

# Neural-Symbolic Solver for Math Word Problems with Auxiliary Tasks

Jinghui Qin<sup>1</sup>, Xiaodan Liang<sup>1,2</sup>, Yining Hong<sup>3</sup>, Jianheng Tang<sup>1</sup> and Liang Lin<sup>1,2\*</sup>

<sup>1</sup> Sun Yat-sen University

<sup>2</sup> Dark Matter AI Inc.

<sup>3</sup> University of California, Los Angeles

qinjinghui@mail2.sysu.edu.cn, {xdliang328, sqrt3tjh}@gmail.com,  
yininghong@cs.ucla.edu, linliang@ieee.org

## Abstract

Previous math word problem solvers following the encoder-decoder paradigm fail to explicitly incorporate essential math symbolic constraints, leading to unexplainable and unreasonable predictions. Herein, we propose Neural-Symbolic Solver (NS-Solver) to explicitly and seamlessly incorporate different levels of symbolic constraints by auxiliary tasks. Our NS-Solver consists of a problem reader to encode problems, a programmer to generate symbolic equations, and a symbolic executor to obtain answers. Along with target expression supervision, our solver is also optimized via 4 new auxiliary objectives to enforce different symbolic reasoning: a) self-supervised number prediction task predicting both number quantity and number locations; b) commonsense constant prediction task predicting what prior knowledge (e.g. how many legs a chicken has) is required; c) program consistency checker computing the semantic loss between predicted equation and target equation to ensure reasonable equation mapping; d) duality exploiting task exploiting the quasi duality between symbolic equation generation and problem’s part-of-speech generation to enhance the understanding ability of a solver. Besides, to provide a more realistic and challenging benchmark for developing a universal and scalable solver, we also construct a new large-scale MWP benchmark CM17K consisting of 4 kinds of MWPs (arithmetic, one-unknown linear, one-unknown non-linear, equation set) with more than 17K samples. Extensive experiments on Math23K and our CM17k demonstrate the superiority of our NS-Solver compared to state-of-the-art methods<sup>1</sup>.

## 1 Introduction

Deep neural networks have achieved remarkable successes in natural language processing recently. Although neural models have demonstrated performance superior to humans on some tasks, e.g. reading comprehension (Rajpurkar et al., 2016; Devlin et al., 2019; Lan et al.), it still lacks the ability of discrete reasoning, resulting in low accuracy on math reasoning. Thus, it is hard for pure neural network approaches to tackle the task of solving math word problems (MWPs), which requires a model to be capable of natural language understanding and discrete reasoning. MWP solving aims to automatically answer a math word problem by understanding the textual description of the problem and reasoning out the underlying answer. A typical MWP is a short story that describes a partial state of the world and poses a question about an unknown quantity or multiple unknown quantities. To solve an MWP, the relevant quantities need to be identified from the text. Furthermore, the correct operators along with their computation order among these quantities need to be determined. Therefore, integrating neural networks with symbolic reasoning is crucial for solving MWPs. Inspired by the recent amazing progress on neural semantic parsing (Liang et al., 2017a) and reading comprehension (Chen et al., 2019), we address this problem by neural-symbolic computing.

Recently, many researchers (Wang et al., 2017; Huang et al., 2018; Wang et al., 2018b, 2019; Xie and Sun, 2019; Chiang and Chen, 2019), inspired by an encoder-decoder framework (Cho et al., 2014), apply neural networks to solve MWPs by learning the mapping function between problems and their corresponding equations, and achieve remarkable successes. The encoder uses a neural network to represent a problem as a real-valued vector, and the decoder uses another neural network to

---

\*Corresponding Author

<sup>1</sup>The code and the new CM17k dataset are available at <https://github.com/QinJinghui/NS-Solver>.

generate an equation or expression token by token. The main difference among previous methods is the way to decode expressions or equations. However, they only follow the encoder-decoder paradigm but lacking the ability to explicitly incorporate essential math symbolic constraints (e.g. commonsense constants, formulation regularization), leading to unexplainable and unreasonable predictions. Besides, most of them only focus on arithmetic MWPs without any unknown, preventing them from generalizing to various types of MWPs, such as equation set problems.

To address the above issues, we propose a novel **Neural-Symbolic Solver** (NS-Solver), which explicitly and seamlessly incorporates different levels of symbolic constraints by auxiliary learning tasks. Our NS-Solver consists of three main components, a *problem reader* to encode the math word problems into vector representations, a *programmer* to generate the symbolic grounded equations, which are executed to produce answers, and a *symbolic executor* to obtain final results. In addition to the supervised training objective between generated symbolic grounded equations and ground-truth equations, our solver is also optimized by four novel auxiliary objectives that enforce four levels of problem understanding and symbolic reasoning. First, we apply **number prediction task** to predict both the number quantity and number location in the problem in a self-supervised manner. Second, we deploy **commonsense constant prediction task** to predict what prior commonsense knowledge (e.g. how many legs a chicken has) is required for our solver. Third, we propose **program consistency checker** to compute the semantic loss between the predicted program and ground-truth equation to ensure reasonable equation mapping. Finally, we also propose a novel **duality exploiting task** that exploits the quasi duality between symbolic grounded equation generation and the problem’s part-of-speech generation to enhance the understanding ability of our solver. There are some key advantages of our solution. First of all, the above four auxiliary tasks can produce additional training signals, which improves the data efficiency in training and makes our solver more robust. Second, using the predicted constant to constrain the target symbolic table can reduce the search space greatly, which means that our solver can generate correct symbolic grounded equations easier and better. Third, the auxiliary tasks have been proven

to help reduce the domain gap between seen and unseen MWPs (Sun et al., 2019, 2020), thus improving the reasoning ability of our solver.

Besides, beyond the current large-scale high-quality MWP benchmark that only includes one type of problems, we also construct a large-scale challenging Chinese MWPs dataset CM17K, which contains 4 types of MWPs (arithmetic MWPs, one-unknown linear MWPs, one-unknown non-linear MWPs, equation set problems) with more than 17K samples, to provide a more realistic and challenging benchmark for developing a universal and scalable math solver. Extensive experiments on public Math23K and our proposed CM17k demonstrate the superiority of our NS-Solver compared to state-of-the-art methods in predicting final results while ensuring intermediate equation rationality.

