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

Keyword Filtering

Keyword filtering is used to remove the noise in the K-means clustering. Take the example of Motion problems. We first identify the cluster that includes all motion problems and some other noise problems, and then compute the word frequencies for this cluster. We select the 50 most frequent words found by K-means and manually filter them to construct the keyword list, such as "hours, city, travel, car, kilometers, speed", etc. The problems that don't include any one of the keywords are deemed not to belong to this type and thus be discarded. The Motion cluster has 1732 problems before keyword filtering and 28 problems are removed, and we manually validate that the removed problems indeed don't belong to the Motion type. Similarly, we double-check the filtered problems for other problem types as well.

Dataset Distribution

Here we show some dataset distributions on problem length about each problem types in Figure 2. We also show the world cloud for each of the problem types in Figure 1.

Relation Extraction Details

Predicate	Definition
Equal ($F(v_i)$, $F(v_j)$)	$F(v_i) = F(v_j)$
More_than($F(v_i)$, $F(v_j)$, n)	$F(v_i) = F(v_j) + n$
Less_than($F(v_i)$, $F(v_j)$, n)	$F(v_i) = F(v_j) - n$
Times_of($F(v_i)$, $F(v_j)$, n)	$F(v_i) = F(v_j) * n$

Table 1: Definitions of Predicates

In Table 1, we provide a detailed definition of the functions. In Table 2, we provide an example of relations represented by predicates in first-order logic.

Named Entity Recognition Model

We adapt the **BIOS** labeling method to pre-process our dataset. We use the bert-Chinese-base[1] pre-trained model. We run the model for 4 epochs. The learning rate is set as 3e-5. The length is 128 and batch size is 32. The model is trained on 1 NVIDIA 2080Ti GPU.

[1]https://github.com/google-research/bert/blob/master/multilingual.md

Seq2Seq Model

For Seq2Seq, we adopt the Seq2Seq model in (Wang et al. 2018) but without equation normalization. However, rather than take the whole problem as input and output a single equation that represents the relations of the whole problem, we encode the relations of a problem separately. We concatenate the nodes and attributes to the relation text. Table 3 shows some examples of the Seq2Seq input and output. Specifically, we use two-layer Bi-LSTM cells with 512 hidden units as decoder. In addition, we use Adam optimizer with learning rate 1e-3, $\beta_1 = 0.9$, and $\beta_2 = 0.99$. We run for 100 epochs, the batch size is 64 and the dropout rate is 0.5.

Parsing Constraints

List of dependency parsing soft constraints:

1. An attribute marked OBJ / DEP / NUMMOD / NMOD:TMOD (or is a child of a unit marked OBJ/DEP) of a tree with the agent node marked NSUBJ.
2. An attribute marked OBJ / DEP / NUMMOD / NMOD:TMOD (or is a child of a unit marked OBJ/DEP) of a tree with the verb of the event node being the root (except for verbs like "is", "has"...).
3. An event node is a sub-tree of which an agent word is the root.
4. An agent node is placed in the sub-tree of which the verb of the event node is the root.
5. An agent node is before a RELCL word and is not after a RELCL word.
6. An agent node is NMOD/POSS of another noun word.

References

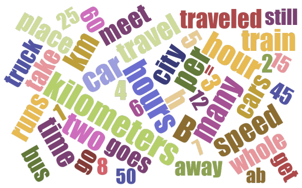
Wang, L.; Wang, Y.; Cai, D.; Zhang, D.; and Liu, X. 2018. Translating Math Word Problem to Expression Tree. In *EMNLP*.

Problem	Nodes&Attributes	Predicates
Ming reads a story book. The first day he reads (1/6) of the whole book. The second day he reads 24 pages. And the third day he reads 150% of the total number of page read in the previous two days. There's (1/4) left of the whole book, so how many pages does this book have?	{{world: {a story book (v_w): {quantity: x}}, agent (v_a): {Ming}, event: {The first day he reads (1/6) of the whole book (v_{e1}): {quantity: x_1}}, {The second day he reads 24 pages (v_{e2}): {quantity: 24}}, {the third day he reads 150% of the total number of page read in the previous two days (v_{e3}): {quantity: x_3}}} }}	Times_of(Quantity(v_{e1}), Quantity(v_w), (1/6)), Times_of(Quantity(v_{e3}), Sum(Quantity(v_{e1}), Quantity(v_{e2})), 150%), Times_of(Left(Quantity(v_w), Quantity(v_{e1}), Quantity(v_{e2}), Quantity(v_{e3}), (1/4)))

