Towards the Disappearing Truth: Fine-Grained Joint Causal Influences Learning with Hidden Variable-Driven Causal Hypergraphs in Time Series

Kun Zhu, Chunhui Zhao*
College of Control Science and Engineering, Zhejiang University, Hangzhou, China
{kunzhu, chhzhao}@zju.edu.cn

Abstract

Causal discovery under Granger causality framework has yielded widespread concerns in time series analysis task. Nevertheless, most previous methods are unaware of the underlying ***causality disappearing problem***, that is, certain weak causalities are less focusable and may be lost during the modeling process, thus leading to biased causal conclusions. Therefore, we propose to introduce joint causal influences (i.e., causal influences from the union of multiple variables) as additional causal indication information to help identify weak causalities. Further, to break the limitation of existing methods that implicitly and coarsely model joint causal influences, we propose a novel hidden variable-driven causal hypergraph neural network to meticulously explore the locality and diversity of joint causal influences, and realize its explicit and fine-grained modeling. Specifically, we introduce hidden variables to construct a causal hypergraph for explicitly characterizing various fine-grained joint causal influences. Then, we customize a dual causal information transfer mechanism (encompassing a multi-level causal path and an information aggregation path) to realize the free diffusion and meticulous aggregation of joint causal influences and facilitate its adaptive learning. Finally, we design a multi-view collaborative optimization constraint toguarantee the characterization diversity of causal hypergraph and capture remarkable forecasting relationships (i.e., causalities). Experiments are conducted to demonstrate the superiority of the proposed model.

Introduction

Causal discovery on time series aims to interpret the causal structure from the observed data, which is help to gain insight into the underlying mechanisms of certain systems (Sugihara et al. 2012; Wang et al. 2020; Xu et al. 2019). Among various causal discovery methods, Granger causality (Granger 1969) is a popular and widely accepted paradigm (Appiah 2018; Charakopoulos et al. 2018; Chen and Zhao 2022), which has a high interpretability because it explains causalities as remarkable forecasting relationships. At present, some studies (Schwab et al. 2019; Song et al. 2023; Cheng et al. 2023) have recognized the shortcoming of linear assumption in traditional Granger causality methods. Accordingly, they have introduced deep neural networks to extend traditional Granger causality methods into nonlinear scenarios to encourage promising results.

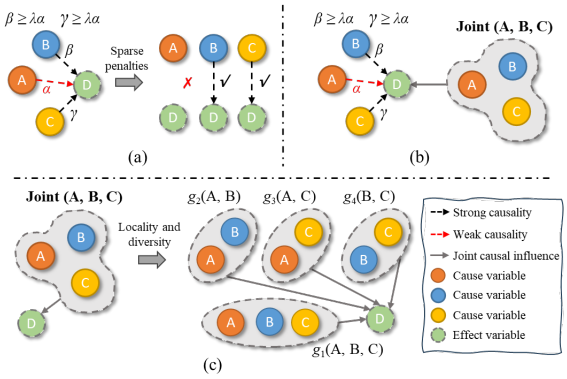


Figure 1: An example of pair-wise causalities and joint causal influences. (a) causality disappearing problem. (b) joint causal influences from variables A, B, and C. (c) locality and diversity of joint causal influences.

In these studies, a popular paradigm is to regard the weights of neural networks as causalities and design some sparsity penalties to gradually shrink unimportant weights (non-causalities) to zeros during training iterations, thereby finding remarkable causalities. However, in real scenarios, causalities between variables may be strong or weak. When the sparsity penalty is too strong or the number of training iterations is overmuch, the weights for certain weak causalities (relatively small) may be reduced to zeros along with non-causalities, thereby inducing the ***causality disappearing problem***, i.e., certain weak causalities are harder to focus on and may even be omitted in the modeling process. For example, genes, nutrition, and sleep time can all affect a person's height. Among them, genes may have a greater impact on height, while the impacts of other causes (i.e., nutrition and sleep time) are relatively weak, less prominent, and easier to be overlooked.

A general example of causality disappearing problem is shown in Figure 1(a), we assume that A, B, C are three

*Corresponding author: chhzhao@zju.edu.cn

cause variables of variable D and the corresponding causal coefficients α, β, γ (i.e., the strength of causality) satisfy $\beta \geq \lambda\alpha$ and $\gamma \geq \lambda\alpha$ (λ is a hyper-parameter to distinguish strong or weak causalities). It indicates that the causality between A and D is relatively weak (i.e., the causal influence of A on D (red dashed line) is much lower than that of B and C (black dashed lines)). In the modeling process, the causality between A and D may be difficult to discover or even be missed under sparse penalty, ultimately leading to blocked causal results. In particular, the strength of causality in this study is a relative concept and thus we use the magnitude of causal coefficient to measure the strength of causality.

The key to address the causality disappearing problem is to capture relatively weak causalities. In this study, we attempt to sense weak causalities from a new perspective, i.e., ***joint causal influences***, which differ from common separate causal influence (pair-wise or low-order causalities) and are produced by the union of multiple variables (high-order causalities) (Williams and Beer 2010; Chen et al. 2022). For example, as shown in Figure 1(b), variable A can have a separate causal influence on D alone (A → D), and it may also have joint causal influences on D through uniting with B and C (Joint(A, B, C) → D). In other words, complete causal influences between two variables can be divided into two parts: separate causal influence and joint causal influences (Van Leeuwen et al. 2021). Based on the above cognition, a feasible idea to solve the causality disappearing problem is to capture joint causal influences (Joint(A, B, C) → D), thereby providing additional causal indication information for pair-wise weak causalities (causality between A and D) and facilitating its identification.

