Noun-MWP: Math Word Problem Meet Noun Answers

TAEHUN CHA, JAEHEUN JUNG, AND DONGHUN LEE

DEPARTMENT OF MATHEMATICS, KOREA UNIVERSITY

Authors

CHA, Taehun

- Ph.D. student in Mathematics @ Korea University
- Research interest : Natural language processing, AI in finance
- Contact : cth127@korea.ac.kr

JUNG, Jaeheun (정재헌)

- Ph.D student in mathematics @ Korea University
- Research interest: Training algorithm, Geometry & Algebra for neural network, Reinforcement learning
- Contact: wodsos@korea.ac.kr

LEE, Donghun (이동헌)

- Ph.D. in Computer Science @ Princeton University
- M.S. in Computational Biology @ Carnegie Mellon University
- B.A. in Biochemistry @ Columbia University

I have an unusual course of academic life :)

What is the Noun-MWP?

MWP(Math Word Problem)

April's discount flowers was having a sale where each flower was 3 dollars. If Zoe bought 8 roses and 2 daisies how much did she spend?

$$3 \times (8 + 2)$$

- "Mathematical exercise where significant background information on the problem is presented in ordinary language." (Wiki)
- With numerical answer
- Known to be solved with shallow heuristic (Patel et al., 2021)

What is the Noun-MWP?

MWP(Math Word Problem)

April's discount flowers was having a sale where each flower was 3 dollars. If Zoe bought 8 roses and 2 daisies how much did she spend?

$$3 \times (8 + 2)$$

Extractive QA

In 1517, the seventeen-year-old King sailed to Castile. There, his Fle mish court . . . . In May 1518, Charles traveled to Barcelona in Arag on. Where did Charles travel to first, Castile or Barcelona?

$$argmin\{Castile: 1517, Barcelona: 1518\}$$

- Extracting the most relevant text span from a text given a question
- DROP dataset (Dua et al., 2019) requires simple mathematical reasoning like comparison between numbers.

What is the Noun-MWP?

MWP(Math Word Problem)

April's discount flowers was having a sale where each flower was 3 dollars. If Zoe bought 8 roses and 2 daisies how much did she spend?

$$3 \times (8 + 2)$$

Extractive QA

In 1517, the seventeen-year-old King sailed to Castile. There, his Flemish court . . . . In May 1518, Charles traveled to Barcelona in Aragon. Where did Charles travel to first, Castile or Barcelona?

$$argmin\{Castile: 1517, Barcelona: 1518\}$$

Noun-MWP(Ours)

Sihyeon collected 64 stamps, and Soma collected 1 fewer stamps than Sihyeon. Junwoo collected 7 bundles of 10 stamps each. Who among the three has collected the fewest stamps?

$$argmin\{Sihyeon: 64, Soma: 64 - 1, Junwoo: 7 \times 10\}$$

- Noun-MWP is a class of MWPs whose answer is a noun-substring of the input.
- It can be seen as extractive QA problems requiring mathematical reasoning.

Dataset

- Collected 604 problems from elementary school level textbooks and TUNiB dataset (Keum et al., 2022)
- Labeled expressions manually
  - Sanity check by applying rule-based solver to expressions and match it with answers from the source textbook.)
- Constructed 5 folds as conventional MWP dat aset did.
- Part of our dataset can be found from:

https://github.com/invigorator96/NounMWP

Ratio of Problems by Required Operations				
$+$	$-$	$\times$	$\div$	Simple Assignment
20.9%	8.4%	18.4%	9.3%	55.8%

Ratio of Problems by Expression Length				
≤ 8	$9 \sim 10$	$11 \sim 12$	$13 \sim 14$	≥ 15
46.5%	39.4%	9.6%	4.1%	0.3%

