Solving Math Word Problems with Pre-trained Language Models

Presenter: Qun Liu (刘群)

Huawei Noah's Ark Lab

2021世界计算大会，2021-09-17，湖南长沙
Generate & Rank: A Multi-task Framework for Math Word Problems

Jianhao Shen\textsuperscript{1\dagger}, Yichun Yin\textsuperscript{2}, Lin Li\textsuperscript{3}, Lifeng Shang\textsuperscript{2}, Xin Jiang\textsuperscript{2}, Ming Zhang\textsuperscript{1\ast}, Qun Liu\textsuperscript{2}

\textsuperscript{1}Department of Computer Science, School of EECS, Peking University
\textsuperscript{2}Huawei Noah’s Ark Lab
\textsuperscript{3}Huawei HiSilicon

\{jhseshn, mzhang_cs\}@pku.edu.cn
\{yinyichun, lilin29, shang.lifeng, jiang.xin, qun.liu\}@huawei.com

Accepted by Findings of EMNLP 2021
Content

Introduction to Math Word Problem (MWP) Solving

MWP Solving Approaches: A Brief Survey

Generate & Rank: A Multi-task Framework for Solving MVPs

Conclusion
Content

Introduction to Math Word Problem (MWP) Solving

MWP Solving Approaches: A Brief Survey

Generate & Rank: A Multi-task Framework for Solving MVPs

Conclusion
Math Word Problem (MWP)

- Input: a math problem described in natural language, with a question about an unknown quantity
- Output: an expression that solves the problem

<table>
<thead>
<tr>
<th>Original MWP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem</strong></td>
</tr>
<tr>
<td><strong>Solution</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number-mapped MWP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem</strong></td>
</tr>
<tr>
<td><strong>Solution</strong></td>
</tr>
</tbody>
</table>
The Evolution of MWP Solvers
Datasets for MWPs

Statistics of arithmetic word problem datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th># problems</th>
<th># single-op</th>
<th># multi-op</th>
<th>operators $O$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA1</td>
<td>134</td>
<td>112</td>
<td>22</td>
<td>${+, -}$</td>
</tr>
<tr>
<td>IXL</td>
<td>140</td>
<td>119</td>
<td>21</td>
<td>${+, -}$</td>
</tr>
<tr>
<td>MA2</td>
<td>121</td>
<td>96</td>
<td>25</td>
<td>${+, -}$</td>
</tr>
<tr>
<td>AI2</td>
<td>395</td>
<td>327</td>
<td>68</td>
<td>${+, -}$</td>
</tr>
<tr>
<td>IL</td>
<td>562</td>
<td>562</td>
<td>0</td>
<td>${+, -}$</td>
</tr>
<tr>
<td>CC</td>
<td>600</td>
<td>0</td>
<td>600</td>
<td>${+, -, \times, \div}$</td>
</tr>
<tr>
<td>SingleEQ</td>
<td>508</td>
<td>390</td>
<td>118</td>
<td>${+, -, \times, \div}$</td>
</tr>
<tr>
<td>AllArith</td>
<td>831</td>
<td>634</td>
<td>197</td>
<td>${+, -, \times, \div}$</td>
</tr>
<tr>
<td>MAWPS-S</td>
<td>2,373</td>
<td>1,311</td>
<td>1,062</td>
<td>${+, -, \times, \div}$</td>
</tr>
<tr>
<td>Dolphin-S</td>
<td>7,070</td>
<td>115</td>
<td>6,955</td>
<td>${+, -, \times, \div}$</td>
</tr>
<tr>
<td>Math23K</td>
<td>23,162</td>
<td>3,131</td>
<td>20,031</td>
<td>${+, -, \times, \div}$</td>
</tr>
</tbody>
</table>
Content

Introduction to Math Word Problem (MWP) Solving

MWP Solving Approaches: A Brief Survey

Generate & Rank: A Multi-task Framework for Solving MVPs

Conclusion
Rule-based Approaches (1960-2010)

▶ Production Rule:
  ▶ A set of conditions to be met
  ▶ A set of actions to carry out

NAME: NowProp
CONDITIONS: 1. Does STM contain a proposition?
             2. Does that proposition have the predicate NOW?
ACTIONS: 1. Put the proposition in the specification slot of the
          text base.
          2. Put (TIME:PRESENT) in the specification slot of
             the problem model.