However, accurately modeling joint causal influences remains a challenging task, because there may be various joint causal influences across all variables (diversity) and each joint causal influence may be produced by only partial variables (locality). For example, as shown in Figure 1(c), the joint causal influences Joint(A, B, C) is not directly equivalent to g_1(A, B, C), but may be any combination of g_1(A, B, C), g_2(A, B), g_3(A, C), and g_4(B, C), where $g_i(\ )$ denotes various joint causal influence functions. Unfortunately, in the latest studies, the modeling for joint causal influences is still relatively elementary because its locality and diversity has not yet been explored. Specifically, these studies merely construct some stacked neural networks to implicitly consider the single joint causal influence in which all variables are involved (i.e., g_1(A, B, C)), which induces coarse-grained and biased modeling for joint causal influences and is insufficient to provide dependable causal indication information.

Therefore, we propose a novel hidden variable-driven causal hypergraph neural network (CHGNN) to break the limitation of implicit and coarse-grained modeling, and explicitly learn diverse fine-grained joint causal influences, so as to deliver reliable causal indication information for identifying weak causalities. Specifically, we firstly generate a series of hidden variables to construct a causal hypergraph for explicitly characterizing various fine-grained joint causal influences. Then, a dual causal information transfer mechanism (DCM) is customized to realize the free diffusion and meticulous aggregation of joint causal influences and facilitate its adaptive learning, ultimately generating a fine-grained causal graph. Finally, a multi-view collaborative optimization constraint is designed to ensure the diversity of causal hypergraph and mine remarkable forecasting relationships (i.e., causalities).

The main contributions in this study are listed as follows:

- To the best of authors, it is the first time to summarize the causality disappearing problem in causal discovery task on time series, and we further propose to introduce joint causal influences to solve this problem.
- Unlike most existing studies that implicitly and coarsely consider joint causal influences, we propose a hidden variable-driven causal hypergraph to explicitly model fine-grained joint causal influences, thus providing reliable causal indication information to sense weak causalities.
- We design a multi-view collaborative optimization constraint, which enhances the characterization diversity of joint causal influences, while facilitating the screening and filtering of non-critical forecasting relationships, ultimately improving the accuracy and significance of causalities identification.

Related Work

Causal Discovery on Time Series Lately, causal discovery, identifying the causalities among variables, has become a challenge in the time series analysis task. Among various causal discovery methods (Runge et al. 2019; Zheng et al. 2018; Iseki et al. 2019), Granger causality methods have drawn momentum due to its high interpretability, which aims to uncover the remarkable forecasting relationships and regards them as causalities. Traditional Granger causality methods usually assume linear time series dynamics and construct the vector autoregressive model (Tank et al. 2022). However, they have poor generalization performance because they cannot deal with nonlinear scenarios. Thus, some studies introduce the deep neural networks (Zhao 2022; Chai et al. 2022) into Granger causality framework to help infer nonlinear causalities. For example, Tank et al. (2022) leveraged a neural Granger causality framework to mine the causality based on sparse weight parameters in neural network layers (i.e., autoregressive multilayer perceptron (MLP) and long-short term memory networks (LSTM)). Marcinkevic̆s and Vogt (2021) extended the self-explaining neural networks to effectively quantize the causal effects and used a time-reversed method to eliminate some spurious causalities. However, these studies mostly fail to notice the causality

