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

NOAH'S ARK LAB

HUAWEI

Generate & Rank: A Multi-task Framework for Math Word Problems

Jianhao Shen[1†], Yichun Yin[2], Lin Li[3], Lifeng Shang[2], Xin Jiang[2], Ming Zhang[1*], Qun Liu[2]

[1]Department of Computer Science, School of EECS, Peking University

[2]Huawei Noah's Ark Lab

[3]Huawei HiSilicon

{jhshen, mzhang_cs}@pku.edu.cn

{yinyichun, lilin29, shang.lifeng, jiang.xin, qun.liu}@huawei.com

Accepted by Findings of EMNLP 2021

Content

Content

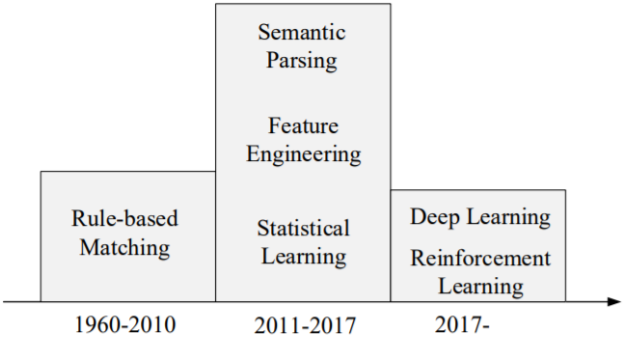
Math Word Problem (MWP)

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

Original MWP	
Problem	A project is completed in 25 days by 12 workers. If it takes 20 days to complete, how many workers will it take?
Solution	*25 * 12 / 20*
Number-mapped MWP	
Problem	A project is completed in *NUM0* days by *NUM1* workers. If it takes *NUM2* days to complete, how many workers will it take?
Solution	*NUM0 * NUM1 / NUM2*

HUAWEI NOAH'S ARK LAB

The Evolution of MWP Solvers



Datasets for MWPs

Statistics of arithmetic word problem datasets.

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

HUAWEI NOAH'S ARK LAB

Content

Introduction to Math Word Problem (MWP) Solving

MWP Solving Approaches: A Brief Survey

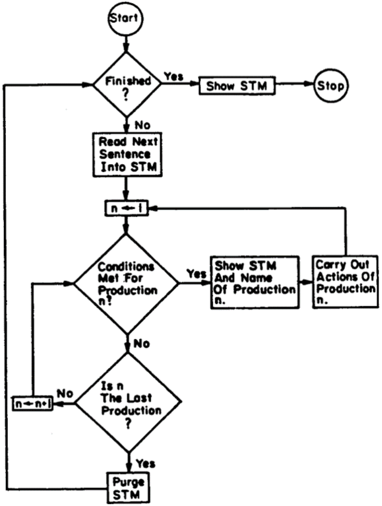
Generate & Rank: A Multi-task Framework for Solving MVPs

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.

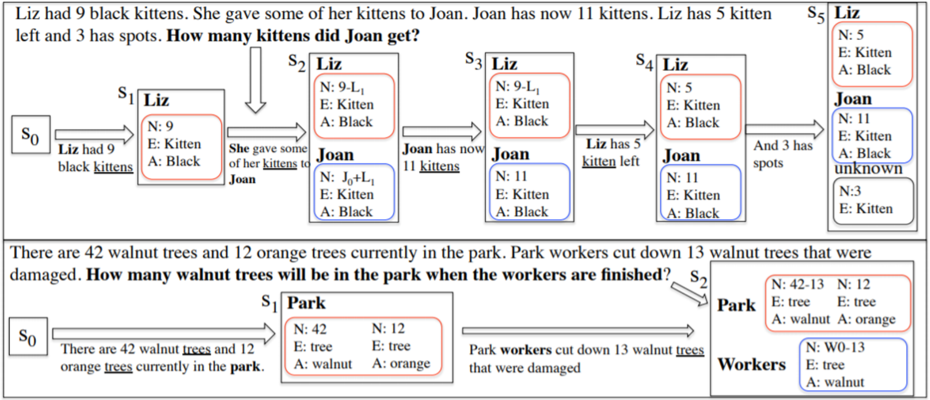


C. R. Fletcher, Understanding and solving arithmetic word problems: A computer simulation, Behavior Research Methods, Instruments, & Computers 17(5), 1985

