Causal Discovery using Model Invariance through Knockoff Interventions

Anonymous Authors[1]

Abstract

Cause-effect analysis is crucial to understand the underlying mechanism of a system. We propose to exploit model invariance through interventions on the predictors to infer causality in a non-linear multivariate system of time series. We model non-linear interaction in time series using DeepAR and then expose the model to different environments using knockoff intervention to test model invariance. Knockoffs are in-distribution null variables generated without knowing the response. We test model invariance where we show that the distribution of the response residual does not change significantly upon interventions on non-causal features. We use synthetically generated time series to evaluate and compare our approach with other causality methods. Overall our proposed method outperforms other widely used methods.

1. Introduction

Discovering causal graph in a system is vital to understand system's dynamics and its response to changes in environment. The methods which extract causal information from observational data are crucial for the broad field of artificial intelligence (AI) (Jordan, 2019). In this work we use deep networks to learn nonlinear interactions in multivariate time series and test model invariance through knockoff interventions for causal inference. We build our work on the assumption that the causal mechanism in a real-world system does not change across diverse environments (Peters et al., 2016; Heinze-Deml et al., 2018; Pfister et al., 2019). More specifically, the conditional distribution of the response given its causal predictors remains invariant for all observations over different settings.

Deep networks best suit to learn nonlinear interactions in a multivariate system. We use DeepAR (Salinas et al., 2020) to model nonlinear relations in multivariate time series. DeepAR characterizes conditioning on observed variables and estimate latent features from time series that might help in the identification of spurious associations. To expose trained models to unseen settings for causal inference, we can intervene by omitting or replacing model predictors. Since, deep networks can not handle missing input or out-of-distribution (OOD) data, we propose to use the theoretically well-established Knockoffs framework (Barber et al., 2020) to generate interventional environments. Knockoff copies are statistically null-variables, in-distribution and as uncorrelated as possible with the originals. To highlight the effectiveness of Knockoffs in our method, we compare it with other types of interventions such as mean-distribution, OOD, uniform distribution. Using our approach, we discover full causal graph assuming causal sufficiency and stationarity in our data. We compare our work with VAR Granger causality (VAR-GC), VARLiNGAM and PCMCI (Runge et al., 2019) using synthetic dataset. We show that our method outperforms these methods in nonlinear causal discovery.

The remaining of this paper is organized as follows. In Section 2, we cover the related work. We provide methodological background in Section 3. In Section 4, we describe our proposed method. Results are presented and discussed in Section 5. In Section 6 we conclude our work.

2. Related Work

The methods in (Peters et al., 2016; Heinze-Deml et al., 2018; Pfister et al., 2019; Gamella & Heinze-Deml, 2020; Martinet et al., 2021) use invariant causal prediction (ICP) for causal inference assuming that data samples are drawn from heterogeneous environments and can identify the ancestors of the response. These methods are limited in a sense that they estimate causal relationship of the candidate variable with the target only rather than discovering full causal graph as done in our work through multivariate target modelling. Moreover, they search for subset of predictors which remains unchanged throughout environments while we estimate a consistent causal set using model invariance across interventional environments over multiple forecast windows. ICP then outputs the intersection of all subsets of invariant predictor which forms a strict subset of all direct causes or may even be empty. This is because disjoint sets of predictors can be invariant, yielding an empty intersection. In (Mogensen et al., 2022), the authors introduce and characterize minimally invariant sets of predictors, that is, invariant sets S for which no proper subset is invariant. They propose to consider the union S_{IAS} of all minimally invariant sets, where IAS stands for invariant ancestry search.

