Deep Learning-Based Dynamic Granger Causality Analysis on Multivariate Time Series

Authors: Liu Bo, Hong Shenda, Li Hongyan, Li Hongyan

Date: 2022-11-08T00:00:00+00:00

Abstract

An important aspect of complex dynamical systems research is discovering the interaction relationships among components that exist in the form of multivariate time series. Compared to correlation, causality provides deeper insights into the essential mechanisms of systems and plays a significant role in the study of complex dynamical systems by going beyond simple correlation analysis. Granger causality analysis provides a powerful time series structure discovery framework and has been widely applied across various fields since its inception, holding significant research importance and practical value. However, in highly complex nonlinear dynamical systems such as human physiological systems, where causal relationships are established on multivariate, nonlinear, and dynamic time series, this poses a substantial challenge for traditional Granger causality analysis methods. To address this, this paper proposes a deep learning-based dynamic Granger causality analysis method called DNNGC to discover Granger causal relationships among complex, dynamic, and nonlinear multivariate time series. DNNGC leverages the modeling capability of deep learning for complex nonlinear relationships to fully capture the nonlinear dependencies among multivariate temporal data. It achieves dynamic causal analysis by incorporating time periods of existence into the original Granger causality framework and improves its predictive structure to reduce model space complexity from O(n2) to O(n), thereby enhancing feasibility for analyzing large-scale time series in real-world systems. Furthermore, this paper designs a neural network architecture tailored for Granger causality and provides a detailed discussion of DNNGC's experimental results under three typical forms of Granger causality existence. Finally, this paper applies DNNGC to analyze interactions among various physiological systems during human sleep, uncovering potential dynamic Granger causality patterns in the human body, which holds significant practical implications.

Full Text

Preamble

Dynamic Granger Causality Analysis of Multivariate Time Series Based on Deep Learning

Liu Bo[1,2], Hong Shenda[3], Li Hongyan[1,2]
[1]School of Intelligence Science and Technology, Peking University, Beijing 100871, China
[2]Key Laboratory of Machine Perception (Peking University), Ministry of Education, Beijing 100871, China
[3]National Institute of Health Data Science, Peking University, Beijing 100871, China

Abstract

A fundamental challenge in studying complex dynamical systems is discovering the interaction patterns among system components represented as multivariate time series. Causality provides deeper insights into system mechanisms than correlation, playing a crucial role in understanding complex dynamical systems beyond simple correlation analysis. Granger causality analysis offers a powerful framework for time series structure discovery and has been widely applied across various domains since its inception, holding significant research value and practical importance. However, in highly complex nonlinear dynamical systems such as human physiological systems, causal relationships are embedded in multivariate, nonlinear, and dynamic time series, posing substantial challenges to traditional Granger causality analysis methods.

To address these challenges, this paper proposes a deep learning-based dynamic Granger causality analysis method called DNNGC to discover Granger causal relationships among complex, dynamic, and nonlinear multivariate time series. DNNGC leverages the capacity of deep learning to model complex nonlinear relationships, thereby fully capturing the nonlinear dependencies in multivariate temporal data. By augmenting traditional Granger causality with existence time periods, DNNGC enables dynamic causal analysis. Furthermore, we improve its predictive structure to reduce model space complexity from $O(n^2)$ to $O(n)$, enhancing feasibility for large-scale time series analysis in real-world systems. Additionally, we design a neural network architecture tailored for Granger causality analysis and present detailed experimental results under three typical forms of Granger causality. Finally, we apply DNNGC to analyze causal interactions among physiological systems during human sleep, revealing potential dynamic Granger causal relationships with significant practical implications.

Keywords: deep learning, causal inference, information flow, multivariate time series, dynamic Granger causality

1. Introduction

The activities of components in complex dynamical systems are typically recorded and stored as multivariate time series, which reflect the operational processes of these systems [1]. Investigating the dynamically evolving interaction patterns within these sequences can reveal and illuminate the underlying mechanisms and structures governing system behavior. For instance, electroencephalograms (EEG) capture and reflect neuronal activity in the brain, with existing research demonstrating that information flow correlates with intrinsic brain network architecture [2]. Analyzing brain functional network structures from multivariate time series using various interaction analysis methods enhances our understanding of how the brain executes specific functions [3-4].

