Wanqi Zhou, Shuanghao Bai, Shujian Yu, Qibin Zhao, Badong Chen*

Key Contributions

Construct a Time Series Forecasting Neural Network *Figure 1.* Motivation and our proposed neural Granger causality method. To learn the true Granger causality, we need to estimate the **i** Construct a Time Series Forecasting Neural Network $\begin{array}{ccc} & & & & \end{array}$ al Notwork As illustrated in (Hoffman et al., 2019), we can rewrite the Ed. 4 and 2 me Series Forecasting Neural Network input variables *x*. Tr(*·*) represents the trace function. Ultiincontraction. Trace function α represents the trace function α

Residual MLP-based Model relationship: *X*⁴ *X*² *X*¹ ! *X*³ ! *X*4. To comprehend this relationship, current neural Granger causality methods need to construct and the same number of models as the dimensions of the importance of the importance of each variable and $\frac{1}{2}$ **Regularizer-based Model**
Acobian Regularization or a constant or a dimensional output space. *z* is the output with respect to **The input-output Jacobian matrix effects** over the increases to compute \mathbf{r} and \mathbf{r} and increases to computational overhead that increases to computational over the increases of \mathbf{r}

For the first issue, to develop a unified neural Granger causality framework with the multiple of parameters to be et al., 2023; Zeng et al., 2023; Z
Seng et al., 2023; Zeng et al., 20 Frobienus norm $\begin{array}{|c|} \hline \end{array}$ Let W be the set of parameters to be optimized for the neural network f proposed methods, i.e., JRNGC-L1, JRNGC-L1, JRNGC-F on five-and-control of the five-and-control of the five-andysis method. Before that, it is considered to emphasize that, it is considered to expect the considered to emphasize that we can c t of narameters to he ontimized for the neural case of the university to the series permitted for the model of reducing computational costs by orders of magnitude. Above all, our method can be named JRNGC-L1, JRNGC-F ed for the set of parameters to be optimized for the neural network f 1 $\overline{}$ *D* 1 $\overline{}$ *N* Above all, our method can be named JRNGC-L1, JRNGC-F Let W be the set of parameters to be optimized for the neural network f

function of JRNGC-L1 can be formulated as follows: the formulated as follows: the formulated as follows: the f
The following as follows: the following as follows: the following as follows: the following as follows: the fol

The advantage of adding residual layers was marginal for advantage of adding residual for a ding residual for
The advantage of adding residual for a ding residual for a ding residual for a ding residual for a ding residu

the convenience of illustration.

Table 5. Comparative performance on CausalTime benchmark datasets. We highlight the best and the second best in bold and with @*x^j* $\overline{}$

underlining, respectively. variable to compute the performance on eausal rime continuation datasets. We inginight the cost and all
derlining, respectively.

dimensional output space. *z* is the output with respect to

I network (NN) with shared hidden layer? causality constraint processes

Jacobian Regularizer cobian matrix allows us to find the relationships between

Preliminaries Given a *D*-dimensional multivariate time series x =

Definition We regularize the L_1 norm or squared *Frobienus norm* of the input-output Jacobian matrix: dependence of the innut-output lacobian matri \mathbb{R} is a following is a following in \mathbb{R} 1. With the past information of each time series x*^t*⌧:*t*¹ as input, we forecast their future x*^t*

- To our best knowledge, this is the first work to harness a **Single NN model** with shared hidden layers for multivariate Granger causality analysis. It is necessary to incorporate the number of number of number of \mathcal{L} γ \cdot γ io our best knowledge, this is the firs is accomplished by applying *L*¹ or *L*² norm regularizer to Weights of the first layer causality analysis.
- We propose a novel neural network framework to learn **Granger causality by incorporating an input-output Jacobian regularizer** in the training objective. case of universe of universe prediction, where the weight $\frac{1}{2}$ quare be expressed as a direct relationship from \mathbf{r}
- Our method can not only obtain the summary Granger causality but also the full-time Granger causality. \overline{C} coucolity but alco the full time Cropa may be need the neural control of the set of
- Extensive experiments show our method can outperform state-of-the-art baselines. ing full-time Granger causality. This full-time Granger causality. This is one wishes to consider the construction of α 3. Method