The method in (Peters et al., 2016) works with non-stationary data assuming that the environments are known in order to exploit the invariance. It is argued that without knowledge of the environments, the causal inference becomes difficult. Moreover, estimating the environments from data in naive manner i.e. clustering, change point detection, etc., and subsequently applying the existing methodology may lead to a loss of the methods coverage, guarantee or yield less powerful results. The authors of (Pfister et al., 2019) do not estimate the environments but instead utilizes the existing non-invariances in observations. In our method, we generate knockoff representation of the predictors to represent different environments and test model invariance. The work in (Ahmad et al., 2021) uses deep networks with knockoff intervention to estimate Granger causality in nonlinear time series. The authors of (Tank et al., 2021; Romano et al., 2020) uses deep networks combined with sparsity-inducing penalties on the weights to estimate causal significance of the predictors. However, it is difficult to retrieve clear information from weights of the deep networks that can be used for interpretability (Luo et al., 2020). Attention-based networks have also been used, where network parameters are utilized to determine time delay in the cause-effect relationship (Nauta et al., 2019). The PCMCI (Runge et al., 2019) method uses linear and nonlinear conditional independent (CI) tests to infer causality in multivariate time series. The extended version of cause-effect variational autoencoders (CEVAE) (Louizos et al., 2017) integrates the domain-specific knowledge for nonlinear causal inference in ecological time series (Trifunov et al., 2019). Time-frequency based analysis has been done to handle spurious associations in time series (Shadaydeh et al., 2019). However, they use spectral representation of linear VAR-GC and hence might struggle in case of nonlinearity.

3. Methodological Background

In this section we introduce the necessary background on DeepAR (Salinas et al., 2020) and the Knockoffs framework (Romano et al., 2020).

3.1. DeepAR

DeepAR is a powerful method suitable for probabilistic forecasting of high dimensional nonlinear non-stationary time series. It uses LSTM-based autoregressive recurrent neural networks (Graves, 2013) and learns nonlinear interactions from the history of all the related time-series. Few key attributes of DeepAR compared to other forecasting methods are: It extracts hidden features in the form of seasonal patterns and trends in the data, and characterizes conditioning on the entire set of variables during inference. Moreover, in this method, lesser hand-crafted feature engineering is needed to capture complex, group-dependent behavior.

3.2. Knockoffs

The Knockoffs framework is developed to estimate feature importance using CI testing with controllable false discovery rate (Barber & Candès, 2015). Given a set of observed variables $Z = (Z_1, \ldots, Z_n)$ in an environment with known distribution P_Z, and a predictive model, the knockoffs of the observed variables, defined as $\widetilde{Z} = (\widetilde{Z}_1, \ldots, \widetilde{Z}_n)$, represents different environments while maintaining the in-distribution and uncorrelation property of the original environment variable. Moreover, Knockoffs are null-variables, which means they are generated without looking at model response. Furthermore, knockoff variables should satisfy the pairwise exchangeability condition (Barber & Candès, 2015)

$$(Z_1, \widetilde{Z_2}, \widetilde{Z_1}, Z_2) \stackrel{d}{=} (Z_1, Z_2, \widetilde{Z_1}, \widetilde{Z_2}) \tag{1}$$

for any subset $\mathbb{A} \subseteq 1, \ldots, n$, here $\stackrel{d}{=}$ represents equal distributions. The $(Z, \widetilde{Z})_{swap(\mathbb{A})}$ is obtained from $(Z, \widetilde{Z})$ by swapping the entries Z_j and $\widetilde{Z_j}$ for each $j \in \mathbb{A}$.

The knockoff mechanism can be thought of as generating a probability distribution $P_{\widetilde{Z}|Z}(.|z)$ which is the conditional distribution of $\widetilde{Z}$ given $Z = z$ chosen such that the obtained joint distribution of $(Z, \widetilde{Z})$ which is equal to

$$P_Z(z) P_{\widetilde{Z}|Z}(\widetilde{z}|z),$$

is pairwise $(z_j, \widetilde{z}_j)$ symmetric and satisfy the exchangeability condition in (1). For variables with Gaussian distribution $P_Z = \mathcal{N}_n(\mathbf{0}_n, \Sigma)$, (Candes et al., 2017) shows that knockoffs $\widetilde{\mathbf{Z}}_{i,*}$ can be drawn from the conditional distribution