C. R. Fletcher, Understanding and solving arithmetic word problems: A computer simulation, Behavior Research Methods, Instruments, & Computers 17(5), 1985

- Identify entities, quantities and operators

Mohammad Javad Hosseini et al., Learning to Solve Arithmetic Word Problems with Verb Categorization, EMNLP 2014
Some useful tricks:

- Decode with predefined rules
- Significant number identification
- Equation normalization

Yan Wang et al., Deep neural solver for math word problems. EMNLP 2017
Lei Wang et al., Translating a Math Word Problem to an Expression Tree. EMNLP 2018
Template-Based Solvers with Recursive Neural Networks

- Coarse-to-fine generation
  - Generate template first: \((n_1 \text{ <op> } n_3) \text{ <op> } n_2\)
  - Infer missing ops: \((n_1 - n_3) / n_2\)

Lei Wang et al., Template-Based Math Word Problem Solvers with Recursive Neural Networks. AAAI 2019
Template-Based Solvers with Recursive Neural Networks

- The benefit of templates
  - Template generation is less challenging than direct generation of whole expression
  - With template we can encode structure information as well

<table>
<thead>
<tr>
<th>Our Approach</th>
<th>MAWPS</th>
<th>Math23K</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-RNN</td>
<td>66.8</td>
<td>66.9</td>
</tr>
<tr>
<td>- EN</td>
<td>63.9</td>
<td>61.1</td>
</tr>
<tr>
<td>- Bi-LSTM</td>
<td>31.1</td>
<td>34.1</td>
</tr>
<tr>
<td>- Self-Att</td>
<td>66.3</td>
<td>65.1</td>
</tr>
</tbody>
</table>

Lei Wang et al., Template-Based Math Word Problem Solvers with Recursive Neural Networks. AAAI 2019
A Goal-Driven Tree-Structured Neural Model for MWPs

- Mathematical expressions are tree-structured
- Each node has a goal (to solve a sub-question)

**Problem:** Robin was making baggies of cookies with 6 cookies in each bag. If she had 23 chocolate cookies and 25 oatmeal cookies, how many baggies could she make?

**Solution Expression:** \((23 + 25) \div 6\)  
**Solution:** 8

Zhipeng Xie and Shichao Sun, A Goal-Driven Tree-Structured Neural Model for Math Word Problems. IJCAI 2019
A Goal-Driven Tree-Structured Neural Model for MWPs

Zhipeng Xie and Shichao Sun, A Goal-Driven Tree-Structured Neural Model for Math Word Problems. IJCAI 2019
Graph-to-Tree Learning for Solving Math Word Problems

► Enrich text encoding with graph features

Jipeng Zhang et al., Graph-to-Tree Learning for Solving Math Word Problems. ACL 2020
Graph-to-Tree Learning for Solving Math Word Problems

- **Graph Construction:**
  - Quantity Cell
    - Quantity
    - Associated nouns
    - Adjectives
    - Units and rates
  - Undirected edges between quantity and other nodes in each cell
  - Directed edges between quantities pointing from larger to smaller numbers

<table>
<thead>
<tr>
<th></th>
<th>MAWPS</th>
<th>Math23K</th>
<th>Math23K*</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTS</td>
<td>82.6</td>
<td>75.6</td>
<td>74.3</td>
</tr>
<tr>
<td>Graph2Tree</td>
<td>83.7</td>
<td>77.4</td>
<td>75.5</td>
</tr>
</tbody>
</table>

Jipeng Zhang et al., Graph-to-Tree Learning for Solving Math Word Problems. ACL 2020
SMART: A Situation Model for Algebra Story Problems via Attributed Grammar

- The process of human solving algebra story problems
  - first hallucinate a situation model
  - Perform arithmetic reasoning
  - Update the situation model and repeat

Yining Hong et al., SMART: A Situation Model for Algebra Story Problems via Attributed Grammar, AAAI 2021
SMART: A Situation Model for Algebra Story Problems via Attributed Grammar

- A situation model is represented as an Attribute Grammar

\[ G = (S, V, A, E, R) \]

- \( S \) is the start symbol.
- \( V = \{S, \text{World}, \text{Agents}, \text{Agent}, \text{Events}, \text{Event}\} \)
- \( A = \{\text{rate, amount, total}\} \)
- \( E = \{e: e \text{ is a valid equation on attributes.}\} \)
- \( R = \{S \rightarrow \text{World} \}
  \rightarrow \text{Agents}
  \rightarrow \text{Agents Agent | Agent}
  \rightarrow \text{Events}
  \rightarrow \text{Events Event | Event}\} \)

- Attribute A
  - Rate: “A per B” or “each A has B”
  - Amount: a measurement of units of rate (e.g., hour)
  - Total = rate*amount

- Event E
  - Constraints on the attributes
SMART: A Situation Model for Algebra Story Problems via Attributed Grammar

Problem
A car travels from city A to city B. It travels for 6 hours on the first day, with a speed of 65 kilometers an hour. On the second day, it travels (1/3) of the distance he travels the first day. Then it arrives. What is the distance between city A and city B?

