Multilingual Deep Neural Math Word Problem Solver

Ramya Sree Boppana
rboppana6

Shrija Mishra
smishra309

Sheryl Ratnam
sratnam6

1 Motivation

While computers have the ability to execute complex tasks and carry out computation at scale, they continue to underperform in the domain of natural language processing and mathematical reasoning. Designing algorithms for automatically solving mathematical word problems (MWP) is a challenging research topic that dates back to the 1960s. However, most of the literature is dedicated to solving MWPs exclusively in English, and few models are equipped to handle word problems in other languages. In this paper, we intend to present a multilingual (Chinese/English) MWP question answering model by exploiting large-scale datasets.

One of the key reasons why this research problem has not progressed is because of the wide semantic gap in parsing natural language into machine-understandable logic. In this study, we narrow our focus to related work on deep learning (DL) techniques that address this gap. Deep Neural Solver (DNS) pioneered the application of DL-based solvers to arithmetic word problems by using a RNN-based SEQ2SEQ model in combination with a similarity-based retrieval model(Wang et al., 2017)]. DILTON is another such DL-based architecture that predicts the math operation to be performed by utilizing GRUs and LSTMS and processing the problem in two separate parts – the contextual problem information and the main query (Mehta et al., 2017). On the other hand, Wang et. al. propose an equation normalization method and compare different SEQ2SEQ models to develop an ensemble learning method that achieves a higher level of accuracy (Wang et al., 2018).

While each of these proposed models return significantly better performance, none of the aforementioned papers explicitly address language agnostic representations of word problems. Aghajanyan et. al. attempt to unite universal (task agnostic) representations with multilingual (language agnostic) representations by formalizing Universal Grammar as an optimization problem (Aghajanyan et al., 2018). This framework is particularly relevant given that various deep learning models in this domain are language dependent.

2 Goal

We address the problem of automatically solving arithmetic word problems. The input to our system is the problem text that has the background story, which describes the known quantities, and a query, which describes an unknown quantity to be found. The goal of the system is to predict the operations to be performed ('-','+','*','/,') between the known numerical quantities to solve for the unknown quantity. Figure 1 shows an example problem (Zhang et al., 2019).

Word Problem
Oceanside Bike Rental Shop charges 17 dollars plus 7 dollars an hour for renting a bike.Tom paid 80 dollars to rent a bike. How many hours did he pay to have the bike checked out?
Equation
17+(7*x)=80
Solution
x=9

Figure 1: An example of arithmetic word problem

We intend to first develop individual models for each of the languages (Chinese and English). We aim to improve the accuracy of these subsystems. Later, we aim for our larger overall goal of decoupling the language from the problem by either learning language agnostic representations or by using an ensemble of the individual language models.

Some of the early approaches to MWP solvers are Rule-based, Statistic-based, and Tree-based methods which require human intelligence to extract effective features from the problem text. More recently, several attempts have been made to apply DL based methods for math word problem solving which have the ability to learn an effective feature representation in a data-driven manner without human intervention. Overall, the current status of MWP solvers still has great room for improvement especially after the development of large-scale datasets such as Dolphin18K (Huang et al., 2016) in English and Math23K (Wang et al., 2017) in Chinese. One of the number of possible research directions in which we would like to make a new contribution is to take the advantage of such multilingual large datasets to develop a language agnostic deep neural math word solver.

3 Plan

During the course of 6 weeks, we aim to achieve an encouraging model that will enable us solve some of the intricate math word problems. We dedicate the first week to perform a thorough literature review and data collection, where we explore large and diverse datasets available. While there are many datasets that have been developed for this specific research question, we intend to use the Math23K dataset as a starting point (Wang et al., 2017). More importantly, this dataset contains Chinese MWPs, which allows us the flexibility to test multilingual models. The MWPs are of elementary school level difficulty, and contain 23,162 problems. For our English-language dataset, we plan to use Dolphin-S, which contains 7,070 problem in total (Huang et al., 2016). We will reach out to the respective researchers to obtain access to the datasets. In addition to these existing datasets, we posit that a multilingual parallel corpus would be a great addition to the project.

Once we have collected this data, we will move to data pre-processing. We will do data cleaning, and other standard pre-processing procedures including removing stop words, tagging operators and operands, filtering out problems other than arithmetic if applicable and so on. Here, we will spend most of our time and effort in initial data analysis/exploration and observing the useful summary statistics that will help us take decisions in the later stages.

During the next 4 weeks, we will work on model development by experimenting with various models. We intend to use state of the art deep neural networks that can handle input sequences like the problem text in case of MWP. We start with the basic RNNs, a class of neural networks that allow previous outputs to be used as inputs though hidden states and incorporate feedback connections through LSTMs. Attentions and Transformers may be a useful addition for our problem to emphasize the parts of the problem text that describes known and unknown quantities. We will use Colab/Google Cloud for GPU resources to train/test the models. We will split model development into 2 phases. In the first phase (first 2 weeks- aligns with midterm report deadline), we will develop language specific models. In the next phase (latter 2 weeks), we will concentrate on techniques like ensembling for performance improvement and also on making the system language agnostic.

Overall, the system pipeline comprises of three different modules. The first is an embedding module in which we experiment with pre-trained embeddings like word2vec, glove, fasttext and language agnostic embeddings like LASER (Artetxe and Schwenk, 2019) and USE (Yang et al., 2019). The next module will be a sequence auto encoder that encodes the given problem text in the form of a mathematical expression to be solved. The final module will calculate the output from the expression.

During the final week, we will concentrate on analyzing the results like performance variation with different embeddings, models, and parameters. We will try to see any trends in the results and analyze why these trends make sense.

Accuracy pertaining to the true value of the word problem result will be used as our performance metric. Since the available language agnostic embeddings are evaluated on retrieval tasks, we are not sure how they will perform on our MWP task. If we figure out during our experiments that these are not suitable for our task, we will then concentrate on language specific models and use techniques like ensemble learning in order to get a more robust models for individual languages.

We expect all group members to be involved in each stage of our research project to ensure a fair division of labor. All members will participate in the literature review process, and each member will be tasked with obtaining the data and performing pre-processing procedures. For model development, each of the three members will be deploying and evaluating a different model - RNN, LSTM, and Transformer. We will then compile our findings and perform an overarching analysis of the results.

References

Armen Aghajanyan, Xia Song, and Saurabh Tiwary. 2018. Towards language agnostic universal representations. *arXiv preprint arXiv:1809.08510*.

Mikel Artetxe and Holger Schwenk. 2019. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610.

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.

Purvanshi Mehta, Pruthwik Mishra, Vinayak Athavale, Manish Shrivastava, and Dipti Misra Sharma. 2017. Deep neural network based system for solving arithmetic word problems. In *Proceedings of the IJCNLP 2017, System Demonstrations*, pages 65–68.

Lei Wang, Yan Wang, Deng Cai, Dongxiang Zhang, and Xiaojiang Liu. 2018. Translating a math word problem to an expression tree. *arXiv preprint arXiv:1811.05632*.

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.

Yinfei Yang, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, Steve Yuan, Chris Tar, Yun-Hsuan Sung, et al. 2019. Multilingual universal sentence encoder for semantic retrieval. *arXiv preprint arXiv:1907.04307*.

Dongxiang Zhang, Lei Wang, Luming Zhang, Bing Tian Dai, and Heng Tao Shen. 2019. The gap of semantic parsing: A survey on automatic math word problem solvers. *IEEE transactions on pattern analysis and machine intelligence*.