HUAWEI NOAH'S ARK LAB

Semantic Parsing (2011-2017)

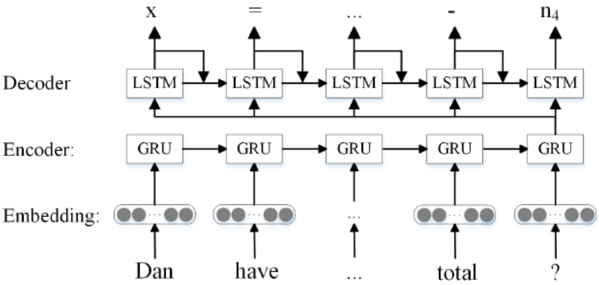
- Identify entities, quantities and operators



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

HUAWEI NOAH'S ARK LAB

Seq2Seq: Deep Neural Solver for Math Word Problems



- Some useful tricks:
  - Decode with predefined rules
  - Significant number identification
  - Equation normalization

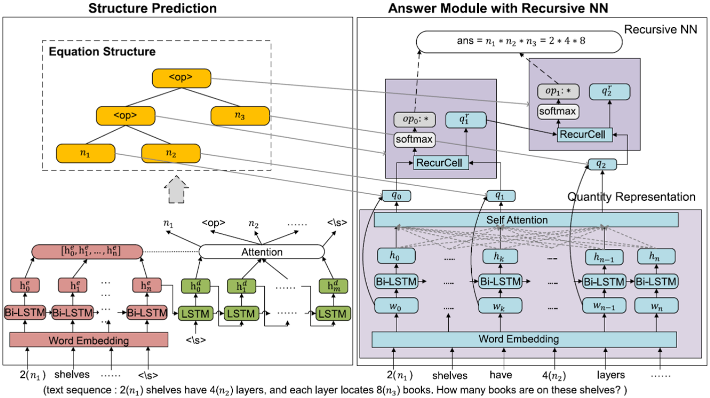
	Acc w/o EN (%)	Acc w/ EN (%)
DNS	58.1	60.7
Bi-LSTM	59.6	66.7
ConvS2S	61.5	64.2
Transformer	59.0	62.3
Ensemble	66.4	**68.4**

Yan Wang et al., Deep neural solver for math word problems. EMNLP 2017
Lei Wang et al., Translating a Math Word Problem to a Expression Tree. EMNLP 2018

HUAWEI NOAH'S ARK LAB

Template-Based Solvers with Recursive Neural Networks

- Coarse-to-fine generation
  - Generate template first: (n_1 <op> n_3) <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

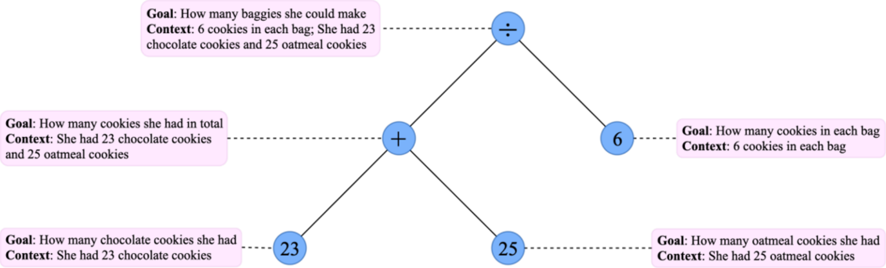
		MAWPS	Math23K
Our Approach	T-RNN	**66.8**	**66.9**
	- EN	63.9	61.1
	- Bi-LSTM	31.1	34.1
	- Self-Att	66.3	65.1