Inference

Situation Model

Learning

Yining Hong et al., SMART: A Situation Model for Algebra Story Problems via Attributed Grammar, AAAI 2021
SMART: A Situation Model for Algebra Story Problems via Attributed Grammar

Yining Hong et al., SMART: A Situation Model for Algebra Story Problems via Attributed Grammar, AAAI 2021
Learning by Fixing: Solving Math Word Problems with Weak Supervision

- Labeled equations are difficult to get
- Weak supervision: we only have answers rather than equations
- Basic idea of learning by fixing:
  - Initialize a generation model
  - Generate equations and check their answers
  - Fix the wrong equations, add both fixed equations and correct equations to training data and update the model

Yining Hong et al., Learning by Fixing: Solving Math Word Problems with Weak Supervision, AAAI2021
Yining Hong et al., Learning by Fixing: Solving Math Word Problems with Weak Supervision, AAAI2021
Learning by Fixing: Solving Math Word Problems with Weak Supervision

- Top-down 1-step fixing
  - Start from the root node
  - Replace the operator and if the new expression gives the correct answer, we get a 1-step solution
  - Inference the expected value of left/right node
  - If the left/right node is leaf and the expected value is in vocabulary, get a 1-step solution
  - Fix left/right node recursively
  - If 1-step fixing failed, randomly change a node and retry

Yining Hong et al., Learning by Fixing: Solving Math Word Problems with Weak Supervision, AAAI2021
Learning by Fixing: Solving Math Word Problems with Weak Supervision

<table>
<thead>
<tr>
<th></th>
<th>Weakly-Supervised</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2seq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>REINFORCE</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>MAPO</td>
<td>10.7</td>
<td></td>
</tr>
<tr>
<td>LBF-w/o-M</td>
<td>44.7</td>
<td></td>
</tr>
<tr>
<td>LBF</td>
<td>43.6</td>
<td></td>
</tr>
<tr>
<td>GTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>REINFORCE</td>
<td>15.8</td>
<td></td>
</tr>
<tr>
<td>MAPO</td>
<td>20.8</td>
<td></td>
</tr>
<tr>
<td>LBF-w/o-M</td>
<td>58.3</td>
<td></td>
</tr>
<tr>
<td>LBF</td>
<td>59.4</td>
<td></td>
</tr>
</tbody>
</table>

Yining Hong et al., Learning by Fixing: Solving Math Word Problems with Weak Supervision, AAAI2021
Content

Introduction to Math Word Problem (MWP) Solving

MWP Solving Approaches: A Brief Survey

Generate & Rank: A Multi-task Framework for Solving MVPs

Conclusion
Motivation

- Previous works formalize MWP as a generation task (like translation)
- However,
  - Math expressions are sensitive to minor mistakes
  - Maximizing generation likelihood doesn’t learn to distinguish minor mistakes explicitly
  - The performance degrades fast as expression gets longer
- Thus, we propose a multi-task framework for MWP
  - Introduce a new ranking task
  - Use a pre-trained model — BART
Pre-trained Language Models

- Recently pretrain-then-finetune is a new trend in NLP tasks
  - Pre-train on large corpus with self-supervised tasks
  - Fine-tune on downstream tasks
Family of Pretrained Language Models
BART

- Bidirectional and Auto-Regressive Transformers

BART pre-training

Finetune BART for text classification
Generate and Rank: A Multi-task Framework for MWPs

Task #1: Generating

1. Multi-task Training
   - Encoder
   - Decoder
   - Problem

Task #2: Ranking

2. Expression Online Updating
   - Expression Bank
   - Generate
   - Disturb

Generating Loss + Ranking Loss

En/Decoder Shared BART
Expression Ground-truth
Expression Candidates
Score
Ranker
Decoder
Encoder
Expression
Problem
Generate and Rank framework

- Generator: Finetune BART on MWP seq2seq task
- Ranker: Sequence pair classification task
  - Feed problem into encoder and expression into decoder
- Joint training: Share encoder and decoder

Task #1: Generating

Task #2: Ranking
Expression Bank

- Model-based Generation
  - Use beam search to produce top-K expressions
- Tree-based Disturbance
- Online updating
  - Update the expression bank at each training epoch