## 2 Related Work

**Deep learning-based MWP Solvers.** Numerous methods have been proposed to tackle the MWP solving task, ranging from rule-based methods (Bakman, 2007; Yuhui et al., 2010), statistical machine learning methods (Kushman et al., 2014; Zhou et al., 2015; Roy and Roth, 2015, 2016; Mitra and Baral, 2016; Huang et al., 2016; Roy and Roth, 2018), semantic parsing methods (Shi et al., 2015; Koncelkedziorski et al., 2015; Huang et al., 2017; Liang et al., 2018a), to deep learning methods (Ling et al., 2017; Wang et al., 2017, 2018b; Huang et al., 2018; Wang et al., 2018a; Xie and Sun, 2019; Wang et al., 2019; Zhang et al., 2020a,b; Qin et al., 2020; Shen and Jin, 2020; Wu et al., 2020; Chen et al., 2021; Hong et al., 2021a,b). However, most deep learning-based methods only follow the encoder-decoder framework without explicitly incorporating essential math symbolic constraints, resulting in some unexplainable and unreasonable predictions. Besides, most of them only focus on arithmetic MWPs, preventing them from generalizing to various types, such as equation set problems.

**Neural-Symbolic Computing.** Neural-symbolic computing has greatly promoted the development of semantic parsing. Jia and Liang (2016); Dong and Lapata (2016); Zhong et al. (2017) applied neural sequence-to-sequence and sequence-to-tree models to semantic parsing with full supervision. Liang et al. (2017b, 2018b) have advanced the state-of-the-art in weakly supervised semantic parsing on knowledge graphs and tabular databases. Al-

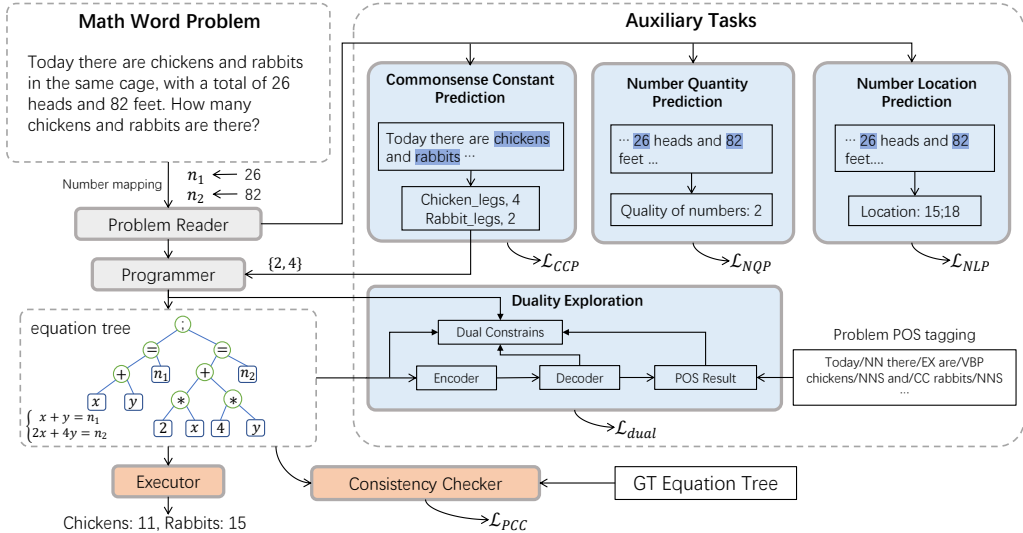


Figure 1: An overview of our NS-Solver. When a problem preprocessed by number mapping and replacement is entered, our problem reader encodes the problem text into context representation. Then our programmer generates a tree-structured symbolic grounded program explicitly. Finally, a symbolic grounded program will be executed to produce answers by the executor. In our NS-Solver, we apply four auxiliary tasks to enhance its problem understanding and symbol reasoning ability for generating better programs.

though most of the successes of semantic parsing are limited to structured data sources, it is not expensive for MWPs since it is easy to crawl lots of problems with annotated equations and answers. Therefore, MWP solving can benefit from supervised neural-symbolic computing.

**Self-Supervised Learning.** Self-supervised auxiliary tasks have been widely used in the fields of natural language understanding (Devlin et al., 2019; Lan et al.). Devlin et al. (2019) applied two self-supervised auxiliary tasks, masked LM and next sentence prediction, to improve the understanding ability of BERT by pretraining. ALBERT (Lan et al.) introduces sentence-order prediction task to address the ineffectiveness of the next sentence prediction task in BERT. Hendrycks et al. (2019) show that self-supervised learning can improve model robustness and uncertainty.

**Dual Learning.** Dual learning, first proposed by He et al. (2016), is a reinforcement training process that jointly trains a primal task and its dual task. Then Xia et al. (2017) considered it as a way of supervised learning and designed a probabilistic regularization term to exploit the duality. It has been widely applied in various fields, such as machine translation (He et al., 2016), sentiment classification (Xia et al., 2017), question answering (Tang et al., 2017), visual question answering (Li et al., 2018), machine reading comprehension (Xiao et al., 2018), and code generation (Wei et al., 2019). To the best of our knowledge, we are the first to ex-

exploit the duality in MWPs. Different from previous works, we design a quasi dual learning method between symbolic grounded equation generation and problem’s part-of-speech generation to enhance the understanding ability by easing the difficulty of generating problems from symbolic equations.

### 3 Neural-Symbolic Solver

In this section, we present the design of the proposed NS-Solver. Its backbone mainly consists of a *problem reader* that encodes the math word problems into vector representations, a *programmer* to generate the symbolic grounded programs in prefix order, and a *symbolic executor* to obtain final results. The overview of our NS-Solver is visualized in Fig. 1. We first introduce the backbone of our NS-Solver in section 3.1, and then we introduce other auxiliary tasks in section 3.2.

#### 3.1 Backbone

**Problem Reader.** Given a problem text  $P = \{x_i\}_{i=1}^n$  processed by number template replacement which maps numeric values in a problem to number templates (e.g., 26 and 82 to  $n_1$  and  $n_2$  in Fig. 1), the problem reader encodes each token  $x_i$  in the problem text into an embedding  $e_i$ . In this work, we deploy a two-layer bidirectional GRU to encode each token  $x_i$  into an embedding  $e_i = \vec{h}_i + \overleftarrow{h}_i$  where  $\vec{h}_i$  and  $\overleftarrow{h}_i$  are from forward and backward GRUs, respectively. Besides, our prob-

lem encoder also outputs a problem representation  $\mathbf{g}_0 = \overrightarrow{\mathbf{h}}_n + \overleftarrow{\mathbf{h}}_0$  as the initial hidden state of our programmer, where  $\overrightarrow{\mathbf{h}}_n$  and  $\overleftarrow{\mathbf{h}}_0$  are the last hidden state of forward and backward GRUs, respectively.

