Supplementary Materials for Paper "Root Cause Analysis for Microservices based on Causal Inference: How Far Are We?"

Luan Pham
RMIT University
Melbourne, Australia
luan.pham@rmit.edu.au

Huong Ha
RMIT University
Melbourne, Australia
huong.ha@rmit.edu.au

Hongyu Zhang
Chongqing University
Chongqing, China
hyzhang@cqu.edu.cn

A MORE INFORMATION ABOUT CAUSAL DISCOVERY METHODS FOR TIME SERIES

In recent years, causal discovery methods have attracted attention from researchers owing to their ability to infer causal relationships from the data [12]. Among these, a large number of methods are dedicated to time-series data with applications in various domains such as healthcare, manufacturing, and energy [1]. Below, we describe some representative causal discovery methods that are commonly used in causal inference-based RCA for microservice systems with metrics data. Furthermore, we also include a recent proposed causal discovery method, which is not yet explored in RCA literature.

Granger Algorithm [5]. The Granger algorithm is based on the concept of *Granger causality*, which assumes that a time series causes another time series if the former provides statistically significant information about the future values of the latter. To derive the dependency between two time series x and y, two autoregression models are constructed: one fitted using y and one fitted using both x and y. Then, a statistical test is performed to see if the predictions of y that are based on past values of x and y are better than the predictions of y that are based on its own past values.

Peter-Clark (PC) Algorithm [11]. The PC algorithm is one of the most popular causal discovery methods for time series data. It starts with a complete undirected graph and then checks for the dependencies and conditional independencies for all pairs of nodes to generate an undirected graph (*skeleton graph*). Finally, a series of rules [11] is applied to the skeleton graph to orient the direction of the undirected edges. Developed from the PC algorithm, the PCMCI algorithm [7] is able to work with time-lagged causal relations and has also been shown to be effective.

Fast Causal Inference (FCI) Algorithm [11]. Similar to the PC algorithm, the FCI algorithm starts with an undirected graph and then sequentially removes the independent or conditionally independent edges with its set of rules. These rules are designed to ensure the FCI algorithm gives asymptotically correct causal graphs even in the presence of confounders. This is an advantage over the PC algorithm, as PC assumes there are no confounding factors that affect the time series.

Greedy Equivalence Search (GES) Algorithm [3]. The GES algorithm starts with a graph without edges and greedily and sequentially adds the edges and their directions in a way that minimizes the Bayesian Information Score (BIC) [8] or the Z score of hypothesis testing. Built on this base method, FGES (Fast GES) [6] constructs the causal graph in a way that is similar to FCI, i.e., relying on the collider causal structure when orienting the edges, making it more computationally efficient.

NOTEARS-Low-Rank (NTLR) Algorithm [4]. NTLR is a gradient-based causal discovery method that extends NOTEARS [13] with low-rank graphs. Leveraging the low-rank assumption and existing techniques, NTLR adapts causal structure learning methods to offer several useful insights into interpretable graphical conditions. It also introduces novel adaptations addressing scenarios where traditional methods may encounter limitations.

LiNGAM Algorithm [9]. The LiNGAM algorithm formulates the relationships between the time series using a linear and acyclic structural equation model with non-Gaussian errors. The causal relationships between the time series can be inferred by checking whether the residuals and predictors of the structural equation model are independent. Similar to PC, LiNGAM also assumes there are no hidden confounders affecting the time series. There are two popular types of LiNGAM: ICALiNGAM [9], which employs the independent component analysis, and DirectLiNGAM [10], which uses regression analysis to build the structural equation model.

B HYPERPARAMETER TUNING

In Section 4.1, when evaluating the performance of causal discovery methods, apart from using default values of the hyperparameters, we also evaluate these methods with tuned hyperparameters. To tune the hyperparameters, we use the Bayesian Information Criterion (BIC) score [2, 8] with cross-validation. Specifically, we first split all the metrics data into 2 parts: 2/3 for training and 1/3 for evaluation. We then create a search space consisting of multiple combinations of hyperparameters. For each combination, we run the causal discovery method on the training set and compute the BIC score on the evaluation set. Finally, we choose the combination of hyperparameters that yields the lowest BIC score (indicating better accuracy) and run the causal discovery method on the entire metrics data with this combination.

C ADDITIONAL EXPERIMENTAL RESULTS

C.1 Efficiency of Causal Discovery Methods

In Section 4.3, we evaluate how efficient causal discovery and RCA methods are. Here, Table S1 reports the running time (in seconds) of nine representative causal discovery methods on all six synthetic datasets. This is to answer the first part of the evaluation in Section 4.3, which is to evaluate the efficiency of causal discovery methods.

C.2 Performance of Causal Discovery Methods with Different Input Data Lengths

In Section 4.4, we evaluate the performance of the causal discovery and RCA methods with different input data lengths.

