

---

# Project Proposal: Discover and Exploit Causality in the Environment via a Combination of Probabilistic Graphical Models and Deep Reinforcement Learning Algorithms

---

Jincheng Zhou    Yang Li    Jiaxin Cheng    Tianmu Lei  
University of Southern California  
{jinchenz, yli546, jiaxinch, tianmule}@usc.edu

## 1 Goal of the proposed project

Devise and implement a model-based reinforcement learning algorithmic framework capable of discovering the causal relations in the environment by explicitly learning a Structured Causal Model (SCM) and using the learned causal model to guide its policy search.

## 2 Motivation

The ability to discover causal relations in the environment is vital for human to survive as the truly intelligent creatures on earth. Without this capability, it is not possible for the mankind to probe the nature and form sophisticated scientific theories such as mathematics, physics and chemistry that define the modern civilization. This goes the same with artificial intelligent agents. Being able to understand its environment by learning the causal mechanism in the surroundings, such an agent can efficiently learn a better policy by reflecting on its past actions through counterfactual reasoning [1]. Besides, because it has the causal models in mind, the agent can also explain its actions and strategies through causal inference, and such an explicitly encoded causal model can be later transferred and utilized by other agents operating in similar environments. Some argues that such a capability is necessary for any agent to be considered as a "general intelligence" [2].

However, discovering and exploiting the causal structure in the environment is a crucial challenge for intelligent agents. Current state-of-the-art reinforcement learning algorithms that rely on deep neural networks are experts in discovering correlations in the environment, but correlations does not imply causation. Without an explicit structural component that serves as the encoding of causal relations in the environment, such agents not only suffer from terrible sample efficiency, but are notoriously bad with respect to the explainability of their behaviors and transferability as well.

Inspired by the ground-breaking work in [1] and the theoretical foundation laid in [3], as well as some preliminary but interesting experiments conducted in [4], we believe that it is feasible to devise an algorithmic framework that combines state-of-the-art Deep RL algorithms with Causal Bayesian Networks (a.k.a. Structured Causal Models), which would endow the agent with the ability to understand causality.

## 3 Problem Formulation

- Input: Any reinforcement learning environment, either synthetic (artificial environment with randomly generated Causal Bayesian Network as its latent causal mechanism), control-theoretic (CartPole, LunarLanding, etc.), or gaming (Atari games that are commonly used as the benchmark environment for RL algorithms)

- Output:
  1. An agent that performs well in the given environment
  2. A Structured Causal Model that the agent learned during its interactions with the environment.

## 4 Method

The general framework will be a combination of state-of-the-art RL algorithms (PPO, DDPG, SAC, etc.) with a mechanism that provides functionalities to perform inference (either exact or approximate) as well as learning on a Structured Causal Model. Learning and inference is quite easy on the part of SCM since past literature on probabilistic inference already provides us with efficient and theoretically sound algorithms to learn a Probabilistic Graphical Model.

However, there are multiple options with respect to the specific design of the RL agent and how deep neural networks play a part in this game.

The simplest option is to consider the agent as a model-based reinforcement learning agent, where it learns the SCM as its model of the environment. However, it is unclear in the current stage how the agent could learn such a SCM and using use this SCM as part of its policy generation.

The second option, inspired by [4], is to use a meta-reinforcement learning agent, powered by Recurrent Neural Network, that learns an optimal meta-policy which will guide the agent when and how to interact with the environment so as to obtain the causal model with maximum efficiency. In other words, this meta-RL agent will learn *how* to learn the causal relations.

Another option, is to apply the framework described in [5], which generalizes reinforcement learning problems as inference problems in specially constructed Probabilistic Graphical Models. This framework is sometimes also termed "maximum entropy reinforcement learning." Because the Structured Causal Model is defined by a Causal Bayesian Network, which is essentially a Probabilistic Graphical Model, this may potentially results in an architecture that coherently integrate the SCM with the policy network.

There are also other aspects during the learning process which involves special designs and to which the learned SCM may contribute. For example, [1] describes how a Structured Causal Model can be used by an off-policy algorithm to guide its policy search and demonstrates the benefits, though this method comes with strong assumptions and certain limitations. We may try to integrate this method as part of our algorithmic framework so that it becomes more powerful.

We will go about trying all these ideas and eventually come up with an elegant and self-contained framework.