**Programmer.** The programmer takes the output of the problem reader as input and the problem representation as the initial hidden state, and then decodes a problem as a sequence of tokens  $\{y_i\}_{i=1}^m$  which are organized as a prefix equation tree. In this work, we deploy a tree-structured decoder (Xie and Sun, 2019) with attention mechanism (Bahdanau et al., 2015) as the backbone of our programmer and modify them with UET representation (Qin et al., 2020) to support more symbols for multiple types of MWPs. In our programmer, the symbolic table consists of four parts. For each problem, the problem-specific symbolic table contains math operators (+, -, \*, /, ^, =, ;), unknown variable ( $x$  and  $y$ ), a series of commonsense constants (1, 3.14, etc) predicted by the Commonsense Constant Prediction Task in 3.2, and the problem-specific number templates ( $n_1, n_2, n_3$ , etc). It should be noticed that ; is a special operator with the lowest priority to integrate multiple equation trees as an ensemble equation tree, so that equation set problems can be handled as simple as arithmetic problems.

**Executor.** We deploy sympy<sup>2</sup>, which is a python library for symbolic mathematics, as our symbolic executor for obtaining final results by solving generated equations.

### 3.2 The Design of Auxiliary Tasks

The MWP solving task remains challenging since previous methods did not take full advantage of the rich semantics contained in a problem and lacking the ability to explicitly incorporate essential math symbolic constraints. In this section, we introduce four auxiliary learning tasks to exploit additional training signals obtained from different tasks and exploit the result of the commonsense constant prediction task to explicitly constrain the constant symbolic table, which can reduce the search space for symbolic generation and ease the difficulty of generating correct constant.

**Self-supervised Number Prediction (SNP) Tasks.** If a solver can fully understand the problem semantics, it should be able to identify the quantity of numbers in a problem (i.e., to count how many numeric values are in the problem) and

their corresponding locations in the problem text accurately. For example, if the solver can understand the problem in Fig. 1, it should be able to predict there are two numbers(26 and 82) in the problem, and their positions are 15 and 18, respectively. Thus, number quantity prediction and number location prediction are two critical self-supervised tasks to help the problem reader fully understand the problem semantics and measure the ability of problem understanding of a solver. Both two number prediction tasks take the mean of the problem encoder’s outputs  $\{e_i\}_{i=1}^n$  as their input and apply a single-layer feed-forward neural network to compute the distribution of number quantity and number locations. The training objectives of two tasks for each problem are formulated as:

$$\begin{aligned} \mathcal{L}_{NQP} &= - \sum_{i=1}^Q qt_i \log p(q_i|P), \\ \mathcal{L}_{NLP} &= - \sum_{i=1}^L lt_i \log p(l_i|P). \end{aligned} \quad (1)$$

where  $L_{NQP}$  and  $L_{NLP}$  denote the loss for the Number Quantity Prediction (NQP) task and Number Location Prediction (NLP) task, respectively.  $Q$  and  $L$  are the maximum possible quantities of number and maximum possible number locations for a problem at the dataset level.  $qt_i$  and  $lt_i$  represent the ground-truth value on  $i$ -th index of the output probability distribution of NQP and NLP, respectively.

**Commonsense Constant Prediction (CCP) Task.** Commonsense constants are important for solving some MWPs while most previous methods only consider the constants 1 and 3.14, which are not enough for a solver to solve problems that need other commonsense constants. However, attaching a lot of constants to the problem-specific symbolic table will enlarge the search space, increasing the difficulty of generating rational symbolic equations. Therefore, we propose a commonsense constant prediction task to predict what prior commonsense knowledge (e.g. a chicken has 2.0 legs and a rabbit has 4.0 legs for the problem in Fig. 1) is required for the solver to solve a problem according to the problem context. In this way, we can reduce the search space greatly, thus improving the performance of our solver. Similar to the number prediction tasks, the commonsense constant prediction task takes the mean of the problem

<sup>2</sup><https://www.sympy.org/>



encoder’s output  $\{e_i\}_{i=1}^n$  as their input and apply a single-layer feed-forward neural network to compute the distribution of number quantity and number locations. The training objective for each problem is formulated as:

$$\mathcal{L}_{CCP} = - \sum_{i=1}^C ct_j \log p(c_i|P). \quad (2)$$

where  $C$  is the total number of constants in the symbolic table and  $ct_i$  represents the true value on  $i$ -th index of the output probability distribution. Since it is impossible for the commonsense constant prediction task to achieve 100% accuracy, in addition to the predicted constants, we add three extra constants that are not predicted but with the highest probability into the symbolic table, making a better trade-off between the size of the search space and prediction accuracy.

**Program Consistency Checker (PCC).** Although a problem can be solved by multiple equivalent but different equations, the predicted equations should be consistent with label equations as much as possible in the supervised learning setting. Therefore, we propose a program consistency checker to check the symbolic program consistency and regularize the model by computing semantic loss between the predicted symbolic program and ground-truth equation to ensure the reasonable symbolic equation mapping. Let  $\hat{y}_i$  and  $y_i$  represent the predicted symbol and ground-truth symbol,  $p_i$  represents the probability of  $\hat{y}_i$ , the semantic loss is obtained by computing a distance between the predicted distribution and ground-truth distribution as:

$$\mathcal{L}_{PCC} = -\log \sum_i \prod_{\hat{y}_i=y_i} p_i \prod_{\hat{y}_i \neq y_i} (1 - p_i). \quad (3)$$

**Duality Exploiting (DE) Task.** Many previous works (He et al., 2016; Xia et al., 2017; Xiao et al., 2018; Wei et al., 2019) have shown promising results by dual learning framework. Although intuitively, MWP solving and MWP generation are related to each other, i.e., the input of MWP solving is the output of MWP generation, and vice versa, it is very hard for the MWP generation task to generate good enough problems only by the equations without any topic information. Therefore, we propose a duality exploiting task to enhance the understanding ability of our solver by exploiting the quasi duality between symbolic grounded equation generation and the problem’s part-of-speech

generation. Given a pair of a problem and its corresponding equations  $(P, T)$ , and  $P'$  is the part-of-speech of  $P$ <sup>3</sup>, the training objective of the duality exploiting task is formulated as:

$$\mathcal{L}_{dual} = [\log \hat{p}(P') + \log p(T|P) - \log \hat{p}(T) - \log p(P'|T)]^2. \quad (4)$$

where  $\hat{p}(P')$  and  $\hat{p}(T)$  are marginal distributions, which can be modeled by their LSTM (Hochreiter and Schmidhuber, 1997)-based language models, respectively. Besides, we deploy a tree-structure encoder inspired by GTS (Xie and Sun, 2019) to encode equations in prefix for POS generation.

