Time Series Causal Discovery Using a Hybrid Method

Saima Absar
University of Arkansas
sa059@uark.edu

Lu Zhang
University of Arkansas
lz006@uark.edu

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10825912

***Abstract*—In this paper, we introduce a novel framework, Neural-HATS for inferring causal structures in time series data using a hybrid method. Neural-HATS uniquely combines conditional independence (CI) testing with continuous optimization-based learning methods to enhance causal discovery. Specifically, it leverages an attention-based encoder-decoder architecture with Kernel Conditional Independence (KCI) testing to enable direct CI tests between time series. These CI test results are integrated into continuous optimization algorithms, enhancing both causal inference accuracy and the effectiveness of continuous optimization models. Experimental evaluations demonstrate that Neural-HATS achieves improved causal graph accuracy.**

***Index Terms*—causal discovery, hybrid method, time series, attention-based encoder-decoder, continuous optimization**

I. INTRODUCTION

The theory of causality is a fundamental part of scientific research and is essential for advancing solutions in various fields like earth science, commerce, politics, healthcare, etc. [4], [7], [15]. Causal discovery, which aims to identify underlying causal structures in data, traditionally relies on randomized controlled trials (RCTs). However, due to their cost and ethical concerns, RCTs are often impractical [16]. With the surge in digital data in recent days, data-driven methods employing machine learning and artificial intelligence have become vital tools for advancing causal analysis and uncovering cause-and-effect relationships.

The causal discovery methods fall into two primary types: constraint-based and score-based. Constraint-based methods use conditional independence tests to infer causal directions but can be computationally intensive [2], [6]. In contrast, score-based methods optimize a score function to select the best causal structure. Recent score-based methods employ continuous optimization and deep learning, but they typically require large datasets, which may be limiting in practical applications [12], [16].

Most causal discovery research centers on static data, yet many real-world datasets are time series, requiring methods suited to temporal structures [3], [7]. While causal discovery for time series has advanced, constraint-based methods are limited due to the challenges of Conditional Independence (CI) testing in temporal data. Traditional CI tests, including kernel-based ones like KCI [17], aren't directly applicable to time-dependent structures, and adaptations like PCMCI [13] often rely on assumptions such as stationarity or sparse data points per time step.

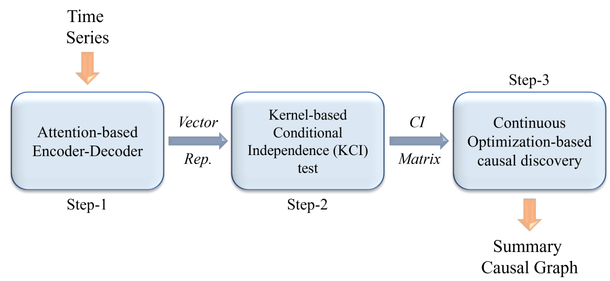


Fig. 1: Overview of the proposed hybrid temporal causal discovery architecture.

Granger causality, foundational in time series causal inference, asserts that one series is causally related to another if its past values improve the prediction of the other's future values, assuming causality precedes effect and lacks hidden confounders [5], [14]. While initially linear [8], [9], Granger causality now extends to non-linear settings through frameworks like TCDF [11], DYNOTEARS [12], GVAR [10], NTiCD [1], etc.

In this work, we introduce a framework for CI testing in time series using an attention-based encoder-decoder architecture with LSTMs to generate vector representations that capture conditional dependencies. These vectors are then passed through a multilayer perceptron (MLP) to predict future states. After end-to-end training, the encoder-decoder produces representations that can be used for CI testing via KCI. Thus, we propose an algorithm that performs CI testing through time series prediction. By leveraging low-order CI tests, we construct a CI matrix, which is then incorporated as a regularization constraint in the loss function of score-based optimization methods for causal structure discovery. This approach efficiently guides the optimization process, avoiding exhaustive testing. Our experiments demonstrate that CI tests enhance the performance of state-of-the-art continuous optimization algorithms. This method combines the strengths of deep neural networks and Kernel Conditional Independence (KCI) testing, removing assumptions about lag structures, stationarity, or linearity. Moreover, it synergizes the efficiency of CI testing with the flexibility of continuous optimization, enabling our framework to benefit from ongoing advancements in both fields.