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

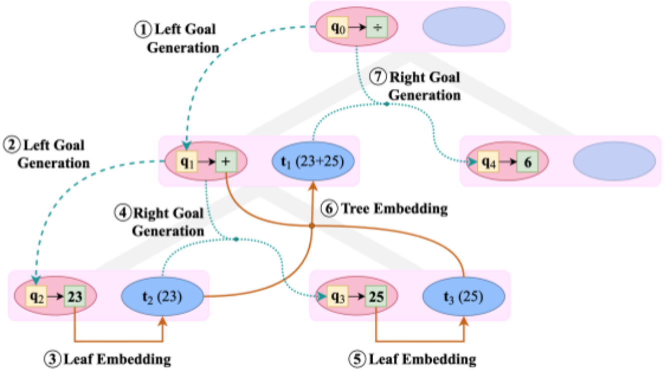
- Mathematical expressions are tree-structured
- Each node has a goal (to solve a sub-question)

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

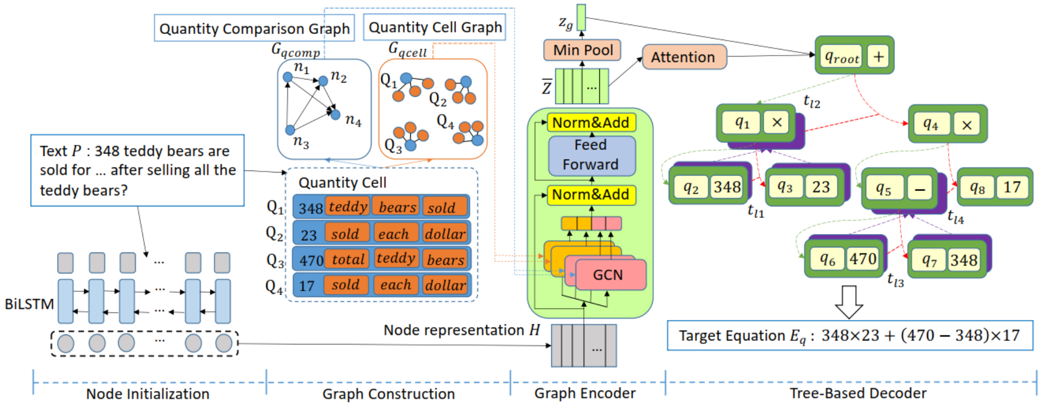


Model	Accuracy(%)
Hybrid model w/ SNI [Wang *et al.*, 2017]	64.7
Ensemble model w/ EN [Wang *et al.*, 2018a]	68.4
GTS model w/o Subtree Embedding	70.0
GTS model	**74.3**

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

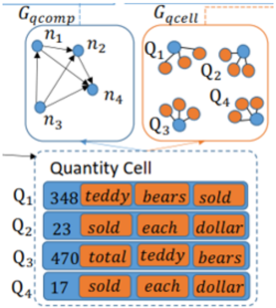


Jipeng Zhang et al., Graph-to-Tree Learning for Solving Math Word Problems. ACL 2020

HUAWEI NOAH'S ARK LAB

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



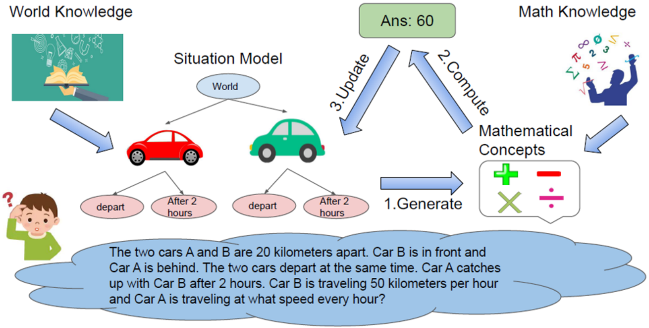
	MAWPS	Math23K	Math23K*
GTS	82.6	75.6	74.3
Graph2Tree	**83.7**	**77.4**	**75.5**