When simple correlation or independence measures are employed to characterize interactions among multivariate time series, inherent limitations in direction identification and the influence of common driver variables may lead to indirect interactions being misidentified as direct correlations [3], resulting in redundant or spurious information within complex systems. In contrast, causality [5-6] provides a deeper explanation of "who causes whom," articulating the direction of information flow between variables and enabling the discovery of essential system mechanisms beyond simple correlation analysis [6]. For example, human physiological systems continuously interact, generating vast multivariate time series. Figure 1 Figure 1: see original paper illustrates multivariate time series composed of electrical signals from four body regions, while (b) and (c) present correlation analysis and dynamic causal analysis results during sleep, respectively. The blue lines represent ground-truth relationships: electrooculogram (EOG) causes electrocardiogram (ECG) and chin electromyography (CHIN EMG), while ECG causes brain activity (EEG). However, correlation analysis identifies all variables as correlated, producing false relationships represented by gray lines in (c) and leading to erroneous conclusions.

For highly complex multivariate nonlinear systems like the human body, causal relationships vary dynamically over time. Physiologists have discovered that deep sleep primarily involves interactions within brain systems, whereas light sleep exhibits increased whole-body system interactions [7]. As shown in Figure 1, this dynamic nature manifests as causal relationships that appear or disappear over time, potentially exhibiting time-varying patterns as illustrated in Figure 1(c). Analyzing dynamic Granger causality from these multivariate time series would significantly enhance our understanding of fundamental system mechanisms. Since a malfunction in any variable's corresponding system could trigger disease in the entire body, knowledge of dynamic causal relationships among physiological components would enable rapid limitation of potential causes to those variables exhibiting causal influence on the faulty variable during the relevant time period, thereby saving precious time in medical treatment.

Granger causality (GC) [8] provides a powerful framework for time series structure discovery by quantifying the extent to which one time series' past improves prediction of another's future evolution, thereby defining "causality" from a predictive perspective. Since its proposal, Granger causality has been widely applied in economics [9], finance [10], and increasingly in neuroscience research [11]. A typical application employs Granger causality as a primary method for analyzing effective connectivity [12] from multivariate time series to investigate how one neural system influences another [13-14].

However, applying traditional Granger causality to causal analysis of multivariate time series in complex systems faces several challenges: (1) traditional GC analyzes only two variables while ignoring other sequences in the system; (2) it requires strong prior knowledge to assume linear relationships, making it difficult to analyze complex nonlinear dependencies; and (3) it statically analyzes causality over the entire time series, failing to capture potential dynamic causal relationships in complex systems.

Deep neural networks can represent complex nonlinear interactions between inputs and outputs. Long Short-Term Memory (LSTM) networks [15-17] have demonstrated impressive performance in predicting multivariate time series [18], and prediction is crucial for determining Granger causality. Moreover, as a data-driven non-parametric approach, deep neural networks can achieve effective models without prior knowledge, making them highly suitable for modeling complex nonlinear relationships in unknown systems [19-20].

2. Related Work

Granger causality traces back to Wiener's statistical concept of causality [21]: if knowing the past of time series Y improves prediction of X (compared to using only X's past), then Y causes X. Nobel laureate Granger [8] formally proposed this time series causality concept within the context of linear vector autoregressive (VAR) models for stochastic processes: if using all available information (representing all relevant variables' history up to time $t-1$) better predicts X_t than using information excluding Y_t, and this result is statistically significant, then Y_t causes X_t—the now-standard definition of Granger causality. Simply put, if prediction of X_t using only its own history improves after incorporating Y_t's history, indicating information flow from Y_t to X_t, then Y_t is said to have Granger causality on X_t.

Traditional Granger causality suffers from several prominent issues: (1) it requires strong prior knowledge to predefine VAR models; (2) it employs linear functions, making nonlinear relationship analysis difficult; (3) it considers only bivariate causality without accounting for potential confounding factors, hindering correct analysis of multivariate causality in complex systems; and (4) it statically analyzes causality across entire time series.

Researchers have proposed numerous improved models to address these limitations [22]. Conditional GC introduces conditional variables [23] to distinguish direct from indirect causality, while penalized models such as lasso [24] and group-lasso [25] address sparse regression problems, extending Granger causality to high-dimensional systems. Nonlinear GC improves upon linear functions using carefully selected kernel functions, such as radial basis functions (RBF) [26] and reproducing kernel Hilbert spaces (RKHS) [27], to handle causality in nonlinear systems. However, these methods still have shortcomings: (1) they require some prior knowledge to predefine model types, which is difficult for unknown systems whose underlying processes are typically unknown in advance; and (2) they struggle to design sufficiently complex functional models to fit the highly nonlinear mechanisms of complex dynamical systems.

