Supplementary Material: Detailed Methodology and Experimental Insights on Classifier Approach

Zhaozhao Ma[*,†]
[*] *Zhejiang University*
[†] *Georgia Institute of Technology*
zhaozhaoma@gatech.edu

Shujian Yu[‡,§]
[‡] *Vrije Universiteit Amsterdam*
[§] *UiT - The Arctic University of Norway*
s.yu3@vu.nl

Abstract

In our paper *Cauchy-Schwarz Divergence Transfer Entropy*, we discussed the classifier-based approach using Cauchy-Schwarz Divergence Transfer Entropy (CS-TE) and tested the method with synthetically generated nonlinear data. In this supplementary document, we will provide a more detailed explanation of the nonlinear data generation strategy, as well as a description of the generation, testing, and performance of linear data.

1 Nonlinear data synthesis

1.1 Nonlinear training data

In our paper, we used a nonlinear vector autoregressive (NVAR) model to synthesize 7,500 time series pairs of length 256, each labeled with a causal label: $X \to Y$, $X \leftarrow Y$, or *No Causation*.

1.2 Nonlinear test data

We use the training data to train a random forest classifier. Furthermore, we employed different data generation functions to generate 300 pairs data with the same length of 256 to test the classifier:

- For the causal direction is $x \to y$:

$$x_t = 0.5x_{t-1} + 0.9N_x \tag{1}$$

$$y_t = 1.5\exp\left(-(x_{t-1} + x_{t-2})\right) + 0.7\cos\left(y_{t-1}^2\right) + 0.2N_y \tag{2}$$

- For the causal direction is $x \leftarrow y$:

$$y_t = 1.2y_{t-1} + 0.3N_y \tag{3}$$

$$x_t = -1.5\exp(-(y_{t-1} + y_{t-2})^2) + 0.7\cos(x_{t-1}^2) + 0.2N_x \tag{4}$$

- For *No Causation*:

$$x_t = 0.5x_{t-1} + 0.9N_x \tag{5}$$

$$y_t = 1.5\cos\left(y_{t-1}^2\right) + 2.5N_y \tag{6}$$

1.3 Nonlinear test result

The classifier achieved an accuracy of 0.96 on the test dataset (please refer to our paper).

2 Linear data synthesis

2.1 Linear training data

In our experiments, we also tested the linear causal discrimination ability of the classifier combined with CS-TE. We similarly used the VAR model to synthesize 7,500 time series pairs of length 256 with causal labels. For example, Eq. (7) was used to generate data for the relationship $X \to Y$:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \frac{1}{p} \sum_{\tau=1}^{p} \begin{bmatrix} a_\tau & 0 \\ c_\tau & d_\tau \end{bmatrix} \begin{bmatrix} x_{t-\tau} \\ y_{t-\tau} \end{bmatrix} + \begin{bmatrix} N_x \\ N_y \end{bmatrix} \tag{7}$$

2.2 Linear test data

Similarly, we used the training data to train the classifier and attempted to modify the data generation strategy to test its performance. We generated 300 pairs of series and evenly distributed causal labels based on Eq.(7) by adjusting the noise variance $p \in \{0.5, 1.0, 1.5, 2.0\}$, lag order $\in \{1, 2, 3\}$.

2.3 Linear test result

The classifier achieved an accuracy of 0.90 on the linear test dataset.