## 5 Milestones

- Implement a simple model-based RL agent that learns and uses the SCM as part of its policy.
- Test its performance on synthetic environments by comparing it to baseline (random) agents and agents that uses traditional deep RL algorithms (PPO, Actor-Critic, etc.).
- Devise a theoretical way for the agent to identify the important events happened in the environment so that it is able to encode this event as the nodes in the SCM.
- Implement an advanced agent with such event encoding capability
- Tests the agent's performance on more meaningful environments, such as control-theoretic and/or gaming environments, and see whether it learns a meaningful and useful causal model
- Integrate the method described in [1] in the agent's off-policy learning algorithms.
- Tests the agent's performance and compare them to state-of-the-art Deep Reinforcement Learning algorithms.

## 6 Expected Results

We expect that our agent, endowed with the ability to discover causality and exploit it for policy search, will learn a valid and useful causal model of the environment as well as performs no worse than state-of-the-art algorithms while have much better sample efficiency.

## 7 Related Works

[6] [7] [8] [9] [10] [3] [11] [12] [13] [4] [14] [15] [5] [16] [17] [18] [2] [19] [1]

## References

- [1] Lars Buesing, Theophane Weber, Yori Zwols, Nicolas Heess, Sebastien Racaniere, Arthur Guez, and Jean-Baptiste Lespiau. Woulda, coulda, shoulda: Counterfactually-guided policy search. In *International Conference on Learning Representations*, 2019.
- [2] Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, Inc., New York, NY, USA, 1st edition, 2018.
- [3] M. Gonzalez-Soto, L. E. Sucar, and H. J. Escalante. Playing against Nature: causal discovery for decision making under uncertainty. *arXiv e-prints*, page arXiv:1807.01268, Jul 2018.
- [4] Ishita Dasgupta, Jane Wang, Silvia Chiappa, Jovana Mitrovic, Pedro Ortega, David Raposo, Edward Hughes, Peter Battaglia, Matthew Botvinick, and Zeb Kurth-Nelson. Causal Reasoning from Meta-reinforcement Learning. *arXiv e-prints*, page arXiv:1901.08162, Jan 2019.
- [5] Sergey Levine. Reinforcement learning and control as probabilistic inference: Tutorial and review. *CoRR*, abs/1805.00909, 2018.
- [6] Razvan Pascanu, Yujia Li, Oriol Vinyals, Nicolas Heess, Lars Buesing, Sebastien Racanière, David Reichert, Théophane Weber, Daan Wierstra, and Peter Battaglia. Learning model-based planning from scratch. *arXiv e-prints*, page arXiv:1707.06170, Jul 2017.
- [7] Olivier Goudet, Diviyani Kalainathan, Philippe Caillou, Isabelle Guyon, David Lopez-Paz, and Michèle Sebag. Causal Generative Neural Networks. *arXiv e-prints*, page arXiv:1711.08936, Nov 2017.
- [8] Karren D. Yang, Abigail Katcoff, and Caroline Uhler. Characterizing and Learning Equivalence Classes of Causal DAGs under Interventions. *arXiv e-prints*, page arXiv:1802.06310, Feb 2018.
- [9] Sergey Levine. Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review. *arXiv e-prints*, page arXiv:1805.00909, May 2018.
- [10] Yao Liu, Omer Gottesman, Aniruddh Raghu, Matthieu Komorowski, Aldo Faisal, Finale Doshi-Velez, and Emma Brunskill. Representation Balancing MDPs for Off-Policy Policy Evaluation. *arXiv e-prints*, page arXiv:1805.09044, May 2018.
- [11] Thanard Kurutach, Aviv Tamar, Ge Yang, Stuart Russell, and Pieter Abbeel. Learning Plannable Representations with Causal InfoGAN. *arXiv e-prints*, page arXiv:1807.09341, Jul 2018.
- [12] Vincent François-Lavet, Yoshua Bengio, Doina Precup, and Joelle Pineau. Combined Reinforcement Learning via Abstract Representations. *arXiv e-prints*, page arXiv:1809.04506, Sep 2018.
- [13] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, and Maosong Sun. Graph Neural Networks: A Review of Methods and Applications. *arXiv e-prints*, page arXiv:1812.08434, Dec 2018.
- [14] Andrew Melnik, Sascha Fleer, Malte Schilling, and Helge Ritter. Modularization of End-to-End Learning: Case Study in Arcade Games. *arXiv e-prints*, page arXiv:1901.09895, Jan 2019.
- [15] Kailash Budhathoki. Rule discovery for exploratory causal reasoning. In *Conference on Neural Information Processing Systems*, 2018.

- [16] Samuel J Gershman, Kenneth A Norman, and Yael Niv. Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences*, 5:43 – 50, 2015. Neuroeconomics.
- [17] Samuel J Gershman. Reinforcement learning and causal models. 2017.
- [18] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
- [19] Pim de Haan, Dinesh Jayaraman, and Sergey Levine. Causal confusion in imitation learning. 2018.