Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)
Doctoral Consortium Track

# Combining Reinforcement Learning and Causal Models for Robotics Applications

**Arquímides Méndez-Molina**
Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Tonantzintla, Puebla, México
arquimides.mendez@gmail.com

## Abstract

The relation between Reinforcement learning (RL) and Causal Modeling(CM) is an underexplored area with untapped potential for any learning task. In this extended abstract of our Ph.D. research proposal, we present a way to combine both areas to improve their respective learning processes, especially in the context of our application area (service robotics). The preliminary results obtained so far are a good starting point for thinking about the success of our research project.

## 1 Introduction

Both Reinforcement Learning (RL) and Causal Modeling (CM) are indispensable part of machine learning and each plays an essential role in artificial intelligence, however, they are usually treated separately, despite the fact that both are directly relevant to problem solving processes. On the one hand, reinforcement learning has proven to be successful in many sequential decision problems (been robotics a prominent field of application); however, it takes a long time to converge to acceptable policies; and on the other hand, causal graphical models is an area that is recently receiving a lot of attention for its ability to explain physical phenomena, however, it is difficult to build causal models due to the amount of data required and the need for interventions.

At present, the first works focusing on the relationship between these areas are beginning to emerge. The existing works can be divided in three main groups. First, those that use Causal Models as side information to improve Reinforcement Learning Algorithms ($CM \rightarrow RL$). Most of those works are limited to Multi Armed Bandit (MAB) settings [Gershman, 2017; Yu *et al.*, 2019; Dasgupta *et al.*, 2019] and also a completely defined Causal Model is assumed, which is hard to obtain in the real world. In the second group are those that use RL to learn causal relationships of the environment directly from data ($RL \rightarrow CM$). The main limitation of these reduced group of works is that the structure is given, so only parameters are learned [Madumal *et al.*, 2019] or that the agent instead of learning a policy for the task instead it uses reinforcement learning as a search strategy to find the graph that achieves the best reward [Zhu and Chen, 2019]. Finally, there are a more recent group [Gonzalez-Soto *et al.*, 2018; Nair *et al.*, 2019; Kansky *et al.*, 2017] that simultaneously do both tasks: learn causal effects from an agent communicating with the environment, and then optimize its policy based on the learned causal relations ($RL \leftrightarrow CM$). Among the limitations we can mention the use of only observational data and also structural assumptions according to the specific problem.

In this Ph.D. research we focus mainly in this last problem ($RL \leftrightarrow CM$), in particular, in the context of service robots. The main novelty in our proposal is that we do not limit ourselves to observational data, but we are able to perform and use interventions, and also that we use partial causal models at the early stages of the learning process.

## 2 Research Problem

We can state our research problem as follows: How can we provide an intelligent agent with the ability to simultaneously learn causal relationships and perform efficient induction of task policies, based on the experiences obtained during the reinforcement learning process?

Formally, let $G$ be a causal graphical model and let $M = (S, A, T, R)$ a sequential Markov Decision Problem (MDP) whose actions ($A$), states ($S$), and rewards ($R$) are causally related and corresponds to variables in $G$. Let ($RL_{\pi/Q}$) denote the process of learning a policy $\pi$ and a value function ($Q$) for $M$ following a reinforcement learning algorithm. Let $CL_G$ denote the learning process of $G$ following a causal discovery algorithm. Consider a decision maker who does not know the parameters nor the structure of $G$ which control the probabilities of observing a consequence (transition $T$ or reward $R$) given an action $a \in A$. We want to solve how to integrate $CL_G$ during $RL_{\pi/Q}$ such that the system can learn faster/better $G$, $\pi$ and $Q$.

Among the elements that support the feasibility of this research is that in the area of robotics interventions (experiments) can be made, which is prohibitive in other areas, which facilitates causal discovery. Also, if an agent knows the possible consequences of its actions, it can make better selections of them. This is particularly relevant in RL because that knowledge, which can be given by a causal model, can significantly reduce the exploration process and therefore accelerate the learning process (as will be seen in the preliminary results).

### 2.1 Expected Contributions

The expected contributions of this work are: (i) an RL algorithm that efficiently uses a CM to learn faster policies, (ii) an algorithm for causal discovery that uses guided interventions from a robotic agent to learn faster and better models, and (iii) a system that incorporates the above two parts to solve robotics problems.

If we succeed, a very attractive set of potential applications arises, such as: explanation (the agent can explain the reason for its actions using causal models), transfer (the learned causal models can be directly reused between similar tasks), and efficiency (reduce the long learning times required by reinforcement learning).

## 3 Scope and Limitations

The proposed algorithm will not be limited to applications in robotics, but it is necessary that the addressed decision problem allows interventions. We are aware that as the research develops, other limitations will arise.

