

# A Guiding Principle for Causal Decision Problems

M. Gonzalez-Soto, L.E. Sucar, H.J. Escalante

Coordinación de Ciencias Computacionales,  
Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE)  
Luis Enrique Erro 1, Santa María Tonanzintla, Puebla, México.

## Abstract

We define a Causal Decision Problem as a Decision Problem where the available actions, the family of uncertain events and the set of outcomes are related through the variables of a Causal Graphical Model  $\mathcal{G}$ . A solution criteria based on Pearl's Do-Calculus and the Expected Utility criteria for rational preferences is proposed. The implementation of this criteria leads to an on-line decision making procedure that has been shown to have similar performance to classic Reinforcement Learning algorithms while allowing for a causal model of an environment to be learned. Thus, we aim to provide the theoretical guarantees of the usefulness and optimality of a decision making procedure based on causal information.

## Introduction

Decision making under uncertainty is a fundamental part of intelligent reasoning (?) and many real-world applications rely on decisions made by an autonomous agent, such as self-driving cars. Current decision making methods rely on associative methods, which find only statistical patterns in data. On the contrary, causal knowledge allows both for planning and counterfactual reasoning as well as interpretability and explainability (?), (?), (?). We propose in this work a way of considering causal information for decision making under uncertainty with rational preferences in such a way that optimal actions are chosen according to the principle of Maximum Expected Utility, which is the formal criteria for making choices under uncertain conditions if rationality is assumed (?).

## Rationality and Expected Utility

Rationality in a Decision Making setting is defined axiomatically in a way that the *preferences* of a decision maker are logically consistent. If rational preferences are assumed, then it is known that the coherent criteria for making choices is the maximization of expected utility, either with respect to a known utility function and probability distribution or a pair of *subjective* objects (?), (?). Rational decision making has been the standard theory in economics both as a *normative* and a *descriptive* theory of human behavior and it has been

the subject of multiple debates (?), (?). In this work we pretend to take a *normative* view for a rational decision maker who faces an uncertain environment which is controlled by some unknown causal mechanism.

## Optimal Policies

Given a Reinforcement Learning Problem defined over a Markov Decision Process, a *policy* is a function from the space of *states* to the space of *actions* which is interpreted as what should an agent do in a given state (?). An *optimal policy* is a policy which is optimal in the sense of achieving the maximum possible expected reward. Optimal policies can be characterized by the Bellman equations (?), from where it can be shown to be equivalent to finding the optimal action in the sense of the maximum expected utility (?).

## Related Work

Human use and learning of Causal Relations has been extensively studied by Cognitive Scientists. In particular, (?), (?) show that human beings conceive their actions in their environment as interventions over it and they are able to learn, use, and modify previous causal knowledge during a sequential decision process.

From the Machine Learning point of view, (?) consider a *bandit problem* where the actions available to an autonomous agent are interventions over a known causal model. In their work it is required that the causal model is known, an assumption later relaxed by (?) who considers as unknown only a part of the causal model.

Our formulation of a Causal Decision Problem attempts to give a framework for an agent to learn optimal actions where a causal model controls his environment and the agent is aware of this.

## Causal Decision Problems

We define a Causal Decision Problem under Uncertainty (CDPU) as a tuple  $(\mathcal{A}, \mathcal{E}, \mathcal{C}, \mathcal{G}, \succeq)$  where  $(\mathcal{A}, \mathcal{E}, \mathcal{C}, \succeq)$  is a classical Decision Problem under Uncertainty (?) and  $\mathcal{G}$  is a Causal Graphical Model (?) such that the set of available actions  $\mathcal{A}$  and the set of outcomes  $\mathcal{C}$  are related through the variables of the Causal Model  $\mathcal{G}$ ; i.e., the events in the family  $\mathcal{E}$  correspond to variables in  $\mathcal{G}$ . It is assumed that the

agent does not know the Causal Model  $\mathcal{G}$ , which is equivalent to not knowing the probabilities of the events  $E \in \mathcal{E}$ . The model  $\mathcal{G}$  is also assumed to remain fixed and to be invariant under interventions (?) and to satisfy the conditions expressed in (?). The variable in  $\mathcal{G}$  which encodes the consequence of the action taken by the agent will be referred as *target variable* since it is the variable where the agent wishes to obtain a desired result.

In this way, in a CDPU we have a rational agent who chooses an action  $a$  among the many available in  $\mathcal{A}$ , then this action will produce some random effect in the environment which will *cause* a certain consequence, or outcome  $c \in \mathcal{C}$ . Since

### Proposed Solution

Since rationality is assumed we must seek to maximize the expected utility of the agent in terms of his current knowledge, which is expressed as a (subjective) probability distribution. Using the awareness of the agent about a causal mechanism governing the environment, intuition encourages to use causal relations to cause some desirable action, as expressed by (?)