### 3.3 Training Objective

Given the training dataset  $\mathbf{D} = \{(P^1, T^1), (P^2, T^2), \dots, (P^N, T^N)\}$ , where  $T^i$  is the universal expression tree of problem  $P^i$ , we minimize the following loss function for our NS-Solver:

$$\mathcal{L} = \sum_{(P,T) \in \mathbf{D}} [\mathcal{L}_{ent1} + \lambda_1 * \mathcal{L}_{dual} + \lambda_2 * \mathcal{L}_{PCC} + \lambda_3 * (\mathcal{L}_{NQP} + \mathcal{L}_{NLP}) + \lambda_4 * \mathcal{L}_{CCP}]. \quad (5)$$

where

$$\mathcal{L}_{ent1} = -\log \prod_{t=1}^m \text{prob}(y_t|P) \quad (6)$$

where  $m$  denotes the size of  $T$ , and  $y_t$  denotes the  $t$ -th output.  $\{\lambda_i\}_{i=1}^4$  are empirical values that will be detailed in Section 4.2.

For the duality exploiting task, there is another loss for training the branch of the problem’s part-of-speech generation:

$$\mathcal{L}_{POS} = \sum_{(P',T) \in \mathbf{D}} [\mathcal{L}_{ent2} + \lambda_5 * \mathcal{L}_{dual} + \lambda_6 * \mathcal{L}_{PCC'}]. \quad (7)$$

where

$$\mathcal{L}_{ent2} = -\log \prod_{t=1}^n \text{prob}(x_t|T) \quad (8)$$

where  $n$  denotes the size of  $P$ , and  $x_t$  denotes the  $t$ -th output.  $\mathcal{L}_{PCC'}$  is the semantic loss between predicted POS and the ground-truth POS.  $\{\lambda_i\}_{i=5}^6$  are empirical values that will also be detailed in Section 4.2.

<sup>3</sup>We use Jieba (<https://github.com/fxsjy/jieba>) to generate the POS of a problem.

## 4 Experiments

### 4.1 CM17K Dataset

Most public MWP datasets are quite small such as ALG514 or exist some incorrect labels such as Dolphin18K. An exception is the Math23K dataset, which contains 23161 problems labeled well with structured equations and answers. However, it only contains one-unknown linear math word problems, which is not sufficient to validate the ability of a math solver about solving multiple types of MWPs. Therefore, we introduce a new high-quality math word problems dataset, called CM17K, to validate the universality of a solver and provide a more realistic and challenging benchmark for developing a universal and scalable math solver. We collect CM17K from two education websites<sup>4</sup>. These problems are oriented grades 6-12, containing 4 types of MWPs with more than 17K samples, including 6215 arithmetic MWPs, 5193 one-unknown linear MWPs, 3129 one-unknown non-linear MWPs, and 2498 equation set problems. It should be noticed that our dataset is sufficient for validating the universality of math word problem solvers since these problems can cover most cases about MWPs. We label our data with structured equations and answers following Math23K (Wang et al., 2017). We split our CM17K into train/valid/test sets at a ratio of 8:1:1.

	Math23K	CM17K
# Avg PL	28.015	54.365
# Avg EL	6.853	13.853
# Avg TS	5.554	11.834
# Avg Num	2.821	6.383
# Avg SNI	2.668	4.111
# Avg Ops	3.943	4.852
# Avg Constants	0.270	0.327

Table 1: Statistics of Math23K and CM17K. PL, EL, TS, Num, SNI, Ops, and Constants represent problem length, equation length, equation tree size, number of quantities in problems, number of quantities occurred in both problems and their corresponding equations, number of operators in equations, and number of constants only occurred in equations, respectively.

The data statistics of Math23K and CM17K are shown in Table 1. From the statistics, we can see that all statistics of CM17K are larger than Math23K. This shows that our dataset is more challenging and difficult for math word problem solvers. Besides, since CM17K contains more types of MWPs than Math23K, CM17K is more suitable

<sup>4</sup><http://www.zxxk.com/> and <http://www.jyeoo.com/>

for validating the reasoning ability of a solver than Math23K.

### 4.2 Experimental Setup and Training Details

#### 4.2.1 Datasets, Baselines, and Metric

We conduct experiments on Math23K and our CM17K. The main state-of-the-arts to be compared are as follows: **DNS** (Wang et al., 2017) is a universal solver based on the seq2seq model with significant number identification (SNI). **GTS** (Xie and Sun, 2019) is a goal-driven tree-structured MWP solver. **StackDecoder** (Chiang and Chen, 2019) is an universal semantically-aligned math word problems solver. (Zhang et al., 2020a) is an enhanced GTS with teacher-student distillation and multi-decoder ensemble. Again, following prior works (Wang et al., 2017; Chiang and Chen, 2019; Xie and Sun, 2019), we use *answer accuracy* as the evaluation metric: if the calculated value of the predicted equation tree equals to the true answer, it is thought as correct since the predicted expression is equivalent to the target expression.

#### 4.2.2 Implementation Details

We use Pytorch<sup>5</sup> to implement our model on Linux with an NVIDIA RTX2080Ti GPU card. All those words with fewer than 5 occurrences are converted into a special token UNK. The size of word embeddings and all hidden states for other layers are set as 128 and 512, respectively. Our model is optimized by ADAM optimizer (Kingma and Ba, 2015) with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 1e^{-8}$ . The mini-batch size is set as 32. The initial learning rate is set as  $1e^{-3}$  and then decreases to half every 40 epochs. To prevent overfitting, we set dropout rate as 0.5 and weight decay as  $1e^{-5}$ . Finally, we conduct greedy search to generate symbolic equation trees. We set  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ ,  $\lambda_5$ , and  $\lambda_6$  as 0.0005, 0.01, 1.0, 0.005, and 0.1 for both datasets, respectively. We set  $\lambda_4$  as 0.000001 for Math23K while we set  $\lambda_4$  as 1.0 for CM17K. All constants are extracted from the training set. In each epoch, all training data is shuffled randomly and then cut into mini-batches.