Figure 2: Overview of tree-based disturbance.
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Math23K†</th>
<th>Math23K‡</th>
<th>MAWPS‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNS</td>
<td>-</td>
<td>58.1</td>
<td>59.5</td>
</tr>
<tr>
<td>Math-EN</td>
<td>66.7</td>
<td>-</td>
<td>69.2</td>
</tr>
<tr>
<td>T-RNN</td>
<td>66.9</td>
<td>-</td>
<td>66.8</td>
</tr>
<tr>
<td>S-Aligned</td>
<td>-</td>
<td>65.8</td>
<td>-</td>
</tr>
<tr>
<td>Group-ATT</td>
<td>69.5</td>
<td>66.9</td>
<td>76.1</td>
</tr>
<tr>
<td>AST-Dec</td>
<td>69.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GTS</td>
<td>75.6</td>
<td>74.3</td>
<td>82.6</td>
</tr>
<tr>
<td>Graph2Tree</td>
<td>77.4</td>
<td>75.5</td>
<td>83.7</td>
</tr>
<tr>
<td>Multi-E/D</td>
<td>78.4</td>
<td>76.9</td>
<td>-</td>
</tr>
<tr>
<td>mBART</td>
<td>80.8</td>
<td>80.0</td>
<td>80.1</td>
</tr>
<tr>
<td>Generate &amp; Rank</td>
<td><strong>85.4</strong></td>
<td><strong>84.3</strong></td>
<td><strong>84.0</strong></td>
</tr>
</tbody>
</table>

Table 2: Solution accuracy on MAWPS and Math23K. † refers to the result of test set and ‡ denotes the result of 5-fold cross-validation. “-” means that the results are not reported in the original papers.

<table>
<thead>
<tr>
<th>#Op</th>
<th>Pro</th>
<th>AST-Dec</th>
<th>G2T</th>
<th>mBART</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.3</td>
<td>82.7</td>
<td>85.5</td>
<td>90.2</td>
<td>90.8 (+0.6)</td>
</tr>
<tr>
<td>2</td>
<td>52.2</td>
<td>74.5</td>
<td>83.7</td>
<td>88.1</td>
<td>90.2 (+2.1)</td>
</tr>
<tr>
<td>3</td>
<td>19.1</td>
<td>59.9</td>
<td>71.7</td>
<td>71.2</td>
<td>79.1 (+7.9)</td>
</tr>
<tr>
<td>4</td>
<td>6.6</td>
<td>42.4</td>
<td>51.5</td>
<td>53.0</td>
<td>63.6 (+10.6)</td>
</tr>
<tr>
<td>5</td>
<td>3.4</td>
<td>44.1</td>
<td>38.2</td>
<td>41.2</td>
<td>58.8 (+17.6)</td>
</tr>
<tr>
<td>6</td>
<td>0.9</td>
<td>55.6</td>
<td>55.6</td>
<td>55.6</td>
<td>88.8 (+33.2)</td>
</tr>
</tbody>
</table>

Table 5: Accuracy for increasing length of expressions. #Op is the number of operations in expressions. Pro denotes proportion of expressions with different lengths.
An Demonstration

Problem: 从甲地到乙地，如果骑自行车每小时行驶16千米，4小时可以到达，如果乘汽车只需要2小时，汽车每小时行驶多少千米？
求解：16 * 4 / 2 = 32.0

Problem: 小明看一本书，第一天看了全书的(1/5)，第二天比第一天多看了14页，剩下的25页第3天看完，这本书共有多少页？
求解：(1 + 25) / (1 - 0.2 - 0.2) = 64.99999999999999

Problem: 一头大象重3.4吨，一头鲸鱼的重量是大象的5.8倍，鲸鱼比大象重多少吨？
求解：3.4 * (5.8 - 1) = 16.32

Problem: 某农场要收割2300公顷小麦，原计划每天收割60公顷，收割5天后改为每天收割80公顷，还需要多少天才能完成任务？
求解：(2300 - 60 * 5) / 80 = 25.0

http://10.90.91.195:8007/
Content

Introduction to Math Word Problem (MWP) Solving

MWP Solving Approaches: A Brief Survey

Generate & Rank: A Multi-task Framework for Solving MVPs

Conclusion
Content

Introduction to Math Word Problem (MWP) Solving

MWP Solving Approaches: A Brief Survey

Generate & Rank: A Multi-task Framework for Solving MVPs

Conclusion
Thank you!

把数字世界带入每个人、每个家庭、每个组织，构建万物互联的智能世界。

Bring digital to every person, home and organization for a fully connected, intelligent world.

Copyright©2018 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.