

A APPENDIX

A.1 TRAINING AND TESTING

As preprocessing, we obtain non-lexical text by replacing the numbers in the question text with the pre-defined special tokens in our symbolic equation graph and commonsense knowledge graph. The procedures are similar to what Table 7 depicts with the following differences:

- Matching words for unknown variables are first extracted from the gold MWP, and query our private database with the given topic word to construct our commonsense knowledge graph.
- *MaKE* transforms operators η_* into equation graph edge labels (relations), and numeric number φ_* into equation graph nodes v .
- More words in MWP text are replaced with special tokens in commonsense knowledge graph, take the sentence in Table 7 as an example, *wheels* are replaced by one node (*counting entity*) in the corresponding commonsense graph.

During training *MaKE*, the input is commonsense knowledge graph and the equation based symbolic graph, the output target is the delexicalized words sequence. We apply the same word refilling post-processing procedure as described in CVAE method to obtain the final MWP.

A.2 IMPLEMENTATION DETAILS

More implementation details are provided here. We tokenize our training data with BPE method (Shibata et al., 1999) and extend the subword vocabulary with our pre-defined special tokens. We initialize the GGNN parameters with normal distribution $N(0, 0.02)$ and the number of GNN hops is set to 3. The dimension of word embedding is 128 with randomly initialized. We utilize GRU for all RNNs, the hidden state size is 512. The size of sampled latent variable is set to 128, and we apply reparameterization trick during training and inference. We set the teacher forcing probability to 0.5 and train our model using Adam optimizer (Kingma & Ba, 2014) with learning rate scheduling. The batch size is set to 32, the beam search width is set to 5. All hyper-parameters are tuned on the development set.

A.3 BASELINE METHODS DETAILS

Template In addition to neural baselines, we use a problem-specific, template-based generator. The template-based method first finds MWP problem with the same type of input equations in the question bank given the input topic words. For instance, the query equation is $x + y = 6$; $2x - 4y = 6$ and the query topic is *vehicle*. As shown in Table 6, we first delexicalize the input equation pairs with special tokens and save them for post-processing. After query our question bank with the delexicalized equation pairs and the topic word, we obtain the pre-stored MWP template and matching words for unknown variables. Finally we fill the MWP template with the previously saved delexicalized words and obtain the generated MWP.

Table 6: Query process in our template-based-method.

Equation templates: $x+y=\alpha$; $bx-cy=d$	Topic: vehicle
Query text: There are α x_entity and y_entity in the parking lot. Each x_entity has b wheels and each y_entity has c wheels. x_entity has d more total wheels than y_entity . How many x_entity are there?	
Query variables: x_entity : motorcycles, y_entity : cars	
Generated MWP There are 6 motorcycles and cars in the parking lot. Each motorcycles has 2 wheels and each cars has 4 wheels. Motorcycles has 6 more total wheels than cars. How many motorcycles are there?	

CVAE (Zhao et al., 2017) Similar to previous work, we apply a sequence-to-sequence(Seq2seq) generation model and adopt a latent variable to capture the diversity of MWPs. We replace the

hierarchical encoder with a one-layer GRU, and the initial state of the decoder is the combination of a latent variable and the final state of the encoder. As shown in Table 7, we apply the delexicalization process and sequence transformation operations for all the training data, the input sequence includes special tokens, operators and the topic word. After refilling the special tokens with corresponding matching words, we obtain the final MWP.

Table 7: Input sequence for seq2seq method, φ_* are numeric number in equations, and η_* are operators. If there is no valid operator or number for a given special token, we fill it with a pad token.

Input equation: $x+y=6; 2x-4y=6$	Topic: vehicle
General expression formula:	$\eta_0\varphi_0x\eta_1\varphi_1y = \varphi_2\eta_2\varphi_3$ $\eta_3\varphi_4x\eta_4\varphi_5y = \varphi_6\eta_5\varphi_7$
Input sequence expression:	$[\eta_0, \varphi_0, \eta_1, \varphi_1, \varphi_2, \eta_2, \varphi_3, \eta_3, \varphi_4, \eta_4, \varphi_5, \varphi_6, \eta_5, \varphi_7, \text{Topic}]$
Input sequence for given example:	$[\text{pad}, \varphi_0, +, \varphi_1, \varphi_2, \text{pad}, \text{pad}, \text{pad}, \varphi_4, -, \varphi_5, \varphi_6,$ $\text{pad}, \text{pad}, x_entity, y_entity, \text{vehicle}]$
Output MWP: There are φ_2 x_entity and y_entity in the parking lot. Each x_entity has φ_4 wheels and each y_entity has φ_5 wheels. x_entity has φ_6 more total wheels than y_entity . How many x_entity are there?	

