Causal Discovery from Temporal Data

Chang Gong
Institute of Computing Technology, Chinese Academy of Sciences
University of Chinese Academy of Sciences
China
gongchang21z@ict.ac.cn

Di Yao*
Institute of Computing Technology, Chinese Academy of Sciences
China
yaodi@ict.ac.cn

Chuzhe Zhang
School of Mathematical Sciences, Fudan University
China
felixzh01@outlook.com

Wenbin Li
Institute of Computing Technology, Chinese Academy of Sciences
University of Chinese Academy of Sciences
China
liwenbin20z@ict.ac.cn

Jingping Bi*
Institute of Computing Technology, Chinese Academy of Sciences
China
bjp@ict.ac.cn

Lun Du
Microsoft
China
lun.du@microsoft.com

Jin Wang
Megagon Labs
United States
jin@megagon.ai

ABSTRACT

Temporal data representing chronological observations of complex systems can be ubiquitously collected in smart industry, medicine, finance and *etc.* In the last decade, many tasks have been studied for mining temporal data and offered significant value for various applications. Among these tasks, causal discovery aims to understand the underlying generation mechanism of temporal data and has attracted much research attention. According to whether the data is calibrated, existing causal discovery approaches can be divided into two subtasks, *i.e.*, multivariate time-series causal discovery, and event sequence causal discovery. Previous tutorials or surveys have primarily focused on causal discovery from time-series data and disregarded the second ones. In this tutorial, we elucidate the correlation between the two subtasks and provide a comprehensive review of the existing solutions. Moreover, we offer some potential applications and summarize new perspectives for discovering causal relations from temporal data. We hope the audiences can obtain a systematic overview of this topic and inspire some new ideas for their own research.

KEYWORDS

causal discovery, temporal data analysis, relational learning

*Corresponding authors.

KDD '23, August 6–10, 2023, Long Beach, CA, USA.

ACM ISBN 979-8-4007-0103-0/23/08.
https://doi.org/10.1145/3580305.3599552

ACM Reference Format:
Chang Gong, Di Yao, Chuzhe Zhang, Wenbin Li, Jingping Bi, Lun Du, and Jin Wang. 2023. Causal Discovery from Temporal Data. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), August 6–10, 2023, Long Beach, CA, USA.* ACM, Washington, DC, USA, 2 pages. https://doi.org/10.1145/3580305.3599552

1 INTRODUCTION

Temporal data is prevalent across various domains, including neuroscience, finance, bioinformatics, and social networks. The development of sensors and computing devices has led to the emergence of temporal data analysis research, encompassing classification, clustering, forecasting, and causal discovery. Among these tasks, causal discovery from temporal data has become a critical but challenging topic. Learned causal structures can explain the data generation process and aid in designing data analysis methods. The distinction between the two types of temporal data for causal discovery, namely multivariate time series (MTS) and event sequences, is based on whether the data is calibrated. As a result, existing causal discovery solutions can also be categorized into two groups. Recent tutorials in causality either focus on causal inference [4, 18] or only cover the causal discovery from relational data [25]. The objective of this tutorial is to offer a comprehensive and structured review of the current state-of-the-art in causal discovery from temporal data, along with an overview of the latest advancements in the field.

MTS is a typical kind of temporal data in various domains, which represents the calibrated states of multiple variables changing over time. The identification of causal relationships within the context of MTS has the potential to enhance both the interpretability and resilience of data analytic models. However, the definitions of causal relations in MTS are not unique, and there exist three graphical representations of causal structures, *i.e.*, full time causal graph, window causal graph, and summary causal graph. This results in different causal discovery solutions. Existing methods for causal discovery from MTS can be broadly categorized into four groups, namely, constraint-based methods [12, 16], score-based methods [5, 15], functional causal model (FCM)-based methods [22], and Granger causality-based methods [13, 19]. This tutorial also provides an overview of additional methods, such as causal models based on differential equations [20], nonlinear state-space methods [17], and logic-based methods [8].

This tutorial also covers the task of causal discovery from event sequences, which involves learning causal relations from irregularly and asynchronously observed time points. This is an important task because most real-world events do not occur within a fixed time interval. A preliminary concept required for causal discovery from event sequences is the multivariate point process. The corresponding solutions for this task can be classified into three main categories: Granger causality-based approaches [7, 24], constraint-based approaches [2], and score-based approaches [3], which are in line with those used for the MTS task.

In this tutorial, we also explore applications and new perspectives of temporal causal discovery. Learning causal relations from temporal data stands to offer advantages in improving the interpretability and resilience of models employed in diverse fields. Applications can be found in root cause analysis [14], video analysis [9], computational advertising [23], bioinformatics [21], and *etc.*, where the discovered causal relations can be either viewed as initial hypotheses or support a multi-stage approach. Lastly, we provide insights into opportunities, including amortized paradigms [11], supervised paradigms [1], and causal representation learning [10].

2 FURTHER READING

Additional information and supplementary materials can be found within the scope of our survey [6] and Github repository[1].

ACKNOWLEDGMENTS

This work has been supported by the National Natural Science Foundation of China under Grant No.: 62002343, 62077044.

REFERENCES

[1] Danilo Benozzo, Emanuele Olivetti, and Paolo Avesani. 2017. Supervised estimation of granger-based causality between time series. *Frontiers in Neuroinformatics* 11 (2017).

[2] Debarun Bhattacharjya, Karthikeyan Shanmugam, Tian Gao, and D. Subramanian. 2022. Process Independence Testing in Proximal Graphical Event Models. In *CLeaR*.