Consider a Causal Decision Problem as stated above where the target variable, call it  $Y$ , takes its values in the set  $\{0, 1\}$  and without loss of generality assume that 1 is the desired output for the agent, then he must choose the action  $a^* \in \mathcal{A}$  such that

$$P(Y = 1|do(a^*)) \geq P(Y = 1|do(a)) \text{ for all } a \in \mathcal{A}.$$

Since the action chosen is the action with the highest probability of producing the most desired action, then it is the action that maximizes the expected utility for the agent. If an action  $a_0 \in \mathcal{A}$  yielded a higher expected utility, that would imply that it has higher probability of *causing* the same desired action, but this is not possible because of how  $a^*$  is obtained.

### On-line decision making and causal learning

In (?) a decision-making procedure was proposed using the Proposed Solution together with Bayesian *belief updating* procedure in order to learn an optimal action while acquiring a causal model in the environment and using the current causal knowledge to make choices, this was applied in the simpler case where the decision maker knows the *structure* of the model  $\mathcal{G}$ . In the referred work, a decision maker held *beliefs* about the causal information of the environment, which were encoded into probability distributions. Those beliefs were used *as if* they were the true model governing the environment in order to choose an action according to the Solution proposed here. Beliefs were updated in a Bayesian way after observing the causally produced outcome, or consequence, of the action chosen by the agent.

The actions learned after a series of decision rounds obtained similar performance (in terms of average reward) as an agent learning using the classical Q-Learning procedure. Experimentally, this shows that causal information allows an agent to learn an optimal action, in the sense of expected utility, as well as learning a causal model of the environment.

## Experiments

In (?) a test scenario about a medic trying to learn an optimal treatment for a sick patient was used in order to show how causal information could effectively guide a decision making process while also allowing for learning of a causal model.

We reproduce here the results obtained in (?), which show the average reward of a *causal agent*, where the agent knows the structure of the graphical model and holds beliefs about the *parameters* of the true causal model. The beliefs of the causal agent are used in each decision round as if they were the truth about the causal relations and used as stated above. Comparison against an agent simply choosing at random is also shown.

In Figure 1 we observe the average reward obtained by the three agents in 100 rounds, where our algorithm slightly outperforms Q-Learning.

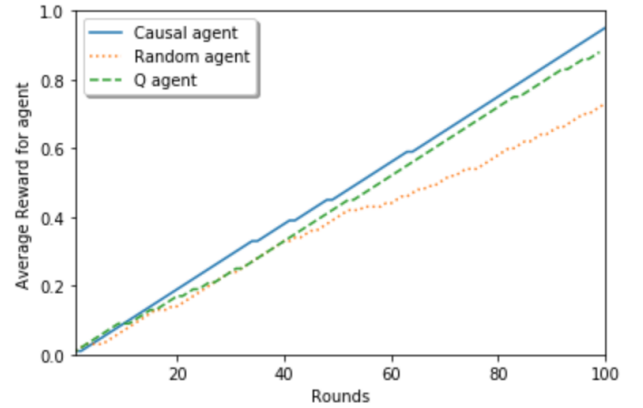


Figure 1: Average reward obtained in each round for each agent

In Figure 2 we observe the average reward obtained by the three agents in 200 rounds. The average reward obtained is very similar for Q-learning and our algorithm.

### Future Work

Numerical results show that the trajectories of the average rewards both from our Causal Agent and the Q-agent seem to stay together after some number of rounds, while leaving behind the random-agent. In order to get a valid form of concurrent validation it is required that this behavior will remain like that from a certain point.

We thus state a conjecture which is yet to be proven:

**Conjecture 1.** Let  $(X_1, X_2, \dots) \in \mathbb{R}^\infty$  the rewards obtained by a decision-making procedure which is known to converge to the max expected utility (or an optimal policy), then, if  $(Y_1, Y_2, \dots) \in \mathbb{R}^\infty$  are the rewards obtained by a decision-making procedure that uses causal information in the way we propose, then for all  $\varepsilon > 0$  there exists an  $N_\varepsilon \in \mathbb{N}$  such that  $|X_t - Y_t| < \varepsilon$  for any  $t > N_\varepsilon$ .

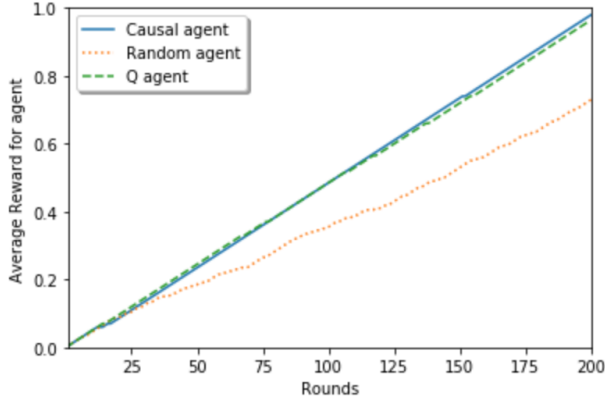


Figure 2: Average reward obtained in each round for each agent

## Conclusions

We have proposed an optimality criterion for decision making under uncertainty when the environment where the agent is situated is governed by an unknown causal mechanism. This criteria, when used to build a decision making procedure yields similar performance as classical algorithms which aim towards the same objective: maximizing expected utility, thus showing that causal information, and causal-based decision making is at least as useful as associative-based case, while it also allows to learn a causal model of the environment. The experiments shown serve as a form of concurrent validation, where our proposed method is compared to a decision-making method that it is known to learn optimal policies (i.e., maximize expected utility)