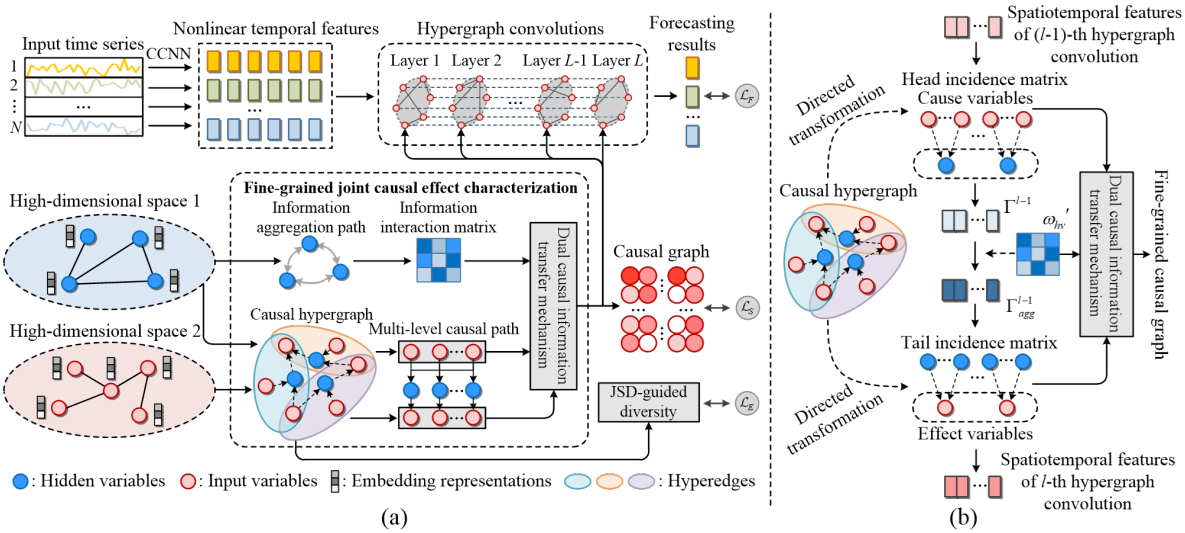


Figure 2: The proposed CHGNN. (a) The framework of CHGNN. (b) The generation process of fine-grained causal graph.

disappearing problem, which may lead to the omitting of some disappearing causalities (i.e., weak causalities), thereby resulting in the biased causal discovery results.

Hypergraph Representation and Modeling Hypergraph is a generalization of the naive graph and a generalized structure for correlation modeling (Zhou et al. 2005; Feng et al. 2019; Jiang et al. 2019). Compared to the naive graph that only models low-order correlations between pairwise nodes (Feng et al. 2020; Zhu and Zhao 2022), hypergraph can model high-order correlations among a set of nodes (i.e., more complex correlations that go beyond the pairwise category), thereby presenting better flexibility and capability in mining deep node interactions. In recent years, hypergraphs have drawn increasing attention and been developed in a variety of fields. For example, Bai et al. (2021) combined hypergraphs and the attention mechanism to adaptively identify the importance of different nodes in the same hyperedge. Gao et al. (2022) proposed a hypergraph neural network and some hyperedge generation methods to accurately model the correlations in multi-modal/multi-type data. Inspired by these studies, we introduce directed hypergraphs into time series causal discovery, which can capture joint causal influences (high-order causalities) while preserving the direction information of causalities.

Methodology

Overview of CHGNN

The framework of CHGNN is shown in Figure 2(a), which consists of two modules: fine-grained joint causal influence characterization, deep spatiotemporal features modeling and causality extracting. In the first module, we construct multiple hidden variables (capturing various joint causal influences) for input variables, and map all variables into high-dimensional spaces to generate their corresponding embedding representations. Then, we use the embedding representations of input variables and hidden variables to generate a causal hypergraph, which can explicitly characterize various joint causal influences. Further, the directed transformation is conducted on the causal hypergraph to gain a multi-level causal path (i.e., cause variables → hidden variables → effect variables), which can ensure the free diffusion of causal influences and guide its ultimate transmission direction. In addition, an information aggregation path (i.e., hidden variables → hidden variables) is constructed to meticulously aggregate the causal information generated by same cause variables in different joint causal influences (different hidden variables), thus facilitating the significance of causalities learning. In the second module, channel-wise convolutional neural network (CCNN) and hypergraph convolution (HGC) are employed to obtain the forecasting results of input variables, while customizing a dual causal information transfer mechanism (DCM) to realize the adaptive learning of joint causal influences and generate the fine-grained causal graph. Finally, a multi-view collaborative optimization constraint is designed to train all modules with an end-to-end manner.

Fine-grained Joint Causal Influence Characterization

High-dimensional Space Generation We first construct M hidden variables $hv_j (1 \le j \le M)$ for N input variables $v_i (1 \le i \le N)$ to capture joint causal influences. Particularly, to facilitate the calculation and subsequent causal hypergraph generation, we set M to $\mu \times N$ (μ is an adjustable hyper-parameter). Then, to meticulously quantify the causalities between input variables and hidden variables, we map input variables and hidden variables into two different high-dimensional spaces respectively and then generate the corresponding embedding representations:

$$VE = \mathrm{Mp}(v_1,\ v_2,\ ...,\ v_N),\ HVE = \mathrm{Mp}(hv_1,\ hv_2,\ ...,\ hv_M) \quad (1)$$

where MP denotes the mapping operation, $VE \in \mathbb{R}^{N \times ED}$ and $HVE \in \mathbb{R}^{M \times ED}$ denote the initial embedding representations that can be adaptively learned during the training process, ED denotes the embedding dimension.

Causal Hypergraph and Multi-level Causal Path To explicitly capture various joint causal influences, the above embedding representations are used to build the causal hypergraph, in which input variables (corresponding to nodes in the causal hypergraph) are regarded as possible cause variables and hidden variables (corresponding to hyperedges in the causal hypergraph) are regarded as possible joint causal influences, as shown in Equation (2).

$$CHG = \tanh(VE' \times (HVE')^{\mathrm{T}}) \quad (2)$$

where tanh is the activation function, $CHG \in \mathbb{R}^{N \times M}$ denotes the causal hypergraph and its specific values can be adaptively learned during training process:

$$CHG(v_i,\ hv_j) = \begin{cases} \gamma \in [-1,\ 0) \cup (0,\ 1],\ v_i \text{ is a cause variable of } hv_j \\ 0,\ v_i \text{ is not a cause variable of } hv_j \end{cases} \quad (3)$$

where $\left|CHG(v_i,\ hv_j)\right|$ denotes the influence strength of i-th input variable in the j-th joint causal influence.

