From Symbolic Logic Reasoning to Soft Reasoning: A Neural-Symbolic Paradigm

Qiming Bao & Michael Witbrock & Jiamou Liu

School of Computer Science, Faculty of Science, University of Auckland

Abstract

Combining deep learning with symbolic reasoning aims to capitalize on the success of both fields and is drawing increasing attention. However, it is as yet unknown to what degree symbolic reasoning can be acquired by end-to-end neural networks. In this paper, we explore the possibility of adapting a neural symbolic reasoner to function as a neural natural-language reasoner, in the hope of more accurately and reliably performing natural-language soft-reasoning tasks. It is noted that natural language is represented using a high-dimensional vector space and reasoning is performed using an iterative memory neural network based on RNNs and an attention mechanism. We use the open source natural-language soft reasoning dataset PARARULE. Our model has been trained in an end-to-end manner to learn whether a logical context contains a given query. We find that the model that converts the character-level embedding to a word-level embedding can achieve high performance (86%) on the PARARULE dataset without relying on a pre-trained language model and the performance rise nearly 1% when we use Gate Attention in the iterative cell.

Research Questions

This paper asks and seeks to answer the following questions.

1) **Can end-to-end neural networks demonstrate the ability to reason with facts and rules over language?** We trained, validated, and tested inference over facts and rules written in natural language called PARARULE [2] and found DeepLogic have a high level of accuracy (86%, Table 1). Also, we make a comparison experiment on Table 1 that DeepLogic and DeepLogic+ are better than RoBERTa-large and similar to RoBERTa-large fine-tuned on RACE dataset [4].
2) **Can the trained end-to-end neural based model solve natural language problems with different reasoning depths?** We found that the model trained on whole depths' dataset was able to solve five rule-based natural language problems with different reasoning depths (80%+ score, Table 1). We have also done subdivision experiments, training the model on datasets of different depths, and verifying on test sets of different depths. The model is trained and tested on datasets with depth=1 and depth=0. The former's The effect is obviously stronger than the latter, which can indicate that the model may have a certain reasoning ability.
3) **Can the end-to-end neural model handle more real-world natural language?** The model also performed well (80%+ score, Table 1) when trained, validated and tested over more natural (crowd-sourced) language. In this part, DeepLogic and RoBERTa-large both show the model's generalization ability to other natural language datasets.
4) **Can the end-to-end neural model find which facts it relies on to determine the answer?** We show that the model can do this to a substantial extent; the model can detect the related facts and rules within four steps (Figure 3).
5) **Can other end-to-end neural based models demonstrate the ability to reason?** We show that two other end-to-end neural based models, the Memory Attention Control Network [3] and the Dynamic Memory Network [6] are also capable of learning these tasks, albeit with lower scores (82% and 79%, respectively, compared to 86%). This suggests that our logic-to-language transfer results are not specific to iterative memory attention networks, although iterative memory attention networks are more likely to learn tasks (Table 1).

Dataset

We inspired our idea from the symbolic logic program, and we find the basic structure has a strong connection between the symbolic logic program and logic represented by natural language.

Symbolic logic deals with how symbols relate to each other. It assigns symbols to verbal reasoning in order to be able to check the veracity of the statements through a mathematical process. You typically see this type of logic used in calculus[1]. The modern development begin with George Boole[2] in the 19th century.

1: Facts	2: Unification	3: 1 Step
$e(l).$	$o(V,V).$	$p(X):-q(X).$
$?e(l).\ 1$	$?o(d,d).\ 1$	$q(a)$
$?i(d).\ 0$	$?o(b,d).\ 0$	$?q(a).\ 1$
		$?q(b).\ 0$

Figure 1: Symbolic Logic Programs from DeepLogic [1]

Questions in PARARULE involve reasoning with rules. The inputs to the model are context (facts + rules) and question. The output is the T/F answer to the question. Here the underlying reasoning for the true fact (Q1) is: Bob is big, therefore rough (rule1) therefore green (rule4). Note that the facts + rules themselves change for different questions in the datasets.

(Input Facts:) Alan is blue. Alan is rough. Alan is young.
Bob is big. Bob is round.
Charlie is big. Charlie is blue. Charlie is green.
Dave is green. Dave is rough.
(Input Rules:) Big people are rough.
If someone is young and round then they are kind.
If someone is round and big then they are blue.
All rough people are green.

Q1: Bob is green. True/false? **[Answer: T]**
Q2: Bob is kind. True/false? **[F]**
Q3: Dave is blue. True/false? **[F]**

Figure 2: An example from PARARULE [2]

Model

This section describes a word-level RNN-based iterative neural network with Gate Attention; for details we refer the reader to the main body of the model from DeepLogic and DMN+. DeepLogic is an iterative memory attention (IMA) network with an end-to-end training method. Besides the multi-hop symbolic logic reasoning, backward chaining and forward chaining models are merged as a comparison experiment, the main difference is we consider word-level embeddings for natural language rather than the use of character-level embedding for symbolic logic.