### 4.3 Answer Accuracy

Following prior works (Wang et al., 2017; Chiang and Chen, 2019; Xie and Sun, 2019), we conduct 5-fold cross-validation on Math23K. For CM17K, we evaluate the performance on the test set. The results are shown in Table 2. From Table 2, we can observe

<sup>5</sup><http://pytorch.org>

that benefiting from the four new auxiliary tasks and neural-symbolic paradigm, our NS-Solver outperforms the baselines on both datasets in terms of answer accuracy. Specifically, for Math23K and CM17K, the accuracy gains of NS-Solver over GTS are 1.37% and 5.93%, respectively. Comparing with TSN-MD, our solver outperforms it by about 0.6% on Math23K. It shows that our model is more feasible for solving multiple types of MWPs. It also shows that our NS-Solver is more effective than other state-of-the-art models on the real-world scenario that needs to solve various MWPs with a unified solver.

Model	Math23K	CM17K
DNS (Wang et al., 2017)	58.1%	15.93%
StackDecoder (Chiang and Chen, 2019)	66.0%	37.24%
GTS (Xie and Sun, 2019)	74.3%	47.12%
TSN-MD (Zhang et al., 2020a)	75.1%	-
NS-Solver (Ours)	<b>75.67%</b>	<b>54.05%</b>

Table 2: Model comparison on answer accuracy

#### 4.4 Comparisons on different subsets

We drill down to analyze the generalization of DNS, GTS, and NS-Solver on different types of MWPs in the test subset of CM17K. Their answer accuracy on different types of MWPs is shown in Table 3. We can observe that our NS-Solver outperforms the other two models by a large margin on all subsets. Specifically, the accuracy gains of our NS-Solver over GTS on four subsets are 3.87%, 9.12%, 6.99%, and 9.44%. This shows that with the help of four auxiliary tasks, our NS-Solver obtains better generalization ability on multiple types of MWPs than baselines.

		arithmetic	one-unknown linear	one-unknown non-linear	equation set
Number		619	526	315	244
DNS	Correct	23	49	67	132
	Accuracy	3.7%	9.32%	21.27%	54.1%
GTS	Correct	255	220	201	128
	Accuracy	41.20%	41.83%	63.80%	52.45%
NS-Solver (Ours)	Correct	279	268	223	151
	Accuracy	<b>45.07%</b>	<b>50.95%</b>	<b>70.79%</b>	<b>61.89%</b>

Table 3: Answer accuracy on CM17K’s test subset.

#### 4.5 Performance on Tree Length

Intuitively, the size of the symbolic equation tree is proportional to the complexity of the mathematical relationship in the problem. The more complex the mathematical relationship is, the more difficult it is to solve the problem. Here, we compare our proposed NS-Solver with GTS on CM17K to show the superiority of our NS-Solver on different equation tree sizes. The answer accuracies for different sizes of expression trees on CM17K test subset are shown in Fig. 2. We can see that there is a tendency

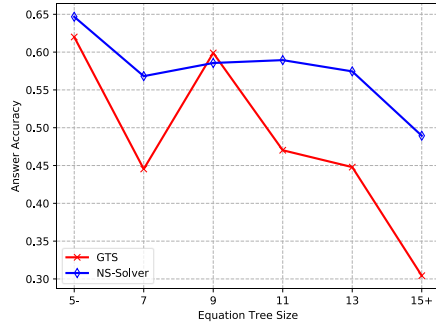


Figure 2: Answer accuracies for different sizes of symbolic equation trees on CM17K.

for answer accuracy to degrade with the growth of the problem complexity measured as the size of the equation tree, and our NS-Solver outperforms GTS on most cases of different equation tree sizes. This shows our NS-Solver can better model the mathematical relationships of the problem than GTS. It can also be noticed that the improvement of our NS-Solver over the GTS is increasing when the problems become more complex.

However, although our model outperforms other methods, there still has room for improvement in semantic understanding and symbolic reasoning since longer equations often match with more complex MWPs which entail more complex math relationships.

##### 4.5.1 Ablation on different auxiliary tasks

We study the contribution of different auxiliary tasks of our NS-Solver. For this purpose, we consider five different combinations: 1) only the backbone [NS-Solver - CCP - SNP - PCC - DE]; 2) backbone + duality exploiting task [NS-Solver - CCP - SNP - PCC]; 3) backbone + duality exploiting task + program consistent checker [NS-Solver - CCP - SNP]; 4) backbone + duality exploiting task + program consistent checker + number prediction tasks [NS-Solver - CCP]; and 5) the proposed NS-Solver [NS-solver]. For each of these combinations, each model was trained for 80 epochs on CM17K and validated on its test subset. The learning rate decreased to half every 20 epochs. The results are provided in Fig. 4.

As one can see, all four auxiliary tasks can improve performance. Specifically, the accuracy gains of DE, PCC, SNP, and CCP are 1.00%, 1.41%, 1.11%, and 1.12%, respectively. Besides, the binary accuracies of the two SNP tasks are 97% (number quantity prediction) and 96.8% (number location prediction). Moreover, the accuracy of our CCP