Neural network-based Granger causality analysis remains in preliminary stages. Attanasio et al. [28] and Montalto et al. [29] applied feedforward neural networks to GC analysis, but their architectures remain relatively simple. Recent work by Alex Tank et al. [30] determines Granger causality through parameters in neural network hidden layers, requiring substantial penalties to drive potentially useless parameters toward zero, which leads to excessively long training times, high complexity, and limited practical significance. Moreover, existing research has not yet integrated neural networks with dynamic Granger causality analysis.

3. Problem Definition

Consider a complex dynamical system containing n variables $(X_1, X_2, \dots, X_n)$. The system's dynamic evolution is observed over a period containing T observation points $(1, 2, \dots, T)$, where the actual time intervals between consecutive observation points may vary and depend on the specific system and time series data. Consequently, each variable forms a time series across the T observation points, with $X_{i,t}$ denoting the value of the i-th variable (time series) at the t-th observation point. Hereinafter, the i-th variable can be understood as the i-th time series.

Definition 1 (System State and Accumulated Information). For any time t representing the t-th observation point, the system state at time t is $\mathcal{U}_t = (X_{1,t}, X_{2,t}, \dots, X_{n,t})^\top \in \mathbb{R}^{n\times k}$, where k is the model order. The multivariate time series constituting the complex dynamical system is defined as $\mathcal{U} = (\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_T)$. The information accumulated by the system before time t is defined as $\mathcal{U}_{t-k:t-1} = (\mathcal{U}_{t-k}, \mathcal{U}_{t-k+1}, \dots, \mathcal{U}_{t-1}) \in \mathbb{R}^{n\times k}$.

Definition 2 (Optimal Prediction). Define $P(X_{j,t}|\mathcal{U}_{t-k:t-1})$ as the optimal prediction of $X_{j,t}$ obtained by predicting using the information $\mathcal{U}_{t-k:t-1}$ accumulated before time t. Define $\mathcal{U}^{i}_{t-k:t-1}$ as the information from which the i-th variable's information has been removed, and $P(X_{j,t}|\mathcal{U}^{i}_{t-k:t-1})$ as the optimal prediction of $X_{j,t}$ obtained using the above information after removing X_i.

Definition 3 (Dynamic Granger Causality). Building upon the concept of Granger causality, we introduce existence time periods to represent that Granger causality exists within specific time intervals. Define the existence time period $\mathrm{ETP} = \{t_1, t_2, \dots, t_L\}$, where L denotes the length of the existence period. Adjacent time points $t_\tau, t_{\tau+1} \in \mathrm{ETP}$ satisfy $t_{\tau+1} - t_\tau \neq 1$, meaning time points within the ETP can be arbitrarily selected from the T observation points, thereby enabling dynamic causal analysis through dynamic time points in the time series. The dynamic Granger causality is defined as the pair (Granger causality, existence time period), where the ETP may be discontinuous. Within the ETP, X_j is judged to have Granger causality on X_i if the following condition holds:

$$\mathrm{var}(P(X_{j,t}|\mathcal{U}_{t-k:t-1})) < \mathrm{var}(P(X_{j,t}|\mathcal{U}^{i}_{t-k:t-1})), \quad \forall t \in \mathrm{ETP}$$

where X_i is the target variable and X_j is the source variable. Dynamic Granger causality achieves dynamic causal analysis by associating Granger causality with its existence time period.

The problem addressed in this paper is: given an ETP specified from all observation points, analyze the Granger causality status among multiple variables in the complex dynamical system during this period. The specific objective is to analyze whether variables X_i and X_j ($i, j \in \{1, \dots, n\}$) exhibit Granger causality within the ETP, using other variables as conditional variables, or whether both directions of causality exist.

Figure 2 [Figure 2: see original paper] illustrates the overall architecture of the proposed deep learning-based dynamic Granger causality analysis method, showing the general process of Granger causality analysis based on deep neural networks.

4. Deep Network Architecture for Granger Causality Analysis

Consider a complex system with n variables $\{X_i\}_{i=1}^{n}$. Within a given time period, let X_j be the target sequence, X_i the source sequence, and $\{X_m\}_{m\neq i,j}$ the conditional variables. The conventional approach for determining whether X_i has Granger causality on X_j builds two models:

$$\begin{aligned} \text{full:} \quad & \hat{X}_{j,t} = f_{\text{full}}(X_{1,t-k:t-1}, \dots, X_{n,t-k:t-1}) + \epsilon_t \\ \text{reduced:} \quad & \hat{X}_{j,t} = f_{\text{reduced}}(X_{1,t-k:t-1}, \dots, X_{i-1,t-k:t-1}, X_{i+1,t-k:t-1}, \dots, X_{n,t-k:t-1}) + \epsilon_t \end{aligned}$$

where $X_{i,t-k:t-1}$ represents the historical information of time series i. Granger causality is determined by comparing the residual variances of the full and reduced models.