$$P_{\widetilde{Z}|Z}(.|\mathbf{Z}_{i,*}) = \mathcal{N}_n((\mathbf{I}_n - S\Sigma^{-1})\mathbf{Z}_{i,*}, 2S - S\Sigma^{-1}S) \tag{2}$$

for any fixed diagonal matrix S satisfying $0 \leq S \leq 2\Sigma$. This leads to the joint distribution of $(\mathbf{Z}_{i,*}, \widetilde{\mathbf{Z}}_{i,*})$ being equal to

$$\mathcal{N}_{2n}\left(\mathbf{0}_{2n}, \begin{bmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{bmatrix}\right),$$

which satisfies the pairwise exchangeablity (1). S represents the difference between the original variable and their knockoff version and the goal is the maximize S so that the term $\Sigma - S$ is minimized which shows the covariance between the original and their knockoff representation.

4. Causality using Model Invariance through Knockoffs

Let $z_i, i = 1, \ldots, N$ be the $N-$variate time series. Each time series $z_{i,t}, t = 1, \ldots, r$ is a realization of length r real-valued discrete stochastic process $Z_i, i = 1, \ldots, N$. Our goal is to estimate the causal relation among all variables

and retrieve a full causal graph. Throughout our work, we assume stationarity and causal sufficiency which implies $Z_i, i = 1, \ldots, N$ is a stationary stochastic process, and the set of observed variables $Z_i, i = 1, \ldots, N$ includes all of the common causes of pairs in Z.

We exploit model invariance for causal discovery in multivariate nonlinear time series where we determine that the underlying causal structure of a system remains invariant irrespective of the interventional environments. We model nonlinear time series with DeepAR and generate knockoff copies of the original data to test model invariance. The invariance of model is tested over multiple forecast windows so that we obtain a consistent causal graph. A model is regarded invariant if the distribution moments of the response residuals i.e μ or σ^2 do not change by intervening on the non-causal predictors. The variables are considered as non-causal if the residual distribution of the model response does not significantly change by exposing the model to their knockoff representation. We retrieve causal graph with a consistent set of causal predictors using hypothesis testing over residual distribution from multiple forecast windows. Any type of distribution comparison test can be associated with our method.

To generate knockoffs intervention data, we implement DeepKnockoffs (Romano et al., 2020), a framework for sampling approximate knockoffs. We generate knockoff copies using a mixture of Gaussian models as implemented in (Gimenez et al., 2019). The procedure consists of first sampling the mixture assignment variable from the posterior distribution. The knockoffs are then sampled from the conditional distribution of the knockoffs given the original variables and the sampled mixture assignment such that the necessary conditions for knockoffs are satisfied. Swapping any knockoff variable with their original version must leave the underlying joint data distribution invariant as in 1. For Gaussian knockoffs, we know that $Z \sim \mathcal{N}(\mu, \Sigma)$. We aim to build knockoffs in such a way that the joint distribution of $(Z, \widetilde{Z}) \sim \mathcal{N}(*, **)$ is Gaussian.

$$* = \begin{bmatrix} \mu \\ \mu \end{bmatrix}, ** = \begin{bmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{bmatrix},$$

Where S is a diagonal matrix such that $** \succeq 0$. S represents the discrepancy between originals and knockoffs. We have the same μ for both original variable and knockoff sample. For covariance matrix, the diagonal entries show autocovariance while the off diagonal value represent covariance of original variable with their knockoff version which should be minimized by maximizing S. We obtain knockoffs $\widetilde{Z}_j, j \subseteq 1, \cdots, n$, for all variables in the multivariate time series and substitute one variable Z_j at a time with $\widetilde{Z}_j$ for each $j \subseteq 1, \cdots, n$ to test model invariance and estimate cause effect relationship.