Solving Noun-MWP

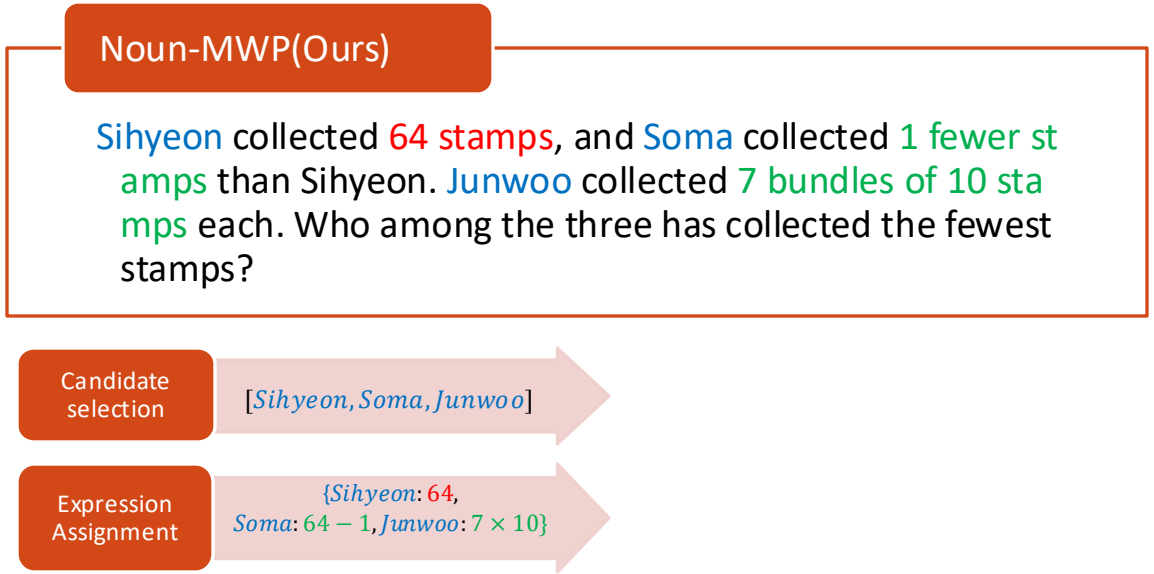
Noun-MWP(Ours)

Sihyeon collected 64 stamps, and Soma collected 1 fewer stamps than Sihyeon. Junwoo collected 7 bundles of 10 stamps each. Who among the three has collected the fewest stamps?

Candidate selection → $[Sihyeon, Soma, Junwoo]$

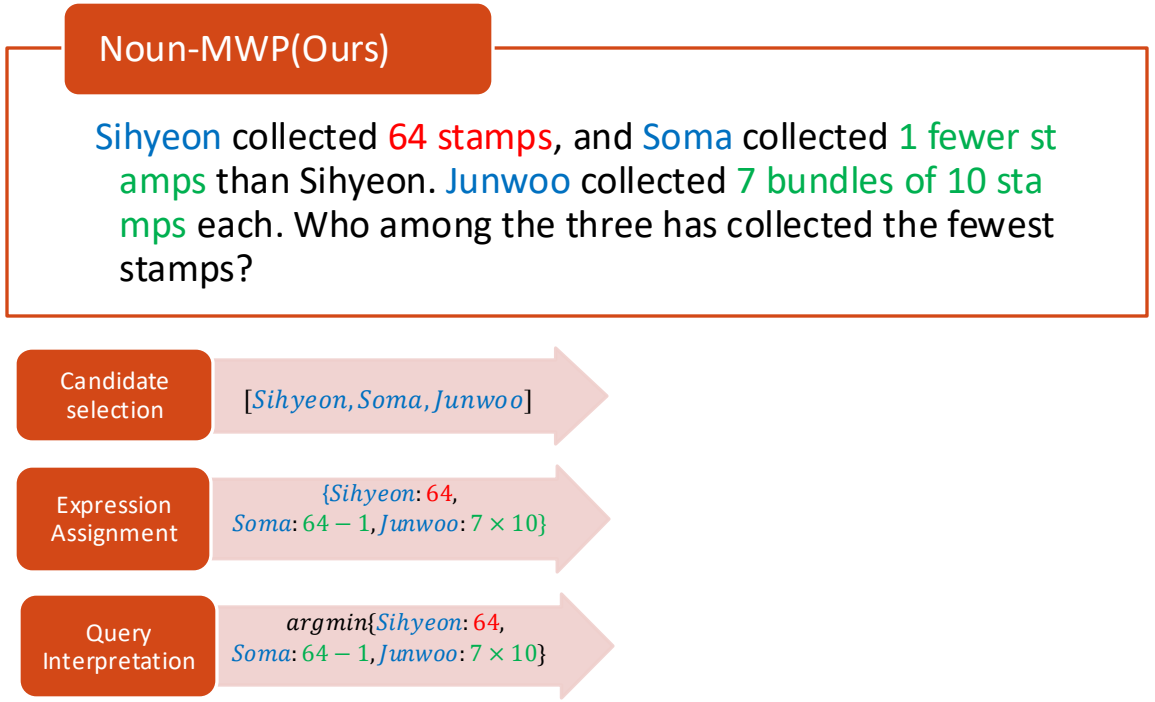
- Extract noun candidates from input text
- This step reduces the search space of answer from whole input text to a few nouns.

