Causality-Driven Reinforcement Learning

Giovanni Briglia

June 2024

Motivation and Objective: Reinforcement learning (RL) is becoming a powerful framework for tackling complex sequential decision-making challenges in diverse fields, thanks to its ability to learn action policies from experience [27]. Advancements in RL techniques have been achieved recently in single and multi-agent scenarios, like, for example, robotics, healthcare [7], traffic management [6] and intelligent transportation systems [12], scheduling [26], Internet of Things [4], and autonomous driving [16]. Nevertheless, significant challenges persist, including:

1. Extensive and inefficient training processes with long simulation times [27, 28]: (deep) RL algorithms typically require substantial data to learn effectively, which can be costly and time-consuming to gather.

2. It is difficult to guarantee safety boundaries for agent behaviour, particularly during early training phases [2] where exploratory actions are typically prevalent. In particular, ensuring that agents act safely and reliably in real-world applications remains a significant challenge.

3. Lack of interpretability: (deep) RL algorithms usually produce impenetrable "black boxes" as policies, making it difficult to understand and interpret their decision-making processes [23].

4. Difficulty in transferring learned knowledge across different environments or tasks [33].

Causal modelling offers a framework for learning and reasoning about how actions impact observed variables *causally* [22]. It aids decision-making by capturing and expressing cause-effect relationships between domain variables within Causal Models (CMs).

In this research, I aim to incorporate CMs into RL to tackle the previously identified challenges, that is, enhancing efficiency and interoperability while ensuring greater safety and generalisation power. CMs, in fact, represent a "white-box" methodology that can help achieve such ambition by enabling reasoning over the dynamics of the environment and the agent-environment interaction.

State of Art: This research project primarily focuses on two key components: RL and CMs. RL problems are typically modelled in terms of the Markov Decision Process (MDP) framework and its derivatives, such as partially observable and multi-agent cases. In its basic form, MDP includes an agent interacting with the environment, where the agent takes actions in given states and observes the resulting rewards and next states. Conversely, a CM is a mathematical framework used to represent and analyze the causal relationships between variables in a system. It aims to elucidate how changes in one or more variables influence others within the system. A widely used causal model is the Structural CM (SCM), often represented as a Directed Acyclic Graph (DAG) along with a set of structural equations. In an SCM, each node in the DAG corresponds to a variable in the system, and directed edges between nodes indicate direct causal influences. CMs can be given or learnt via Causal Discovery (CD) techniques [23, 13, 18, 31, 17], and then exploited for predicting systems states given actions, or planning actions given system states (current and desired), via Causal Inference (CI) [23, 13, 22, 15], that enables estimating how one variable changes when another variable assumes a certain value. I aim to integrate CMs with RL by observing that the causal graph might capture only the causal links relevant to its environment and/or task, not all of them. These "core" mechanisms are applicable across similar environments and/or similar tasks, hence inherently transferable. For instance, such core environment dynamics could focus only on (causal) transitions that directly influence positive and negative rewards, and/or the successful completion of the episode.

The connections and usefulness of CMs in RL research have already been motivated in [10, 3, 24]. Current efforts can be categorized into three main areas [10, 30, 8]: i) using CMs to enhance RL algorithms; ii) leveraging RL trajectories to infer CMs of the environment; iii) simultaneously addressing both by learning CMs through an RL policy while refining that policy using the learned CM. Unlike other frameworks [20, 19, 9, 14, 25, 32, 29], the approach I propose offers the following benefits:

- It can be plugged into any RL algorithm, as it acts as an action filter that may be placed before the action selection step to restrict admissible actions;

- It is adaptable to any environment, as it requires only the standard prior knowledge assumed in RL (the same assumed by RL algorithms): the agent's action space, observation space, and the ability to observe the reward obtained after each action (if any);

- It only focuses on the core environment dynamics. On the one hand, this simplifies CD by reducing the number of considered causal variables and links; on the other hand, it enables greater generalisation as the CM is not tailored to *every* environment dynamic.