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

HUAWEI NOAH'S ARK LAB

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

HUAWEI NOAH'S ARK LAB

SMART: A Situation Model for Algebra Story Problems via Attributed Grammar

- A situation model is represented as an Attribute Grammar

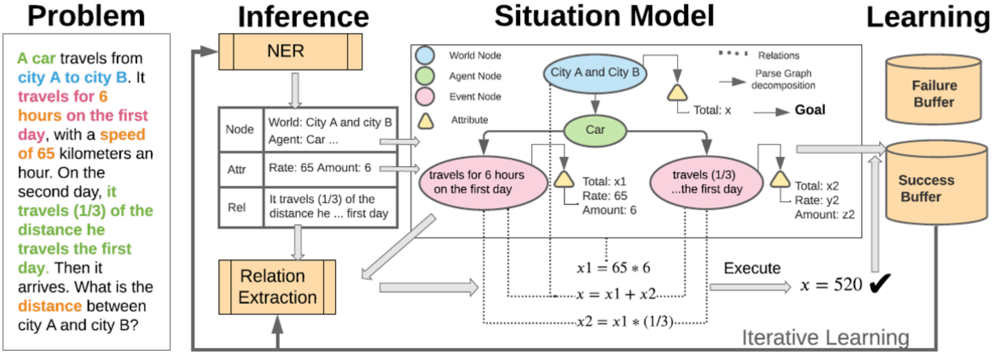
$G = (S, V, A, E, R)$
S is the start symbol.
V = {S, World, Agents, Agent, Events, Event}
A = {rate, amount, total}
E = {e: e is a valid equation on attributes.}
R = {$S \rightarrow$ World
World $\rightarrow$ Agents
Agents $\rightarrow$ Agents Agent | Agent
Agent $\rightarrow$ Events
Events $\rightarrow$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

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

HUAWEI NOAH'S ARK LAB

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

HUAWEI NOAH'S ARK LAB

SMART: A Situation Model for Algebra Story Problems via Attributed Grammar

Model	Overall	Motion	Task	Relation	Price
MathEN	67.8	68.3	70.2	63.3	70.5
GroupATT	67.4	65.2	70.7	63.6	71.5
GTS	76.8	73.2	72.1	76.0	**83.6**
Graph2Tree	76.8	76.9	**79.0**	73.8	78.7
SMART	**79.5**	**79.8**	**79.0**	**77.9**	81.8

Table 5: The answer accuracy on the test set (%).

Model	Overall	Motion	Task	Relation	Price
MathEN	31.7	22.6	28.9	39.9	33.2
GroupATT	35.0	24.0	42.2	42.6	32.7
GTS	45.8	44.5	41.9	49.9	45.3
Graph2Tree	45.1	34.1	47.4	55.1	41.9
SMART	**63.2**	**65.0**	**64.8**	**62.9**	**60.8**

Table 6: The answer accuracy in the OOD evaluation (%). The test set is the 20% longest problems of each type.