Then, we generate a multi-level causal path based on the obtained CHG. Specifically, directed transformation is conducted on CHG to gain head incidence matrix $H_{head} \in \mathbb{R}^{N \times M}$ and tail incidence matrix $H_{tail} \in \mathbb{R}^{N \times M}$, as shown in Equations (4) and (5):

$$H_{head} = CHG + I_N^1 \,||\, ... \,||\, I_N^\mu \quad (4)$$

$$H_{tail} = I_N^1 \,||\, ... \,||\, I_N^\mu \quad (5)$$

where $I_N^i \in \mathbb{R}^{N \times N} (1 \le i \le \mu)$ denotes the identity matrix. Adding $I_N^1 \,||\, ... \,||\, I_N^\mu \in \mathbb{R}^{N \times M}$ into H_{head} is to consider self-influence of input variables.

H_{head} and H_{tail} together constitute the multi-level causal path (i.e., cause variables → hidden variables → effect variables). Specifically, H_{head} represent the first-level causal path (cause variables → hidden variables), it customizes a hyperedge for each hidden variable to receive various causal information from cause variables, thereby ensuring the free diffusion of joint causal influences. H_{tail} represent the second-level causal path (hidden variables → effect variables), it customizes multiple hyperedges for each effect variable to output causal information from hidden variables, thus guiding the final transmission direction of joint causal influences.

Information Aggregation Path As shown in Figure 2(a), we build an information aggregation path among hidden variables to comprehensively consider and integrate the causal information generated by same cause variables in different joint causal influences (i.e., different hidden variables), thereby improving the significance of causalities learning. Specifically, the embedding representations of hidden variables are used to calculate the correlations for obtaining the information interaction matrix $\omega_{hv} \in \mathbb{R}^{M \times M}$, as shown in Equations (6) and (7):

$$\omega_{hv} = \mathrm{ReLU}(\tanh(\lambda \times HVE' \times (HVE')^{\mathrm{T}})) + I_M \quad (6)$$

$$\omega_{hv}' = \mathrm{Sparse}^k(D_{hv}^{-\frac{1}{2}} \omega_{hv} D_{hv}^{-\frac{1}{2}}) \quad (7)$$

where λ is a hyper-parameter to control the saturation rate of tanh, $I_M \in \mathbb{R}^{M \times M}$ denotes the identity matrix, $\omega_{hv}' \in \mathbb{R}^{M \times M}$ denotes the normalized information diffusion matrix, $D_{hv} \in \mathbb{R}^{M \times M}$ denotes the degree matrix of ω_{hv}. $\mathrm{Sparse}^k(\cdot)$ can keep the k largest values of all row vectors in ω_{hv} to achieve the information aggregation among highly correlated hidden variables, because the higher the correlations among hidden variables, the more similar the joint causal influences they characterize, and the more common cause variables they capture.

Deep Spatiotemporal Features Modeling and Causality Extracting

Spatiotemporal Features Modeling In this section, we model the deep spatiotemporal features of multivariate time series to obtain the forecasting results, while capturing remarkable forecasting relationships (i.e., causalities) among all variables. As shown in Figure 2(a), the spatiotemporal features modeling in CHGNN consists of two parts: nonlinear temporal features extraction and hypergraph convolutions for deep causality mining.

(1) Nonlinear Temporal Features Extraction

Assume that $X_{t-T+1:\,t} \in \mathbb{R}^{N \times T}$ denotes the original input time series, where N and T are the number of input variables and time points, respectively. As shown in Figure 2(a), we propose a channel-wise convolutional neural network (CCNN) to regard each variable as an independent channel and perform the nonlinear transformation on the corresponding time series, thereby gaining temporal dependency representation of each variable:

$$NX_{t-T+1:\,t} = \sum_{q=1}^{Q} X_{t-T+1:\,t} * \theta_q \quad (8)$$

where $NX_{t-T+1:\,t} \in \mathbb{R}^{N \times (T-len+1)}$ and $\theta_q \in \mathbb{R}^{1 \times len} (1 \le q \le Q)$ denote all nonlinear temporal features and 1D convolution kernels, respectively. Q is the number of convolution kernels and len is the length of convolution kernels.

(2) Hypergraph Convolutions Mining Deep Causalities

The nonlinear temporal features extracted by CCNN are fed into multiple HGCs to mine the deep spatiotemporal features to output the final forecasting results:

$$F^l = \begin{cases} \mathrm{ReLU}\left(\mathrm{HGCN}\left(F^{l-1}\right)\right), & l = 1, 2, ..., L-1 \\ \mathrm{HGCN}\left(F^{l-1}\right), & l = L \end{cases} \tag{9}$$

where HGC denotes the hypergraph convolution layer, L denotes the number of HGCs. $F^l \in \mathbb{R}^{N \times C_1}$ denotes the extracted spatiotemporal features of l-th HGC, C_1 is the number of units in HGC. In particular, $F^0 = NX_{t-T+1:t} \in \mathbb{R}^{N \times T}$ and $F^L = \hat{x}_{t+1} \in \mathbb{R}^N$ (the forecasting results of input variables).

Most importantly, HGC can mine remarkable forecasting relationships (i.e., causalities) among variables by minimizing forecasting errors, and then generates the fine-grained causal graph, which will be elaborated later.

