Appendix

A Data Preprocessing and Computational Devices

In both semantic parsing and math word problem experiments, we use CoreNLP Tool (Manning et al., 2014) as our external parser to generate constituency trees and dependency parsing trees. For structural brevity, we cut the root node in parsing trees and if nodes in graph are arranged into a "line" without branches, we transform them to a single node. In our experiments, we train our Graph2Tree model on a common single GPU: TITAN Xp. OS version: Ubuntu 16.04.4 LTS, and CUDA version: 8.0.

B Hyperparameter Setup

The Graph2Tree model is trained with Adam optimizer (Kingma and Ba, 2014) with learning rate set to 0.001. We use ReLU as our non-linear function and the greedy search as our inference strategy for decoding. Word embedding layer is initialized with GloVe vectors(300 dimension) from (Pennington et al., 2014). The dropout rate is chosen from $\{0.1, 0.3, 0.5\}$.

For semantic parsing experiments, we have described parameter setting in Section **5.1**.

For math word problem experiments, we here give the details of experiments due to page limitation.

i) On **MAWPS** (Koncel-Kedziorski et al., 2016) dataset, we randomly split the original dataset with the same train/dev/test split ratio following (Robaidek et al., 2018). To reduce vocabulary size, we projected numbers in problem descriptions to indexed markers, e.g., "n1, n2, ...". The batch size we use is 30. The dropout rate is chosen from $\{0.1,0.3\}$. The hop size in GNN is chosen from $\{2,3,4\}$.

ii) On **MATHQA** (Amini et al., 2019) dataset, the train/dev/test split is 29837/4475/2985. We use the annotated formula as our output. The batch size we use is 32. For other hyper-parameters, we use similar settings as in MAWPS.

C Sample Inputs and Outputs

We show the generated results of Graph2Tree model in MAWPS dataset in 1. For generated math equations, we replace original numbers to indexed markers, e.g., $\{1, 2, \ldots\}$.

References

Aida Amini, Saadia Gabriel, Shanchuan Lin, Rik Koncel-Kedziorski, Yejin Choi, and Hannaneh Hajishirzi. 2019. MathQA: Towards interpretable math word problem solving with operation-based formalisms. 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 2357–2367, Minneapolis, Minnesota. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Rik Koncel-Kedziorski, Subhro Roy, Aida Amini, Nate Kushman, and Hannaneh Hajishirzi. 2016. MAWPS: A math word problem repository. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1152–1157, San Diego, California. Association for Computational Linguistics.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word rep-

Problem and math expression
what is 1 % of 2 ? *Ground truth:* 1 * 0.01 * 2 = x *Graph2Tree:* x = (1 * 0.01) * 2
twice a number increased by 2 is 1 . find the number . *Ground truth:* (2.0 * x) + 2 = 1 *Graph2Tree:* (2.0 * x) + 2 = 1
a number is 1 less than the sum of 2 and 3 . what is the number ? *Ground truth:* (2 + 3) - 1 = x *Graph2Tree:* (x - 1) - 1 = (2 + 3) - 3
4 times the sum of a number and 2 is 3 less than 1 times that number . what is the number ? *Ground truth:* 4 * (2 + x) = (1 * x) - 3 *Graph2Tree:* 4 * (2 + x) = (1 * x) - 3
in a class of 2 students , 1 received a grade of a . what percent of the students received a 's ? *Ground truth:* x = (1 / 2) * 100.0 *Graph2Tree:* x = (1 / 2) * 100.0
mrs . hilt bought a yoyo for 1 cents and a whistle for 2 cents . how much did she spend in all for the two toys ? *Ground truth:* x = (1 + 2) *Graph2Tree:* 1 + 2 = x
isabella's hair is 2 inches long . if she gets a haircut and now her hair is 1 inches long , how much of isabella's hair got cut off ? *Ground truth:* x = (2 - 1) *Graph2Tree:* x = (1 - 2)
if sally can paint a house in 1 hours , and john can paint the same house in 2 hour , how many hours will it take for both of them to paint the house together ? *Ground truth:* (1.0 / 1) + (1.0 / 2) = 1.0 / x *Graph2Tree:* ((1.0 / 1) * x) + ((1.0 / 2) * x) = 1.0
alyssa went to 1 soccer games this year , but missed 2 . she went to 3 games last year and plans to go to 4 games next year . how many soccer games will alyssa go to in all ? *Ground truth:* x = 1 + 3 + 4 *Graph2Tree:* x = 1 + 3 + 4
a lawyer bills her clients 1 dollars per hour of service . if a client 's case requires 2 hours to complete , use proportion to calculate how much the client will owe the lawyer in dollars . *Ground truth:* 1 * 2 = x *Graph2Tree:* x = (2 / 1)

Table 1: Sample inputs and outputs of Graph2Tree on MAWPS.

resentation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Benjamin Robaidek, Rik Koncel-Kedziorski, and Hannaneh Hajishirzi. 2018. Data-driven methods for solving algebra word problems. *arXiv preprint arXiv:1804.10718*.