To estimate causal effect of $z_i, i = 1, \ldots, N$ on the response z_j, we test model invariance through its forecast residuals by feeding a knockoff representation $\widetilde{z}_i$ of the original variable z_i. We obtain residuals $e_1, e_2, ..., e_n \sim R$ and $\tilde{e_1}, \tilde{e_1}, ..., \tilde{e_n} \sim \tilde{R}$ over multiple forecast windows for each realization z_j of the stochastic process $Z_j, j = 1, \ldots, N$ with and without intervention on $z_i, i \neq j$. R represents distribution of model residuals for actual outcome and $\tilde{R}$ shows residual distribution after intervention. We compare these two distributions for inferring causality between each two nodes using Kolmogorov–Smirnov test (Smirnov, 1939). The test statistic uses the supremum distance $D_{n,m}$ between R and $\tilde{R}$. Here n, m are the number of samples in the these two distributions.

$$D_{n,m} = \sqrt{\frac{nm}{n+m}} sup|R_n - \tilde{R}_m| \tag{3}$$

The null hypothesis H_0: z_i does not cause z_j, is accepted if the distribution of the residuals is approximately identical across environments $\tilde{R} \overset{d}{\approx} R_i$ with a significant level $\alpha = [5\% \sim 10\%]$. In case the the residual distribution is significantly different $\tilde{R} \overset{d}{\neq} R_i$, the alternate hypothesis H_1: z_i causes z_j is accepted. $D_{n,m}$ tends to have significantly larger value in alternate case. Our method discovers a consistent full causal graph because it is retrieved by testing model invariance over multiple forecast windows.

5. Experimental data and Results

We conducted experiments on synthetically generated time series. The synthetic data model in our experiments has the generic form:

$$Z_t^j = a_j Z_{t-1}^j + \Sigma_i c_i f_i(Z_{t-\tau_i}^i) + \eta_t^j \tag{4}$$

for $j \in 1, ..., N$. The correlation coefficients a_j are selected from $[0.2 \sim 1.0]$. Besides autodependency in the model, there exist linear and exponential functional dependencies $f_i(z)$. We included noise η_t^j, which represents data from uncorrelated, normally distributed noise with zero mean and variances in the range [0.30 $\sim$ 0.9]. The coupling coefficients c_i and lags τ_i between variables are set to [0.2 $\sim$ 1.0] and $[0 \sim 10]$ respectively. The synthetic causal graph contains N=5 variables. To test our method, we generate multiple causal graphs by changing functional dependencies, causal edges, time delays, dynamic noise and correlation coefficients throughout our experiments. To quantify the performance of our method, we calculate the metrics of accuracy, false positive rate (FPR) and F-score.

$$\text{FPR} = \frac{\text{FP}}{\text{FP + TN}} \tag{5}$$

$$\text{F-score} = \frac{\text{TP}}{\text{TP} + 0.5(\text{FP + FN})} \tag{6}$$

Here, TP is the number of causal links that exist in the causal graph and also detected by the method; TN is the number of causal links that neither exist in the causal graph nor detected by the method; FP is the number of causal links that do not exist in the causal graph but detected by the method and FN is the number of the causal links that exist in the causal graph but not detected by the method.

In our experiments we found that methods for generating intervention variables have significant impact on discovering accurate causal graph. To show the effectiveness of proposed knockoff interventions compared to other possible interventions, we tested, using the similar steps described in the previous section, the performance of following intervention methods.

Distribution mean: The predictor z_i is replaced with another variable $\overline{z}_i$, that is comprised of the mean of the original variable and Gaussian noise. $\overline{z}_i = \frac{1}{r}\sum_{t=1}^{r} z_{i,t} + \mathcal{N}(0, \Sigma)$.

