2010 International Symposium on Nonlinear Theory and its Applications
NOLTA2010, Krakow, Poland, September 5-8, 2010

Entropy-Based Measures of Causality and Application to Epilepsy

Dimitris Kugiumtzis

Department of Mathematical, Physical and Computational Sciences,
Faculty of Engineering, Aristotle University of Thessaloniki, Greece
URL: http://users.auth.gr/dkugiu, E-mail: dkugiu@gen.auth.gr

Abstract—Among different measures of coupling and causality, measures based on entropies have gained much attention recently, such as the transfer entropy (TE) and its extensions based on permutation entropies, i.e. the so-called symbolic transfer entropy (STE) and the transfer entropy on rank vectors (TERV). All these measures make use of univariate embedding of each of the two time series. Very recently, we proposed a measure for coupling and causality that is derived from mixed embedding, which relies on information criteria regarding past, current and future states. The components of the mixed embedding vector indicate the presence of information transfer, and a measure is formed to quantify it, called mutual information from mixed embedding (MIME). We compare the measures from univariate embedding, TE, STE and TERV, and the measure from multivariate embedding, MIME. For this, we make simulations on a number of known nonlinear dynamical systems. Further, we apply the four measures to EEG records containing preictal and ictal states. It turns out that MIME is rather robust and conservative in detecting causal effects while the other three measures are positively biased indicating often false causal effects.

keywords: multivariate time series, Granger causality, information flow, mixed embedding, epileptic EEG

1. Introduction

In the study of complex systems, such as climatic processes, financial markets and brain dynamics, it is important to identify and estimate the strength and direction of inter-dependence among the interacting components, measured as multivariate time series. Among different measures of uni- or bi-directed dependence (phase synchronization, Granger causality using prediction models or coherence measures, and local geometric properties in reconstructed state spaces for the driving and driven systems, e.g. see [1, 2]), we concentrate on the class of information-based causality measures [3, 4, 5]. The most popular of this type of causality measures is the transfer entropy (TE) [3]. We also consider the variant of TE using rank vectors instead of sample vectors [4] and the correction of this [6]. The main objective of this work is to compare the transfer entropy measures with a measure we proposed very recently, derived from mutual information conditioned on the components of the driving system present in a mixed embedding [7].

We conduct a simulation study to assess the ability of the measures to detect correctly the direction and strength of coupling, using some known chaotic systems. Then we assess the measures on a real-world application, the investigation of the information flow in brain areas before and after epileptic seizures.

2. Information-based Causality Measures

Let $\{x_t\}$ and $\{y_t\}$, $t = 1, \ldots, n$, denote two simultaneously observed time series derived from the dynamical systems X and Y, respectively. Using the method of delays, the reconstructed points from the two time series are $\mathbf{x}_t = [x_t, x_{t-\tau_x}, \ldots, x_{t-(m_x-1)\tau_x}]$ and $\mathbf{y}_t = [y_t, y_{t-\tau_y}, \ldots, y_{t-(m_y-1)\tau_y}]$, allowing different delay parameters τ_x, τ_y and embedding dimensions m_x, m_y for the systems X and Y, respectively.

Transfer entropy (TE) quantifies the information flow from X to Y by the amount of information explained in Y at T steps ahead by the state of X, accounting for the concurrent state of Y [3]. In terms of the Shannon entropy $H(x) = \sum p(x) \log p(x)$, TE for the causal effect of system X on system Y is defined as

$$\mathrm{TE}_{X\to Y} = H(\mathbf{x}_t, \mathbf{y}_t) - H(\mathbf{y}_t^T, \mathbf{x}_t, \mathbf{y}_t) + H(\mathbf{y}_t^T, \mathbf{y}_t) - H(\mathbf{y}_t). \quad (1)$$

For the estimation of the entropy terms, we use the approach of nearest neighbors [8]. To account for the effect of X on the evolution of Y over a time horizon T, we extend the single one step ahead mapping to the future vector $\mathbf{y}_t^T = [y_{t+1}, \ldots, y_{t+T}]'$ in the definition of TE in eq.(1).