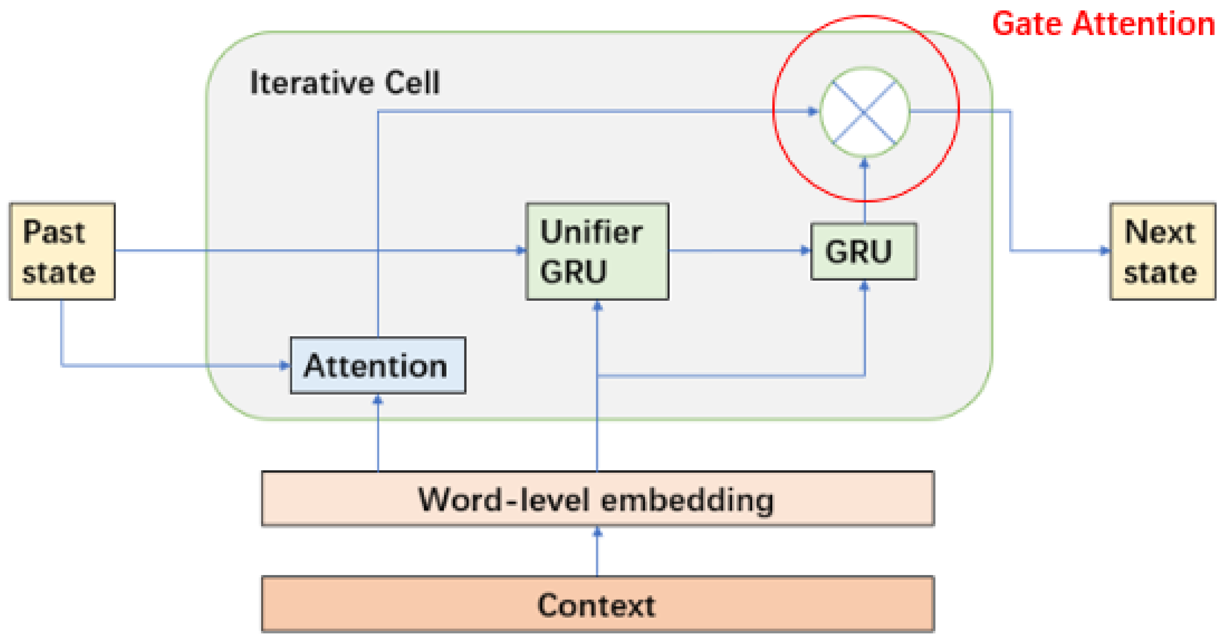

Figure 3: A diagram of the iterative neural cell in the word-level embedding IMA model with Gate Attention (IMA_GloVe_GA).

Results

Using GloVe [5] as the word embedding for the model, Table 1 compares various benchmarks for the end-to-end neural models IMA_GloVe and other varieties (IMA_GloVe_GA) represents DeepLogic model with Gate Attention and we call that model as DeepLogic+, RoBERTa_large and RoBERTa_large fine-tuned on the RACE dataset [4]. We borrow the Gate Attention from DMN+ [6] and Gate Attention has been firstly considered in Dynamic Neural Network which achieves 100% deductive reasoning test accuracy on bAbI task 15.

Table 1: Comparison between DeepLogic, RoBERTa_large, and RoBERTa_large fine-tuned on RACE dataset.

Test↓; Train→	Num Q	IMA_GloVe	IMA_GloVe_GA	RoBERTa_large	RoBERTa_large_RACE
Depth=1	4434	0.861	**0.877**	0.549	**1**
Depth=2	2915	0.853	**0.859**	**0.996**	0.995
Depth=3	2396	0.830	**0.836**	0.478	**0.992**
Depth≤3	9745	0.842	**0.851**	0.491	**0.953**
Depth≤3 + NatLang	20192	**0.810**	0.807	**0.963**	0.938
Depth≤5	13882	0.792	**0.801**	0.491	**0.989**
Depth≤5 + NatLang	24329	0.705	**0.713**	0.494	**0.969**

Conclusions

The contributions of the paper are summarised as follows: (i) migration of a system designed for logical expression embedding to natural language embeddings representing logical expressions and (ii) the first completed migration of a system design from symbolic logical reasoning to natural language soft reasoning. Our implementation is available at `https://github.com/Strong-AI-Lab/A-Neural-Symbolic-Paradigm`.

Forthcoming Research

We plan to build a deep learning model with more evidence of logical reasoning ability. A key point is not only to let the model make classification predictions (True/False) based on existing information and to know how the model accomplishes this. We are carrying out the current research along with these several questions. First, we want to generate a logical reasoning data set, that is, to know what algorithm the dataset is generated based on. We will provide detailed algorithms and codes to generate data. At the same time, we also generate the inference steps needed to understand a problem. This will help us train a model that can generate problem-solving steps. The current progress is available at `https://github.com/Strong-AI-Lab/PARARULE-Plus`

References

[1] Nuri Cingillioglu and Alessandra Russo. Deeplogic: Towards end-to-end differentiable logical reasoning. In *AAAI-MAKE*, 2019.

[2] Peter Clark, Oyvind Tafjord, and Kyle Richardson. Transformers as soft reasoners over language. In *IJCAI*, pages 3882–3890, 2020.

[3] Drew A. Hudson and Christopher D. Manning. Compositional attention networks for machine reasoning. In *ICLR*, 2018.

[4] Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. RACE: Large-scale ReAding comprehension dataset from examinations. In *EMNLP*, pages 785–794, 2017.

[5] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *EMNLP*, pages 1532–1543, 2014.

[6] Caiming Xiong, Stephen Merity, and Richard Socher. Dynamic memory networks for visual and textual question answering. In *ICML*, pages 2397–2406, 2016.

Acknowledgements

Thanks to the authors of the DeepLogic paper [1], for advice and help in understanding and reproducing this work. This has been of great help in completing this research and in planning future research.

[1] https://examples.yourdictionary.com/examples-of-logic.html
[2] https://philosophy.lander.edu/logic/symbolic.html