[3] Debarun Bhattacharjya, Dharmashankar Subramanian, and Tian Gao. 2018. Proximal Graphical Event Models. In *Advances in Neural Information Processing Systems*, Vol. 31. Curran Associates, Inc.

[4] Peng Cui, Zheyan Shen, Sheng Li, Liuyi Yao, Yaliang Li, Zhixuan Chu, and Jing Gao. 2020. Causal inference meets machine learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3527–3528.

[5] Tian Gao, Debarun Bhattacharjya, Elliot Nelson, Miao Liu, and Yue Yu. 2022. IDYNO: Learning Nonparametric DAGs from Interventional Dynamic Data. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA (Proceedings of Machine Learning Research)*, Vol. 162. PMLR, 6988–7001.

[6] Chang Gong, Di Yao, Chuzhe Zhang, Wenbin Li, and Jingping Bi. 2023. Causal Discovery from Temporal Data: An Overview and New Perspectives. arXiv:2303.10112 [cs.LG]

[7] Tsuyoshi Ide, Georgios Kollias, Dzung Phan, and Naoki Abe. 2021. Cardinality-Regularized Hawkes-Granger Model. In *Advances in Neural Information Processing Systems*, Vol. 34. Curran Associates, Inc., 2682–2694.

[8] Samantha Kleinberg. 2011. A Logic for Causal Inference in Time Series with Discrete and Continuous Variables. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*. IJCAI/AAAI, 943–950.

[9] Yunzhu Li, Antonio Torralba, Anima Anandkumar, Dieter Fox, and Animesh Garg. 2020. Causal Discovery in Physical Systems from Videos. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

[10] Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Stratis Gavves. 2022. CITRIS: Causal Identifiability from Temporal Intervened Sequences. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA (Proceedings of Machine Learning Research)*, Vol. 162. PMLR, 13557–13603.

[11] Sindy Löwe, David Madras, Richard Z. Shilling, and Max Welling. 2022. Amortized Causal Discovery: Learning to Infer Causal Graphs from Time-Series Data. In *1st Conference on Causal Learning and Reasoning, CLeaR 2022, Sequoia Conference Center, Eureka, CA, USA, 11-13 April, 2022 (Proceedings of Machine Learning Research)*, Vol. 177. PMLR, 509–525.

[12] Daniel Malinsky and Peter Spirtes. 2018. Causal Structure Learning from Multivariate Time Series in Settings with Unmeasured Confounding. In *Proceedings of 2018 ACM SIGKDD Workshop on Causal Discovery, CD@KDD 2018, London, UK, 20 August 2018 (Proceedings of Machine Learning Research)*, Vol. 92. PMLR, 23–47.

[13] Ricards Marcinkevics and Julia E. Vogt. 2021. Interpretable Models for Granger Causality Using Self-explaining Neural Networks. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

[14] Yuan Meng, Shenglin Zhang, Yongqian Sun, Ruru Zhang, Zhilong Hu, Yiyin Zhang, Chenyang Jia, Zhaogang Wang, and Dan Pei. 2020. Localizing Failure Root Causes in a Microservice through Causality Inference. In *28th IEEE/ACM International Symposium on Quality of Service, IWQoS 2020, Hangzhou, China, June 15-17, 2020*. IEEE, 1–10.

[15] Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. 2020. DYNOTEARS: Structure Learning from Time-Series Data. In *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy] (Proceedings of Machine Learning Research)*, Vol. 108. PMLR, 1595–1605.

[16] Jakob Runge. 2020. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Proceedings of the Thirty-Sixth Conference on Uncertainty in Artificial Intelligence, UAI 2020, virtual online, August 3-6, 2020 (Proceedings of Machine Learning Research)*, Vol. 124. AUAI Press, 1388–1397.

[17] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. 2012. Detecting causality in complex ecosystems. *science* 338, 6106 (2012), 496–500.

[18] Vasilis Syrgkanis, Greg Lewis, Miruna Oprescu, Maggie Hei, Keith Battocchi, Eleanor Dillon, Jing Pan, Yifeng Wu, Paul Lo, Huigang Chen, et al. 2021. Causal inference and machine learning in practice with econml and causalml: Industrial use cases at microsoft, tripadvisor, uber. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 4072–4073.

[19] Alex Tank, Ian Covert, Nicholas J. Foti, Ali Shojaie, and Emily B. Fox. 2022. Neural Granger Causality. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 8 (2022), 4267–4279.

[20] Mark Voortman, Denver Dash, and Marek J. Druzdzel. 2010. Learning Why Things Change: The Difference-Based Causality Learner. In *UAI 2010, Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, USA, July 8-11, 2010*. AUAI Press, 641–650.

[21] Alexander P. Wu, Rohit Singh, and Bonnie Berger. 2022. Granger causal inference on DAGs identifies genomic loci regulating transcription. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*.

[22] Tianhao Wu, Xingyu Wu, Xin Wang, Shikang Liu, and Huanhuan Chen. 2022. Nonlinear Causal Discovery in Time Series. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*. ACM, 4575–4579.

[23] Di Yao, Chang Gong, Lei Zhang, Sheng Chen, and Jingping Bi. 2022. CausalMTA: Eliminating the User Confounding Bias for Causal Multi-touch Attribution. In *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*. ACM, 4342–4352.

[24] Wei Zhang, Thomas Panum, Somesh Jha, Prasad Chalasani, and David Page. 2020. CAUSE: Learning Granger Causality from Event Sequences using Attribution Methods. In *Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Vol. 119. PMLR, 11235–11245.

[25] Elena Zheleva and David Arbour. 2021. Causal inference from network data. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 4096–4097.

[1] https://chaunceykung.github.io/temporal-causal-discovery-tutorial/