In [4], the symbolic transfer entropy (STE) is defined as TE but on rank-points formed by the reconstructed vectors of X and Y. Each sample reconstructed vector, say $\mathbf{w}_t$, in eq.(1) is replaced by the rank-point $\hat{\mathbf{w}}_t = [r_1, r_2, \ldots, r_m]$, where $r_j \in \{1, 2, \ldots, m\}$ are the ranks of the vector components $j = 1, \ldots, m$. Following this sample-point to rank-point conversion, $\mathbf{y}_t^T$ is replaced by the rank point at time $t + T$, $\hat{\mathbf{y}}_{t+T}$, and STE is defined as

$$\mathrm{STE}_{X\to Y} = H(\hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) - H(\hat{\mathbf{y}}_{t+T}, \hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) + H(\hat{\mathbf{y}}_{t+T}, \hat{\mathbf{y}}_t) - H(\hat{\mathbf{y}}_t),$$

where the entropies are computed from the estimated probability mass functions of the rank-points.

In [6], it was shown that instead of replacing $\mathbf{y}_t^T$ with $\hat{\mathbf{y}}_{t+T}$, as done in STE, it is more appropriate to use $\hat{\mathbf{y}}_t^T$ =

$[\hat{y}_{t+1}, \ldots, \hat{y}_{t+T}]$, the ranks of $\mathbf{y}_t^T = [y_{t+1}, \ldots, y_{t+T}]$ in the augmented vector $[\mathbf{y}_t^T, \mathbf{y}_t]$. The proposed measure of transfer entropy on rank vectors (TERV) is

$$\mathrm{TERV}_{X\to Y} = H(\hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) - H(\hat{\mathbf{y}}_t^T, \hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) + H(\hat{\mathbf{y}}_t^T, \hat{\mathbf{y}}_t) - H(\hat{\mathbf{y}}_t).$$

In the following, we present briefly the causality measure of mutual information from mixed embedding (MIME), which relies on a non-uniform embedding scheme for the bivariate time series $\{x_t, y_t\}$, $t = 1, \ldots, n$ with the purpose to explain best the future of Y, as given by the future vector $\mathbf{y}_t^T$ [7]. The rationale of the proposed embedding scheme is that the components of the derived embedding vector, denoted $\mathbf{z}_t$, must be least dependent to each other and able to explain best $\mathbf{y}_t^T$.

The components of the embedding vector $\mathbf{z}_t$ are to be selected from the set of delayed components $\mathbf{Z}_t = \{x_t, x_{t-1}, \ldots, x_{t-L_x}, y_t, \ldots, y_{t-L_y}\}$, where L_x, L_y are the maximum lags for X and Y. The progressive embedding scheme starts with an empty embedding vector $\mathbf{z}_t^0$. Then at a step j, the component $z_t^j \in \mathbf{Z}_t \setminus \mathbf{z}_t^{j-1}$ to be added to $\mathbf{z}_t^{j-1}$ is the one that maximizes the (mutual) information to $\mathbf{y}_t^T$ accounting (conditioning) for the current components in $\mathbf{z}_t^{j-1}$. The criterion for the selection of z_t^j reads

$$\max_{z_t^j} \left\{ I\left(\mathbf{y}_t^T; z_t^j \mid \mathbf{z}_t^{j-1}\right)\right\}.$$

The progressive vector building stops at step j and then $\mathbf{z}_t = \mathbf{z}_t^{j-1}$, if the addition of the new component z_t^j does not improve significantly the mutual information of the future vector and the embedding vector. Thus the stopping criterion is

$$I\left(\mathbf{z}_F; \mathbf{z}_n^{j-1}\right) / I\left(\mathbf{z}_F; \mathbf{z}_n^{j}\right) > A,$$

for a threshold $A \leq 1$. The closer A is to 1 the more relaxed is the stopping criterion allowing for more components to enter in the form of $\mathbf{z}_t$. On the other hand, a smaller A results in lower dimensions of the mixed embedding. Here, we use $A = 0.95$, which was found in [7] to be a good trade-off value.