$$
\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1^{t-\tau}} \cdots \frac{\partial \mathbf{f}}{\partial x_p^{t-1}} \end{bmatrix}
$$
\n
$$
= \begin{bmatrix} \frac{\partial f_1}{\partial x_1^{t-\tau}} \cdots \frac{\partial f_1}{\partial x_1^{t-1}} \cdots \frac{\partial f_1}{\partial x_p^{t-\tau}} \cdots \frac{\partial f_1}{\partial x_p^{t-1}} \\ \vdots \\ \frac{\partial f_D}{\partial x_1^{t-\tau}} \cdots \frac{\partial f_D}{\partial x_1^{t-1}} \cdots \frac{\partial f_D}{\partial x_p^{t-\tau}} \cdots \frac{\partial f_D}{\partial x_p^{t-1}} \end{bmatrix},
$$
\n
$$
\|\mathbf{J}\|_1 = \sum_{i,j} \left| \frac{\partial f_i}{\partial x_j} \right|
$$
\n
$$
\|\mathbf{J}\|_1 = \sum_{i,j} \left| \frac{\partial f_i}{\partial x_j} \right|
$$
\n
$$
\|\mathbf{J}(x)\|_{\mathbf{F}}^2 = \text{Tr}(JJ^{\mathsf{T}}) = \sum_{i \in \mathbb{N}} eJJ^{\mathsf{T}}e^{\mathsf{T}}
$$
\n
$$
= \sum_{i \in \mathbb{N}} \left[\frac{\partial (e \cdot z)}{\partial x} \right]^2,
$$
\n
$$
\text{Frobienus norm}
$$
\nLet W be the

in multivariate multi-model settings, it becomes necessary

represents the specific time and resource requirements the specific time and resource requirements of the specific time and resource requirements of the specific time and resource requirements of the specific time and reso

Incorporate Input-output Jacobian Matrix Regularizer During Training predicting variables and learning causal connections is muapproach and their impact on learning outcomes. The internal control of the internal control of the internal c
This helps hel **Incorporate In** ate time series for the series forecasting model. In this work, we have \mathbb{R}^n **Figure 2.1 The Figure 2018 Incorporate Input-output Jacobian Matrix Regularizer Durin** Frobienus norm of the input-output Jacobian matrix effiet al., 2021, Navarra (L $\overline{0}$), Navarra (Bussmann et al., 2021), Navarra (Bussmann et $\textbf{S}\textbf{S}$ urdi izer During Training $\textbf{S} = \textbf{S}$ Number of \textbf{S} It Jacobian *M*atrix Regularizer *,* (5) **neguiai izer** and tuditions in **E-output Jacobian Matrix Regularize**

*{x*1*, x*2*,xD}*, as shown in Figure 1, our model can be

 η_{proj} random projection dom projection. Utilizing a mini-batch size of *|B|* = 100, n_{proj} random projection S^{D-1} *unit spnere* n_{proj}

$$
\min_{\mathbf{W}} \quad \sum_{t=\tau}^{T} (x_t - f(\mathbf{x}_{
$$

network, proven more capability time series (Dashburg time series (Dashburg time series (Dashburg time series

sparse causal relationships, AUPRC becomes a more reliationships, AUPRC becomes a more reliable becomes a more
AUPRC becomes a more reliable becomes a more reliable becomes a more reliable becomes a more reliable becomes

(ˆ*xi,t xi,t*)

² ⁺ k*J*(*x*)k1*.* (6)

We employ two standard metrics: the area under the receivers α

in the appendix.

 $4.1.1.1₁$

4.1. Metrics

I Extensive experiments show that our JRNGC method achieves SOTA. More details can be found in https://arxiv.org/pdf/2405.08779 Jan Regularizer-based Neural Grande