UniLM (Dong et al., 2019) A pre-trained natural language generation model with transformer encoder and decoder blocks. We fine-tune UNILM on MWP generation task with the same input and output token sequence described in CVAE method.

Transformer (Vaswani et al., 2017) We included a Transformer-based seq2seq model which has proved his success in machine translation tasks. The same input sequence as described in previous method.

MaKE w/o CSKG In order to verify the effectiveness of introducing commonsense knowledge in our task, we generates math problems only given the input of equations. The framework follows Graph2seq learning while get rid of the commonsense knowledge graph.

A.4 ADDITIONAL MWP GENERATION COMPARISON

We provide additional illustrative examples of the MWP generation comparison with unseen equations in Table 8.

A.5 ADDITIONAL DIVERSE MWP RESULTS

We provide additional diverse MWP results in Table 9.

Table 8: Illustrative examples of the MWP generation comparison with unseen equations. () represents the question that the student needs to solve.

Equations: $x=2y$; $4x+6y=56$; Topic: Vehicle	
<i>CVAE</i>	There are many small cars and big cars in the parking lot. There are 6 people in these cars and 56 people in big cars. How many small cars and big cars are there?
<i>UniLM</i>	Doctors have produced a lot of small cars and large cars. There are 0 cars in total. These two types of cars have 56 people. How many of these two cars have a total of 56 people, small cars?
<i>Transformer</i>	There are 56 people in the class to go rowing, and there are 0 cars in total. Among them, the small cars have 4 people each, the big cars have 6 people each, the small cars have (), and the big cars have ().
<i>MaKE</i>	The small car that can carry 4 people and the large car that can carry 6 people, there are 56 people in total, and the number of small cars is twice the number of large cars. Q: There are () large cars.
Equations: $-x+y=20$; $2x-4y=10$; Topic: Poultry	
<i>CVAE</i>	Chickens and rabbits in the same cage, the number of chickens is twice the number of chickens, the number of rabbits and rabbits is twice the number of chickens, and the total number of legs is 10, how many chickens and rabbits are there each?
<i>UniLM</i>	How many chickens and rabbits are there in the 10-legged training centre?
<i>Transformer</i>	There were 20 chickens and rabbits, the total number of legs of the chickens was 10 more than the rabbits, and the chickens had () only.
<i>MaKE</i>	The rabbit has 20 more heads than the chicken and the chicken has 10 more feet than the rabbit. How many chickens and rabbits are there?

Table 9: Additional illustrative example of the diverse MWP generation made by *MaKE*. () represents the question that the student needs to solve.

Equations: $x+y=100$; $20x+30y=2600$; Topic: Buy ticket	
1.	There are two kinds of tickets for a total of 100, one is 20 yuan, the other is 30 yuan, and the total value is 2,600 yuan. How many 20 yuan tickets are there?
2.	There are two kinds of ticket for a football game, one is priced at 30 yuan, and the other is priced at 20 yuan. Zhang Hua bought 100 tickets, which cost him 2600 yuan. He bought () tickets for 30 yuan, bought () tickets for 20 yuan.
3.	Xiao Wang bought 100 tickets for the tour group, some of which were 20 yuan, the other was 30 yuan, and the total ticket price was 2600 yuan. How many tickets were bought for each of the two types.
4.	Our class plan to by 100 tickets, one of which is 20 yuan, the other is 30 yuan, and The total fee is 2,600 yuan. How many tickets should be purchased for each of the two types.
Equations: $-y+x=20$; $2y+4x=66$; Topic: Poultry	
1.	In one farm, there are 20 more cows than ducks, with 66 legs in total, () cows and () ducks.?
2.	Xiaohong's farm has ducks and cows. The number of ducks is 20 less than the number of cows. There are 66 legs in total, Xiaohong have () cows and () ducks.
3.	In a pasture, there are cows and ducks. There are 20 more cows than ducks, and the total number of legs of the cows and ducks is 66. There are () cows?