Neurosymbolic Reinforcement Learning With Sequential Guarantees

Lennert De Smet[1], Gabriele Venturato[1], Giuseppe Marra[1], and Luc De Raedt[1,2]

[1] Department of Computer Science, KU Leuven, Belgium
[2] Center for Applied Autonomous Systems, Örebro, Sweden

Reinforcement learning (RL) is successfully applied in various domains [6, 7, 12, 13, 8], yet it struggles to provide safety and behavioural guarantees [3, 14]. Neurosymbolic AI (NeSy), with its ability to combine logical reasoning and neural perception, has been explored as a potential solution [14, 15, 9]. However, existing NeSy methods, such as probabilistic logic shields [14], focus on single-step guarantees, limiting their effectiveness where multistep reasoning is required. To extend NeSy to efficient sequential reasoning, we introduced *relational* neurosymbolic Markov models (NeSy-MMs) that have been shown promising results on generative tasks [2].

We propose a new framework for neurosymbolic reinforcement learning that incorporates relational NeSy-MMs as internal models for an RL agent. NeSy-MMs allow the agent to reason over multiple time steps and provide safety guarantees throughout the training process. We expect that this integration will provide policies that are resilient to test-time perturbations and adhere to given constraints over time, e.g. safety constraints.

Relational Neurosymbolic Markov Models

Relational NeSy-MMs are sequential probabilistic models over neurally-parametrised discrete-continuous random variables (Figure 1). They are probabilistic reasoning models that use random variables to model symbols, relations, and logical constraints. Neural predicates φ and φ_g map raw inputs (e.g. images) to symbols and vice versa, for *discriminative* and *generative* tasks. For instance, consider a MiniHack [11] game (Figure 2), where the monsters can attack the player. With NeSy-MM we can model the sequences of interactions as well as a safety constraint for the player not being attacked.

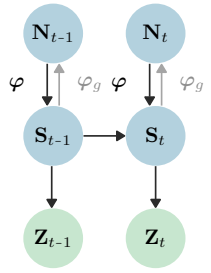

Fig. 1: NeSy-MMs sequentially factorise neural ($\mathbf{N}_t$) and symbolic states ($\mathbf{S}_t$) over time. They can be conditioned on evidence ($\mathbf{Z}_t$).

Because of the sequential structure of NeSy-MMs, part of the world model can be specified by replacing unknown transition functions by neural networks. Finally, NeSy-MMs are relational models, a popular and very expressive representation for representing states in, for instance, databases and planning [10]. Moreover, relational representations facilitate strong generalisation behaviour [4].

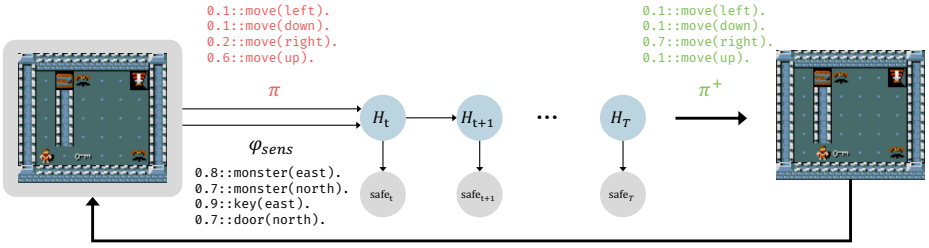

Fig. 2: NeSy-MMs used as neurosymbolic policies that provide safety guarantees. As in a classic RL algorithms, (→) executes an action in the environment, and (←) provides a new observation to the policy. The agent (bottom left) has to reach the staircases (top right). Each NeSy state H_i can contain raw data, or relational symbols. The transition from H_i to H_{i+1} can be fully logical, neural, or a mixture of both. Each state is also conditioned on a safety property, such that the agent is not killed by the monsters.