Uniform distribution: The original variable z_i is replaced with another variable having a uniform distribution which range is defined by the minimum and maximum values of the original variables z_i as suggested in (Jiwoong Im et al., 2021)

Out-of-distribution: Here, the candidate variable z_i that belongs to data distribution D_i is replaced with OOD variable $\overline{z}$ from another Gaussian distribution $\overline{D}$ such that $\overline{D} \overset{d}{\neq} D_i$ (different μ and σ^2). Moreover, $\overline{z}_i$ is selected to be as uncorrelated with the original variable z_i as possible.

Results: We generate multiple causal graphs using synthetic data model and demonstrate the results in terms of FPR and F-score for VAR-GC, VARLiNGAM, PCMCI and DeepAR with all four types of intervention variables represented as DeepAR-K for knockoffs, DeepAR-O for OOD, DeepAR-M for mean and DeepAR-U for uniform distribution in Fig 1 and 2. Moreover in Table 1, we present methods accuracy and F-score calculated from TP, TN, FP and TN in one of our experiments. Overall DeepAR-K performs better than others in accurate causal discovery. Since knockoffs has strong ability to control false discoveries as the framework was specially developed for, we can see low FPR for knockoff interventions as shown in Fig 1 compared to other methods and intervention types. At the same time the detection power of DeepAR-K to identify true causal links is high as illustrated in Fig 2 which could be because of uncorrelated property of the knockoffs with their originals. The closest to our proposed DeepAR-K in terms of achieving high F-score and low FPR are DeepAR-O and DeepAR-U. Since we have non-linear links in our generated causal graphs, methods like VAR-GC, VARLiNGAM and PCMCI struggle to estimate accurate causality among nodes of the graph. VAR-GC suffered the most and obtained a high FPR as it assumes linearity among variables of time series.

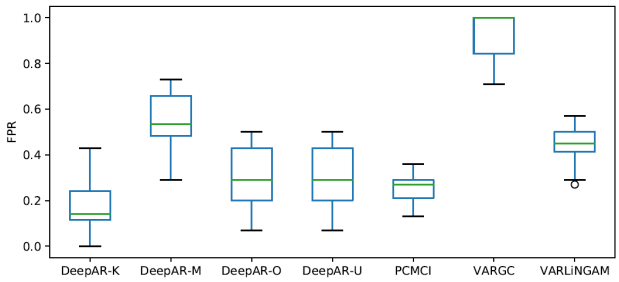

Figure 1. False positive rate of causal inference in synthetic multivariate nonlinear time series by all applied methods, lower is better.

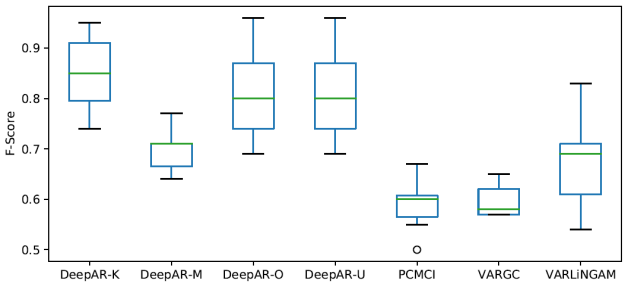

Figure 2. F-Score for estimating causality in synthetic multivariate nonlinear time series by applied methods, higher is better.

Table 1. Accuracy and F-score of causal inference in synthetic multivariate time series by all applied causality methods.

	TP	TN	FP	FN	Accuracy	F-score
VAR-GC	9	3	11	2	0.48	0.58
VARLiNGAM	10	10	4	1	0.80	0.80
PCMCI	6	12	2	5	0.72	0.63
DeepAR-K	**10**	**14**	**0**	**1**	**0.96**	**0.95**
DeepAR-O	11	10	4	0	0.84	0.85
DeepAR-M	10	7	7	1	0.68	0.71
DeepAR-U	10	10	4	1	0.80	0.80

6. Conclusion

We proposed to test model invariance across interventional environments over multiple forecast windows to discover full causal graph. DeepAR is used to learn non-linear relationship in multivariate time series. The model invariance is tested using knockoff interventions. We used synthetic time series data to evaluate our method. The results confirm that the estimation of the environments using Knockoffs to test model invariance for causal inference yield better results in comparison to others. Results also indicate that the proposed method outperforms VAR-GC, VARLiNGAM and PCMCI, however having a high computation time. Further we will test our method on real-world time series and extend the work for non-stationary data.

