NeSy: the key enabler to the next generation of AI?
A neuro-symbolic AI survey

Leonardo Lavalle
1838492
Sapienza, University of Rome
lavalle.1838492@studenti.uniroma1.it

1 Introduction

The current advancements in Artificial Intelligence have achieved remarkable effects, not only within research communities but also across popular media channels, especially in computer vision and natural language understanding (think about the media and political impact that ChatGPT has had in recent months). Nonetheless, influential thinkers have expressed their concerns regarding interpretability and accountability of AI, saying that it will have a negative impact both on humanity and on the future of the planet [9], unless the course is reversed. Another article [11], points out how this exponential progress in AI, in particular in deep learning, has come with a voracious appetite for computing power.

Taking again as an example ChatGPT: I don't want to imagine how many GPUs they (OpenAI) have exploited, how many hours of training they have performed, how much energy they have utilized and how much heat they have produced contributing to pollution and to global warming. In the just cited report, it's been showed that progress across a wide variety of applications is strongly reliant on increases in computing power and that this way of proceeding is rapidly becoming economically, technically and environmentally unsustainable. We cannot increase computational capacities indefinitely, we need to employ more data efficient and robust paradigms in order to overcome this issues.

We need Neuro-symbolic AI (NeSy). In [8] Gary Marcus stigmatizes recent trend to largely emphasized ever-larger training sets and more and more computing. Transformer architectures embodies this way of conceiving learning. In the article he stated that their success, even if they are immensely impressive as statistical inference engines, they are a long way from being a sound basis for robust intelligence. To reach that, he said a knowledge-driven and reasoning-based approach is needed (NeSy). *Is NeSy the key enabler to the next generation of AI?*

NeSy only began to be a formalized field of study since the 1990 and gained systematic research in the early 2000s. During the 2010s, NeSy received relatively less attention, as deep learning techniques started to achieve remarkable success across a wide range of AI tasks. But recently, due to motivations explained previously, NeSy is experiencing its renaissance in the research community. Besides that we are now going to dive into the two main reasons why researchers started to study and developed NeSy approaches.

1.1 Human Cognition

What is human intelligence? The question is as difficult as fascinating and many scientists coming from different disciplines tried to give a answer. Intelligence can be treated as the integration of two fundamental cognitive abilities: *learning* (ability to learn from experience) and *reasoning* (ability to reason from what has been learned). Like many times in AI, the first NeSy systems were ideated with the aim of emulating these two human mind's capabilities and combining them in an collaborative and coherent way. The debate about what is the real essence of human cognition is still open and very animated. Some cognitive scientists argue that human thinking depends upon symbol manipulation and that the mind should be considered as a "*Symbolic Logic Machinery*". On the other hand, human intelligence has a physical basis in the brain, which is composed of numerous neurons, so it seems reasonable to simulate the anatomy of the nervous system with artificial neurons (*Biologic Neural Network*). This dichotomy, as stated in [12], can be also observed in the debate regarding *Deductive vs Inductive* reasoning or in the one of *Continuity Principle vs Compositionality Principle* in terms of information processing, and has historically divided the development of the field of AI.

But Kahneman's *fast and slow thinking* theory changed everything. Basically humans' decisions are supported and guided by the cooperation of two main kinds of capabilities: the so called *system 1* which provides tools for intuitive, imprecise, fast, and often unconscious decisions ("*thinking fast*") and *system 2* which handles more complex situations where logical and rational thinking is needed to reach a complex decision ("*thinking slow*").

System 1 and *system 2* need each other and since their characteristics are strikingly similar to those of the *connectionist* approach and the *symbolic* approach to AI, an increasing number of AI researchers have started to reconsider the relationship between them and acknowledge the significance of NeSy [3].

1.2 Symbolism and Connectionism

Rather than taking the motivation from the desire of understanding and modeling human cognition, the study of NeSy is also driven by a more practical one: we require to combine *symbolism* and *connectionism* AI paradigms because they complement each other with respect to their advantages and disadvantages.