Problem	Generated Symbolic Equation	Auxiliary Task
<p><b>Case 1:</b> 学校买来 <math>\text{NUM}(n_0 [5])</math> 盒羽毛球，每盒 <math>\text{NUM}(n_1 [12])</math> 个，共用 <math>\text{NUM}(n_2 [240])</math> 元，平均每个羽毛球多少元？(The school bought <math>\text{NUM}(n_0 [5])</math> boxes of badminton, each box of <math>\text{NUM}(n_1 [12])</math>, sharing <math>\text{NUM}(n_2 [240])</math> yuan, how much is the average price of each badminton?)</p>	<p>SNS-solver - CCP - NP - PCC - DE (Ours):  <math>x = n_2 / n_1</math> (error)</p> <p>SNS-solver - CCP - NP - PCC (Ours):  <math>x = n_2 / (n_0 * n_1)</math> (correct)</p>	+Duality Exploiting (DE)
<p><b>Case 2:</b> 小杰与同学们去南岳山玩，每人车票费是 <math>\text{NUM}(n_0 [22])</math> 元，他们总共花了 <math>\text{NUM}(n_1 [154])</math> 元车票费，他们买了几张票？(Xiaojie went to Nanyueshan with his classmates. The ticket per person was <math>\text{NUM}(n_0 [22])</math> yuan. They spent a total of <math>\text{NUM}(n_1 [154])</math> yuan. How many tickets were they bought?)  <b>Groundtruth:</b> <math>x = n_1 / n_0</math></p>	<p>SNS-solver - CCP - NP - PCC (Ours):  <math>x = n_1 / (n_0 * 1.0)</math> (correct)</p> <p>SNS-solver - CCP - NP (Ours):  <math>x = n_1 / n_0</math> (correct)</p>	+Program Consistency Checker (PCC)
<p><b>Case 3:</b> 妈妈想给 <math>\text{NUM}(n_0 [1])</math> 间长 <math>\text{NUM}(n_1 [7])</math> 米，宽 <math>\text{NUM}(n_2 [4])</math> 米的房间铺上地砖，每平方米的地砖价钱是 <math>\text{NUM}(n_3 [60])</math> 元，那么铺好地砖至少要花多少钱？(Mother wants to lay a floor tile in <math>\text{NUM}(n_0 [1])</math> room with a length of <math>\text{NUM}(n_1 [7])</math> meters and a width of <math>\text{NUM}(n_2 [4])</math> meters. The price per square meter of floor tiles is <math>\text{NUM}(n_3 [60])</math> yuan. So how much does it cost to lay the floor tiles?)</p>	<p>SNS-solver - CCP - NP (Ours):  <math>x = n_3 * n_1 * n_2 / 10000</math> (error)</p> <p>SNS-solver - CCP (Ours):  <math>x = n_3 * n_1 * n_2</math> (correct)</p>	+Number Prediction (NP)
<p><b>Case 4:</b> 小胖家装修新房了，准备在客厅铺上地砖，客厅是长方形的地面，长 <math>\text{NUM}(n_0 [5])</math> 米，宽 <math>\text{NUM}(n_1 [6])</math> 米，他选中了边长为 <math>\text{NUM}(n_2 [40])</math> 厘米的正方形地砖，他至少要购买多少块这样的地砖？(The chubby family has renovated a new house and is ready to lay floor tiles in the living room. The living room is a rectangular floor with a length of <math>\text{NUM}(n_0 [5])</math> meters and a width of <math>\text{NUM}(n_1 [6])</math> meters. He chose a square floor tile with a side length of <math>\text{NUM}(n_2 [40])</math> cm. How many pieces of floor tiles should he buy at least?)</p>	<p>SNS-solver - CCP (Ours):  <math>x = n_0 * n_1 / ((n_2 / 100) * (n_2 / 10))</math> (error)</p> <p>SNS-solver (Ours):  <math>x = n_0 * n_1 / ((n_2 / 100) * (n_2 / 100))</math> (correct)</p>	+Commonsense Constant Prediction (CCP)
<p><b>Case 5:</b> 甲、乙 <math>\text{NUM}(n_0 [2])</math> 地相距 <math>\text{NUM}(n_1 [200])</math> 千米，快车速度为 <math>\text{NUM}(n_2 [120])</math> 千米每小时，慢车速度为 <math>\text{NUM}(n_3 [80])</math> 千米每小时，慢车从甲地出发，快车从乙地出发。如果 <math>\text{NUM}(n_0 [2])</math> 车同时出发，相向而行，出发后几时 <math>\text{NUM}(n_0 [2])</math> 车相遇？(The distance between <math>\text{NUM}(n_0 [2])</math> locations A and B is <math>\text{NUM}(n_1 [200])</math> kilometers, the speed of express train is <math>\text{NUM}(n_2 [120])</math> kilometers per hour, and the speed of slow train is <math>\text{NUM}(n_3 [80])</math> kilometers per hour. If the <math>\text{NUM}(n_0 [2])</math> cars depart at the same time and travel towards each other, when will the <math>\text{NUM}(n_0 [2])</math> cars meet after departure?)</p>	<p>GTS:  <math>n_2 * x = n_1 + n_3 * x</math> (error)</p> <p>SNS-solver (Ours):  <math>n_2 * x + n_3 * x = n_1</math> (correct)</p>	+All above four tasks

Figure 3: Typical cases. Note that the results are represented as infix order which is more readable than prefix order. The programs generated by NS-Solver are also translated into human-readable equations. Constants and number symbols are labelled in red and cyan, respectively.

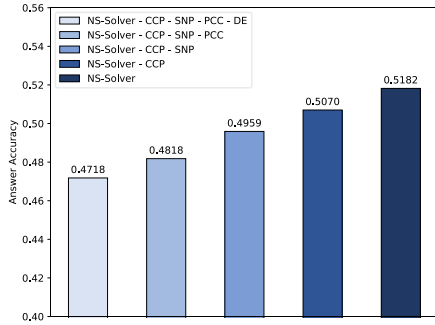


Figure 4: Ablation Study on different auxiliary components. ‘-’ represents we remove the component.

task is 97.8%. This shows that our auxiliary tasks can enhance our NS-Solver to enforce better problem understanding and symbol reasoning. Overall, our proposed NS-Solver achieves the best answer accuracy.

#### 4.6 Case Study

We also present the results of our NS-Solver with different combinations of four auxiliary tasks in Fig. 3. Benefiting from explicitly exploiting the probabilistic correlation between two quasi dual tasks to regularize the training process in our duality exploiting (DE) task, our [NS-solver - CCP - SNP - PCC] can generate correct equations by understanding the problem better while [NS-solver - CCP - SNP - PCC - DE] generates error equations, as shown in **Case 1**. With the program consis-

tency checker (PCC) that effectively regularizes the model’s output by constraining the distance between predicted symbols and ground-truth symbols during training, [NS-solver - CCP - SNP] can generate more consistent equations with the ground-truth than [NS-solver - CCP - SNP - PCC], as shown in **Case 2**. With self-supervised number prediction (SNP), [NS-solver - CCP] can generate better results and avoid generating symbols that do not belong to the problem, as shown in **Case 3**. With commonsense constant prediction (CCP), our NS-Solver manages to choose correct constants by constraining the constant symbolic table using predicted results of CCP. As shown in **Case 4**, [NS-solver - CCP] chooses error constant 10 while NS-solver chooses two correct constants. Besides, although GTS and NS-Solver generate the same symbols sometimes, our NS-Solver generates correct equations with the help of our four auxiliary objectives, as shown in **Case 5**. Overall, all four auxiliary tasks can improve our NS-Solver’s understanding and reasoning ability.

