Title

Osama F. Zaki[1,2]

[1]Robotaar, Livingston, EH54 6DD, Scotland, UK, osama.zaki@robotaar.com
[2]Sinai University, ElArish, Sinai, Egypt, osama.farouk@su.edu.eg

Abstract

In this paper a neuro-symbolic architecture is proposed. We hold the view that the hidden layers of a neural network are just data/information processing units with their associated memory. Those layers don't represent any semantics of the application they are processing it, i.e. they are general-purpose machines. They can be built up from basic logic gates, e.g. NAND gates. In most applications a great effort is made to encode the "examples", and to configure the parameters, in a correct and efficient manner. With a lot of training, they can perform a task, sometimes produce excellent outcome, but sometimes they hallucinate. Even worse they cannot explain how and why they have produced this behaviour. This is because they are just trained to do tasks, but they have not learned it. However, besides being a general-purpose machine, they work in parallel and distributed manner, which make them powerful. Syntax and semantics of the application can be encoded at the edges of neural networks, i.e. input and output layers, in form of plain data, taxonomies, ontologies, or knowledge graphs. This understanding of neural networks forces us to device a pragmatic neuro-symbolic architecture that handle those facts.

Keywords

neuro-symbolic, architecture, semantics, pragmatic, neural networks

1. Introduction

The neuro-symbolic integration has started in shy as new direction in AI Since 2005. Several ideas since then have been presented in [1] [2], and reviews have been recently compiled in this area [3] [4]. It has been by recognised by many AI researchers and practitioners the advantages to make the two streams allies together, that is the expressiveness of the representation of the symbolic approach, and the computational power of the neural networks. This compensates the weakness in each approach. The outcome of this would be an explainable AI, which increase the trustworthiness element in AI systems. This area has advanced enough now that allow us to propose an architecture that integrates the two main streams in AI.

2. State of the art

Useful approaches proposed so far are representing symbolic knowledge as a neural network using rule-based and probabilistic databases [5][6], or to fine-tune symbolic rules based on the output of neural learning [7], or model-based integration of reasoning and training [8]. Those promising approaches were also acknowledged by Mileo [9].

Attempts have made to dive into neural networks, with some success, by dis-sectioning [10], or by decomposing [11], so that a logical formalism can be obtained, while other tried by

AAAI Fall Symposium 2024 on "Large Language Models for Knowledge Graph and Ontology Engineering"

distilling knowledge in and out of neural, and non-neural machine learning [12] [13]. In [14] an attempt was made to obtain compositional explanation of neurons in deep representation. All these attempts have not resulted yet in a coherent or clear integration of neuro-symbolic AI.

Mileo [9] suggested a neuro-symbolic cycle that leverages symbolic knowledge, deductive reasoning with human feedback for explainability, which is named extraction-explanation-injection cycle. The main key elements of this cycle are the use of graphs for neural activities (co-activation graph was devised for represent statistical correlations among relevant features of any deep representation in a feed forward neural network), with probabilistic logic rules to represent, causality, complex relational dependencies, or uncertainty. Then vector embedding as a technique to inject extracted knowledge from a trained model into any layer of a non-trained network. However, they suggested an inductive rule learning to adjust the weights of logical formulas by maximising the log likelihood of the training data is better situated to capture the association between logical rules and semantic concepts. And to utilise an approach such as Network Dissection for disentangled representations to capture the association between the semantic concepts and individual units. The involvement of human experts in this cycle iteratively are crucial during the neural-cognitive mapping and the hybrid representation learning processes. They have listed the challenges facing such cycle has for further discussions. Beside those challenges, we think the suggested approach is still far from practicality to prove its validity.

3. Axioms and hypotheses of the pragmatic neuron-symbolic architecture

- Neural Networks are a general-purpose machine can be implemented in HW or SW.
- Semantics are at the edges of neural network and fades as we go deeper.
- Neural networks are not entirely a black box, but scalability comes with complexity.
- To visualise paths from input to output units is intuitively possible.
- Inputs to NN can be plain data, taxonomies, ontologies, knowledge graphs, or natural languages.
- Outputs of NN are mainly a prediction in a form of a decision, class or cluster.
- The success of a neural network relies heavily on encoding/decoding schemes for the input/output units.
- A simple logical formalism is sufficient to describe the mapping between inputs (testing dataset) and outputs, whether it is a correlation, a relationship or a function.
- SW adaptors can be implemented for various types of inputs/outputs.
- Human intervention is mandatory. (a human experts can assign probabilistic value to the correlation, the relationship or the function extracted.)

