

## Noun-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?  $\rightarrow 3 \times (8 + 2)$

### 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?  $\rightarrow \text{argmin}\{\text{Castile: } 1517, \text{Barcelona: } 1518\}$

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

## Dataset

### Korean Noun-MWP dataset

- 604 problems were collected from textbooks
- Manually labeled
- Construct 5 folds

### Problem complexity:

#### Ratio of Problems by Required Operations

+	-	×	÷	Simple Assignment
20.9%	8.4%	18.4%	9.3%	55.8%

#### Ratio of Problems by Expression Length

≤ 8	9 ~ 10	11 ~ 12	13 ~ 14	≥ 15
46.5%	39.4%	9.6%	4.1%	0.3%

## Method

### Candidate selection

[Sihyeon, Soma, Junwoo]

- Extract candidates from the context

### Expression Assignment

{Sihyeon: 64, Soma: 64 - 1, Junwoo: 7 × 10}

- Generate math expressions
- Get correspondence between candidates ⇔ answers

### Query Interpretation

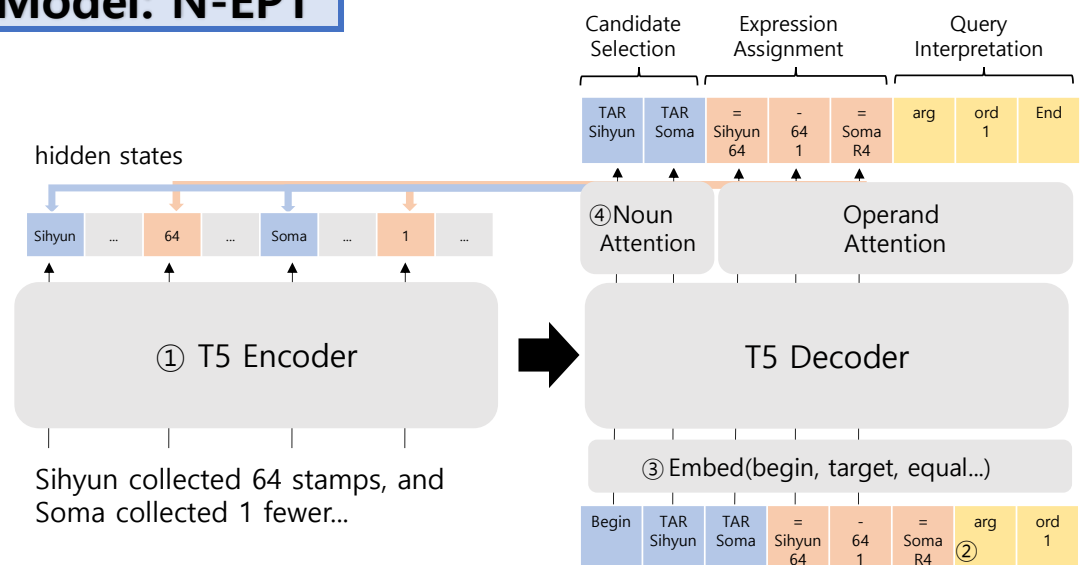
$\text{argmin}\{\text{Sihyeon: } 64, \text{Soma: } 64 - 1, \text{Junwoo: } 7 \times 10\}$

- Get a representation of query
- Find the most suitable answer candidate

### Implementation

- N-EPT: Modified from MWP solver, EPT

## Model: N-EPT



### Key changes from EPT

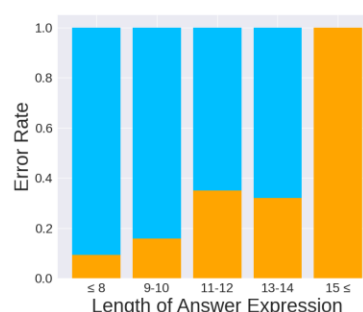
- ① ALBERT to T5
- ② Additional operators
- ③ Pretrained embedding
- ④ Add noun attention module

## Results

	N-EPT	KE-T5	KLUE-RoBERTa
Small	81.67%	—%	45.67%
(Std.Err)	(2.11%)	(-%)	(4.52%)
Base	<b>84.33%</b>	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%)

- Extractive QA model cannot solve Noun-MWP well
- Generative model (T5) struggles more

## Discussion



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

- Hard to solve long expression problems
- Errors usually happen in the candidate selection step

## Conclusion

We propose

- New task, Noun-MWP, math word problem with noun answers
- New dataset, Korean Noun-MWP, consists of 604 problems
- New solver, N-EPT by modifying EPT, and evaluated on our Noun-MWP dataset