Fine-grained Causal Graph Generation Each HGC in CHGNN contains a dual causal information transfer mechanism (DCM), which consists of the multi-level causal path and information aggregation path mentioned before. DCM can guarantee the free diffusion and meticulous aggregation of joint causal influences and achieve its adaptive learning, thereby gaining the generation of fine-grained causal graph. As shown in Figure 2(b), taking the l-th HGC as an example, DCM consists of three stages, which is detailed as follows:

(1) Cause Variables → Hidden Variables

According to the spatiotemporal features of (l-1)-th HGC $F^{l-1} \in \mathbb{R}^{N \times C_1}$ and various joint causal influences characterized by the head incidence matrix H_{head}, hidden variable features $\Gamma^{l-1} \in \mathbb{R}^{M \times C_1}$ can be obtained:

$$\Gamma^{l-1} = D_{head}^{-1} H_{head}^{\mathrm{T}} F^{l-1} \Theta^l \tag{10}$$

$$D_{head} = \mathrm{diag}(\mathrm{csum}(H_{head})) \tag{11}$$

where csum(·) can obtain the sum of each column, diag(·) can transform a vector to diagonal matrix, $\Theta^l \in \mathbb{R}^{C_1 \times C_1}$ denotes the weight parameters of l-th HGC, C_1 the number of units in HGC, $D_{head} \in \mathbb{R}^{M \times M}$ denotes the degree matrix of H_{head}, indicating that it uses the average aggregation for hypergraph convolution.

(2) Hidden Variables → Hidden Variables

Based on the ω_{hv}', the information interaction among highly correlated hidden variables can be realized. Specifically, it can enhance and aggregate the causal information generated by same cause variables in different joint causal influences to obtain refined hidden variable features $\Gamma_{agg}^{l-1} \in \mathbb{R}^{M \times C_1}$, as shown in Equation (12).

$$\Gamma_{agg}^{l-1} = \omega_{hv}' \Gamma^{l-1} \tag{12}$$

(3) Hidden Variables → Effect Variables

The refined hidden variable features can be transferred to the corresponding effect variables using the tail incidence matrix H_{tail}, and the spatiotemporal features of l-th HGC $F^l \in \mathbb{R}^{N \times C_1}$ can be obtained:

$$F^l = D_{tail}^{-1} H_{tail} \Gamma_{agg}^{l-1} \tag{13}$$

$$D_{tail} = \mathrm{diag}(\mathrm{rsum}(H_{tail})) \tag{14}$$

where rsum(·) can obtain the sum of each row, $D_{tail} \in \mathbb{R}^{N \times N}$ denotes the degree matrix of H_{tail}.

Finally, five components involved in DCM can be fused to generate the fine-grained causal graph $A \in \mathbb{R}^{N \times N}$, as shown in Equation (15).

$$A = D_{tail}^{-1} H_{tail} \omega_{hv}' D_{head}^{-1} H_{head}^{\mathrm{T}} \tag{15}$$

Multi-view Collaborative Optimization Constraint

In this section, we design a multi-view collaborative optimization constraint $\mathcal{L}_{mcoc}$ to achieve the end-to-end training of CHGNN, which contains three parts:

JS Divergence-guided Diversity Loss For each effect variable, we select its corresponding all hyperedges (multiple joint causal influences) and maximize the sum of the JS divergence of pair-wise hyperedges, thereby ensuring the characterization diversity of joint causal influences, as shown in Equations (16) - (18).

$$H'_{head}(i, j) = \mathrm{softmax}(H_{head}(i, j)) = \frac{exp(H_{head}(i, j))}{\sum_{i=0}^{N} exp(H_{head}(i, j))} \tag{16}$$

$$JS(a) = \sum_{j=1}^{\mu} \sum_{z=j+1}^{\mu} D_{JS}(H'_{head}[:, a + j \times N] \,||\, H'_{head}[:, a + z \times N]) \tag{17}$$

$$\mathcal{L}_{JS} = \frac{1}{\mu} \sqrt{\sum_{a=1}^{\mu} (JS(a))^2} \tag{18}$$

where softmax is the activation function. $H_{head}(i, j)$ denotes the value in the i-th row $(1 \le i \le N)$ and j-th column $(1 \le j \le M)$ in H_{head}. $JS(a)$ denotes the sum of JS divergence for a-th effect variable $(1 \le a \le N)$. μ is an adjustable hyper-parameter that satisfies $\mu \times N = M$.

