Reasoning in Artificial Neural Networks

Vasin Srisupavanich

***Abstract*—Over the past decade, deep learning systems have enjoyed tremendous success in the area of computer vision and natural language understanding. However, these systems still struggle in tasks which require deliberate thinking and reasoning process. Recently, many approaches to extends the capability of the neural networks have emerged. This paper reviews several recent approaches which takes inspiration from human's cognitive process and symbolic view of artificial intelligence to solve reasoning tasks, and outline future research directions in this challenging problem.**

***Index Terms*—deep learning, artificial neural networks, reasoning, attention, memory, graph neural networks, neural-symbolic**

I. Introduction

THE ability to reason is one of the most important features of human intelligence. It allows us to understand abstract concepts and make complex decisions. Human excels at tasks that require deliberate thinking, such as planning and symbol manipulation, while the current state of machines are limited to simpler pattern matching problems. Incorporating reasoning ability to machines has been a long standing goal, but a very difficult challenge in the field of Artificial Intelligence (AI). Solving this problem would mean a significant step toward artificial general intelligence, which will ultimately benefits humankind. This paper reviews recent approaches in building artificial neural networks (ANN) that can learn to reason, and overview the current state of the art results and applications from deep learning systems with reasoning capability.

II. Background

According to [1], reasoning refers to the ability to "algebraically manipulating previously acquired knowledge in order to answer a new question". Historically, the approach to create an AI system capable of reasoning has been from a symbolic point of view. In a symbolic AI, knowledge is represented as symbols, rules are handcrafted by human, and reasoning is the process of inference. However, these systems have been overshadowed by the success of deep learning in various domains in the past decades.

Despite the astonishing power of the neural networks, they tend to learn statistical mapping between input and output, rather than the true causal relations. With the goal to improve deep learning system beyond pattern matching, many researchers have tried to combine symbolic AI with neural networks, which became a subfield called Neural-symbolic. Apart from that, researchers also take inspiration from neuroscience. As human reasoning involves extracting knowledge from memory and paying attention to specific part of information, this has resulted in an extension of neural network in the form of memory and attention mechanism.

III. Main Approaches

A. Attention Mechanism

Attention mechanism was first introduced in 2014 for a machine translation task [2]. Since then this mechanism has became an important tool for deep learning in various applications. This idea is loosely motivated by how human biological system works. For instance, human visual attention allows us to focus on specific region with high resolution, while ignoring other irrelevant information. In the context of machine translation, attention model enables the machine to focus on specific words at a time rather than the full sentence.

In a neural machine translation model (NMT), the architecture consists of an encoder-decoder (seq2seq) structure. An encoder, typically a recurrent neural networks (RNN), learn to encode a source sentence into a fixed length vector. Then the decoder network output the encoded vector into another language. Figure 1(a) shows the traditional encoder-decoder architecture. The apparent problem of this architecture is that the model tends to forget relevant information in a long sentence. With the addition of attention model (figure 1(b)), this problem is mitigated, as the decoder can learn to attend to different parts of the source sentence. The attention weights, which are from a feed forward neural networks, are jointly train along with the encoder-decoder networks.

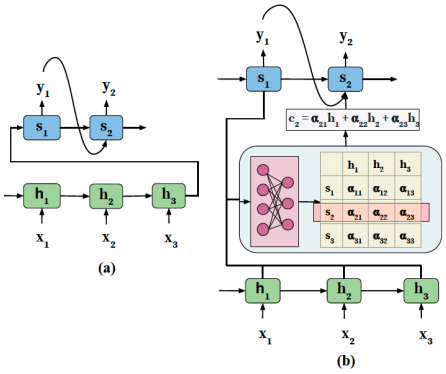

Fig. 1. NMT architecture (a) traditional (b) with attention model. Figure from [3]

In another influential paper by Xu et al [4], attention model is applied to generate caption from images. In this task, convolutional neural networks (CNN) is used as an encoder to extract features from raw images. Then a long short-term memory network (LSTM) is used to generate the words, conditioned on the attention weights. Figure 2 demonstrate how the model learn to attend to specific part of the image with the corresponding word. Also proposed in [4], the attention models can be classified in two types: soft and hard attention. In soft attention, the model is smooth and differentiable end

to end, and the context vector is computed by the weighted average of the whole image. In contrast, in hard attention, the context vector is computed from stochastically sampled patch of image. The hard attention model can be faster when making inference, however, they are difficult to optimized, and require reinforcement learning to train.