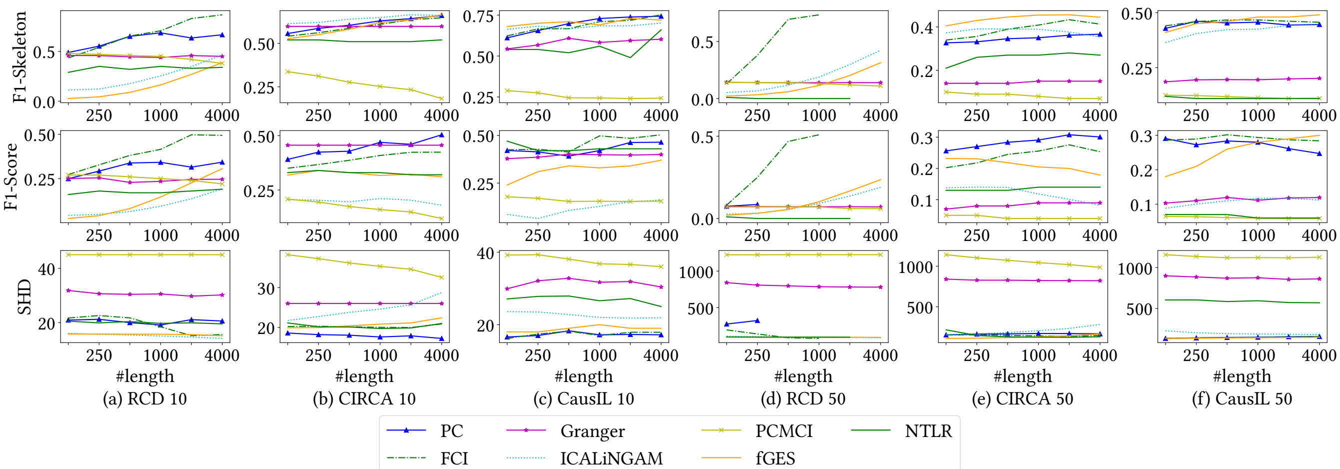
First, we show Figure S2, which reports the performance of seven causal discovery methods (Granger, PC, FCI, ICALiNGAM, PCMCI and fGES) on the six synthetic datasets with input data lengths

Table S1. Running time (in seconds) of nine causal discovery algorithms on synthetic datasets in default settings.

	CIRCA10	RCD10	CausIL10	CIRCA50	RCD50	CausIL50
PC	0.38	0.27	0.3	3.47	2.01	38.6
FCI	0.51	0.28	0.3	8.19	2.32	336.6
Granger	1.95	2.53	4.4	70.78	57.09	118.8
ICALiNGAM	0.75	0.33	0.6	13.58	2.3	16.1
DirectLiNGAM	0.63	0.51	0.7	32.13	26.65	41.8
GES	3.24	0.58	4.6	84813*	136.19	5543.5**
fGES	1.92	0.37	2.2	34.42	5.28	67.4
PCMCI	4.8	5.95	17.2	255.25	542.77	1052.7
NTLR	45.53	14.49	368.1	932.61	-	2406.42

(*) results are reported from 1 case, which takes nearly 24 hours.

(**) results are partially obtained from 2/100 cases due to exceeding the 1 hour/case time-out constraint for 2 consecutive times (1h20m and 1h40m, respectively).



(*) PC, FCI, and NTLR results on RCD50 were partially obtained due to OOM errors during execution and exceeding the time limit.

Figure S2. Performance of seven causal discovery methods on six synthetic datasets with different data lengths. A log-scale is used for the x-axis [1].

varying from 125 to 4000 data points. This figure is used to answer the first part of this evaluation, which is to evaluate how causal discovery methods perform with different input data lengths.

REFERENCES

[1] Charles K. Assaad, Emilie Devijver, and Eric Gaussier. 2022. Survey and Evaluation of Causal Discovery Methods for Time Series. *Journal of Artificial Intelligence Research* 73 (2022).

[2] Konstantina Biza, Ioannis Tsamardinos, and Sofia Triantafillou. 2020. Tuning causal discovery algorithms. In *International Conference on Probabilistic Graphical Models*. PMLR, 17–28.

[3] David Maxwell Chickering. 2002. Learning Equivalence Classes of Bayesian-Network Structures. *Journal of Machine Learning Research* 2 (2002), 445–498.

[4] Zhuangyan Fang, Shengyu Zhu, Jiji Zhang, Yue Liu, Zhitang Chen, and Yangbo He. 2023. On Low-Rank Directed Acyclic Graphs and Causal Structure Learning. *IEEE Transactions on Neural Networks and Learning Systems* (2023).

[5] C.W.J. Granger. 1980. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* 2 (1980), 329–352.

[6] Joseph D. Ramsey, Madelyn Glymour, Ruben Sanchez-Romero, and Clark Glymour. 2017. A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International Journal of Data Science and Analytics* 3, 2 (2017), 121–129.

[7] Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. 2019. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances* 5, 11 (2019).

[8] Gideon Schwarz. 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6, 2 (1978), 461 – 464.

[9] Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvarinen, and Antti Kerminen. 2006. A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research* 7, 72 (2006), 2003–2030.

[10] Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvärinen, Yoshinobu Kawahara, Takashi Washio, Patrik O. Hoyer, and Kenneth Bollen. 2011. DirectLiNGAM: A Direct Method for Learning a Linear Non-Gaussian Structural Equation Model. *Journal of Machine Learning Research* 12 (2011), 1225–1248.

[11] P. Spirtes, C. Glymour, and R. Scheines. 1993. *Causation, Prediction, and Search* (1st ed.). MIT press.

[12] Matthew J. Vowels, Necati Cihan Camgoz, and Richard Bowden. 2022. D'Ya Like DAGs? A Survey on Structure Learning and Causal Discovery. *Comput. Surveys* 55, 4, Article 82 (2022).

[13] Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. 2018. Dags with no tears: Continuous optimization for structure learning. *Advances in neural information processing systems* 31 (2018).