Collection Sparse Loss To ensure the significance of causalities, we conduct a collection sparsity constraint on the causal graph $A \in \mathbb{R}^{N \times N}$. Specifically, each row in A is regarded as a collection, using the L1 penalty within each collection and L2 penalty between collections:

$$\mathcal{L}_S = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \left| A(i, j) \right| \right)^2 \tag{19}$$

Forecasting Error Minimizing Loss To obtain better forecasting relationships to represent causalities under the Granger causality framework, we minimize the forecasting errors of CHGNN, as shown in Equation (20).

$$\mathcal{L}_F = \frac{1}{N} \left\| x_{t+1} - \hat{x}_{t+1} \right\|_2^2 \tag{20}$$

Model	VAR-30		Joint-VAR-10		Joint-VAR-20		LORENZ-96		fMRI	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
LG	92.18	75.61	81.24	76.98	81.59	71.80	48.42	15.94	71.55	31.52
VGAM	77.63	47.06	76.95	56.80	60.11	35.65	89.87	44.18	53.73	12.10
cMLP	92.80	78.66	67.18	46.59	59.68	32.60	99.71	97.98	67.80	26.45
cLSTM	95.03	83.33	78.06	65.68	64.53	38.16	**99.98**	**99.83**	86.31	42.65
TCDF	79.86	44.75	85.66	80.12	69.44	41.59	95.73	76.55	91.34	52.45
eSRU	85.29	60.08	85.83	80.99	76.30	61.35	99.27	95.43	82.27	38.00
GVAR	84.28	59.22	86.69	83.66	82.82	69.99	99.68	98.10	89.74	56.96
CHGNN	**96.03**	**88.91**	**93.10**	**90.14**	**86.94**	**80.81**	99.80	98.72	**92.54**	**82.41**

Note: The experimental results are the mean values after 5 executions, respectively. The best values are marked in bold, and the second-best values are marked with underline.

Table 2: Averaged AUROC and AUPRC on five synthetic datasets.

where x_{t+1}, $\hat{x}_{t+1} \in \mathbb{R}^N$ denote real and predicted values at time point t+1, respectively. $\|\cdot\|_2$ denotes the L2-norm.

Finally, the above three losses are integrated to gain the $\mathcal{L}_{mcoc}$, as shown in Equation (21).

$$\mathcal{L}_{mcoc} = \eta_1 \mathcal{L}_F - \eta_2 \mathcal{L}_{JS} + \eta_3 \mathcal{L}_S \tag{21}$$

where $0 \le \eta_1,\ \eta_2,\ \eta_3 \le 1$ are the hyper-parameters.

Experiments

Synthetic and Real Datasets

We conduct the experiment on six datasets, including three benchmark synthetic datasets (VAR-30, Lorenz-96, fMRI), two specific synthetic datasets (Joint-VAR-10 and Joint-VAR-20) and one real dataset (real gas turbine fault data (GTF)). Detailed information is provided in the Appendix.

Baseline Models and Evaluation Metrics

We compare the proposed CHGNN with seven baseline models (list in Table 1) to demonstrate its effectiveness:

Only consider pairwise causalities
Lasso Granger (LG) (Arnold et al. 2007)
VARLiNGAM (VGAM) (Hyvärinen et al. 2010)
Coarse-grained modeling for joint causal influences
TCDF (Nauta et al. 2019)
eSRU (Khanna and Tan 2020)
GVAR (Marcinkevic˘s and Vogt 2021)
cMLP (Tank et al. 2022)
cLSTM (Tank et al. 2022)

Table 1 Baseline models.

The detailed parameters of CHGNN can be referred to the Appendix. In addition, since the real causal graph is Boolean matrix, a threshold truncation operation is required on the causal graph (obtained by Equation (15)) to make each element be either 0 or 1. To eliminate the influence of subjective assumptions in the performance evaluation, we choose two metrics that do not require a predetermined threshold, i.e., area under the receiver operating characteristic curve (AUROC) and precision-recall curve (AUPRC). We only consider off-diagonal elements of causal graph in the evaluation process.

Comparison with Baseline Models

Table 2 shows the causal discovery results on the five synthetic datasets, some phenomenon can be summarized as follows: the proposed CHGNN exhibits superior results other baseline models. LG and VGAM show unsatisfactory generalization ability, as they perform well in linear scenarios (i.e., VAR-30), mediocrely in mixed linear-nonlinear scenarios (i.e., Joint-VAR-10 and Joint-VAR-20), and poorly in nonlinear scenarios (i.e., Lorenz-96 and fMRI). cMLP and cLSTM show great results in the high-dimensional scenario (i.e., VAR-30), but have a significant performance degradation in scenarios with causality disappearing problem (i.e., Joint-VAR-10 and Joint-VAR-20), because they mainly rely on sparsity constraints that may lead to the omission of some weak causality. Comparatively, TCDF, eSRU, and GVAR have better results due to their specially designed neural network architecture. Moreover, GVAR exhibit more competitive AUPRC among them, because TCDF and eSRU is struggled to shrink weight coefficients to exact zeros (Marcinkevic˘s and Vogt 2021), thus leading to more redundant causalities. However, their models still have suboptimal performance because they only implicitly and coarsely model joint causal influences, which is not sufficient to help address the causality disappearing problem. Apparently, our CHGNN significantly outperform all baseline models, especially AUPRC, indicating that CHGNN tends to gain more concise and accuracy causalities.

To further illustrate the superiority of CHGNN in solving causality disappearing problem, Figure 3 visualizes the predicted and real causal graphs of CHGNN and some baseline models. Causalities marked by red boxes represent weak causalities. It can be observed that weak causalities are indeed difficult to identify, while CHGNN presents better sensing ability for them. This phenomenon demon-

strates that capturing fine-grained joint causal influences is necessary to identify weak causalities.