Solving Noun-MWP



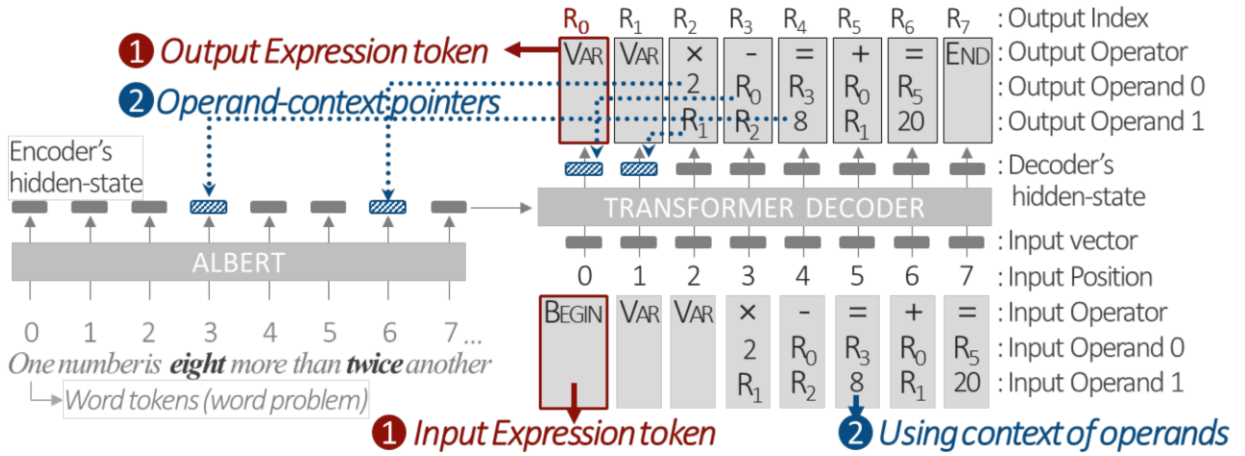
- Generate numerical values and assign them to candidate nouns.
- This step contains arithmetic operations as classical MWP does.

Solving Noun-MWP



- Interpret query sentence
- (e.g. *Who among the three has collected the fewest stamps?*)
- Find the candidate satisfying the query.