Model	BERT + Tree Decoder (Xie and Sun, 2019)	NS-Solver + BERT
CM17K	55.0%	<b>60.68%</b>

Table 4: Generalization to different backbone



#### 4.7 Extends to other backbone

To show that our auxiliary tasks can be adapted to other backbones, we replace GTS’s encoder with BERT (**BERT + Tree Decoder**) and NS-Solver’s encoder with BERT (**NS-Solver + BERT**), where we adopt a Chinese BERT-base pre-trained with whole word masking (Cui et al., 2020). We conduct experiments on CM17K. The results are shown in Table 4. We can observe that with auxiliary tasks, our **NS-Solver + BERT** still can outperform **BERT + Tree Decoder**, which shows that our auxiliary tasks’ strong generalization.

### 5 Conclusion

In this work, we propose Neural-Symbolic Solver (NS-Solver) to explicitly and seamlessly incorporate different levels of symbolic constraints by four auxiliary tasks. Our NS-Solver consists of a problem reader to encode problems, a programmer to generate a symbolic grounded program, and a symbolic executor to obtain final results. In addition to supervised learning with target expression, our solver is also optimized via four new auxiliary objectives that enforce four levels of symbolic reasoning. Besides, we also construct a new dataset CM17K containing 4 types of MWPs with more than 17K samples, which provides a more realistic and challenging benchmark for developing a universal and scalable math solver. Extensive experiments on Math23K and CM17K demonstrate the superiority of our NS-Solver compared to state-of-the-art methods in answer accuracy while ensuring intermediate equation rationality.

### 6 Ethical Impact

We collected CM17K from two online education websites, which is only used for academic research, and the copyright belongs to the original websites. This work may inspire research in the field of numerical reasoning.

**Acknowledgements** This work was supported in part by National Key R&D Program of China under Grant No.2020AAA0109700, National Natural Science Foundation of China (NSFC) under Grant No.U19A2073, No.61976233 and No. 61836012, the Natural Science Foundation of Guangdong Province under Grant No. 2017A030312006, Guangdong Province Basic and Applied Basic Research (Regional Joint Fund-Key) Grant No.2019B1515120039,

Shenzhen Fundamental Research Program (Project No.RCYX20200714114642083 and No.JCYJ20190807154211365), Zhijiang Lab’s Open Fund (No.2020AA3AB14), CSIG Young Fellow Support Fund, and Guangdong Provincial Key Laboratory of Information Security Technology.

### References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Yefim Bakman. 2007. Robust understanding of word problems with extraneous information. *Computing Research Repository*, arXiv:math/0701393.
- Jiaqi Chen, Jianheng Tang, Jinghui Qin, Xiaodan Liang, Lingbo Liu, Eric P. Xing, and Liang Lin. 2021. GeoQA: A geometric question answering benchmark towards multimodal numerical reasoning. *arXiv preprint arXiv:2105.14517*.
- Xinyun Chen, Chen Liang, Adams Wei Yu, Denny Zhou, Dawn Song, and Quoc V Le. 2019. Neural symbolic reader: Scalable integration of distributed and symbolic representations for reading comprehension. In *International Conference on Learning Representations*.
- Ting-Rui Chiang and Yun-Nung Chen. 2019. Semantically-aligned equation generation for solving and reasoning math word problems. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2656–2668. Association for Computational Linguistics.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. [Learning phrase representations using RNN encoder-decoder for statistical machine translation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.
- Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. 2020. [Revisiting pre-trained models for Chinese natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 657–668, Online. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of](#)

- deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Li Dong and Mirella Lapata. 2016. [Language to logical form with neural attention](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33–43, Berlin, Germany. Association for Computational Linguistics.
- Di He, Yingce Xia, Tao Qin, Liwei Wang, Nenghai Yu, Tie-Yan Liu, and Wei-Ying Ma. 2016. Dual learning for machine translation. In *Advances in neural information processing systems*, pages 820–828.
- Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. 2019. Using self-supervised learning can improve model robustness and uncertainty. In *Advances in Neural Information Processing Systems*, pages 15637–15648.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Yining Hong, Qing Li, Daniel Ciao, Siyuan Huang, and Song-Chun Zhu. 2021a. Learning by fixing: Solving math word problems with weak supervision. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*.
- Yining Hong, Qing Li, Ran Gong, Daniel Ciao, Siyuan Huang, and Song-Chun Zhu. 2021b. Smart: A situation model for algebra story problems via attributed grammar. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI-21*.
- Danqing Huang, Jing Liu, Chin-Yew Lin, and Jian Yin. 2018. Neural math word problem solver with reinforcement learning. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 213–223. Association for Computational Linguistics.
- Danqing Huang, Shuming Shi, Chin-Yew Lin, and Jian Yin. 2017. Learning fine-grained expressions to solve math word problems. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 805–814. Association for Computational Linguistics.
- Danqing Huang, Shuming Shi, Chin-Yew Lin, Jian Yin, and Wei-Ying Ma. 2016. How well do computers solve math word problems? large-scale dataset construction and evaluation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 887–896. Association for Computational Linguistics.
- Robin Jia and Percy Liang. 2016. [Data recombination for neural semantic parsing](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12–22, Berlin, Germany. Association for Computational Linguistics.
- Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *international conference on learning representations*.
- Rik Koncelkedziorski, Hannaneh Hajishirzi, Ashish Sabharwal, Oren Etzioni, and Siena Dumas Ang. 2015. Parsing algebraic word problems into equations. *Transactions of the Association for Computational Linguistics*, 3:585–597.
- Nate Kushman, Yoav Artzi, Luke Zettlemoyer, and Regina Barzilay. 2014. Learning to automatically solve algebra word problems. In *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 271–281.
- Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*.
- Yikang Li, Nan Duan, Bolei Zhou, Xiao Chu, Wanli Ouyang, Xiaogang Wang, and Ming Zhou. 2018. Visual question generation as dual task of visual question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6116–6124.
- Chao-Chun Liang, Yu-Shiang Wong, Yi-Chung Lin, and Keh-Yih Su. 2018a. [A meaning-based statistical English math word problem solver](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 652–662, New Orleans, Louisiana. Association for Computational Linguistics.
- Chen Liang, Jonathan Berant, Quoc Le, Kenneth D. Forbus, and Ni Lao. 2017a. [Neural symbolic machines: Learning semantic parsers on Freebase with weak supervision](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 23–33, Vancouver, Canada. Association for Computational Linguistics.
- Chen Liang, Jonathan Berant, Quoc Le, Kenneth D. Forbus, and Ni Lao. 2017b. [Neural symbolic machines: Learning semantic parsers on Freebase with weak supervision](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 23–33, Vancouver, Canada. Association for Computational Linguistics.