Fig. 2. Visualization of the attended region conditioned on the corresponding word. Figure from [4]

B. Memory Augmented Neural Networks

Similar to human, in order to retrieve the information necessary for the desired tasks, a machine needs to maintain some memory that needs to be efficiently organized and queried. This is particularly essential for complex tasks, such as multi-hop reasoning. Traditional RNN and its variant LSTM can memorize information in the hidden states, however, they are limited to only short term dependencies. Recent approaches to alleviate this issues involves explicit memory representation and the use of external memory, as proposed in Neural Turing Machine [5] and Memory Networks [6].

A Neural Turing Machine (NTM) contains two major components: a neural network controller and a memory bank (Figure 3). A controller can be any type of neural networks, feed forward or RNN, and is responsible for read and write operations on the memory matrix. The read and write operations are done selectively by soft attention mechanism, making every components fully differentiable, and thereby the whole model can be trained using gradient descent. The experiments reported in [5] have shown that NTM significantly outperforms traditional LSTM in tasks, such as copying and sorting.

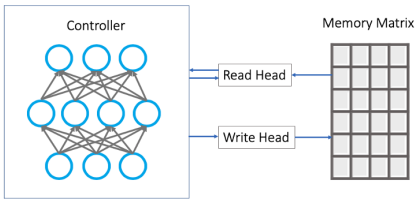

Fig. 3. Architecture of Neural Turing Machine

Another recent neural networks model, utilizing both attention and memory, designed to facilitate explicit reasoning is the MAC networks [7]. In contrast to NTM, the MAC networks don't utilizing a global external memory. Instead, each node (MAC cell) is recurrent, and has its own memory and control state. In contrast to traditional neural networks, the separation between control and memory encourages the networks to learn the computational process and reasoning operations, rather than to approximate direct transformation between the input and output. Figure 4 shows the result of the MAC networks from a visual reasoning task using CLEVR data set [8]. The MAC networks was able to achieve 98.94% accuracy, halving the error rate from the best prior model.

Fig. 4. Example of MAC networks performing novel visual reasoning task. Figure from [7]

C. Neural-Symbolic

In contrast to the purely sub-symbolic systems mention in the previous approaches, neural-symbolic is the subfield that try to integrate symbolic AI with the deep learning system. The idea is that the symbolic part world help the system logically reason about symbols, which is what traditional neural networks are weak at, while the neural network part would allow the system to learn, which the symbolic AI is not capable of.

In a visual question answering (VQA) task, recent neural-symbolic system, NS-CL [9] employed deep representation learning for visual recognition and language understanding, while the reasoning part is solved by symbolic program execution. The framework proposed in NS-CL contains three separate modules: a visual module where the objects are detected and vector representations are extracted, a semantic parser where the question text is parsed into a tree of predefined domain specific language (DSL) of executable program, and a symbolic program executer which take in the parsed program and vector representation of objects to derive the answer.

It has been shown that this method requires significantly less amount of training data to accurately answer questions, and was able to generalize to new scenes and questions better than traditional neural networks based method. Another added benefit of the neural-symbolic system is that the results are fully interpretable, as the execution trace is visible. Figure 5 shows the result of this model on the VQS dataset, along with the generated symbolic program. On the other hand, one significantly limitation is that the symbolic functions (DSL) need to be pre-defined. As shown in [9], functions such as filter, count and query are hard-coded into the system, thus this method wouldn't be able to scale in the real-world system.

D. Graph Neural Networks

Recently, graph neural networks (GNN) has been gaining popularity in various domains, particularly in tasks where data are represented as graph with complex relationship and interdependency between objects. A graph is a data structure consisted of vertices (nodes) and edges. Depending on the

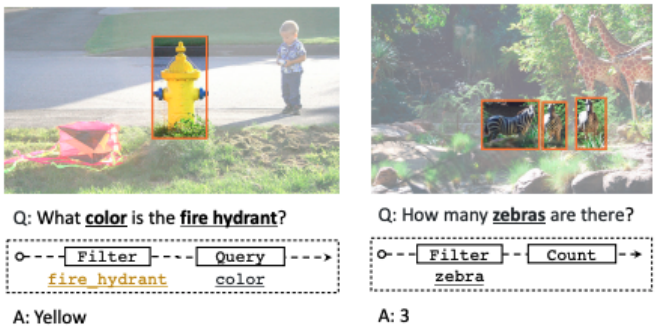

Fig. 5. Result of NS-CL on the VQS dataset. Figure from [9]

relationship between two objects, the edge can either be directed or undirected. In GNN, the node, edge and output can be flexibly represented in different types depending on the tasks. For instance, when the input is the image, the node would represent the patch of the image, and when the text is the input, the node would represent a sequence of words.