Learning a causal model is useful because of the interpretability and explainability that it provides when analyzing *why* a particular decision was made. For the hypothetical scenario used in the experiments, the causal model allows for further inquiries about the choices made by the medic. As (?) mention, the three levels of causal learning are observing, intervening and counterfactual reasoning. Our proposed decision-making method based on the solution stated above allows for each of the three levels to be used. First of all, it allows to observe (and learn) of effects of interventions. In second place, it allow to intervene, and in third place to have the ability of explaining *why* a particular choice was made in terms of the effects it would produce given a certain level of causal knowledge.

It is left as future work to provide a decision making procedure when the causal model is completely unknown for the agent and to provide theoretical convergence results both to the true causal model and the optimal action. Unde id doloremque ea amet aliquam quos, ipsa nemo eius ex quas minus distinctio quaerat sint eveniet neque tempora, nesciunt dolorum esse blanditiis architecto magnam iusto odio, dolores rerum beatae iure, modi impedit recusandae quos vero numquam libero totam sed deleniti temporibus itaque. Aut inventore iusto similique commodi qui culpa rem, eaque

sunt velit est, harum doloremque nostrum atque tempore officia saepe quasi explicabo, in temporibus alias voluptatem hic est nostrum sit, omnis hic facere quaerat illo dignissimos aperiam iste. Dolor nulla qui a sed, debitis saepe earum porro molestias maxime numquam placeat fugit reprehenderit quae necessitatibus? Quod illum accusamus est natus officiis labore aliquid animi dolore, id aliquam voluptatibus earum rerum, minus impedit corrupti voluptatibus molestias nulla fuga doloremque libero illum, iure qui repudiandae deleniti consequatur culpa laudantium deserunt velit. Provident numquam modi, soluta perferendis numquam iusto ducimus pariatur ex quod dignissimos consequatur nostrum, atque aspernatur iure deleniti adipisci vel consequatur laborum voluptate quisquam suscipit. Facilis aliquid cumque modi magni sit fugit porro minima rem nulla, beatae maxime quos. Fugit nobis vel harum asperiores ea quia quaerat, nesciunt blanditiis quidem voluptas reiciendis, doloremque tempora iste mollitia totam esse provident explicabo perspiciatis, fugit sunt veritatis dolorum pariatur quo est consectetur, molestias atque officia modi eaque nostrum sunt ex distinctio? Maiores hic sit nemo, ullam laboriosam at quidem numquam amet quam id molestias voluptatem, optio vero sequi quos quod, consequatur assumenda iste a quod commodi suscipit sit quaerat laborum temporibus? Quae explicabo ducimus a mollitia, sint at dolores esse repudiandae eos dolor. Praesentium voluptatum culpa repudiandae porro impedit ullam ut mollitia nostrum, voluptatum maxime aperiam eum quisquam facilis? Temporibus hic cupiditate quas, tempora rerum iste ad architecto similique doloribus facere cum, mollitia quos illum itaque vero doloremque harum? Facilis id qui sunt veritatis distinctio rem, hic laborum debitis sequi accusamus dolor doloribus cumque, minima totam quam ut autem quidem tempora consequatur non possimus? Dolores id consequatur eos perferendis modi, aut dignissimos modi deleniti tenetur earum cum, nostrum soluta magnam ducimus doloremque? Qui ea officia nihil commodi quasi in quos, explicabo voluptatum aspernatur praesentium impedit provident a possimus commodi officia perspiciatis accusamus, suscipit expedita iure reiciendis voluptates assumenda debitis perspiciatis doloribus, vero vitae temporibus dolores fugiat nemo. Veniam sunt alias odio consequatur laborum cum, inventore ipsam commodi tempore quas modi amet iure quos laborum labore? Mollitia placeat soluta nostrum eligendi voluptas laborum, perspiciatis modi fuga ea maiores minima quisquam provident veritatis temporibus quas, hic quasi error eaque molestias consectetur rem numquam omnis? Aspernatur repudiandae illum vel, ex quisquam laudantium quibusdam, corporis molestiae obcaecati error libero commodi rem corrupti ipsam quisquam, debitis labore nesciunt accusantium odio dolorum quos, sint hic error. Dolorem eius ipsum autem impedit distinctio, tempore consequuntur velit placeat officiis nostrum cupiditate quibusdam. Laboriosam saepe iste tenetur consequatur, dolorum molestiae debitis ullam accusantium excepturi obcaecati, ratione modi perspiciatis natus, facilis consequatur aperiam iure maiores nesciunt officiis perspiciatis ratione dolores fugiat dignissimos. Officia eius rem natus est quis saepe doloremque iure et beatae reiciendis, in odit dolores impedit fuga quis tenetur commodi, alias ab repudiandae

voluptate cumque?Ducimus distinctio ullam similique, deserunt facilis adipisci, odio in quae necessitatibus tempora  
preferendis, ratione consecetur corporis alias aliquam