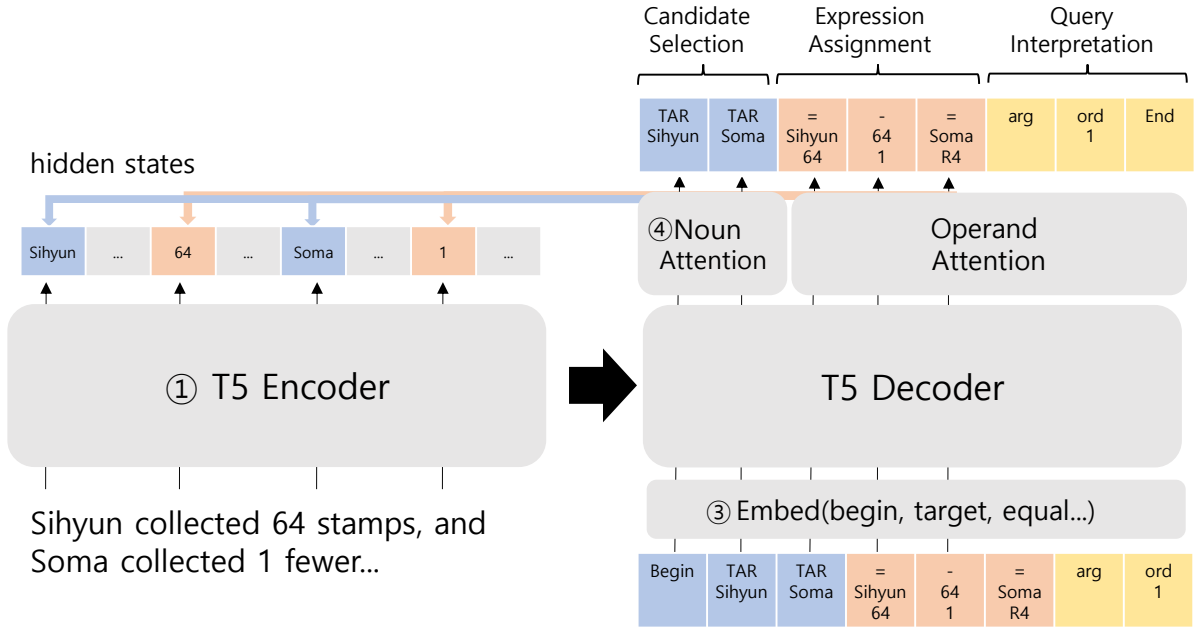
Model: N-EPT



Original EPT for classical MWPs

- Use ALBERT as a backbone
- Predict operator(+, -, *, /) and operand triples
- Predict operand in input sentence with operand attention module

Model: N-EPT



- Use T5 as a backbone
- Add operators(arg, ord, find) to handle Noun-MWP
- Predict candidate noun in input sentence with noun attention module
- Use pretrained embedding to embed operator/operand tokens.

Model: N-EPT

Additional operators

- **TAR**(target): In contrast to special token VAR(variable) in original EPT, it's an unary operator paired with a candidate noun in the input string.
- **arg**: Special command to generate list of (noun, value) pairs after the expression assignment.
- **ord**(order): Binary list operators choosing the n-th smallest element of the list.
- **find**: Binary list operators choosing an element indexed by n.

Model: N-EPT

- Problem: "Arin drank 3/7 of the whole bottle of milk, and Eunhye drank the rest. Which of the two drank less milk?"
- Expression:

Label	Expression	
R_0	$[TAR, Arin]$	Candidate selection
R_1	$[TAR, Eunhye]$	
R_2	$[=, R_0, 3/7]$	Expression assignment
R_3	$[-, 1, R_0]$	
R_4	$[=, R_1, R_3]$	
R_5	$[arg]$	Query interpretation
R_6	$[ord, R_5, 1]$	
R_7	$[END]$	

- 'TAR' assigns candidate nouns.
- 'arg' generate a list of 'TAR's.
- 'ord' select the smallest element of the generated list.

Results: Answer accuracy

	N-EPT	KE-T5	KLUE-RoBERTa
Small	81.67%	–%	45.67%
(Std.Err)	(2.11%)	(-%)	(4.52%)
Base	**84.33%**	6.80%	46.17%
(Std.Err)	(2.44%)	(9.58%)	(5.21%)
Large	82.50%	28.17%	47.67%
(Std.Err)	(0.75%)	(2.44%)	(3.22%)

- Evaluation : 5-fold cross-validation
- N-EPT achieves over 80% answer accuracy
  - With every encoder-decoder size.
- Comparison: SOTA QA models
  - Fine-tuned to our dataset
  - Fails to solve the Noun-MWP well
  - Note: If candidate selection is correct, the expected accuracy for random choice is 44.87%

Results: Error cases

	count	percent
Candidate Selection	55	58.5%
Expression Assignment	22	23.4%
Query Interpretation	17	18.1%
Total	94	100.0%

Problem(EN)
There are 53 Chinese students and 60 American students in international children's schools. When 10 students are gathered in a group, 5 Japanese students are left in 5 groups, and 9 Korean students are left in 4 groups. Write down which country has the second largest number of students.

True expressions

Candidate selection:	X_0:Chinese, X_1: American, X_2: Japanese, X_3: Korean
expression assignment:	$X_0 = 53$, $X_1 = 60$, $X_2 = 5 \times 10 + 5$, $X_3 = 4 \times 10 + 9$
Query interpretation:	(arg),(ord,R_i,-2)
Answer:	Japanese