979-8-3503-6248-0/24/$31.00 ©2024 IEEE

II. Method

Consider a multivariate time series $\mathcal{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ comprising d variables, each with a consistent length of n. In this paper, we aim to uncover causal relationships in multivariate time series $\mathcal{X}$ and represent them in terms of a summary causal graph. Our hybrid approach involves three main steps:

1) Encode hidden information from the time series into vector representations using an attention-based encoder-decoder framework.
2) Utilize the KCI to test for conditional independence among variables and form a conditional independence matrix based on the results.
3) Integrate this matrix into a continuous optimization causal discovery method as a regularization term, to derive the summary causal graph.

An overview of our proposed approach is shown in Fig. 1. We introduce an algorithm for conditional independence (CI) testing, called Neural Hybrid Approach for Time Series causal discovery (Neural-HATS). This algorithm employs a multi-layer long-short-term memory (LSTM) network with a self-attention mechanism to capture hidden (vector) representations within time series through time series prediction. These representations are then utilized in the kernel conditional independence (KCI) test to identify the conditional independence relationships. Conditional independence between two time series α, β given a subset of time series $\mathbf{C}$ is defined as follows: β is conditionally independent of α given $\mathbf{C}$ if for any time point t, the past of $\mathbf{C}$ until time t gives the same predictable information about β as the past of both α and $\mathbf{C}$ until time t, denoted by $\alpha \not\to \beta \mid \mathbf{C}$. Based on this, we identify all conditional independence relations for low degrees (1 or 0) and construct a CI matrix as below:

$$M_{\alpha,\beta} = \begin{cases} 1 & \exists C \in V \setminus \{\alpha, \beta\}, \quad \alpha \not\to \beta | C, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

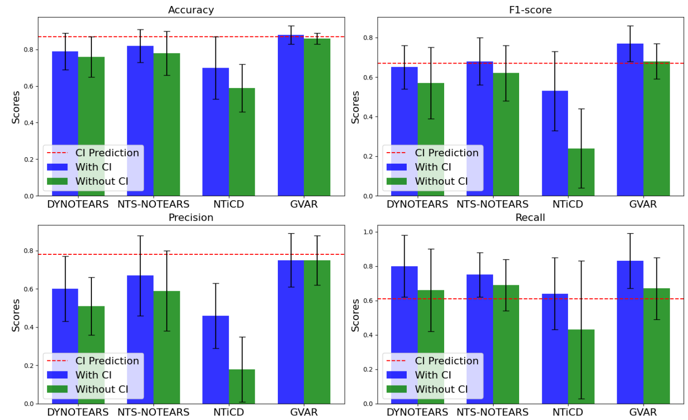
In the context of a general continuous optimization-based causal discovery method, where $\mathcal{L}$ represents the loss function, we incorporate our CI matrix as a regularization term, formulated as follows:

$$\min_{A} \mathcal{L} = \mathcal{L} + R(\theta) + \lambda_{CI} ||M \circ A||_F^2 \tag{2}$$

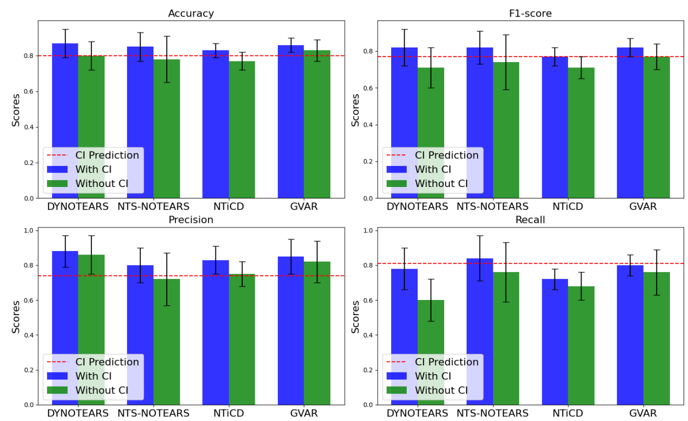
Here $\circ$ denotes the Hadamard product, and λ_{CI} is the regularization parameter that balances the impact of the CI matrix. The term $||M \circ A||_F^2$ serves to reduce the influence of penalized elements in A by M. Solving Eq. (2) via continuous optimization methods enables inference of a more accurate summary causal graph A.