4. Objectives of the pragmatic neuron-symbolic architecture

The architecture should be modular and loosely coupled. It should utilise knowledge graphs as a main tool for the symbolic representation. It should employ logical formalism with probabilistic

rules capabilities. It should employ deductive reasoning algorithms to check consistency and to introduce new concepts or attach instances to concepts. It should integrate inductive machine learning algorithms that can generalise. The neuro-symbolic architecture should be applicable for different neural networks architectures.

Human expert intervention is a key in this architecture. We can identify three tasks for the human expert: A) to derive some sort of fuzzy axioms when preparing training and validation dataset, 2) to validate and verify knowledge obtained from the trained model, and 3) to code the suitable SW adaptors.

5. Conclusion

In this short paper we present a philosophical yet a pragmatic neuro-symbolic architecture that utilises technologies which mostly already exist and tackling the integration of neuro-symbolic by focusing on the semantics presented at the edges of the neural networks, or more accurately, embedding the neural network architecture into the knowledge base. In a sense, this is capturing how human brain is processing whatever holds in the mind, updating and modifying its own knowledge as he learns in his life.

References

[1] T. R. Besold, Neural-symbolic learning and reasoning: A survey and interpretation, in: P. Hitzler (Ed.), Neuro-Symbolic Artificial Intelligence: The State of the Art, IOS Press, 2021, pp. 1 – 51. doi:10.3233/FAIA210348.

[2] P. Hitzler (Ed.), Compendium of Neurosymbolic Artificial Intelligence, IOS Press, 2023.

[3] C. Michel–Delétie, M. K. Sarker, Neuro-symbolic methods for trustworthy ai: a systematic review, Neurosymbolic Artificial Intelligence (2024).

[4] A. Sheth, K. Roy, M. Gaur, Neurosymbolic ai - why, what, and how, arXiv (2023). doi:3.

[5] M. V. M. França, G. Zaverucha, A. S. d'Avila Garcez, Fast relational learning using bottom clause propositionalization with artificial neural networks, Machine Learning 94 (2014). doi:doi:10.1007/s10994-013-5392-1.

[6] W. Cohen, F. Yang, K. Mazaitis, Tensorlog: A probabilistic database implemented using deep-learning infrastructure, J. Artif. Intell. Res (2020).

[7] L. Serafini, I. Donadello, A. d'Avila Garcez, Learning and reasoning in logic tensor networks: theory and application to semantic image interpretation, in: A. Seffah, B. Penzenstadler, C. Alves, X. Peng (Eds.), Proceedings of the Symposium on Applied Computing, 2017. doi:doi:10.1145/3019612.3019642.

[8] S. Tran, A. d'Avila Garcez, Deep logic networks: Inserting and extracting knowledge from deep belief networks, in: IEEE Trans. Neural Networks Learn. Syst., 2018. doi:doi:10.1109/TNNLS.2016.2603784.

[9] A. Mileo, Towards a neuro-symbolic cycle for human-centered explainability, Neurosymbolic Artificial Intelligence (2024).

[10] D. Bau, B. Zhou, A. Khosla, A. Oliva, A. Torralba, Network dissection: Quantifying

interpretability of deep visual representations, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017. doi:doi:10.1109/CVPR.2017.354.

[11] K. Fischer, A. René, C. Keup, M. Layer, D. Dahmen, M. Helias, Decomposing neural networks as mappings of correlation functions, Phys. Rev. Research (2022).

[12] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, arXiv (2015).

[13] S. Fukui, J. Yu, M. Hashimoto, Distilling knowledge for non-neural networks, in: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2019. doi:doi:10.1109/APSIPAASC47483.2019.9023120.

[14] J. Mu, J. A. and, Compositional explanations of neurons, in: Proceedings of the 34th International Conference on Neural Information Processing System, 2020.