This conventional approach requires building two models to test causality from X_i to X_j, and another two models for the reverse direction, totaling $2n^2$ models for multivariate problems—prohibitively complex for real-world systems. Therefore, this paper improves the predictive structure of Granger causality by employing joint modeling to analyze all potential causal effects of a single variable simultaneously, reducing the number of models to $n+1$.

The improved deep neural network architecture for Granger causality analysis is shown in Figure 2. At time t, the system state is $\mathcal{U}_t = (X_{1,t}, X_{2,t}, \dots, X_{n,t})^\top$, and the accumulated information before time t is $\mathcal{U}_{t-k:t-1}$ as defined in Equation (4). Rather than inputting individual variable histories $X_{i,t-k:t-1}$ as in Equations (2)-(3), the entire system state $\mathcal{U}_{t-k:t-1}$ is used as a lagged input, representing the whole system's past. Let $\mathcal{U}^i_{t-k:t-1}$ denote the information with variable X_i removed, used to assess whether source variable X_i has Granger causality on other variables. The output predicts the entire system state $\mathcal{U}_t$ at time t, not just $X_{j,t}$.

Specifically, we construct the following models:

$$\begin{aligned} \text{full:} \quad & \hat{\mathcal{U}}_t = f_{\text{full}}(\mathcal{U}_{t-k:t-1}) + \epsilon_t \\ \text{reduced}_i: \quad & \hat{\mathcal{U}}_t^{(i)} = f_{\text{reduced}i}(\mathcal{U}^i_{t-k:t-1}) + \epsilon_t \end{aligned}$$

In the causality inference module of Figure 2, we determine whether source variable X_i has Granger causality on target variable X_j by extracting X_j's optimal predictions from both models. Using a sliding window approach from time t_{ep} to $t_{ep} + L - 1$, we obtain corresponding predictions of X_j from both models. All predictions form prediction sequences $\hat{X}_j^{\text{full}}$ and $\hat{X}_j^{\text{reduced}_i}$, from which we compute their error variances against the true time series X_j:

$$\begin{aligned} \text{err}_{\text{full},j} &= \text{var}(\hat{X}_j^{\text{full}} - X_j) \\ \text{err}_{\text{reduced}i,j} &= \text{var}(\hat{X}_j^{\text{reduced}_i} - X_j) \end{aligned}$$

A single reduced$_i$ model can assess Granger causality from X_i to all n variables in the system. By varying the source variable removed in reduced$_i$ models, DNNGC requires only $n + 1$ models—one shared full model and n reduced models—to determine Granger causality between any pair of variables with directionality.

Based on Equations (5)-(6), we construct a deep neural network architecture suitable for Granger causality analysis, as shown in the deep neural network module of Figure 2. The architecture uses an LSTM-FC network, with recurrent neural network LSTM as the main model and Fully Connected (FC) layers for output dimension adjustment. This architecture is exceptionally well-suited for Granger causality problems: each training input is $\mathcal{U}_{t-k:t-1}$, where each time step receives a system state, perfectly accommodating the lag structure of Granger causality without disrupting per-timestep state information.

LSTM's complex parameter update mechanism enables it to establish sophisticated nonlinear dependencies with past information. Building upon RNN, LSTM introduces a second hidden state variable c_t, called the cell state, which contains information from past system states and updates parameters through gating mechanisms. The specific update formulas are:

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\
\tilde{c}_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
$$

where i_t, f_t, and o_t represent input, forget, and output gates, respectively, controlling how each component of the cell state updates and transfers to the hidden state h_t used for prediction. The forget gate f_t controls how much past cell state influences future states, while the input gate i_t controls how current observations affect the new cell state.

LSTM establishes relationships with long-term and complex dependencies in time series during parameter updates. For analyzing dependencies at time t, if the forget gate approaches 1, information from time $t-1$ flows entirely into time t, creating a dependency. If the forget gate remains 1, information from distant past moments continuously propagates to time t, establishing dependencies with far-past states. When skipping certain time points, parameters can be adjusted such that the computed c_t and h_t equal their previous values. In the Granger causality context, this flexible architecture can model complex nonlinear dependencies between time series.