III. Experiments

In this section, we perform comprehensive experiments to assess Neural-HATS across two datasets: non-linear synthetic data generated with the Vector Autoregressive model, and a real-world dataset, Netsim, which simulates Functional



(a) Synthetic data.



(b) Real data (Netsim).

Fig. 2: The performance of different baselines with and without the incorporation of Neural-HATS.

Magnetic Resonance Imaging (FMRI). Figure 2 illustrates Neural-HATS' performance as a hybrid method for four baseline methods: DYNOTEARS, NTS-NOTEARS, GVAR, and NTiCD. The plot shows results from these baseline models before and after integrating Neural-HATS. Additionally, we present the performance obtained by converting the CI matrix directly to the adjacency matrix with red dashed lines. The results show that Neural-HATS generally outperforms all base models in terms of accuracy, precision, recall, and f1-score.

IV. Conclusion

This paper presents Neural-HATS, a method that merges conditional independence (CI) tests with continuous optimization for temporal causal discovery. We propose an attention-based encoder-decoder framework that generates vector embeddings from time series data, enabling kernel-based conditional independence (KCI) testing. The CI tests on these embeddings produce a CI matrix, which is incorporated into a continuous optimization approach yielding more accurate causal graphs. Our model operates without assumptions of linearity or acyclicity and demonstrates significant performance enhancements when combined with four state-of-the-art score-based methods, validated on synthetic and real-world datasets.

REFERENCES

[1] Saima Absar, Yongkai Wu, and Lu Zhang. Neural time-invariant causal discovery from time series data. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2023.

[2] Saima Absar and Lu Zhang. Discovering time-invariant causal structure from temporal data. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 2807–2811, 2021.

[3] Charles K Assaad, Emilie Devijver, and Eric Gaussier. Survey and evaluation of causal discovery methods for time series. *Journal of Artificial Intelligence Research*, 73:767–819, 2022.

[4] Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10:524, 2019.

[5] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[6] Ruocheng Guo, Lu Cheng, Jundong Li, P Richard Hahn, and Huan Liu. A survey of learning causality with data: Problems and methods. *ACM Computing Surveys (CSUR)*, 53(4):1–37, 2020.

[7] Uzma Hasan, Emam Hossain, and Md Osman Gani. A survey on causal discovery methods for temporal and non-temporal data. *arXiv preprint arXiv:2303.15027*, 2023.

[8] Craig Hiemstra and Jonathan D Jones. Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*, 49(5):1639–1664, 1994.

[9] Ling Luo, Wei Liu, Irena Koprinska, and Fang Chen. Discovering causal structures from time series data via enhanced granger causality. In *AI 2015: Advances in Artificial Intelligence: 28th Australasian Joint Conference, Canberra, ACT, Australia, November 30–December 4, 2015, Proceedings 28*, pages 365–378. Springer, 2015.

[10] Ričards Marcinkevičs and Julia E Vogt. Interpretable models for granger causality using self-explaining neural networks. *arXiv preprint arXiv:2101.07600*, 2021.

[11] Meike Nauta, Doina Bucur, and Christin Seifert. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1):19, 2019.

[12] Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. Dynotears: Structure learning from time-series data. In *International Conference on Artificial Intelligence and Statistics*, pages 1595–1605. PMLR, 2020.

[13] Jakob Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075310, 2018.

[14] Ali Shojaie and Emily B Fox. Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, 9:289–319, 2022.

[15] Kshitiz Tiwari and Lu Zhang. Implications of minimum description length for adversarial attack in natural language processing. *Entropy*, 26(5):354, 2024.

[16] Matthew J Vowels, Necati Cihan Camgoz, and Richard Bowden. D'ya like dags? a survey on structure learning and causal discovery. *ACM Computing Surveys*, 55(4):1–36, 2022.

[17] Kun Zhang, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:1202.3775*, 2012.