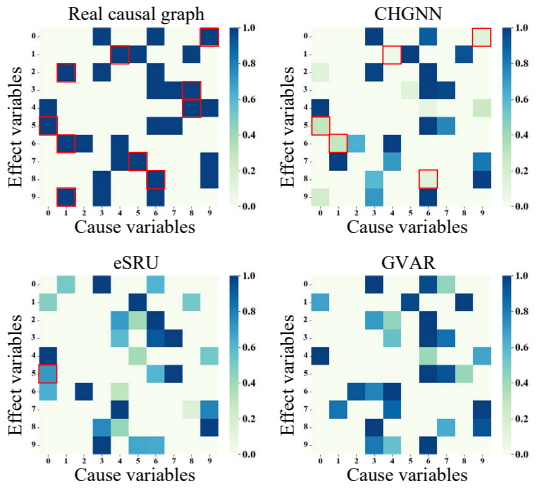


Figure 3: Visualization of predicted and real causal graphs on Joint-VAR-10, preserving 3 cause variables with largest coefficients for each effect variable.

In addition, the detailed results on real dataset GTF can be referred to the Appendix.

Specific Analysis

Ablation Study

We explore the importance of some components in CHGNN, so the following four variants are constructed:

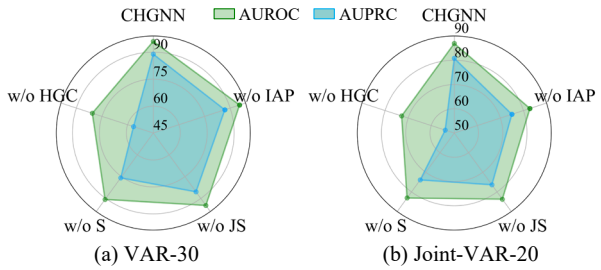
(1) w/o IAP: deleting information aggregation path.

(2) w/o JS: deleting JS divergence-guided diversity loss.

(3) w/o S: deleting collection sparse loss.

(4) w/o HGC: replacing all hypergraph convolutions with common graph convolutions (Kipf and Welling 2017).

The corresponding results on two datasets are shown in Figure 4. Some intuitive conclusion can be obtained: the neglect of information aggregation path can degrade the causal discovery results (compare CHGNN with w/o IAP); the characterization diversity of joint causal influences is essential for CHGNN (compare CHGNN with w/o JS); the collection sparse loss can ensure the great performance of CHGNN (compare CHGNN with w/o S); the hypergraph convolutions in the causal hypergraph are more suitable for modeling joint causal influences than common graph convolutions (compare CHGNN with w/o HGC).



(a) VAR-30 (b) Joint-VAR-20

Figure 4: Ablation results of CHGNN and its four variants.

Visualization of Captured Joint Causal Influences

We visualize part hyperedges (corresponding to hidden variables) in the causal hypergraph of CHGNN, as shown in Figure 5. Taking Joint-VAR-10 for an example, in the real causal network (Figure 7 (a)), v_3, v_6, and v_9 are the cause variables for effect variable v_0; the causality $v_9 \rightarrow v_0$ (blue dashed line) is weak and the causalities $v_3 \rightarrow v_0$ and $v_6 \rightarrow v_0$ (red lines) are strong; v_3 and v_9 can generate a joint causal influence to v_0; v_6 and v_9 can generate a joint causal influence to v_0. Figure 7 (b) shows joint causal influences captured by two hidden variables pointing to v_0. Hidden variable hv_0 successfully captures the joint causal influence from v_6 and v_9 and hidden variable hv_1 successfully capture the joint causal influence from v_3 and v_9. In summary, the above results demonstrate the effectiveness of CHGNN to capture fine-grained joint causal influences, thereby enabling accuracy sensing for weak causalities.

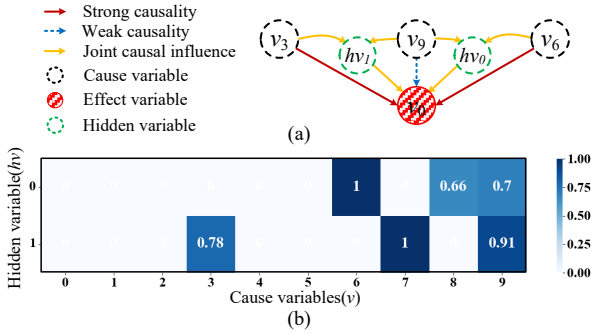


Figure 5: Visualization of captured and real joint causal influences on Joint-VAR-10. (a) real causal influences. (b) the captured joint causal influences, preserving 3 cause variables with largest coefficients for each hidden variable.

Conclusion

In this study, we explicitly mine fine-grained joint causal influences as additional causal indication information, which can effectively address the overlooked causality disappearing problem in the causal discovery task of time series. Specifically, the proposed CHGNN can sense the locality of joint causal influences through a customized causal hypergraph and capture diverse joint causal influences through the proposed multi-view collaborative optimization constraint, ultimately facilitating its explicit and fine-grained modeling. To validate the superiority of CHGNN, we compare CHGNN with other models (i.e., seven baseline models and some variants from CHGNN) on synthetic and real datasets. It can be concluded that CHGNN presents better causal discovery performance and possesses better sensing ability for weak causalities and imperative components in CHGNN can mutually boost each other to enhance the causal discovery ability of CHGNN. Furthermore, in real application scenarios, CHGNN also displays the promising potential to identify complex causalities.