5. Granger Causality Inference

In practice, when investigating whether time series i has Granger causality on time series j, a sufficiently complex deep neural network with adequate training time may overfit the data. Even after removing a source variable X_i that truly has Granger causality on target X_j, the reduced$_i$ model's predictions may not be substantially affected. This phenomenon is particularly evident in relatively simple datasets. Two primary approaches address overfitting: reducing parameter dimensions or decreasing the effective size of each dimension [31]. Since parameter space dimensions are fixed in our problem, we decrease effective parameter size through techniques like regularization and early stopping. Regularization introduces an additional hyperparameter λ controlling parameter penalty strength, but λ is difficult to interpret and tune. Since our objective is to compare which model (full or reduced$_i$) predicts better, and λ directly affects final predictions, determining the optimal value among infinite possibilities is challenging—different values may even yield opposite conclusions, making regularization unsuitable. Early stopping halts training when overfitting is detected on a validation set, but our problem context lacks a validation set, rendering it inappropriate as well.

To amplify the prediction performance gap between full and reduced$_i$ models, we approach from gradient descent dynamics: if time series j causally influences

i, the full model containing j's information should fit i more easily and achieve better predictions earlier in training compared to the reduced$_i$ model lacking j's information. Therefore, we consider prediction errors after each training epoch and determine Granger causality by comparing early-training performance between models.

Specifically, during training of each full and reduced$_i$ model, after each complete forward and backward pass through the entire dataset (one epoch), we compute the residual variance between predictions and true values. All epoch-wise residual variances form residual vectors:

$$\mathrm{Error}_{\mathrm{full}} = (\mathrm{err}^1_{\mathrm{full}}, \mathrm{err}^2_{\mathrm{full}}, \dots, \mathrm{err}^{N_{\mathrm{epoch}}}_{\mathrm{full}})$$
$$\mathrm{Error}_{\mathrm{reduced}i} = (\mathrm{err}^1_{\mathrm{reduced}i}, \mathrm{err}^2_{\mathrm{reduced}i}, \dots, \mathrm{err}^{N_{\mathrm{epoch}}}_{\mathrm{reduced}i})$$

where $i = 1, 2, \dots, n$ and N_{epoch} denotes the number of training epochs. The criteria for determining whether source variable X_i has Granger causality on target X_j are: (1) across all epochs, the full model's residual vector column j is smaller than the reduced$_i$ model's column j more frequently than it is larger; and (2) for each epoch, the mean of the first epoch residuals in the full model's column j is smaller than that of the reduced$_i$ model. These conditions are formalized as:

$$\mathrm{NUM}(\mathrm{Error}_{\mathrm{full}}[:, j] < \mathrm{Error}_{\mathrm{reduced}i}[:, j]) > \mathrm{NUM}(\mathrm{Error}_{\mathrm{full}}[:, j] > \mathrm{Error}_{\mathrm{reduced}i}[:, j])$$
$$\mathrm{MEAN}(\mathrm{Error}_{\mathrm{full}}[1 : \mathrm{epoch}, j]) < \mathrm{MEAN}(\mathrm{Error}_{\mathrm{reduced}i}[1 : \mathrm{epoch}, j]), \quad \forall \mathrm{epoch} \in \{1, \dots, N_{\mathrm{epoch}}\}$$

where NUM() counts satisfying elements, $[:, j]$ denotes column j, and MEAN() computes running means. Finally, statistical tests determine significance. If time series i has Granger causality on j, we assign $\mathrm{Causality}[j, i] = 1$; otherwise, 0. Repeating this analysis for all pairs yields the final causality matrix.

The DNNGC method pseudocode is presented in Algorithm 1.

6. Experimental Results and Analysis

We validate DNNGC's effectiveness and performance through simulation experiments and explore dynamic Granger causality in human physiological networks during deep and light sleep using real sleep data.

6.1 Simulation Experiments

Based on the complexity of Granger causality patterns, we consider three typical forms: chain, tree, and ring structures, and investigate how causality length affects performance. Overall results demonstrate that our method accurately discovers underlying Granger causality in systems (multivariate time series) exhibiting these three patterns.

