Semantically-Aligned Equation Generation for Solving and Reasoning Math Word Problems

Ting-Rui Chiang and Yun-Nung (Vivian) Chen

https://github.com/MiuLab/E2EMathSolver
Math Word Problem

Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?

Reasoning & Solving

\[ x = 10 - 1 \times 5 \div 0.5 \]
Prior Work

Non-neural approaches

- Template-based
  (Kushman et al., Upadhyay and Chang)

\[ x = (? + ?) \times ? - ? \]

Rely on hand-crafted features!

Deep learning

- Seq2Seq
  (Wang et al., Ling et al.)

\[ Problem \]

\[ x = (1+2) \times 3 - 4 \]

\[ generate \]

\[ x = (1+2) \times 3 - 4 \]

Does not use the structure of math expression.

Our model is end-to-end and structural!
Overview of the Proposed Model

Encoder

Decoder

Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?

\[ x = 10 - 1 \times 5 \div 0.5 \]
Look Again at the Problem

Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?
Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?

\[ x = \left( 10 - 1 \times 5 \right) \div 0.5 \]
Idea: Bridging Symbolic and Semantic Worlds

Symbolic World

\[
\begin{align*}
2 & > -3 \\
\pi & \approx 3.14 \\
\sqrt{2} & = 1 + 2 \cdot 3 \\
\frac{1}{(1 - 2) + 3} & = 101_2 = 5_{10}
\end{align*}
\]

Semantic World
Preprocess

Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?
Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?
Each notebook takes $0.5 and ...
Semantic Generation for Unknown $x$

Each notebook takes $0.5$ and...
Each notebook takes $0.5$ and each pen takes $1$. Tom has $10$. How many notebooks can he buy after buying 5 pens?
Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?

\[ x = (10 - \text{Price of a pen} \times 5) \]

Number of pens bought.

Price of a pen.
Equation Generation in Postfix

Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?

\[ x \ 10 \ 1 \ 5 \times - \ 0.5 \div = \]
Equation Generation by Stack Actions

- Stack is used
- The decoder generates stack actions.
- An equation is generated with actions on stack.

$$x = 10 - 1 \times 5 \div 0.5$$

Decoder
Action Selection in Each Step

Encoder

Decoder

stack action
\{+, -, \times, \div, =, Push\}

classifier
Equation Generation by Stack Actions

Target Equation: $x = 10 - 1 \times 5 \div 0.5$

Generated Actions:

Action: push
Equation Generation by Stack Actions

Target Equation: \( x = 10 - 1 \times 5 \div 0.5 \)

Generated Actions: \( x \ 10 \ 1 \ 5 \)

Action: push
Equation Generation by Stack Actions

Target Equation: \( x = 10 - 1 \times 5 \div 0.5 \)

Generated Actions: \( x \ 10 \ 1 \ 5 \)

Action: \( \times \)
Equation Generation by Stack Actions

Target Equation: \( x = 10 - 1 \times 5 \div 0.5 \)

Generated Actions: \( x \ 10 \ 1 \ 5 \ \times \ 0.5 \ \div = \)

After many steps...

\[ x = (10 - 1 \times 5) \div 0.5 \]
Training Process

- Target equation is given.
- Trained as Seq2Seq.

Each notebook takes $0.5 and each pen takes $1. Tom has $10. How many notebooks can he buy after buying 5 pens?
Experiments

- Dataset: Math23k
- In Chinese
- 23000 math word problems.
- Operators: +, -, ×, ÷
Results

<table>
<thead>
<tr>
<th></th>
<th>Retrieval</th>
<th>Template</th>
<th>Generation</th>
<th>Ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc.</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>Retrieval</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>BLSTM</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>Self-Attention</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Seq2Seq w/SNI</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td>Proposed</td>
<td>65</td>
<td>70</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td>Hybrid</td>
<td>65</td>
<td>70</td>
<td>70</td>
<td>65</td>
</tr>
</tbody>
</table>

$\approx 8\%$

$> 1\%$
Ablation Test

Char-Based
Word-Based
Word-Based -Semantic
≈ 3%

Word-Based -Gate

Word-Based -Gate -Attention
≈ 2.5%

Word-Based -Gate -Attention -Stack
Self-Attention for Qualitative Analysis

Each notebook takes $0.5$ and ...
Self-Attention for Qualitative Analysis

Encoder

Each notebook takes $0.5$ and ...
Attention for Operand Semantics

The attention focuses on:

- Informative verbs
  - “gain”, “get”, “fill”, etc.

- Quantifier-related words
  - “every”, “how many”, etc.
Conclusion

Three main contributions

- **Approach:** equation generation with stack
- **Originality:** automatic extraction of operand semantics
- **Performance:** a SOTA end-to-end neural model on Math23k
Code Available @
https://github.com/MiuLab/E2EMathSolver