It has been shown that GNN is well suited for relational reasoning task, due to the strong relational inductive bias [10]. In contrast to CNN, where the inductive bias is the locality of the receptive field, and RNN where the bias is the sequentiality of data, GNN can express arbitrary relationship among entities. Recent work in [11] studied what task a neural network can learn to reason. The result revealed that the performance of the neural networks increase with more algorithmic alignment in the reasoning process, and the reason that GNN generalized better than other types of neural networks in many reasoning tasks is because the underlying reasoning process of GNN resemble dynamic programming.

To illustrate the use of the graph networks, a recent model called Neural State Machine (NSM) [12] utilized graph structure to represent the scene of an image in a VQA task. This model works in two stages: modeling and inference. In the modeling stage, the image is decomposed into a probabilistic graph that capture the semantic of the visual scene. The node corresponds to object within the image, while the edge represents both spatial and semantic relation. In the inference stage, the sequential reasoning process is performed over the graph by iteratively traversing its node guided by the question. Figure 6 shows the overall process with the generated scene graph.

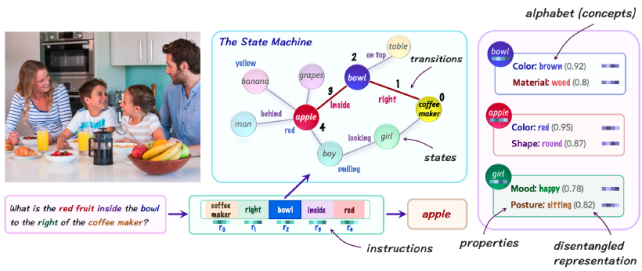

Fig. 6. Overall process of Neural State Machine. Figure from [12]

IV. Discussion

We have seen several approaches in building a neural networks with reasoning capability. Each approach has its own strength and weakness. While the attention mechanism allow the model to focus on specific part of information, this alone cannot be used to solve the reasoning task. Similar to working memory in the human brain, adding explicit memory to the neural networks give the model the ability to utilize the previous knowledge, however, the networks with global memory can be difficult to train. In a neural-symbolic system, the results from the networks are more interpretable, unlike a black box system in a traditional ANN, but this come with a cost of fixed symbolic function, and can be difficult to scale into the real problem. In contrast to neural-symbolic system, a graph neural networks can be used to recreate the symbolic structure, while maintaining the end to end differentiability. Nevertheless, many problems and processes cannot be easily represented as graph. There are still a lot of open questions in how to alleviate each of the weakness and combine each of the approach. However, I am convince that with the use of attention, memory, and the right inductive prior, this is the right step towards making a machine that is more intelligent.

References

[1] L. Bottou, "From machine learning to machine reasoning," *Machine learning*, vol. 94, no. 2, pp. 133–149, 2014.

[2] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[3] S. Chaudhari, G. Polatkan, R. Ramanath, and V. Mithal, "An attentive survey of attention models. arxiv 2019," *arXiv preprint arXiv:1904.02874*.

[4] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*, 2015, pp. 2048–2057.

[5] A. Graves, G. Wayne, and I. Danihelka, "Neural turing machines," *arXiv preprint arXiv:1410.5401*, 2014.

[6] J. Weston, S. Chopra, and A. Bordes, "Memory networks," *arXiv preprint arXiv:1410.3916*, 2014.

[7] D. A. Hudson and C. D. Manning, "Compositional attention networks for machine reasoning," in *International Conference on Learning Representations (ICLR)*, 2018.

[8] J. Johnson, B. Hariharan, L. van der Maaten, L. Fei-Fei, C. L. Zitnick, and R. Girshick, "Clevr: A diagnostic dataset for compositional language and elementary visual reasoning," in *CVPR*, 2017.

[9] J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu, "The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=rJgMlhRctm

[10] P. Battaglia, J. B. C. Hamrick, V. Bapst, A. Sanchez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. E. Dahl, A. Vaswani, K. Allen, C. Nash, V. J. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu, "Relational inductive biases, deep learning, and graph networks," *arXiv*, 2018. [Online]. Available: https://arxiv.org/pdf/1806.01261.pdf

[11] K. Xu, J. Li, M. Zhang, S. S. Du, K.-i. Kawarabayashi, and S. Jegelka, "What can neural networks reason about?" *arXiv preprint arXiv:1905.13211*, 2019.

[12] D. Hudson and C. D. Manning, "Learning by abstraction: The neural state machine," in *Advances in Neural Information Processing Systems*, 2019, pp. 5901–5914.