Simulation data are generated using the Lorenz-96 model [32], a strongly nonlinear system described by the following differential equations:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad i = 1, 2, \dots, p$$

where $x_{t(-1)} = x_{t(p-1)}$, $x_{t0} = x_{tp}$, $x_{t(p+1)} = x_{t1}$, and F is a forcing parameter determining nonlinearity and chaos. A system with n time series has n^2 possible Granger causal relationships. A good causal discovery method should maximize true causal discoveries while controlling false positives [5]. We therefore use Accuracy as the evaluation metric:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Chain-Structured Granger Causality Chain structures are characterized by each time series causing at most one other series or being caused by at most one series. The experimental data structure is shown in Figure 3 Figure 3: see original paper, where blue lines indicate Granger causality. Using identical network parameters for full and reduced$_i$ models, we test chain lengths of 1, 2, 4, and 8 with varying hidden unit counts. Results in Figure 3(b) show that accuracies for lengths 1 and 2 reach 100% multiple times, with minimum accuracy exceeding 75%. Lengths 4 and 8 achieve average accuracies around 85% and 80%, respectively. Minimum accuracies occur at extreme hidden unit counts: too few units prevent adequate nonlinear modeling, while too many cause overfitting, reducing the error gap between full and reduced models and decreasing accuracy.

Tree-Structured Granger Causality Tree structures feature each time series causing at most one other series but potentially being caused by any number of series. Experiments follow the same settings as Section 6.1.1. Results in Figure 4 [Figure 4: see original paper] show 100% accuracy for width 2, with maximum accuracies of 95% and 92% for widths 4 and 8, respectively. Performance is sensitive to hidden unit counts, with large accuracy fluctuations. Minimum accuracies again occur at extreme hidden unit values.

Ring-Structured Granger Causality Ring structures are more complex, containing both chain and tree patterns, where each time series can cause or be caused by any other series, as shown in Figure 6 Figure 6: see original paper. Setting $p = 10$ and $F = 10$ in Lorenz-96 with sampling rate $\Delta t = 0.1$, we generate 10-variable multivariate time series with 1000 continuous time points (Figure 5 [Figure 5: see original paper]). The true causality pattern is shown in Figures 6(a) and 6(c). This complex dataset demonstrates our method's performance on highly nonlinear multivariate time series.

Experiments use a 2-layer deep neural network with 256 hidden units per layer, learning rate 0.01, lag $k = 10$, and $N_{\text{epoch}} = 200$. After 300 training epochs (near convergence), we apply the criteria from Section 5 using the first 300 epochs. The direct output is Causality $\in \mathbb{R}^{10\times 10}$, visualized in Figure 6(b) (red boxes indicate errors). Results achieve 90.0% accuracy, improving to 94% when requiring only the mean condition (Equation 12).

6.2 Sleep Physiological Analysis

Human physiological systems exhibit multivariate, complex nonlinear, and dynamic characteristics, representing a typical real-world complex dynamical system. Applying our method to discover causal relationships during sleep analyzes its performance on real systems with potential medical significance.

Physiologists have identified stable EOG-ECG interactions during light sleep that disappear during deep sleep [7]. We first validate this finding at the causality level, then analyze how this interaction varies across sleep stages, and finally discover causal patterns among physiological systems during deep and light sleep.

Data are from the ISRUC-SLEEP Dataset [33], comprising approximately 8-hour polysomnography (PSG) recordings from 118 subjects at CHUC Hospital's Sleep Medicine Center (2009-2013), sampled at 200 Hz. The dataset includes 6 EEG, 2 EOG, 1 ECG, 1 chin EMG, and 1 leg EMG channel (11 total), with approximate sensor locations shown in Figure 9 [Figure 9: see original paper] and data examples in Figure 1(a).

First, we analyze Granger causality between EOG and ECG during deep and light sleep separately as an evaluation benchmark. Using the DNNGC metric from Equation (9), values >0 indicate Granger causality from source to target, with larger values indicating stronger relationships. Experiments use all 118 subjects' sleep data with a 2-layer network (6000 hidden units per layer) and batch size 128. Results in Figure 7 [Figure 7: see original paper] show EOG $\rightarrow$ ECG causality exists in both sleep stages, but the error improvement is $10\times$ greater during light sleep, indicating substantially stronger and more stable causality.

For a single subject's whole-night signals, we analyze dynamic Granger causality in 30-second intervals. Figure 8 [Figure 8: see original paper] shows EOG $\rightarrow$ ECG causality exists in 55.1% of light sleep intervals versus only 28.4% during deep sleep. The results also reveal that during non-REM periods (0.5-2.5h and 4.5-7h), EOG $\rightarrow$ ECG causality persists, typically transitioning between existence and non-existence with deep/light sleep changes. During wake periods (0-0.5h and 2.5-4.5h), the causality changes frequently and dramatically, indicating a more chaotic state.

