Finding the causes of failure: Neuralwave 2024

Mark Sobolev
Faculty of Informatics
Università della Svizzera Italiana
Lugano, Switzerland

Marco Gabriel
Faculty of Informatics
Università della Svizzera Italiana
Lugano, Switzerland

Abstract

We proposed a solution to a causal discovery in a multiple dataset context setting with a prior structural knowledge. Given the different measurements set, we produced an directed graph of causal influences and a ranking of an influencing variables. In addition we implemented a way to precisely constraint our causal model to reflect the changed mode. We developed a complex evaluation metric set for the unsupervised settings. Finally we designed a way to combine different approaches to causal discovery to produce the most reliable approach to generalised causal discovery setting.

1 Objective

We've been tackling a problem, where we were given several datasets with different latent context, for those we needed to provide the directed acyclic graph, which could be interpreted casually. Namely the datasets were structured as a set of columns, which were grouped in a several blocks, with an assumption of locality between elements of a block. We needed to discover various causal related structures about the given data:

1. The underlying causal graph
2. The root causes of the observed change
3. Visualisation of the graph, the root causes and their influence

1.1 Causal discovery

Causal discovery is a classic task, which also is extremely challenging. The causality discovery is used mainly in biology (Sachs, 2005) The task is to infer a Directed Acyclic Graph which minimizes a specific metric (which is trying to model causality) by observing the given data. The classic algorithms included using constrained-based approach to construct a graph (glymour,2001: 5.4.2); the more novel ones usually vary in their assumptions (e.g. linearity and non-Gaussianity for DirectLiNGAM (Shimizu et al 2011)), but propose also more flexible ways to learn and evaluate the results (Zheng et al 2018). The complexity of the causal discovery relates to several facts: searching over the space of DAG is known to be NP-hard (Chickering 1996).

A recent major breakthrough in the field can be characterised by the introduction of a reformulation of a task to a continuous optimisation problem (Zheng, 2018); this introduced a family of new gradient-optimisational algorithms (Wang et al, 2021; Yu et al. 2019), however some skepticism is also present (Kaiser, 2021).

Preprint. Under review.

1.2 Evaluation of causal graph in unknown setting

As the task in a general setting should be solved without any ground truth given (except structural prior knowledge), one should come up with an algorithm for this. The task is not well researched, however there are some general strategies for doing such. First, you can divide your data to several sets, and try to evaluate your DAGs on the test data using Markov blanket (Pearl, 1986) , following (Wang, 2025). Second, a set of formal criterions for the graph evaluation can be suggested, such as generalizability (Yang, 2019), and graph falsifying metrics (Eulig, 2023). Both ways were computationally infeasible for us, so we've developed 4 intermediate strategies for the evaluation:

1.2.1 Peer dataset

The first way to evaluate the algorithms was by providing a helper dataset (i.e. the most popular nowadays IHDP (Shalit, 2017)), and evaluating datasets on that. As usually the causal discovery is done by using synthetic datasets (Cheng, 2022), it should be a good estimate for an implementation.

1.2.2 Distribution change evaluation

The second way was done by incorporating the fact that we had two different distributions for two datasets, thus the algorithm was to train the model on the first, prior dataset, and try to minimize the drop of causal estimation (with a helper regression algorithm e.g. XGBoost (Chen & Guestrin, 2016)).

1.2.3 Likelihood of the graph by the inner metrics

The third way is the comparison of the inner statistical metrics of different algorithms, e.g. p-values of the DirectLinGam [2]. By this way the algorithms themselves report the quality of the output.

1.2.4 Predicting the data with the causal graph

This approach is based on training the regression and just comparing the output by the usual regression metrics, namely MSE or MAPE.

2 Approach

2.1 Methodology

The process is divided into three stages. First, we do not have access to any ground-truth. In the second stage, we get access to the first 15 rows of the target adjacency-matrix. In the last stage, we get additional access to the lower diagonal adjacency-matrix. We use this ground-truth as prior knowledge in the following algorithms as well as to evaluate our results.

In every stage, we fit various algorithms to the given data and compared their results. We outline these algorithms in the following sections.

2.2 Algorithms

2.2.1 LiNGAM

LiNGAM estimates structural equation models or linear causal Bayesian networks, assuming the data is non-Gaussian. First, to apply DirectLiNGAM[2], we merge the datasets while introducing a new feature to differentiate the origins. Secondly, we use MultiGroupDirectLiNGAM: [6] on the separate datasets.