Acknowledgments

This work is partly supported by the National Natural Science Foundation of China (No. 62125306), and National Key R&D Program of China (2019YFC1908100, 2022YFB3304703).

References

Appiah, M. O. (2018). Investigating the multivariate Granger causality between energy consumption, economic growth and CO2 emissions in Ghana. *Energy Policy*, 112, 198-208.

Arnold, A., Liu, Y., & Abe, N. (2007). Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 66-75

Bai, S., Zhang, F., & Torr, P. H. (2021). Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110, 107637.

Chai, Z., Zhao, C., Huang, B., & Chen, H. (2021). A deep probabilistic transfer learning framework for soft sensor modeling with missing data. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 7598-7609.

Charakopoulos, A. K., Katsouli, G. A., & Karakasidis, T. E. (2018). Dynamics and causalities of atmospheric and oceanic data identified by complex networks and Granger causality analysis. *Physica A: Statistical Mechanics and its Applications*, 495, 436-453.

Chen, H., Du, K., Yang, X., & Li, C. (2022). A review and roadmap of deep learning causal discovery in different variable paradigms. *arXiv preprint arXiv: 2209.06367*.

Chen, J., & Zhao, C. (2022). Multi-lag and multi-type temporal causality inference and analysis for industrial process fault diagnosis. *Control Engineering Practice*, 124, 105174.

Cheng, Y., Yang, R., Xiao, T., Li, Z., Suo, J., & He, K. et al. (2023). CUTS: neural causal discovery from irregular time-series data. In *Proceedings of the 11th International Conference on Learning Representations*, pp. 1-24.

Feng, L., Zhao, C., Li, Y., Zhou, M., Qiao, H., & Fu, C. (2020). Multichannel diffusion graph convolutional network for the prediction of endpoint composition in the converter steelmaking process. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-13.

Feng, Y., You, H., Zhang, Z., Ji, R., & Gao, Y. (2019). Hypergraph neural networks. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, pp. 3558-3565.

Gao, Y., Feng, Y., Ji, S., & Ji, R. (2022). HGNN+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3), 3181-3199.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424-438.

Hyvärinen, A., Zhang, K., Shimizu, S., & Hoyer, P. O. (2010). Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11(5), 1709-1731.

Iseki, A., Mukuta, Y., Ushiku, Y., & Harada, T. (2019). Estimating the causal effect from partially observed time series. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, pp. 3919-3926.

Jiang, J., Wei, Y., Feng, Y., Cao, J., & Gao, Y. (2019). Dynamic Hypergraph Neural Networks. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 2635-2641.

Khanna, S., & Tan, V. Y. (2019). Economy statistical recurrent units for inferring nonlinear granger causality. In *Proceedings of the 7th International Conference on Learning Representations*, pp. 1-19.

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations*, pp. 1-14.

Marcinkevičs, R., & Vogt, J. E. (2021). Interpretable models for granger causality using self-explaining neural networks. In *Proceedings of the 9th International Conference on Learning Representations*, pp. 1-19.

Nauta, M., Bucur, D., & Seifert, C. (2019). Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1), 19.

Schwab, P., Miladinovic, D., & Karlen, W. (2019). Granger-causal attentive mixtures of experts: Learning important features with neural networks. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, pp. 4846-4853.

Song, P., Zhao, C., & Huang, B. (2023). MPGE and RootRank: A sufficient root cause characterization and quantification framework for industrial process faults. *Neural Networks*, 161, 397-417.

Sugihara, G., May, R., Ye, H., Hsieh, C. H., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 338(6106), 496-500.

Tank, A., Covert, I., Foti, N., Shojaie, A., & Fox, E. B. (2022). Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8), 4267-4279.

Van Leeuwen, P. J., DeCaria, M., Chakraborty, N., & Pulido, M. (2021). A framework for causal discovery in non-intervenable systems. *Chaos*, 31(12), 123128.

Wang, Y., Menkovski, V., Wang, H., Du, X., & Pechenizkiy, M. (2020). Causal discovery from incomplete data: a deep learning approach. *arXiv preprint arXiv: 2001.05343*.

Williams, P. L., & Beer, R. D. (2010). Nonnegative decomposition of multivariate information. *arXiv preprint arXiv: 1004.2515*.

Xu, C., Huang, H., & Yoo, S. (2019). Scalable causal graph learning through a deep neural network. In P*roceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1853-1862.

Zhao, C. (2022). Perspectives on nonstationary process monitoring in the era of industrial artificial intelligence. *Journal of Process Control*, 116, 255-272.

Zheng, X., Aragam, B., Ravikumar, P. K., & Xing, E. P. (2018). Dags with no tears: Continuous optimization for structure learning. In *Proceedings of the 32nd Annual Conference on Neural Information Processing Systems*, pp. 9492-9503.

Zhou, D., Huang, J., & Schölkopf, B. (2005). Learning with hypergraphs: Clustering, classification, and embedding. In *Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pp. 1601-1608.

Zhu, K., & Zhao, C. (2022). Dynamic graph-based adaptive learning for online industrial soft sensor with mutable spatial coupling relations. *IEEE Transactions on Industrial Electronics*, 70(9), 9614-9622.