Project proposal : Autoregressive VAE for Causal Modeling

Tom Marty [1 2]

Abstract

Causal modeling is an active area of research that consist in inferring the functional components of a causal model from multi-intervention data. A causal model is made of independent mechanisms that govern the interactions between the different variables in the model. Recovering these causal mechanisms poses significant challenges because of the inherent confounding bias that can arise from the different interventional datasets considered. In this project, we propose to use a Variational auto-encoder with discrete latent in order to jointly estimate the interventional regimes and interventional targets of the data, required to deconfound the estimation of the independent causal mechanisms.

1. Motivations

2. Background

In this section, we introduce some of the basic concepts that will be used in the rest of the report. We will first introduce the concept of Structural Causal Models (SCM), the concept of intervention on these structure, then quickly present the usual assumptions made in the causal modeling literature. If the reader is already familiar with these concepts, we recommend skipping directly to Section 3.

2.1. Structural Causal Models

A Structural Causal Model (SCM) mathematically formalize the cause-effect relationships between the different random variables in a system. Precisely, an SCM is a triplet $\mathcal{S}(\mathcal{G}, \mathbb{P}_{\boldsymbol{\epsilon}}, \mathcal{F})$ that defines the data-generating process of N endogenous variables $\mathcal{X} = \{X_1, X_2, \cdots, X_N\}$ and N independent exogenous noise terms $\boldsymbol{\epsilon} = \{\epsilon_1, \epsilon_2, \cdots, \epsilon_N\}$. It is composed of three terms :

[1]Université de Montréal, Québec, Canada [2]Mila, Quebec AI Institute. Correspondence to: Tom Marty <tom.marty@mila.quebec>.

IFT 6269 project report - Université de Montréal

- A Directed Acyclic Graph [1] $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that represents the causal structure of the model. Each node X_i in the graph represent a random variable and each edge $X_i \rightarrow X_j$ represent a direct causal relationship between the two variables.
- A distribution over N independent exogenous noise terms $\boldsymbol{\epsilon} = \{\epsilon_1, \epsilon_2, \cdots, \epsilon_N\}$ that represent the unobserved factors influencing the endogenous variables.
- A set of N mechanism $\mathcal{F} = \{f_1, f_2, \cdots, f_N\}$ that represent the functional form of the causal relationships between the variables in the model. Each variable X_i is a function of its parents $\text{Pa}(X_i)$ defined by the DAG and some exogenous noise ϵ_i sampled from $\mathbb{P}_{\boldsymbol{\epsilon_i}}$:

$$X_i = f_i(\text{Pa}(X_i), \epsilon_i) \quad \epsilon_i \sim \mathbb{P}_{\epsilon_i} \tag{1}$$

This formulation induces a set of conditional probability distributions $p(X_i|\text{Pa}(X_i))$ such that the joint distribution over the variables X can be expressed as a product of these conditional distributions:

$$p(X) = \prod_{i=1}^{N} p(X_i|\text{Pa}(X_i)).$$

For the rest of this report, we will consider the variables X_i and ϵ_i to be univariate.

2.2. interventions and targets

An important concept in the field of Causality is the notion of **intervention**. An intervention $\mathcal{I}$ is a modification of the data-generating process of a causal model $\mathcal{S}$ that consist in modifying a set of mechanisms $f_i \in \mathcal{F}_{\mathcal{I}}$:

$$f_i(\text{Pa}(X_i), \epsilon_i) \rightarrow f_i^{\mathcal{I}}(\text{Pa}(X_i), \epsilon_i) \quad \forall i \in \mathcal{F}_{\mathcal{I}} \tag{2}$$

For a given intervention, the **targets** refers to the set of variables in the causal model whose mechanisms are altered by the intervention. There are different types of interventions that can be considered, we usually consider *single* intervention where the set $\mathcal{F}_{\mathcal{I}}$ is reduced to a single altered mechanism, we refer the reader to Pearl (2009) for a more detailed discussion on interventions.

[1]*commonly called a DAG

interventional regime Another important notion is the notion of interventional regime, which caracterise a set of samples $D_{\mathcal{I}_k}$ that were generated under the same intervention $\mathcal{I}_k$. Additionally, when no intervention is applied, the data is said to be generated under the observational regime $D_{\mathcal{O}}$.

2.3. Common assumptions considered

Markovian mechanisms

Faithfulness

Structural minimality

Faithfulness

3. Setting

In causal modeling, given a dataset $\mathcal{D}$ generated from an SCM $\mathcal{S} = (\mathcal{G}, \mathbb{P}_{\epsilon}, \mathcal{F})$, the goal is usually to retrieve full knowledge about $\mathcal{S}$. In *Causal Discovery*, the challenge is to infer the graph $\mathcal{G}$, which is a particularily harduous task because of the inherent difficulties arising from discrete optimization (a graph a is discrete structure) under constraint (e.g. Acyclicity). In *Causal Modelling*, the goal is to learn the mechanisms $\mathcal{F}$. Depending on what is to be learned, different assumptions have to be made on the data-generation process in order to identify the causal model. These assumptions can be summarized as follows:

1. unknown/known interventional regime
2. unknown/known single/multiple interventional targets
3. unknown/known graph structure
4. restricting assumptions on the mechanisms (e.g. Additive noise, linear mechanisms, gaussian noise...)

