Causal Structure Learning: a Bayesian approach based on random graphs

Mauricio Gonzalez Soto, Ivan Feliciano-Avelino
L. Enrique Sucar, Hugo J. Escalante Balderas

Instituto Nacional de Astrofísica, Óptica y Electrónica.

Introduction

- We adopt the *manipulationist* interpretation of Causality:

  A manipulation of a cause will result in a manipulation of the effect
- We are working in Causal Decision Making (CDM).
- We attempt to learn a causal graphical structure while using it to make good choices.
- What does a good decision mean? See [Gonzalez-Soto et al., 2019]

Previous work on CDM

[Joyce, 1999] • lay the foundations.

[Pearl, 2009] • assumes known causal information.

[Lattimore et al., 2016] • assume a known causal graphical model.

[Sen et al., 2017] • assumes as known part of the model.

[Gonzalez-Soto et al., 2018] • assumes a known graphical structure.

Ours • the model is not known.

Contributions

(a) We use graphs in order to model the uncertainty about the existence of causal relationships within a given set of variables.

(b) We adopt a Bayesian point of view in order to capture a causal structure via interaction with a causal environment.

(c) We conduct experiments to show the usefulness of our method to learn a causal structure as well as the optimal action.

Background

- A random graph can be thought of as a dynamic object which starts as a set of vertices and succesive edges are added at random according to some probability law.
- Our Bayesian modeling consists of:

Source of uncertainty

Existence or not of a causal relationship between a given pair of variables.

Probabilistic model

Probability of the occurrence of an edge in a random graph.

Related Work

- An *active learning* approach can be used to learn causal models from data, where one starts with some initial graph and then select the instances in the data that allow to add and orient edges in order to end with a fully oriented graph.
- Active learning algorithms for causal discovery can be found in [Tong and Koller, 2001, Murphy, 2001, Meganck et al., 2006, He and Geng, 2008, Hauser and Bühlmann, 2012, Ness et al., 2017, Rubenstein et al., 2017].

Methodology

- Let a rational decision maker (or agent) consider the following set of variables $\mathcal{X} = \{X_1, \ldots, X_n\}$, and they are causally connected by a fixed causal graphical model $\mathcal{G}$.
- The agent knows that she can only intervene one variable, e.g. X_1, and does so in order to alter the value of some identified *reward variable*, e.g. X_n.
- The agent knows a *causal ordering* of the variables.
- We will form a random graph using the *beliefs*, p_{ij}, that the decision maker has about the existence of a causal link between nodes in a graph.
- Formally, let $p_{ij} \in [0, 1]$ be the *belief* that the agent has over a causal relation (directed link) existing between the variable with index i and the variable with index j.
- This is, the decision maker has belief $p_{ij} \in [0, 1]$ that $X_i \to X_j$.
- Let G an initial *random* DAG formed as follows: the node set is $N = \{1, \ldots, n\}$ and a there exists a link between i and j with probability p_{ij}.
- The agent makes an intervention a^* over the possible values that X_1 can take within the resulting graph G.
- Once the action is taken, a full realization $X_1 = x_1, \ldots, X_n = x_n$ is observed.
- In order to update the initial beliefs p_{ij}, according to what has been observed, we use Bayes Theorem as follows:

$$p'_{ij} \propto p(X_1{=}a^*,\ldots,X_i{=}x_i,\ldots,X_j{=}x_j,\ldots X_n{=}x_n|\text{current graph})p_{ij}.$$

- Then, we update the model generating a new graph according to p'_{ij}.

Implementation

Require: The maximum number of iterations k.

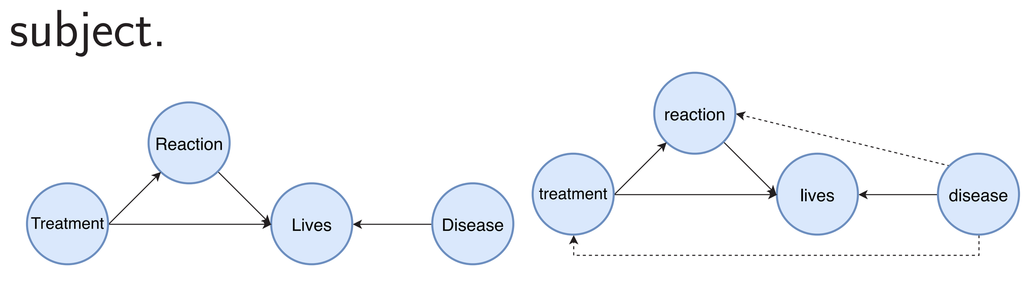
Initialize p_{ij} randomly.
for $t = 1, \ldots, k$ **do**
  Initialize G with link node i to node j with probability p_{ij}.
  Take action $a*$ for graph G and probabilities given by a count of the observations.
  Update beliefs

$$p^{t+1}_{ij} \propto p(X_1{=}a^*,\ldots,X_i{=}x_i,\ldots,X_j{=}x_j,\ldots,X_n{=}x_n|\text{current graph})p^t_{ij}.$$