Symbolism. Many early AI systems (from the 1950s), were built upon symbolistic models. With this approach, cognitive processes can be achieved through the manipulation of symbols, employing a sequence of rules and logical operations on the symbolic representation. They are excellent at principled and exhibit inherently high explainability. Nevertheless, are difficult to train and are weak when applied outside their designated domain.

Connectionism. Known by its most successful technique, deep neural networks. It's the paradigm behind the vast majority of recent successful AI systems. They are very good at discovering statistic patterns from raw data and are robust against noisy data. On the other hand, these methods are *data hungry* and *black boxes*. They lack comprehensibility and is almost impossible to understand why decisions are made. This is the main problem: people have to deal with systems that make potentially catastrophic decisions, which are challenging to comprehend, laborious to correct and consequently difficult to place trust in. For this reasons many neural systems are not employed in decision-critical applications such as medical diagnosis or autonomous driving.

As a result, the integration of neural and symbolic approaches seems to be a natural step towards more a powerful, trustworthy and robust AI: *NeSy is the key enabler to the next generation of AI!*

2 Approaches

There exist only one taxonomy for NeSy approaches in the literature and is the scheme proposed by Kautz [6]. I will not present it in detail because quite confusing and untangled, but I will instead illustrate the different NeSy methods based on *how* knowledge is represented (Knowledge Representation) and *where* this symbolic knowledge is embedded in the neural network (Knowledge Embedding).

Knowledge Representation. There are many ways in which symbolic knowledge can be represented. I'll follow the categorization used in [12] (see Figure 1). The simplest way is through *knowledge graphs*, typically directed labeled graphs formed by entities as nodes and relations between entities as edges. *Propositional Logic* provides a flexible declarative language for formalizing knowledge about facts and dependencies, playing an important role for the integration of prior knowledge into neural networks. Due to its simplicity, many early works consider symbolic knowledge in the form of propositional logic, including [13]. *First Order Logic* is an extension to propositional logic and more powerful. It allows us to quantify over objects thanks to universal and existential quantifiers. To capture the full expressive power of first-order logic, *fuzzy logic* is employed [4] to transform prior knowledge into additional training objectives. Some other approaches (e.g. [1]) adopt Prolog, a logic *programming language*, for knowledge representation. There exist other types of knowledge representation in addition to those mentioned above: mathematic expressions and, as it exactly happens in [14], specific symbolic sequences generated from some informal symbolic systems with self-defined rules (*symbolic expression* category in Figure 1).

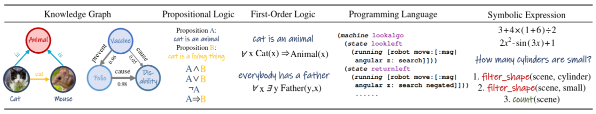

Figure 1: Illustrative overview of symbolic knowledge representations in NeSy.

Knowledge Embedding. This other categorization dimension gives us a more profound understanding of the integration of symbolic knowledge and neural networks in modern NeSy systems. Several solutions straightforwardly embed symbolic knowledge into the structure of training *data*, as it happens for example in [13]. In this paper, propositional logic rules are leveraged as prior symbolic knowledge to improve performance of deep models. The formulas, transformed in graphs, are embedded through a Graph Convolutional Network (see Figure 2). These logic embeddings are then used to form a logic loss that guides neural network training; the loss encourages the network to be consistent with prior knowledge. This methods, in some tasks as Visual Relation Prediction, where the goal is to predict relations between objects in images, significantly outperforms not-symbolic models.

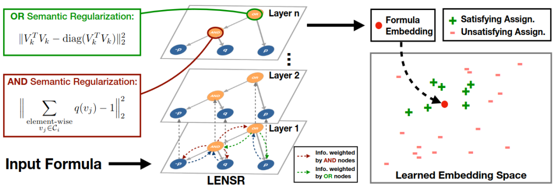

Figure 2: Logic graphs representing formulae are projected onto a manifold. This embedding space is used to form logic losses that regularize deep neural networks for a target task.