Research Questions: The primary objective of this project is to create a theoretical and practical framework for incorporating causal knowledge into both model-free and model-based (deep) RL algorithms, in single and multi-agent systems. Specifically, this research seeks to answer the following critical questions:

1. Can we enhance the efficiency, safety, and generalization capabilities of arbitrary model-free RL algorithms by incorporating in action selection strategies the CMs of core environmental dynamics?
2. Can this approach be applied to any combination of target environment, RL algorithm, and task?
3. What is the minimal CM necessary to achieve a measurable improvement?

Methodology: To address the proposed research questions, I aim to:

1. Broaden and complete an already ongoing comprehensive literature review of CD and CI algorithms, highlighting their applications and use cases, and identifying scenarios where certain algorithms outperform others. Based on this review, I plan to develop my own CD and CI library and compare it with existing ones.
2. Develop a formal definition of a "minimal CM" and "core environment dynamics", along with related properties, to highlight their role in enhancing RL.
3. Provide both formal and empirical evidence demonstrating how these concepts influence the safety, efficiency, and generalization capabilities of (deep) RL algorithms.
4. Develop a framework within a modular software library designed to support virtually any model-free RL algorithm and operating environment, with the assumption being that the environment allows the agent to observe the obtained reward (typical in RL).
5. Test this framework to evaluate how much and under which conditions the CM can improve sample efficiency, and whether this than an impact on RL performance.
6. Test this framework to assess the generalization capabilities of the algorithms when reusing the same CM learnt in one environment in other similar environments.

Overall, the idea is that (deep) RL algorithms should focus on solving the specific problem optimally, while CMs take care of efficiency, safety, and generalisability.

The evaluation of the proposed causal (deep) RL algorithms will follow the experimental design principles recommended by [5, 11, 21, 1]. These guidelines propose conducting 10+ simulations for the same task using different seeds. The Inter-Quartile Mean will be employed to assess each performance metric, thereby mitigating the impact of outliers. Metrics of interest include: cumulative and average reward per episode; number of actions to finish episodes; time to complete the episode; success/failure when available for the task at hand. Causality-enhanced RL algorithms will be compared to their "vanilla" versions to assess improvement.

Conclusion: In the long term, my aspirations reach beyond academia. My dedication to leverage AI for societal benefit drives my commitment to continuous learning and innovation. I am particularly focused on developing trustworthy AI, improving the explainability and safety of decision-making processes to positively impact society.

I eagerly anticipate the opportunity to discuss my application further; I have the passion, ability, and tenacity to reach my goals within your excellent program, and I hope you will give me the opportunity to demonstrate it. Thank you for considering my application.

Sincerely,

Giovanni Briglia

References

[1] Rishabh Agarwal et al. "Deep reinforcement learning at the edge of the statistical precipice". In: *Advances in neural information processing systems* 34 (2021), pp. 29304–29320.

[2] Dario Amodei et al. "Concrete problems in AI safety". In: *arXiv preprint arXiv:1606.06565* (2016).

[3] James Bannon et al. "Causality and batch reinforcement learning: Complementary approaches to planning in unknown domains". In: *arXiv preprint arXiv:2006.02579* (2020).

[4] Wuhui Chen et al. "Deep Reinforcement Learning for Internet of Things: A Comprehensive Survey". In: *IEEE Communications Surveys /& Tutorials* 23.3 (2021), pp. 1659–1692.

[5] Cédric Colas, Olivier Sigaud, and Pierre-Yves Oudeyer. "Gep-pg: Decoupling exploration and exploitation in deep reinforcement learning algorithms". In: *International conference on machine learning*. PMLR. 2018, pp. 1039–1048.

[6] Ioan-Sorin Comşa et al. "Towards 5G: A reinforcement learning-based scheduling solution for data traffic management". In: *IEEE Transactions on Network and Service Management* 15.4 (2018), pp. 1661–1675.

[7] Antonio Coronato et al. "Reinforcement learning for intelligent healthcare applications: A survey". In: *Artificial Intelligence in Medicine* 109 (2020), p. 101964.

[8] Zhihong Deng et al. "Causal reinforcement learning: A survey". In: *arXiv preprint arXiv:2307.01452* (2023).