- Chen Liang, Mohammad Norouzi, Jonathan Berant, Quoc V Le, and Ni Lao. 2018b. Memory augmented policy optimization for program synthesis and semantic parsing. In *Advances in Neural Information Processing Systems*, pages 9994–10006.
- Wang Ling, Dani Yogatama, Chris Dyer, and Phil Blunsom. 2017. Program induction by rationale generation: Learning to solve and explain algebraic word problems. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 158–167. Association for Computational Linguistics.
- Arindam Mitra and Chitta Baral. 2016. Learning to use formulas to solve simple arithmetic problems. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2144–2153. Association for Computational Linguistics.
- Jinghui Qin, Lihui Lin, Xiaodan Liang, Rumin Zhang, and Liang Lin. 2020. Semantically-aligned universal tree-structured solver for math word problems. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3780–3789.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. [SQuAD: 100,000+ questions for machine comprehension of text](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas. Association for Computational Linguistics.
- Subhro Roy and Dan Roth. 2015. Solving general arithmetic word problems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1743–1752. Association for Computational Linguistics.
- Subhro Roy and Dan Roth. 2016. Unit dependency graph and its application to arithmetic word problem solving. In *Thirtieth AAAI Conference on Artificial Intelligence*, pages 3082–3088.
- Subhro Roy and Dan Roth. 2018. Mapping to declarative knowledge for word problem solving. *Transactions of the Association for Computational Linguistics*, 6:159–172.
- Yibin Shen and Cheqing Jin. 2020. [Solving math word problems with multi-encoders and multi-decoders](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2924–2934, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Shuming Shi, Yuehui Wang, Chin-Yew Lin, Xiaojiang Liu, and Yong Rui. 2015. Automatically solving number word problems by semantic parsing and reasoning. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1132–1142. Association for Computational Linguistics.
- Yu Sun, Eric Tzeng, Trevor Darrell, and Alexei A Efros. 2019. Unsupervised domain adaptation through self-supervision. *arXiv preprint arXiv:1909.11825*.
- Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. 2020. Test-time training with self-supervision for generalization under distribution shifts. In *International Conference on Machine Learning*, pages 9229–9248. PMLR.
- Duyu Tang, Nan Duan, Tao Qin, Zhao Yan, and Ming Zhou. 2017. Question answering and question generation as dual tasks. *arXiv preprint arXiv:1706.02027*.
- Lei Wang, Yan Wang, Deng Cai, Dongxiang Zhang, and Xiaojiang Liu. 2018a. Translating a math word problem to a expression tree. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1064–1069. Association for Computational Linguistics.
- Lei Wang, Dongxiang Zhang, Lianli Gao, Jingkuan Song, Long Guo, and Heng Tao Shen. 2018b. Mathdqn: Solving arithmetic word problems via deep reinforcement learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, pages 5545–5552.
- Lei Wang, Dongxiang Zhang, Zhang Jipeng, Xing Xu, Lianli Gao, Bing Tian Dai, and Heng Tao Shen. 2019. Template-based math word problem solvers with recursive neural networks. In *Thirty-Third AAAI Conference on Artificial Intelligence*, pages 7144–7151.
- Yan Wang, Xiaojiang Liu, and Shuming Shi. 2017. Deep neural solver for math word problems. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 845–854. Association for Computational Linguistics.
- Bolin Wei, Ge Li, Xin Xia, Zhiyi Fu, and Zhi Jin. 2019. Code generation as a dual task of code summarization. In *Advances in Neural Information Processing Systems*, pages 6559–6569.
- Qinzhao Wu, Qi Zhang, Jinlan Fu, and Xuanjing Huang. 2020. [A knowledge-aware sequence-to-tree network for math word problem solving](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7137–7146, Online. Association for Computational Linguistics.
- Yingce Xia, Tao Qin, Wei Chen, Jiang Bian, Nenghai Yu, and Tie-Yan Liu. 2017. Dual supervised learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3789–3798. JMLR. org.
- Han Xiao, Feng Wang, Jianfeng Yan, and Jingyao Zheng. 2018. Dual ask-answer network for machine reading comprehension. *arXiv preprint arXiv:1809.01997*.

- Zhipeng Xie and Shichao Sun. 2019. A goal-driven tree-structured neural model for math word problems. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5299–5305. International Joint Conferences on Artificial Intelligence Organization.
- Ma Yuhui, Zhou Ying, Cui Guangzuo, Ren Yun, and Huang Ronghuai. 2010. Frame-based calculus of solving arithmetic multi-step addition and subtraction word problems. In *International Workshop on Education Technology and Computer Science*, volume 2, pages 476–479.
- Jipeng Zhang, Roy Ka-Wei Lee, Ee-Peng Lim, Wei Qin, Lei Wang, Jie Shao, and Qianru Sun. 2020a. [Teacher-student networks with multiple decoders for solving math word problem](#). In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4011–4017. International Joint Conferences on Artificial Intelligence Organization. Main track.
- Jipeng Zhang, Lei Wang, Roy Ka-Wei Lee, Yi Bin, Yan Wang, Jie Shao, and Ee-Peng Lim. 2020b. Graph-to-tree learning for solving math word problems. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3928–3937.
- Victor Zhong, Caiming Xiong, and Richard Socher. 2017. Seq2sql: Generating structured queries from natural language using reinforcement learning. *arXiv preprint arXiv:1709.00103*.
- Lipu Zhou, Shuaixiang Dai, and Liwei Chen. 2015. Learn to solve algebra word problems using quadratic programming. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 817–822. Association for Computational Linguistics.