Finally, integrating data from 188 subjects and analyzing each 1.5-hour sleep cycle [34], we filter deep and light sleep stages to discover causal patterns among physiological systems. Figure 9 shows more numerous and complex causal interactions during light sleep compared to deep sleep.

7. Conclusion

This paper addresses limitations of existing Granger causality analysis methods by proposing a deep learning-based dynamic Granger causality analysis method and a tailored neural network architecture. By integrating Granger causality with deep learning, we reduce model space complexity and enable modeling of complex nonlinear relationships. Experiments on Lorenz-96 simulated data across three typical causality patterns demonstrate effectiveness and accuracy. Application to real human sleep physiological multivariate time series validates physiological findings and explores dynamic Granger causality among physiological systems throughout the night, revealing potential medical value.

Future work will extend this method to downstream tasks such as time series clustering [35-36] and anomaly detection [37], exploring specific application scenarios while further refining the approach.

References

[1] Gao Tengfei, Liu Yongyan, Tang Yunbo, et al. A massively parallel bayesian approach to factorization-based analysis of big time series data[J]. Journal of Computer Research and Development, 2019, 56(7): 1567-1577. (高腾飞, 刘勇琰, 汤云波, 等. 面向时间序列大数据海量并行贝叶斯因子化分析方法 [J]. 计算机研究与发展, 2019, 56(7): 1567-1577.)

[2] Stam C J, van Straaten E C. Go with the flow: use of a directed phase lag index (dPLI) to characterize patterns of phase relations in a large-scale model of brain dynamics[J]. Neuroimage, 2012, 62(3): 1415-1428.

[3] Sheikhattar A, Miran S, Liu Ji, et al. Extracting neuronal functional network dynamics via adaptive granger causality analysis[J]. Proceedings of the National Academy of Sciences, 2018, 115(17): E3869-E3878.

[4] Sameshima K, Baccala L A. Brain connectivity: an overview[M]//Methods in brain connectivity inference through multivariate time series analysis. USA: CRC Press, 2014: 1-9.

[5] Li Yan, Wang Ting, Liu Wanwei, et al. ICIC_{Target}: a novel discovery algorithm for local causality network of target variable[J]. Journal of Computer Research and Development, 2016, 53(7): 1544-1560. (李岩, 王挺, 刘万伟, 等. ICIC_{Target}: 目标节点的局部因果关系网络的发现算法 [J]. 计算机研究与发展, 2016, 53(7): 1544-1560.)

[6] Runge J, Nowack P, Kretschmer M, et al. Detecting and quantifying causal associations in large nonlinear time series datasets[J]. Science Advances, 2019,

5(11): eaau4996.

[7] Bashan A, Bartsch R, Kantelhardt J W, et al. Network physiology reveals relations between network topology and physiological function[J]. Nature Communications, 2012, 3, 702.

[8] Granger C W J. Investigating causal relations by econometric models and cross-spectral methods[J]. Econometrica, 1969, 37(3): 424-438.

[9] Chiou-Wei S Z, Chen C F, Zhu Zhen. Economic growth and energy consumption revisited—evidence from linear and nonlinear granger causality[J]. Energy Economics, 2008, 30(6): 3063–3076.

[10] Hong Yongmiao, Liu Yanhui, Wang Shouyang. Granger causality in risk and detection of extreme risk spillover between financial markets[J]. Journal of Econometrics, 2009, 150(2): 271–287.

[11] Seth A K, Barrett A B, Barnett L. Granger causality analysis in neuroscience and neuroimaging[J]. Journal of Neuroscience, 2015, 35(8): 3293-3297.

[12] Friston K J. Functional and effective connectivity in neuroimaging: A synthesis[J]. Human Brain Mapping, 1996, 2(1-2): 56-78.

[13] Dhamala M, Rangarajan G, Ding M. Analyzing information flow in brain networks with nonparametric granger causality[J]. Neuroimage, 2008, 41(2): 354-362.

[14] Chén O Y, Cao Hengyi, Reinen J M, et al. Resting-state brain information flow predicts cognitive flexibility in humans[J]. Scientific Report, 2019, 9, 3879.

[15] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9 (8): 1735-1780.

[16] Graves A. Supervised sequence labelling[M]//Supervised sequence labelling with recurrent neural networks. Berlin: Springer, 2012: 5-13.

[17] Xu Jinghang, Zuo Wanli, Liang Shining, Wang Ying. Causal relation extraction based on graph attention networks[J]. Journal of Computer Research and Development, 2020, 57(1): 159-174. (许晶航, 左万利, 梁世宁, 王英. 基于图注意力网络的因果关系抽取 [J]. 计算机研究与发展, 2020, 57(1): 159-174.)