end for
return The beliefs p_{ij}

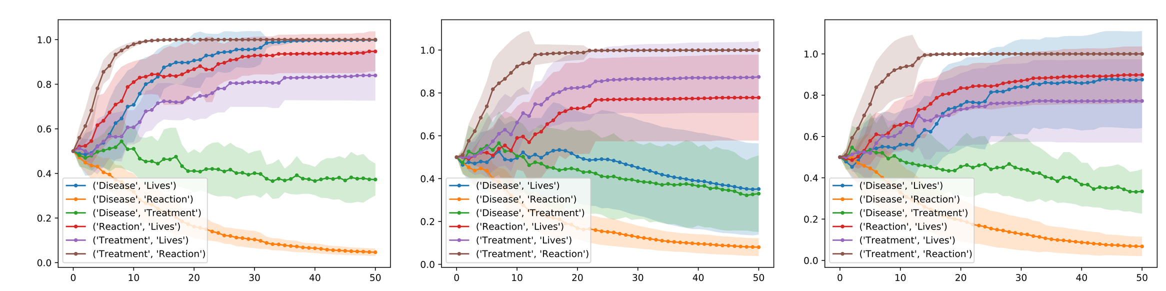
Disease-Treatment Problem

Consider a patient who can have one of two possible diseases. A doctor can treat the disease with either treatment A or treatment B, both of which carry some risk. Whether a patient is cured or not depends on the disease she or he is dealing with, the given treatment, and a possible negative reaction that the latter may have on the subject.



(a) Ground truth (b) $\mathcal{G}$ with all the possible edges given by the beliefs

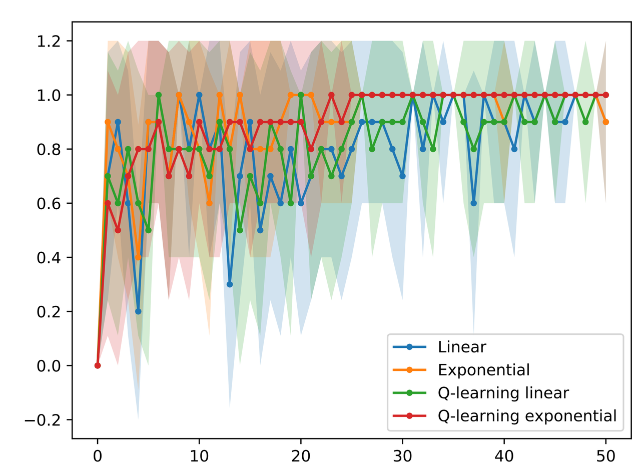
Results



(c) Random action selection policy (d) ϵ-greedy with exponential decay (e) ϵ-greedy with linear decay

Figure 1. Beliefs over time.

How good is the agent at making decisions?



Conclusions and future work

- We have presented a purely Bayesian random graph-based methodology which is able to learn a causal structure from interventions.
- Our approach is flexible and easily implementable as shown by the experiments, which also reflect a good performance on the attacked scenario.
- In this work we have assumed the existence of a causal ordering of the variables. In future work this assumption should be relaxed.

References

[Gonzalez-Soto et al., 2018] Gonzalez-Soto, M., Sucar, L. E., and Escalante, H. J. (2018). Playing against nature: causal discovery for decision making under uncertainty. In *Machine Learning for Causal Inference, Counterfactual Prediction and Autonomous Action (CausalML) Workshop at ICML 2018.*

[Gonzalez-Soto et al., 2019] Gonzalez-Soto, M., Sucar, L. E., and J., E. H. (2019). Von neumann-morgenstern and savage theorems for causal decision making. In *New in ML Forum co-located with NeurIPS 2019.*

[Hauser and Bühlmann, 2012] Hauser, A. and Bühlmann, P. (2012). Two optimal strategies for active learning of causal models from interventions. In *Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, pages 123–130.

[He and Geng, 2008] He, Y.-B. and Geng, Z. (2008). Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9(Nov):2523–2547.

[Joyce, 1999] Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

[Lattimore et al., 2016] Lattimore, F., Lattimore, T., and Reid, M. D. (2016). Causal bandits: Learning good interventions via causal inference. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 1181–1189. Curran Associates, Inc.

[Meganck et al., 2006] Meganck, S., Leray, P., and Manderick, B. (2006). Learning causal bayesian networks from observations and experiments: A decision theoretic approach. In *International Conference on Modeling Decisions for Artificial Intelligence*, pages 58–69. Springer.

[Murphy, 2001] Murphy, K. P. (2001). Active learning of causal bayes net structure.

[Ness et al., 2017] Ness, R. O., Sachs, K., Mallick, P., and Vitek, O. (2017). A bayesian active learning experimental design for inferring signaling networks. In Sahinalp, S. C., editor, *Research in Computational Molecular Biology*, pages 134–156, Cham. Springer International Publishing.

[Pearl, 2009] Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition.

[Rubenstein et al., 2017] Rubenstein, P. K., Tolstikhin, I., Hennig, P., and Schoelkopf, B. (2017). Probabilistic active learning of functions in structural causal models. *arXiv preprint arXiv:1706.10234.*

[Sen et al., 2017] Sen, R., Shanmugam, K., Dimakis, A. G., and Shakkottai, S. (2017). Identifying best interventions through online importance sampling. In *International Conference on Machine Learning*, pages 3057–3066.

[Tong and Koller, 2001] Tong, S. and Koller, D. (2001). Active learning for structure in bayesian networks. In *International joint conference on artificial intelligence*, volume 17, pages 863–869. LAWRENCE ERLBAUM ASSOCIATES LTD.