Generalize better to longer problems

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

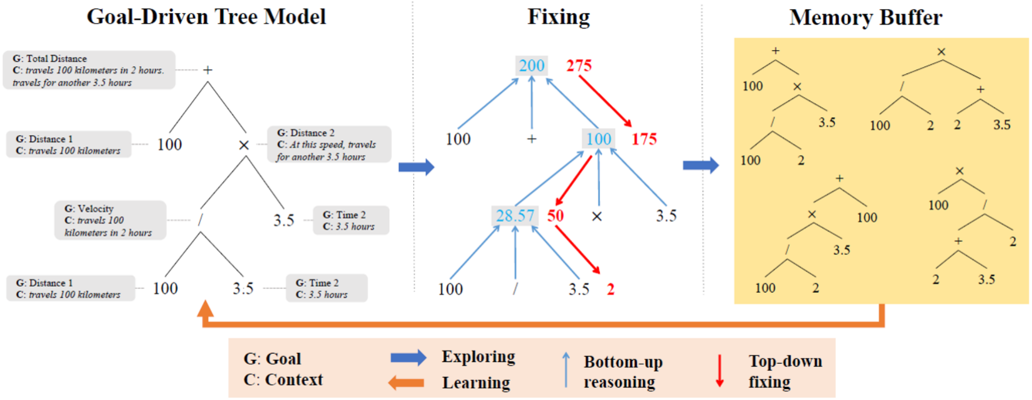
HUAWEI NOAH'S ARK LAB

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

Learning by Fixing: Solving Math Word Problems with Weak Supervision

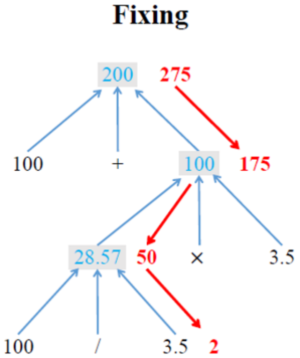


Yining Hong et al., Learning by Fixing: Solving Math Word Problems with Weak Supervision, AAAI2021

HUAWEI NOAH'S ARK LAB

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

	Weakly-Supervised	
Seq2seq	REINFORCE	1.2
	MAPO	10.7
	LBF-w/o-M	44.7
	LBF	43.6
GTS	REINFORCE	15.8
	MAPO	20.8
	LBF-w/o-M	58.3
	LBF	**59.4**