There's another way of embedding knowledge that worth mentioning. According to Kautz taxonomy [6], *Type 5* NeSy systems embed symbolic knowledge into the distributed representation, by means of training objectives that are specialized to the knowledge. Logic Tensor Networks [2] (better explained in Section 3), are the most representative and prominent approach of this nature. They translate first-order logic formulae as fuzzy relations on real numbers for neural computing. In this way, thanks to gradient based sub-symbolic learning, logic rules are embedded into the network learning objective.

Despite the recent progress, it is still hard to achieve a compact NeSy system that has both strong logic reasoning and expressive statistic learning abilities. In the sense of Kautz's vision, the realization of *Type 6* symbolic systems, which encompass these combined abilities, remains unfortunately a distant prospect.

2.1 Application Scenarios

In this section I will introduce some of the most interesting application fields of NeSy systems and, in my opinion, the ones which can express their full potential with the help of symbolic knowledge.

Visual Question Answering. VQA addresses visual perception in addition to language understanding, and I firmly think that symbolic reasoning in such a task is needed and can really improve the performances of connectionist models. I acknowledge the fact that there exist transformer-based models such as ViLT [7] which achieves astonishing results, but if we want to extend VQA to real-world scenarios and deeply understand how the decision process works, a NeSy approach is necessary. A step in this direction has been made by [14], where the neural and symbolic parts are not mixed together but simply focus on different but complementary tasks (*Type 3* categorization according to Kautz [6]). As the title of the paper recalls, the method employs deep neural networks for the visual recognition and language understanding part and a disentangled reasoning part using symbolic program execution (it's used the last type of knowledge representation illustrated in Figure 1).

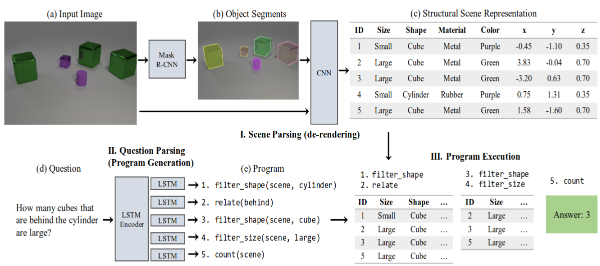

Figure 3: Architecture of the method implemented in [14].

As we can see in Figure 3 the model has three components: a scene parser, a question parser and a program executor. The visual module uses Mask R-CNN to generate segment proposals and predicts the categorical labels such as color, material, size, and shape, and a ResNet-34 to extract the spacial attributes such as pose and 3D coordinates. The question parser, using a bidirectional LSTM, maps an input question in natural language to a latent program (see Figure 3e). The program executor is a collection of functional modules in Python which host all logic operations behind the questions in the dataset (CLEVR) and outputs the final answer to the question. Besides the great results on CLEVR, this approach reduces the need of training data, it requires minimal computational and memory cost and more importantly, gives full transparency to the reasoning process and we are thus able to interpret and diagnose each execution step.

Robotics and Control. Safety and data efficiency are critical in constructing robotics systems and decision-making is also quite challenging. Therefore, symbolic additions could be a life saver for this area. In [10], a NeSy approach is proposed for producing robust neuro-symbolic programs to improve the autonomous driving design. Compared with neural network-based driving design, this method is more stable, safer and provides an interpretable symbolic decision-making process.

Common Sense. Common sense is something humans have by default and something AI methods inherently lack and aim to acquired. In recent years, numerous researchers have concentrated their efforts in this direction hoping to create a more refined form of intelligence. It is necessary in various applications, with conversational scenarios being particularly reliant on common sense to engage more natural and broad-ranging conversations. In [1] the authors tried to tackle this problem with the help of a neuro-symbolic theorem prover. In a conversational setting we have a speaker which tries to be understood by a system. Usually, when we talk to people, we use an "imprecise" language in the sense that most of the time we omit commonsense *presumptions*. For example, if I say to my system (that could be Alexa for example) *"If tomorrow there's public transportation strike, wake me up early"*, it needs to automatically infer the implicit presumption that I wish to be woken up early because if the bus is unavailable I will spend more time going to the university by walk. These are connections and mental associations that we make naturally and automatically, but unfortunately machines don't have this capability. Then, the aim of the work was to let the system understand such "imprecisely" stated commands of the human speaker and discover their hidden commonsense presumptions. The system has already a small knowledge base of commonsense facts and completes it as it interacts with the user. Basically, the neuro-symbolic theorem prover using Prolog as logic programming language and LSTM as neural model, infers commonsense presumption by extracting a chain of commonsense knowledge that "explains" the speaker's input. Formally, this reasoning chain is a proof tree to prove what the user said to the system (that's why "*theorem prover*"). Without going into further technical details, I think this approach, with all its limits, is a nice and first attempt that goes in the right direction.