## 4 Preliminary Results

We decided first to explore how a given causal knowledge could be used to learn new tasks more efficiently. For this, we proposed and tested an action selection algorithm guided by causal knowledge in a simple navigation task that can be consulted for more details in [Méndez-Molina *et al.*, 2020][1]. As our baseline we implemented a vanilla version of the Q-learning algorithm and we compared it with our version which we denominate Causal Q-learning. Although the problem attacked is simple because all the causes we have are direct and observable, the experimental results show that using causal models in the Q-learning action selection step leads to a jump-start reward and faster convergence in all the experiments. At present, we are working on how to use interventional data in combination with observational data generated by an RL agent for Causal Discovery. We are implementing an adaptive strategy based on [Eberhardt, 2007] that tells the agent what is next better intervention.

## 5 Directions for the Remaining Work

We continue working on how to define and narrow down causal discovery using RL. It is very likely that the models discovered by the RL agent will not be complete, yet still useful. We need to define a strategy that combines the stages of causal discovery and causal inference with the stages of exploration and exploitation in reinforcement learning. Our preliminary idea is in some way inspired by Dyna-Q [Sutton and Barto, 1998]. In this algorithm a model-based system is used to produce simulated experience from which a model-free system could learn. In our case the model would be a causal model. Some important challenges are how to: (i) effectively and simultaneously couple RL and CM with partial models and (ii) reduce the number of data required by both parts.

[1]This work was presented in the Causal Reasoning Workshop at the Mexican International Conference on Artificial Intelligence 2019

### 5.1 Evaluation

We plan to evaluate our work using different strategies according to each phase. For the first phase $(CM \rightarrow RL)$: the RL algorithm that used the causal model will be compared against a classical RL algorithm. The proposed algorithm will be considered successful if it converges faster or obtains higher rewards for a defined number of episodes. For the second phase $(RL \rightarrow CM)$: the algorithm's ability for recovering the skeleton and V-structures of the ground truth causal structures will be assessed. Here we will used either domains where there is a ground truth available or where there is an expert that can validate the model. Finally for the final phase $(RL \leftrightarrow CM)$: a set of tasks of increasing complexity in the field of robotics will be design in order to test the algorithm. The results obtained using the algorithm will be compared against traditional reinforcement learning algorithms and other related approaches proposed in the literature. Also, we pretend to explore the transfer learning ability of our algorithm by using the causal model generated for a simple task as input to another more complex task.

## References

[Dasgupta *et al.*, 2019] Ishita Dasgupta, Jane X. Wang, and et. al. Causal reasoning from meta-reinforcement learning. *CoRR*, abs/1901.08162, 2019.

[Eberhardt, 2007] Frederick Eberhardt. *Causation and intervention*. PhD thesis, Department of Philosophy, Carnegie Mellon University, 2007.

[Gershman, 2017] Samuel J. Gershman. Reinforcement learning and causal models. *The Oxford handbook of causal reasoning*, 1:295, 2017.

[Gonzalez-Soto *et al.*, 2018] Mauricio Gonzalez-Soto, Luis E. Sucar, and Hugo J. Escalante. Playing against nature: causal discovery for decision making under uncertainty. *arXiv preprint arXiv:1807.01268*, 2018.

[Kansky *et al.*, 2017] Ken Kansky, Tom Silver, and et al. Schema networks: Zero-shot transfer with a generative causal model of intuitive physics. In *Proc of the 34th Int Conf on ML*, pages 1809–1818, 2017.

[Madumal *et al.*, 2019] Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. Explainable reinforcement learning through a causal lens. *arXiv preprint arXiv:1905.10958*, 2019.

[Méndez-Molina *et al.*, 2020] Arquímides Méndez-Molina, Ivan Feliciano Avelino, Eduardo F. Morales, and Luis E. Sucar. Causal based q-learning. *Research in Computing Science*, 149:95–104, 2020.

[Nair *et al.*, 2019] Suraj Nair, Yuke Zhu, Silvio Savarese, and Li Fei-Fei. Causal induction from visual observations for goal directed tasks. *arXiv preprint arXiv:1910.01751*, 2019.

[Sutton and Barto, 1998] Richard S. Sutton and Andrew G. Barto. *Reinforcement learning - an introduction*. Adaptive computation and machine learning. MIT Press, 1998.

[Yu *et al.*, 2019] Chao Yu, Yinzhao Dong, and et al. Incorporating causal factors into reinforcement learning for dynamic treatment regimes in HIV. *BMC Med. Inf. & Decision Making*, 19-S(2):19–29, 2019.

[Zhu and Chen, 2019] Shengyu Zhu and Zhitang Chen. Causal discovery with reinforcement learning. *CoRR*, abs/1906.04477, 2019.