In this project, we will consider a slightly unusual setting :

We consider a dataset $\mathcal{D} = \{D_{\mathcal{O}}, D_{I_1}, \cdots, D_{I_N}\}$ composed of samples generated under N different interventional regimes. Each regime is obtained by performing a soft intervention on a single mechanism. Each mechanism is intervened only once, meaning that there are N variables and N corresponding interventions. Finally, we consider **that the graph $\mathcal{G}$ is known**. We discuss the relevance of such a specific setting in Section **??**. The subtlety of our setting is that the data is completely shuffled such that for each sample $X \in \mathcal{D}$ neither the interventional regime it belongs to nor the interventional target of this regime is known. Finally, the task consist in modeling each of the mechanisms $f(X_i|\text{Pa}(X_i))$ under both the observational regime and each of the interventional regimes.

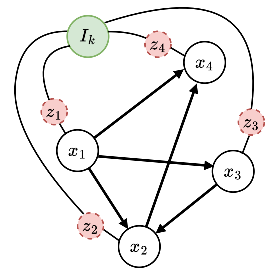

Figure 1. Graphical model of the data-generation process augmented with the unobserved latent variables. I_k represent the interventional regime and the z_i $i \in [1, N]$ are boolean variable that encode the state of each mechanism $p(X_i|\text{Pa}(X_i))$

.

3.1. Relevance

Why is it interesting ? show that with counter example saying retrieving the observational regime exactly is tricky.

Known graph is a pretty strong assumption, but we argue that this problem remains not trivial.

In this setting, it is not trivial to infer each mechanism under observational regime $p_{obs}(X_i|\text{Pa}(X_i))$ of the causal model $\mathcal{M}$, because the straightforward maximisation of data-loglikelihood will yield probabilistic models $\hat{p_{obs}}(X_i|\text{Pa}(X_i))$ that will be biaised by samples for which the true mechanism $p_{obs}(X_i|\text{Pa}(X_i))$ was in fact not in its observational form : $p_{obs} \rightarrow p_{int}$. In order to unbiaise this estimate, we therefore need to account for these possible interventions by also modeling and estimating the state of each mechanism using a boolean variable z_i. Recalling that some samples share some dependencies since they were sampled under the same interventional regime, we can also infer the regime trough a variable I_k and leverage that shared characteristic to better encode the interventions.

We could represents this data-generation process with a latent-variable model where the unknown latent are boolean variables representing the state of each mechanisms (*unintervened / intervened*). In this setting, we are trying to jointly infer the latent variables and the parameters θ of our statistical model $\hat{p_{obs}}$. To this end, we will use the variational-EM algorithm.

Outline First, we will be discussing the validity of the assumptions made in the data-generation process alongside the ones made on the probabilistic model, then we will derive the EM algorithm itself and explain the necessity for the variational variant of EM. Finally we will run our method on a set of controlled experiments, and discuss the results when compared with other methods for causal modeling.

4. Method

VAE is a general framework for solving statistical modeling problem with unobserved latent variables. Basically, the algorithm consist in maximizing a lower-bound of the log-likelihood of the data :

$$\log p(X|\theta) \geq \mathcal{L}(q,\theta) = \mathbb{E}_{q(z)}\left[\log p(X,z|\theta)\right] \tag{3}$$

In our setting, since we are only interested in modeling the observational regime, we can simplify the problem by considering the following lower-bound:

$$\log p_{obs}(X|\theta) \geq \mathcal{L}(q,\theta) = \mathbb{E}_{q(z|X;\theta)}\left[\log p_{obs}(X,z,\theta)\right] \tag{4}$$

$$= \mathbb{E}_{q(z|X;\theta)}\left[\sum_{X^k}\sum_{i=1}^{N} \mathbb{1}_{z_i^k=0} \log\left(p(z_i^k=0)p(X_i^k|\mathrm{Pa}(X_i^k), z_i^k=0,\theta)\right)\right] \tag{5}$$

5. Experiments

5.1. Data-generation process

5.2. Evaluation metrics

6. Conclusion

(Mahajan et al., 2024)

References

Mahajan, D., Gladrow, J., Hilmkil, A., Zhang, C., and Scetbon, M. Zero-Shot Learning of Causal Models, October 2024. URL http://arxiv.org/abs/2410.06128. arXiv:2410.06128.

Pearl, J. *Causality*. Cambridge University Press, 2 edition, 2009.