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

HUAWEI NOAH'S ARK LAB

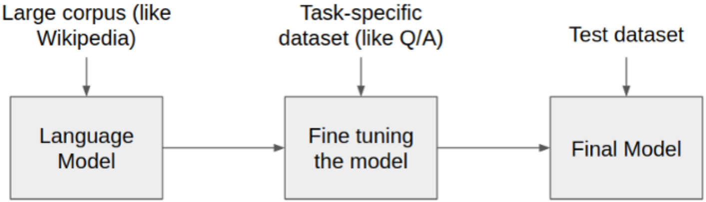
Content

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

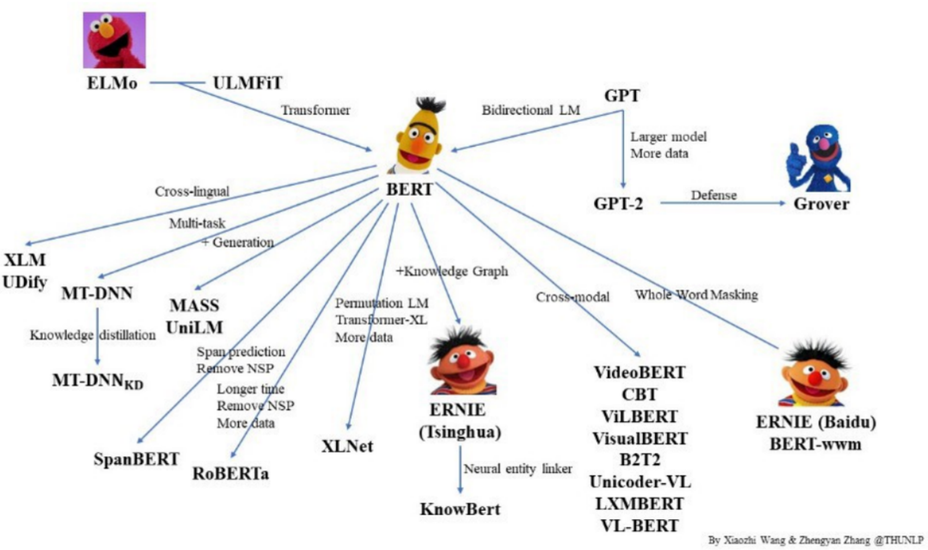
Pre-trained Language Models

- 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



HUAWEI NOAH'S ARK LAB

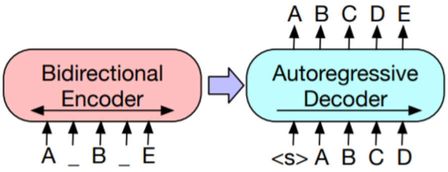
Family of Pretrained Language Models



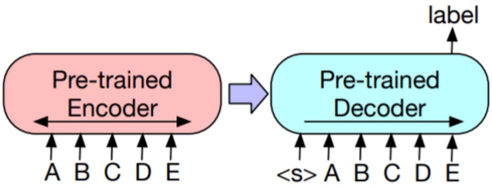
HUAWEI NOAH'S ARK LAB

BART

- Bidirectional and Auto-Regressive Transformers

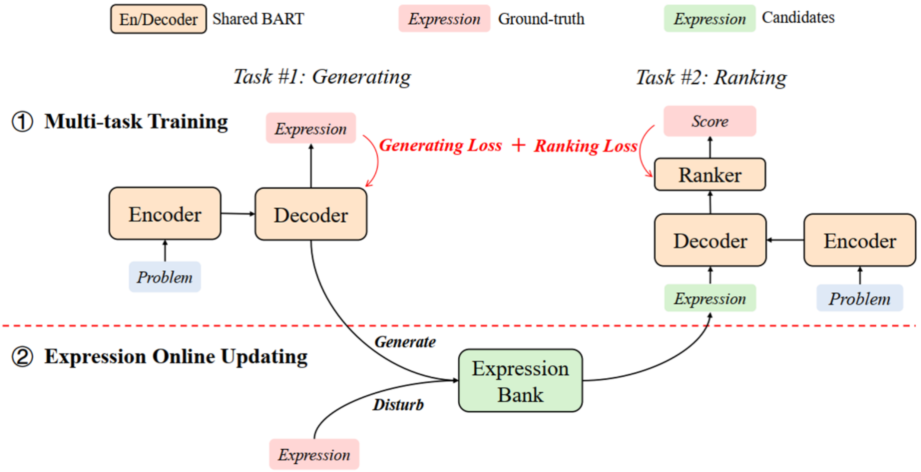


BART pre-training



Finetune BART for text classification

Generate and Rank: A Multi-task Framework for MWPs



HUAWEI NOAH'S ARK LAB

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

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

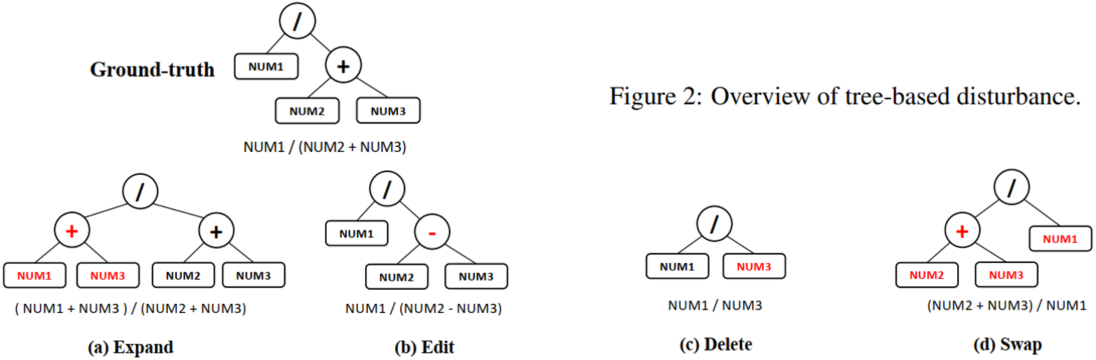


Figure 2: Overview of tree-based disturbance.

HUAWEI NOAH'S ARK LAB

Results

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

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.

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

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

HUAWEI NOAH'S ARK LAB

An Demonstration

小学数学应用题自动求解demo

liuqun 退出登录

题目：

请输入题目

提交

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

示例

History Board

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14页，剩下的25页第3天看完，这本书共有多少页？

求解：(14 + 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

©2021 华为技术有限公司 [语音语义实验室]

http://10.90.91.195:8007/

HUAWEI NOAH'S ARK LAB

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

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.

HUAWEI