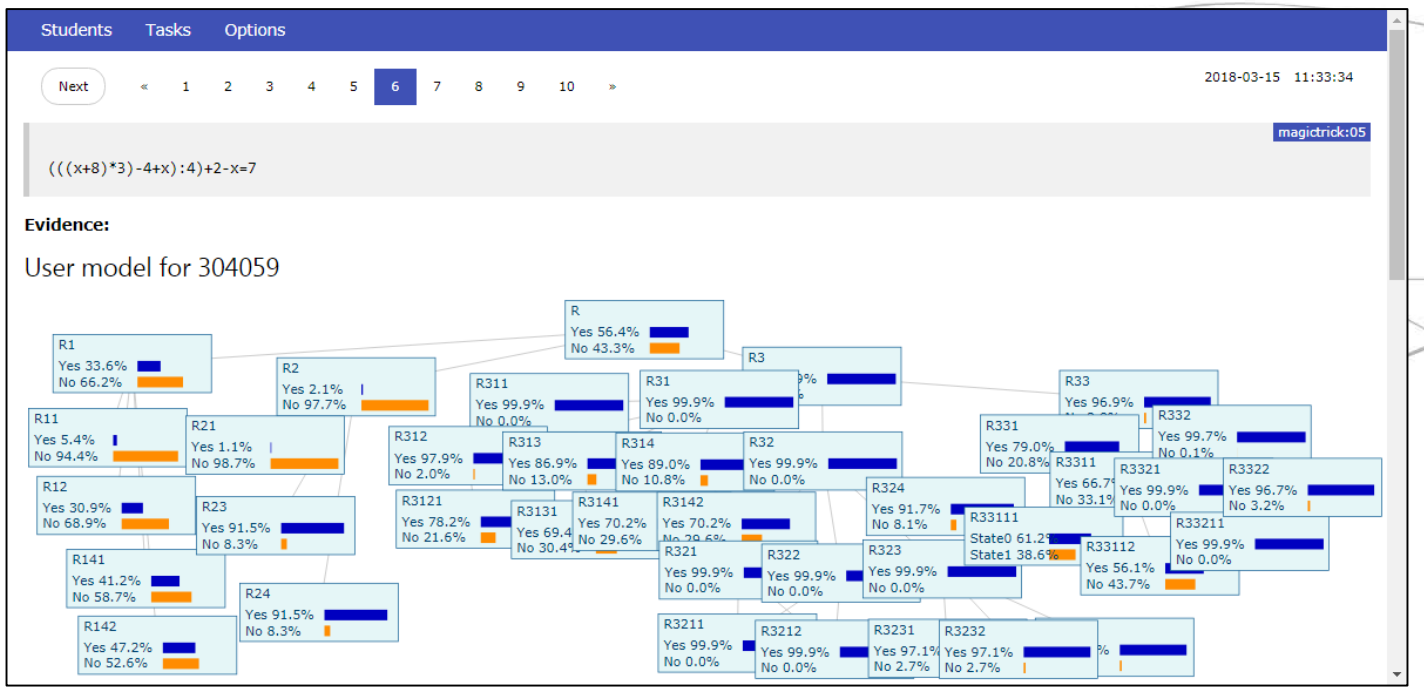
- Large dataset for the Magical trick task
  - Collected between 2011 and 2015
  - On LaboMep platform, developed by Sesamath
  - 2956 student responses (grade 8-10)

- Dataset was used to test the feasibility of assessing free-form input
- High-level solution approach could be recognized for nearly 80%

Grugeon-Allys, B., Chenevotot-Quentin, F., Pilet, J., Prévit, D.: Online automated assessment and student learning: the Pepite project in elementary algebra. In: Ball, L., Drijvers, P., Ladel, S., Siller, H.S., Tabach, M., Vale, C. (eds.) Uses of Technology in Primary and Secondary Mathematics Education: Tools, Topics and Trends, pp. 245–266. Springer (2018)

Datasets for testing (continued)

- We organized four pilots (March-June 2018):
  - Pilots in Germany (1), the Netherlands (2), and France (1)
  - Collected 940 answers by 78 students
  - Ten tasks (some with sub-parts)
- Data used for calibrating assessment software and user models



Domain reasoner

Analysis of input proceeds in 3 phases:

1. Extract mathematical expressions
   - Some natural language-specific pre-processing
   - Compare x+3 with 3x5=15
2. Parse extracted expressions
3. Recognise solution approach
   - View recognition as a parsing problem
   - Recogniser must be flexible enough to deal with imperfections and mistakes