Independent component analysis (ICA-LiNGAM)

2.2.2 FCI

Fast Casual inference (Spirtes, 2001) is an algorithm which employs the fact that one can build the graph from the scratch by using conditional independence, and thus it will grow computationally only if the nodes are having the causal structure behind them. For us it was valuable due to the computational constraints we've been working in for a high-rank task we've been given

2.2.3 VAR-LINGRAM

Var-Lingram (Shimizu, 2006) is a time series extension for a Lingram model.

2.2.4 PC

2.2.5 Structural Agnostic Model (SAM)

We use the SAM implementation of the CausalDiscoveryToolbox (CDT). SAM implements an adversarial game in which a separate model generates each variable, given real values from all others. A discriminator attempts to distinguish between real and generated samples. Finally, a sparsity penalty forces each generator to consider only a small subset of the variables, yielding a sparse causal graph. Lastly, the commonly made causal sufficiency condition (that there exist no hidden common cause to any pair of observed variables) can also be alleviated. each conditional generator (an artificial neural network) plays to produce one of the variables Xi given real observations of a (small) subset of all the others, hypothesized to be its direct causes, and noise. It is orders of magnitudes slower than Lingam, but also a lot better [4].

2.2.6 LLM: o1-preview from OpenAI

We prompt the o1-preview LLM from OpenAI to respond with the TODO

2.3 Frameworks

2.3.1 DoWhy

https://github.com/py-why/dowhy [1] [5]

2.3.2 CausalDiscoveryToolbox

https://github.com/FenTechSolutions/CausalDiscoveryToolbox [3]

2.3.3 gCastle

https://github.com/huawei-noah/trustworthyAI/tree/master/gcastle [7]

3 Results

3.1 Causal Graph

3.2 Root Cause

3.3 Visualisation

4 Challenges

The most challenging issue throughout the process was to evaluate possible solutions without knowing the ground truth.

`http://www.neurips.cc/`

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

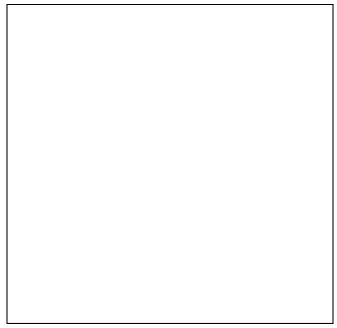

Figure 1: Sample figure caption.

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated...

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2023` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2023}
```

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form "A. Anonymous" and include a copy of the anonymized paper in the supplementary material.

4.1 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.[2]

4.2 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

[1] Sample of the first footnote.
[2] As in this example.

Table 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

4.3 Tables

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

This package was used to typeset Table 1.

References

[1] Patrick Blöbaum et al. "DoWhy-GCM: An Extension of DoWhy for Causal Inference in Graphical Causal Models". In: *Journal of Machine Learning Research* 25.147 (2024), pp. 1–7. URL: http://jmlr.org/papers/v25/22-1258.html.

[2] Takashi Ikeuchi et al. "Python package for causal discovery based on LiNGAM". In: *Journal of Machine Learning Research* 24.14 (2023), pp. 1–8. URL: http://jmlr.org/papers/v24/21-0321.html.

[3] Diviyan Kalainathan and Olivier Goudet. *Causal Discovery Toolbox: Uncover causal relationships in Python*. 2019. arXiv: 1903.02278 [stat.CO]. URL: https://arxiv.org/abs/1903.02278.

[4] Diviyan Kalainathan et al. *Structural Agnostic Modeling: Adversarial Learning of Causal Graphs*. 2022. arXiv: 1803.04929 [stat.ML]. URL: https://arxiv.org/abs/1803.04929.

[5] Amit Sharma and Emre Kiciman. *DoWhy: An End-to-End Library for Causal Inference*. 2020. arXiv: 2011.04216 [stat.ME]. URL: https://arxiv.org/abs/2011.04216.

[6] Shohei Shimizu. "Joint estimation of linear non-Gaussian acyclic models". In: *Neurocomputing* 81 (2012), pp. 104–107. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2011.11.005. URL: https://www.sciencedirect.com/science/article/pii/S0925231211006813.

[7] Keli Zhang et al. *gCastle: A Python Toolbox for Causal Discovery*. 2021. arXiv: 2111.15155 [cs.LG]. URL: https://arxiv.org/abs/2111.15155.