Semantic Image Interpretation. Another fascinating application scenario where NeSy strategies are really useful to improve robustness and performance. Interpreting high-level semantics from visual perception is a task that can really benefit from the exploitation of external symbolic knowledge regarding the relations between visual semantics and structured properties of depicted objects. For instance in [4], *part-of* relations between objects are formalized in the form of first-order logic (and then fuzzy logic) using the LTN framework [2]. The paper regards only one of the main applications in which LTN can really help and thus, in Section 3, I will highlight its strengths. One of the authors of "Logic Tensor Networks for Semantic Image Interpretation" [4], is Luciano Serafini, an AI researcher who also co-wrote [2] and has given a talk to our "Seminars in Artificial Intelligence and Robotics" course at Sapienza just about NeSy. The paper aims to solve Semantic Image Interpretation (SII), that is the task of extracting structured semantic descriptions from images. This means that we classify image's bounding boxes and detect the relevant *part-of* relations between objects, applying LTNs. *Why LTNs?* They can express relational knowledge in FOL serving as constraints on the data-driven learning within tensor networks. In order to make the model learns and help to improve the performance of the Fast R-CNN, prior knowledge about the dependencies between bounding boxes types and *part-of* relations are specified in LTN in the form of logical axioms. The final results prove that not only LTN enhances the neural model, but it also adds robustness to the learning system when errors are present in the labels of the training data (presence of noisy data).

3 Logic Tensor Networks

Logic Tensor Networks (LTN) is a neurosymbolic framework (as already mentioned before) that supports querying, learning and reasoning. The authors of the paper [2] showed that LTN provides a uniform language to represent and compute efficiently many of the most important AI tasks, including Semantic Image Interpretation. Due to its immediate success the framework is available both in pytorch and tensorflow. Luciano Serafini, one of the main authors, said, during his talk to us at DIAG, that in order to combine learning and reasoning we need to *ground* symbols to reality (Symbol Grounding Problem) via information propagation. In LTN this is done by the use of an infinitely-valued *fuzzy logic* called Real Logic which let us to concretely interpret domains as tensors in the real field. In fact the grounding function associates a tensor of real numbers to any logical term and a real number in the interval $[0, 1]$ to any formula. For all the above reasons, it couldn't be possible to use discrete logics where the possible truth value are only 0 and 1. They introduced a new semantics where functions are interpreted as real-valued functions and predicates are interpreted as fuzzy relations on real tensors. The semantics of the *connectives* ($\wedge$, $\vee$, $\Rightarrow$, $\neg$) and *quantifiers* ($\forall$, $\exists$) is defined according to the semantics of first-order fuzzy logic [5]. Other two interesting quantifiers were added: *diagonal* quantification and *guarded* quantification which quantifies over a set of elements that satisfy some condition. Saying that, in this framework setting, functions and predicates are learnable (a gradient descent algorithm is involved) and indeed all fuzzy logic formulas with their connectives and quantifiers must be differentiable (authors made sure of it). The aim of *learning* within the LTN framework is to learn the grounding function constants, functions and predicates. The learning of constants grounding corresponds to the learning of embeddings, the learning of functions grounding corresponds to the learning of regression tasks and the learning of predicates grounding corresponds to the one of classification tasks in machine learning. That's exactly what I meant when talking about combining reasoning with learning. Real Logic can be used to specify a big number of tasks, as *clustering* for example, and by interpreting such a specification in Logic Tensor Networks we can effectively solve the problem.