- For some answers, the analysis fails to produce evidence

Examples

Pour le prouver il faut utiliser x comme nombre quelconque:

x+8=x+8 (x+8)*3=3x+24 3x+24-4=3x+20 3x+20+x=4x+20

(4x+20)/4=x+5 x+5+2=x+7 x+7-x=7

Donc oui on trouve 7 pour n'importe quel nombre de départ

32*25%=8 32-8=24 24*25%=6 24-6=18 18*25%=4,5 18-4,5=13,5

13,5*25%=3,375 13,5-3,375=10,125 10,125*25%=2,53125 10,125-

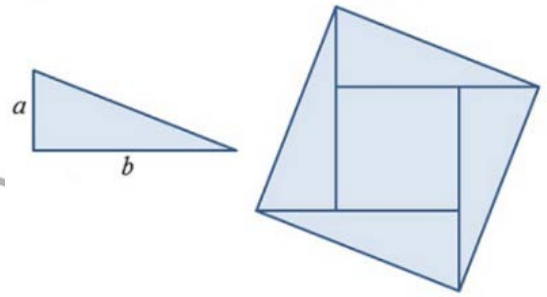
2,25125 7,87375*25%=1,9684374 7,87275-1,9684374=5,.....

Es sind 6 von den figuren in der reihe

je hebt 4 keer de driehoek dus 4(0,5ab)

het vierkant in het midden is $(b-a)^2$

A=2ab+$(b-a)^2$



Phase 1: math extraction

- Most of the natural language is ignored
- MathML presentation language is converted to plain text
- Extraction is specialized for each task (e.g. multi-letter variables)
- Language-specific transformations (mal, plus, quadrat, hoch)
- Disambiguation of symbols (x+3 versus 3x5=15), based on heuristics

1.Puppe: 32cm
2.Puppe: 32* $\frac{75}{100}$ =24cm
3.Puppe:24* $\frac{75}{100}$ =18cm
4.Puppe:18* $\frac{75}{100}$ =13,5cm
5.Puppe:13,5* $\frac{75}{100}$=10,125
6.Puppe:10,125* $\frac{75}{100}$=7,59375
Es gibt 6 Puppen

(a plus b) mal (a plus 5)

On your right hand side you see the first three of a series Matryoshka dolls. The puppets fit into each other, due to a decreasing height. The biggest puppet is 32 cm high. Each next puppet is 25% smaller than the previous one. In this sequence, there are no puppets smaller than 6 cm.

How many puppets are there in this series? Write down your intermediate steps.

Phase 2: parsing expressions

- Surprisingly few expressions cannot be parsed
- Common mistake: unbalanced parentheses
- Incorrect ‘chaining’ of equations (in Dutch: ‘breien’)
- Equations can be checked for equality, which is particularly helpful for spotting mistakes
  - For example, 5+8 = 40

`x+8)*3)-4)+x)/4)+2)-x`

`[(x+8)*3-4+x)]/4+2-x=7`

5 5+8= 12*3= 48-4= 44+5 49*2= 98-5= 93

Open Universiteit
www.ou.nl

Phase 3: strategy recognition

- Approach: consider recognition as a parsing problem
- During parsing, keep track of introduced variables
- Parsing must be flexible enough to:
  - recognize mistakes and imperfections (error-correcting parser)
  - deal with implicit steps
  - continue after basic calculation mistakes
  - identify algebraic misconceptions
  - spot missing parentheses when setting up expression
- Matching with normal forms (associativity, commutativity, basic calculations)

Trade-off between flexibility and computation time:

- calibrate recognizer based on collected data

About the software

- Built on top of the Ideas framework
- Offered as feedback and assessment services
  - Connected with Numworx (the Digital Mathematics Environment)
  - Currently, communication in a very simple xml-format
  - Planned support for QTI standard
- Software is still under development
- To be released under a permissive free software license (Apache 2)
- Developed in Haskell (approximately 15,000 LOC)
  - General support for assessing math story problems
  - Specific support for 10 tasks (approximately 300 LOC per task)

Conclusions

- We have developed a framework for fine-grained cognitive assessment of free-form solutions to math story problems
- Our domain reasoner extracts the mathematics, analyses intermediate steps, and tries to recognize the solution approach
- The solution approach was recognized in nearly 80% of the answers for two small-scale pilot studies

Presented after lunch:

- We use Bayesian task models and a user model for the inference, storage, and update of user knowledge
- In the future, we will organize more extensive evaluations