Table 2: An example of predicates

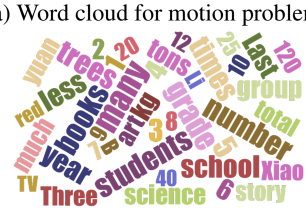
Problem	Seq2Seq Input	Seq2Seq Output
Ming reads a story book. The first day he reads (1/6) of the whole book. The second day he reads 24 pages. And the third day he reads 150% of the total number of page read in the previous two days. There's (1/4) left of the whole book, so how many pages does this book have?	The first day he reads $(1/6)$ of the whole book {{world: {a story book: {quantity: x}}, agent: {Ming}, event: {The first day...the whole book: {quantity: x_1}}}}	$x_1 = x * (1/6)$
	the third day he reads 150% of the total number of pages read in the previous two days.{{world: ..., agent: .., event: {The first day: {quantity: x_1}}, {The second day: {quantity: 24}}, {The third day: {quantity: x_3}}}}	$x_3 = (x_1 + 24) * 150\%$
	There's (1/4) left of the whole book {{world: ..., agent: .., event: {The first day: ..., {The second day: ..., {The third day: ...}}}	$x - (x_1 + 24 + x_3) = x * (1/4)$
A car drives from City A to City B, traveling 217 kilometers in 3.5 hours in the morning. At this speed, it will take another 4.5 hours to reach City B. How many kilometers is the distance between City A and City B?	At this speed {{world: City A and City B{quantity: x}}, agent: car, event: {traveling 217 kilometers in 3.5 hours in the morning: {quantity: 217, unit: 3.5, speed: y_1}}, {it will take another 4.5 hours to reach City B.: {quantity:x_2 unit: 4.5, speed:y_2}} }}	$y_1 = y_2$
	reach city B {{... event: {traveling ... morning: {quantity: 217, unit: 3.5, speed: y_1}}, {it will ... City B.: {quantity:x_2, unit: 4.5, speed:y_2}} }}	$x = 217 + x_2$

Table 3: Examples of Seq2Seq input and output

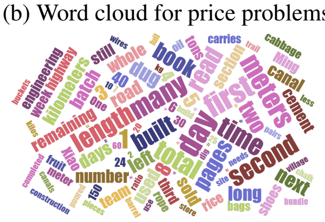


(a) Word cloud for motion problems

(b) Word cloud for price problems

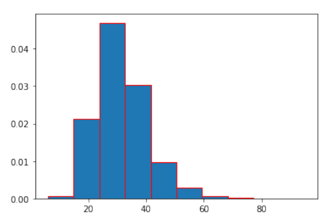


(c) Word cloud for relation problems

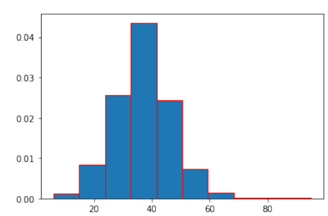


(d) Word cloud for task completion problems

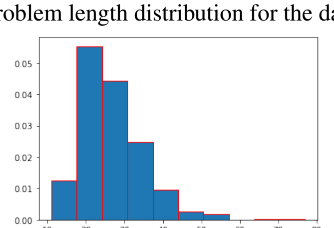
Figure 1: Word cloud figures



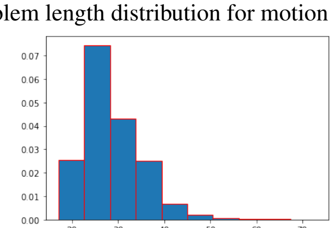
(a) problem length distribution for the dataset



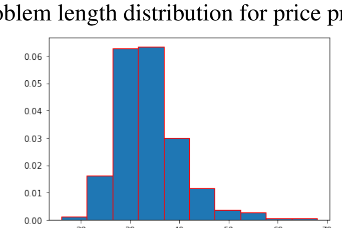
(b) problem length distribution for motion problems



(c) problem length distribution for price problems



(d) problem length distribution for relation problems



(e) problem length distribution for task completion problems

Figure 2: Problem length