[9] Ivan Feliciano-Avelino et al. "Causal based action selection policy for reinforcement learning". In: *Mexican International Conference on Artificial Intelligence*. Springer. 2021, pp. 213–227.

[10] Samuel J Gershman. "Reinforcement learning and causal models". In: *The Oxford handbook of causal reasoning* 1 (2017), p. 295.

[11] Rihab Gorsane et al. "Towards a standardised performance evaluation protocol for cooperative marl". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 5510–5521.

[12] Ammar Haydari and Yasin Yılmaz. "Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey". In: *IEEE Transactions on Intelligent Transportation Systems* 23.1 (2022), pp. 11–32. DOI: 10.1109/TITS.2020.3008612.

[13] Christina Heinze-Deml, Marloes H Maathuis, and Nicolai Meinshausen. "Causal structure learning". In: *Annual Review of Statistics and Its Application* 5 (2018), pp. 371–391.

[14] Xing Hu et al. "Causality-driven hierarchical structure discovery for reinforcement learning". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 20064–20076.

[15] Andrew Jesson et al. "Scalable sensitivity and uncertainty analyses for causal-effect estimates of continuous-valued interventions". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 13892–13907.

[16] B Ravi Kiran et al. "Deep reinforcement learning for autonomous driving: A survey". In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (2021), pp. 4909–4926.

[17] Thuc Duy Le et al. "A fast PC algorithm for high dimensional causal discovery with multi-core PCs". In: *IEEE/ACM transactions on computational biology and bioinformatics* 16.5 (2016), pp. 1483–1495.

[18] Daniel Malinsky and David Danks. "Causal discovery algorithms: A practical guide". In: *Philosophy Compass* 13.1 (2018), e12470.

[19] Arquímides Méndez-Molina, Eduardo F Morales, and L Enrique Sucar. "CARL: A Synergistic Framework for Causal Reinforcement Learning". In: *IEEE Access* 11 (2023), pp. 126462–126481.

[20] Arquímides Méndez-Molina et al. "Causal Based Q-Learning." In: *Res. Comput. Sci.* 149.3 (2020), pp. 95–104.

[21] Andrew Patterson et al. "Empirical design in reinforcement learning". In: *arXiv preprint arXiv:2304.01315* (2023).

[22] Judea Pearl. "Graphical Models for Probabilistic and Causal Reasoning." In: *Computing Handbook, 3rd ed.(1)* (2014), pp. 44–1.

[23] Bernhard Schölkopf. "Causality for machine learning". In: *Probabilistic and causal inference: The works of Judea Pearl*. 2022, pp. 765–804.

[24] Bernhard Schölkopf et al. "Toward causal representation learning". In: *Proceedings of the IEEE* 109.5 (2021), pp. 612–634.

[25] Maximilian Seitzer, Bernhard Schölkopf, and Georg Martius. "Causal influence detection for improving efficiency in reinforcement learning". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 22905–22918.

[26] Chathurangi Shyalika, Thushari Silva, and Asoka Karunananda. "Reinforcement learning in dynamic task scheduling: A review". In: *SN Computer Science* 1.6 (2020), p. 306.

[27] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[28] Hado Van Hasselt, Arthur Guez, and David Silver. "Deep reinforcement learning with double q-learning". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 30. 1. 2016.

[29] Chao-Han Huck Yang et al. "Causal inference q-network: Toward resilient reinforcement learning". In: *Self-Supervision for Reinforcement Learning Workshop-ICLR 2021*. 2021.

[30] Yan Zeng et al. "A survey on causal reinforcement learning". In: *arXiv preprint arXiv:2302.05209* (2023).

[31] Yujia Zheng et al. "Causal-learn: Causal discovery in python". In: *Journal of Machine Learning Research* 25.60 (2024), pp. 1–8.

[32] Wenxuan Zhu, Chao Yu, and Qiang Zhang. "Causal deep reinforcement learning using observational data". In: *arXiv preprint arXiv:2211.15355* (2022).

[33] Zhuangdi Zhu et al. "Transfer learning in deep reinforcement learning: A survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).