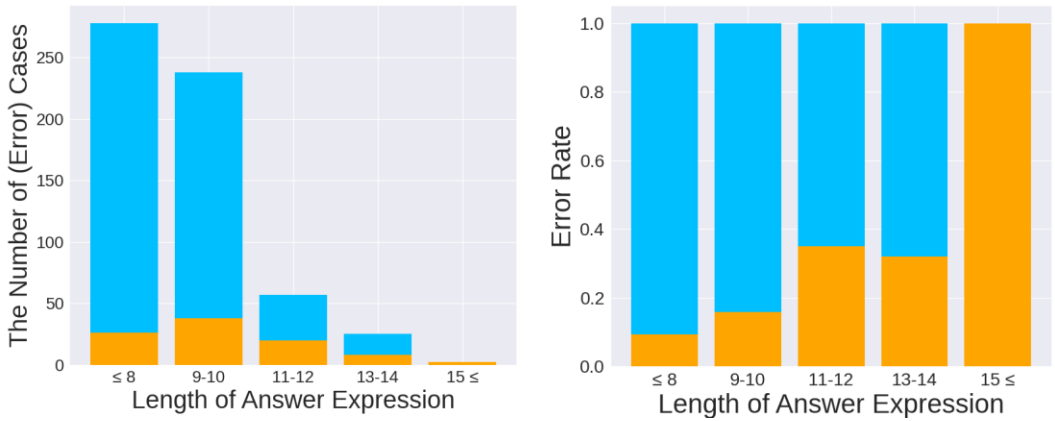
Predicted expressions

Candidate selection:	X_0:Chinese, X_1: American, X_2: Korean
expression assignment:	$X_0 = 53$, $X_1 = 60$, $X_2 = 9$
Query interpretation:	(arg),(ord,R_i,2)
Answer:	Chinese

- Total 94 error cases collected
  - Over all 5-folds
- More than half of error cases were from the candidate selection step
- The model failed to find candidate **Japanese**, successively failed to generate expressions an d query.

Discussion

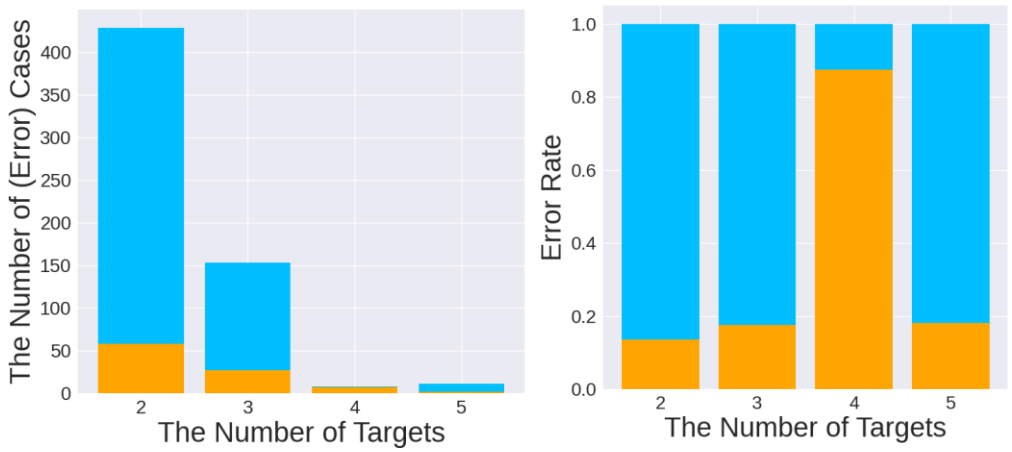
- Impact of Expression Length



- The longer, the harder. (Wu et al., 2020)
- Maybe because of the sparsity of data with long expression in our dataset.

Discussion

- Impact of the Number of Candidates



- The more, the harder.
- Also because of the sparsity of data with many candidates in our dataset.
- Five target case is consisted with 'find' question, which is relatively easy.

Conclusion

- We propose a novel task, Noun-MWP, math word problem with noun answers.
  - It requires understanding of relationship between a noun entity and an arithmetic operation.
  - We construct Korean Noun-MWP dataset consists of 604 problems.
- We propose systematic method to solve Noun-MWP
  - consists of candidate selection, expression assignment, and query interpretation
- and a model, N-EPT by modifying the EPT.
- We evaluated it on custom Noun-MWP dataset and achieved better performance than previous extractive and generative QA models.

Reference

- Are NLP models really able to solve simple math word problems?, Patel et al., NAACL 2021
- DROP: A reading comprehension benchmark requiring discrete reasoning over paragraphs, Dua et al., ACL 2019
- KMWP, korean math word problems, Keum et al., 2022
- Point to the Expression: Solving Algebraic Word Problems using the Expression-Pointer Transformer Model, Kim et al., EMNLP 2020
- A knowledge-aware sequence-to-tree network for math word problem solving, Wu et al., EMNLP 2020