References

Ahmad, W., Shadaydeh, M., and Denzler, J. Causal inference in non-linear time-series using deep networks and knockoff counterfactuals. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 449–454. IEEE, 2021.

Barber, R. F. and Candès, E. J. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5), Oct 2015. ISSN 0090-5364. doi: 10.1214/15-aos1337. URL http://dx.doi.org/10.1214/15-AOS1337.

Barber, R. F., Candès, E. J., and Samworth, R. J. Robust inference with knockoffs. *The Annals of Statistics*, 48(3): 1409–1431, 2020.

Candes, E., Fan, Y., Janson, L., and Lv, J. Panning for gold: Model-x knockoffs for high-dimensional controlled variable selection, 2017.

Gamella, J. L. and Heinze-Deml, C. Active invariant causal prediction: Experiment selection through stability. *Advances in Neural Information Processing Systems*, 33: 15464–15475, 2020.

Gimenez, J. R., Ghorbani, A., and Zou, J. Knockoffs for the mass: new feature importance statistics with false discovery guarantees. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2125–2133. PMLR, 2019.

Graves, A. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

Heinze-Deml, C., Peters, J., and Meinshausen, N. Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6(2), 2018.

Jiwoong Im, D., Cho, K., and Razavian, N. Causal effect variational autoencoder with uniform treatment. *arXiv e-prints*, pp. arXiv–2111, 2021.

Jordan, M. I. Artificial intelligence—the revolution hasn't happened yet. 2019.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Louizos, C., Shalit, U., Mooij, J., Sontag, D., Zemel, R., and Welling, M. Causal effect inference with deep latent-variable models. *arXiv preprint arXiv:1705.08821*, 2017.

Luo, Y., Peng, J., and Ma, J. When causal inference meets deep learning. *Nature Machine Intelligence*, 2(8):426–427, 2020.

Martinet, G., Strzalkowski, A., and Engelhardt, B. E. Variance minimization in the wasserstein space for invariant causal prediction. *arXiv preprint arXiv:2110.07064*, 2021.

Mogensen, P. B., Thams, N., and Peters, J. Invariant ancestry search. *arXiv preprint arXiv:2202.00913*, 2022.

Nauta, M., Bucur, D., and Seifert, C. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1):312–340, 2019.

Peters, J., Bühlmann, P., and Meinshausen, N. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):947–1012, 2016.

Pfister, N., Bühlmann, P., and Peters, J. Invariant causal prediction for sequential data. *Journal of the American Statistical Association*, 114(527):1264–1276, 2019.

Romano, Y., Sesia, M., and Candès, E. Deep knockoffs. *Journal of the American Statistical Association*, 115(532): 1861–1872, 2020.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11):eaau4996, 2019.

Salinas, D., Flunkert, V., Gasthaus, J., and Januschowski, T. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

Shadaydeh, M., Denzler, J., García, Y. G., and Mahecha, M. Time-frequency causal inference uncovers anomalous events in environmental systems. In *German Conference on Pattern Recognition*, pp. 499–512. Springer, 2019.

Smirnov, N. V. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, 2(2):3–14, 1939.

Tank, A., Covert, I., Foti, N., Shojaie, A., and Fox, E. B. Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021. ISSN 1939-3539. doi: 10.1109/tpami.2021.3065601. URL http://dx.doi.org/10.1109/TPAMI.2021.3065601.

Trifunov, V. T., Shadaydeh, M., Runge, J., Eyring, V., Reichstein, M., and Denzler, J. Nonlinear causal link estimation under hidden confounding with an application to time series anomaly detection. In *German Conference on Pattern Recognition*, pp. 261–273. Springer, 2019.