[18] Sun Chenxi, Hong Shenda, Wang Jingyu, et al. A systematic review of deep learning methods for modeling electrocardiograms during sleep[J]. Physiological Measurement, 2022, 43(8): 08TR02.

[19] Spadon G, Hong Shenda, Brandoli B, et al. Pay attention to evolution: time series forecasting with deep graph-evolution Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(9): 5368-5384.

[20] Yang Ling, Hong Shenda. Unsupervised time-series representation learning with iterative bilinear temporal-spectral fusion[C]//Proceedings of the 39th International Conference on Machine Learning. New York: ACM, 2022: 25038-25054.

[21] Wiener N. The theory of prediction[M]//Modern Mathematics for the Engineer: First Series. USA: Dover Publications, 1956: 165-190.

[22] Ren Weijie, Han Min. Survey on causality analysis of multivariate time series[J]. Acta Automatica Sinica, 2021, 47(1): 64-78. (任伟杰, 韩敏. 多元时间序列因果关系分析研究综述 [J]. 自动化学报, 2021, 47(1): 64-78.)

[23] Geweke J F. Measures of conditional linear dependence and feedback between time series[J]. Journal of the American Statistical Association, 1984, 79(388): 907-915.

[24] Arnold A, Liu Yan, Abe N. Temporal causal modeling with graphical granger methods[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM, 2007: 66-75.

[25] Yang Guanxue, Wang Lin, Wang Xiaofan. Reconstruction of complex directional networks with group lasso nonlinear conditional granger causality[J]. Scientific Reports, 2017, 7(1): 2991.

[26] Ancona N, Marinazzo D, Stramaglia S. Radial basis function approach to nonlinear granger causality of time series[J]. Physical Review E, 2004, 70(5): 056221.

[27] Marinazzo D, Pellicoro M, Stramaglia S. Kernel-granger causality and the analysis of dynamical networks[J]. Physical Review E, 2008, 77(5): 056215.

[28] Attanasio A, Triacca U. Detecting human influence on climate using neural networks based granger causality[J]. Theoretical and Applied Climatology, 2011, 103(1-2): 103-107.

[29] Montalto A, Stramaglia S, Faes L, et al. Neural networks with non-uniform embedding and explicit validation phase to assess granger causality[J]. Neural Networks, 2015, 71:159-171.

[30] Tank A, Covert I, Foti N, et al. Neural granger causality[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(8): 4267-4279.

[31] Lutz P. Early stopping - but when?[M]//Neural Networks: Tricks of the Trade. Berlin: Springer, 2012: 53-67.

[32] Lorenz E N, Emanuel K A. Optimal Sites for supplementary weather observations simulation with a small model [J]. Journal of American Meteorological Society, 1998, 55(3): 399-414.

[33] Sirvan K, Sousa T, Santos J M, et al. ISRUC-Sleep: a comprehensive public dataset for sleep researchers[J]. Computer Methods and Programs in Biomedicine, 2016, 124: 180-192.

[34] Carskadon M A, Dement W C. Normal human sleep: an overview[M]//Principles and Practice of Sleep Medicine. St. Louis: Elsevier Saunders, 2011: 16-26.

[35] Du Guowang, Zhou Lihua, Yang yudi, et al. Deep multiple auto-encoder-based multi-view clustering[J]. Data Science and Engineering, 2021, 6:323-338.

[36] Xu Chenxiao, Hao Huang, Yoo S. A deep neural network for multivariate time series clustering with result interpretation[C]//Proceedings of 2021 International Joint Conference on Neural Networks, Piscataway: IEEE, 2021: 1-8.

[37] Huang Xunhua, Zhang Fengbin, Fan Haoyi, Xi Liang. Multimodal adversarial learning based unsupervised time series anomaly detection[J]. Journal of Computer Research and Development, 2021, 58(8): 1655-1667. (黄训华, 张凤斌, 樊好义, 等. 基于多模态对抗学习的无监督时间序列异常检测 [J]. 计算机研究与发展, 2021, 58(8): 1655-1667.)

Corresponding Author: Li Hongyan, E-mail: leehy@pku.edu.cn

Author Contributions:

Liu Bo: Proposed algorithmic ideas, designed and conducted experiments, wrote the manuscript;
Li Hongyan: Provided overall guidance, supervised and revised the manuscript;
Hong Shenda: Proposed research topics, supervised experimental design and data analysis, supervised and revised the manuscript.

Note: Figure translations are in progress. See original paper for figures.