Clustering is a form of unsupervised learning where the data is characterized by constraints alone. LTN can formulate such constraints and after defining domains, variables, predicates, axioms and grounding function, and training the model, we can see the clustering problem solved in Figure 4.

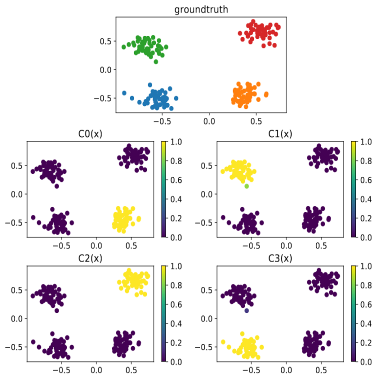

Figure 4: LTN solving a clustering problem by constraint optimization.

4 Conclusions

Combining the strengths of neural networks and symbolic reasoning, neuro-symbolic systems offer a unique and powerful framework for addressing complex problems. Although a strong NeSy system is still far from achieved and there are still challenges to be overcome, such as scaling the integration of knowledge representations to larger domains, I'm optimistic about its future and believe NeSy is a promising direction. This family of approaches can certainly start a path towards a more responsible and sustainable way of producing AI for the benefit of the planet and towards a more explainable and interpretable one. NeSy has all the credentials for becoming the key enabler to the next generation of AI!

References

[1] Forough Arabshahi, Jennifer Lee, Mikayla Gawarecki, Kathryn Mazaitis, Amos Azaria, and Tom Mitchell. 2021. Conversational neuro-symbolic commonsense reasoning.

[2] Samy Badreddine, Artur d'Avila Garcez, Luciano Serafini, and Michael Spranger. 2022. Logic tensor networks. *Artificial Intelligence*, 303:103649.

[3] Grady Booch, Francesco Fabiano, Lior Horesh, K Kate, Jonathan Lenchner, N Linck, Andrea Loreggia, K Murugesan, N Mattei, F Rossi, and Biplav Srivastava. 2021. Thinking fast and slow in ai.

[4] Ivan Donadello, Luciano Serafini, and Artur d'Avila Garcez. 2017. Logic tensor networks for semantic image interpretation.

[5] Petr Hájek. 1998. *Metamathematics of Fuzzy Logic*. Dordrecht, Boston and London: Kluwer Academic Publishers.

[6] Henry A. Kautz. 2022. The third ai summer: Aaai robert s. engelmore memorial lecture. *AI Magazine*, 43(1):105–125.

[7] Wonjae Kim, Bokyung Son, and Ildoo Kim. 2021. Vilt: Vision-and-language transformer without convolution or region supervision.

[8] Gary Marcus. 2020. The next decade in ai: Four steps towards robust artificial intelligence.

[9] Stuart Russell, Sabine Hauert, Russ Altman, and Manuela Veloso. 2015. Robotics: Ethics of artificial intelligence. *Nature*, 521:415–8.

[10] Jiankai Sun, Hao Sun, Tian Han, and Bolei Zhou. 2021. Neuro-symbolic program search for autonomous driving decision module design. In *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, pages 21–30. PMLR.

[11] Neil C. Thompson, Kristjan Greenewald, Keeheon Lee, and Gabriel F. Manso. 2022. The computational limits of deep learning.

[12] Wenguan Wang, Yi Yang, and Fei Wu. 2022. Towards data-and knowledge-driven artificial intelligence: A survey on neuro-symbolic computing.

[13] Yaqi Xie, Ziwei Xu, Mohan S. Kankanhalli, Kuldeep S. Meel, and Harold Soh. 2019. Embedding symbolic knowledge into deep networks.

[14] Kexin Yi, Jiajun Wu, Chuang Gan, Antonio Torralba, Pushmeet Kohli, and Joshua B. Tenenbaum. 2019. Neural-symbolic vqa: Disentangling reasoning from vision and language understanding.