Neurosymbolic Reinforcement Learning by Example

The goal of using NeSy-MMs as RL policies is to obtain formal guarantees within a given time horizon. Previous efforts [14] have focused on providing single-step guarantees by shielding [5] a neural policy with a probabilistic logic program [1]. While effective, this approach does not scale to multistep guarantees because of the #P-hardness of its inference procedure. NeSy-MMs resolve this problem by using unbiased approximate inference techniques instead. Consider again a MiniHack level where the agent is in a room with two monsters and has to reach a goal (Figure 2). The optimal strategy in this case is to take the key and wait to lure the two monsters away from the goal. Only once the monsters are close enough and the agent has the key, it can move through the corridor, open the door, and move safely to the goal before the monsters can catch up. Hence, safely reaching the goal is not something that can be decided by single-step reasoning. Concretely, if the agent is governed by a policy π and a sensor φ_{sens} gives an estimate of the current state of the game, then these will form the input to a NeSy-MM. The NeSy-MM then updates the policy to $\pi^+(a|\square) = \pi(a \mid \text{safe}_{t:T}, \square)$ that incorporates the safety constraints via approximate Bayesian inference. Finally, we want to obtain a policy such that,

$$P_{\pi^+}(\text{safe}_{t:T} \mid \square) \geq \underbrace{P_{\pi^+}(\text{safe}_t \mid \square) \geq P_{\pi}(\text{safe}_{t:T} \mid \square)}_{\text{from [14]}}.$$

This means our NeSy policy is going to be safer than the single time-step shielded policy from [14], that is in turn safer than the unshielded policy, for any time horizon. In the future, we aim to empirically verify this idea and more closely integrate NeSy-MMs into the RL framework by analysing the behaviour of the expected reward in the presence of neurosymbolic policies.

References

[1] De Raedt, L., Kimmig, A., Toivonen, H.: Problog: A probabilistic prolog and its application in link discovery. In: IJCAI. Hyderabad (2007)

[2] De Smet, L., Venturato, G., De Raedt, L., Marra, G.: Neurosymbolic markov models. In: ICML 2024 Workshop on Structured Probabilistic Inference & Generative Modeling (2024)

[3] Garcıa, J., Fernández, F.: A comprehensive survey on safe reinforcement learning. Journal of Machine Learning Research **16**(1), 1437–1480 (2015)

[4] Hummel, J.E., Holyoak, K.J.: A symbolic-connectionist theory of relational inference and generalization. Psychological review **110**(2), 220 (2003)

[5] Jansen, N., Könighofer, B., Junges, S., Serban, A., Bloem, R.: Safe reinforcement learning using probabilistic shields. In: 31st International Conference on Concurrency Theory (CONCUR 2020). Schloss-Dagstuhl-Leibniz Zentrum für Informatik (2020)

[6] Juang, B.H., Rabiner, L.R.: Hidden markov models for speech recognition. Technometrics **33**(3), 251–272 (1991)

[7] Khiatani, D., Ghose, U.: Weather forecasting using hidden markov model. In: 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN). pp. 220–225. IEEE (2017)

[8] Mor, B., Garhwal, S., Kumar, A.: A systematic review of hidden markov models and their applications. Archives of Computational Methods in Engineering **28**, 1429 – 1448 (2020)

[9] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, F.: Deep learning and process understanding for data-driven earth system science. Nature **566**(7743), 195–204 (2019)

[10] Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Pearson, Hoboken, 4th edition edn. (2020)

[11] Samvelyan, M., Kirk, R., Kurin, V., Parker-Holder, J., Jiang, M., Hambro, E., Petroni, F., Kuttler, H., Grefenstette, E., Rocktäschel, T.: Minihack the planet: A sandbox for open-ended reinforcement learning research. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1) (2021)

[12] Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., et al.: Mastering atari, go, chess and shogi by planning with a learned model. Nature **588**(7839), 604–609 (2020)

[13] Van Roy, M., Robberechts, P., Yang, W.C., De Raedt, L., Davis, J.: A markov framework for learning and reasoning about strategies in professional soccer. Journal of Artificial Intelligence Research **77**, 517–562 (2023)

[14] Yang, W.C., Marra, G., Rens, G., De Raedt, L.: Safe reinforcement learning via probabilistic logic shields. In: Elkind, E. (ed.) Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. pp. 5739–5749. International Joint Conferences on Artificial Intelligence Organization (8 2023). https://doi.org/10.24963/ijcai.2023/637, main Track

[15] Zhang, H., Dang, M., Peng, N., Van Den Broeck, G.: Tractable control for autoregressive language generation. In: Krause, A., Brunskill, E., Cho,

K., Engelhardt, B., Sabato, S., Scarlett, J. (eds.) Proceedings of the 40th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 202, pp. 40932–40945. PMLR (23–29 Jul 2023)