The embedding vector from the mixed embedding scheme may contain components from both X and Y, and can be represented in terms of these two sets of components as

$$\mathbf{z}_t = [\mathbf{z}_t^x, \mathbf{z}_t^y] = [x_{t-l_{x1}}, x_{t-l_{x2}}, \ldots, x_{t-l_{xm_x}}, y_{t-l_{y1}}, \ldots, y_{t-l_{ym_y}}].$$

Then the measure MIME is defined as

$$\mathrm{MIME}_{X\to Y} = 1 - \frac{I(\mathbf{y}_t^T; \mathbf{z}_t^y)}{I(\mathbf{y}_t^T; \mathbf{z}_t)} = \frac{I(\mathbf{y}_t^T; \mathbf{z}_t^x \mid \mathbf{z}_t^y)}{I(\mathbf{y}_t^T; \mathbf{z}_t)}.$$

$\mathrm{MIME}_{X\to Y}$ measures the information of Y explained only by components of X in the embedding vector, normalized by the total mutual information (in order to give a value between 0 and 1). If $\mathbf{z}_t$ contains no components from X, then $\mathrm{MIME}_{X\to Y} = 0$ and X has no effect on the future of Y.

In the computation of MIME, first the optimal representation of the driving system X and response system Y in the mixed embedding is found, giving the m_x and m_y components in $\mathbf{z}_t$, respectively. This is a main difference of MIME from the other information-based causality measures, for which m_x and m_y are decided a priori as the embedding dimensions of separate fixed delay embeddings for X and Y, commonly setting $m_x = m_y$ [3, 4, 5]. Indeed we have found that the outcome of TE, STE and TERV depends strongly on the choice of m_x and m_y [2, 6], and MIME overcomes this problem using the progressive embedding scheme at the cost of significantly larger computation time.

3. Simulations on Chaotic Coupled Systems

In the simulations below we always set $\tau_x = \tau_y = 1$ and use 10 neighboring points for the estimation of entropies in TE and MIME.

We first evaluate the measures on the driver-response Henon system given by

$$\begin{aligned} x_{n+1} &= 1.4 - x_n^2 + 0.3x_{n-1} \\ y_{n+1} &= 1.4 - \left(Cy_nx_n + (1-C)y_n^2\right) + 0.3y_{n-1}, \end{aligned}$$

for coupling strength C taken the values $0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$. For each C we generate 100 realizations and compute the measures for $T = 1$. For MIME we set $L_x = L_y = 5$. The results on TE, STE and TERV vary significantly with the choice of m_x and m_y, and we even observe larger values in the wrong coupling direction when m_x is much smaller than m_y. We further show results on TE, STE and TERV only for $m_x = m_y$, which is the choice typically met in the works on these measures. In Fig. 1, we show results for $m_x = m_y = 3$ and for noise-free and noisy time series of small and large length n. We note that regardless of n and the presence of noise all the measures detect well the causal effect $X \to Y$ obtaining positive values, but STE and TERV are also positive for $C = 0$ and in the wrong direction of coupling, whereas TE gets positive only for noisy data and large C, in particular when $n = 4096$ (see Fig. 1d). On the other hand, $\mathrm{MIME}_{Y\to X}$ is always zero and actually this holds for all 100 realizations, as shown in Fig. 1b for the standard deviation (SD) of the measures. MIME has also the smallest SD for the direction $X \to Y$ giving a rather stable and consistent estimation of the causal effect. It should be noted that for very weak coupling ($C = 0.05$), MIME does not detect the driving of X as no components enter the form of the constructed embedding vector, while TERV in particular is larger in the correct direction. As argued in [7], MIME can be more sensitive to weak coupling by increasing the threshold A, but then there is higher probability of having irrelevant components in the form of the embedding vector by chance, which then generates positive $\mathrm{MIME}_{Y\to X}$ as well. Thus using $A = 0.95$, MIME turns out to be a more strict but stable measure of causality.

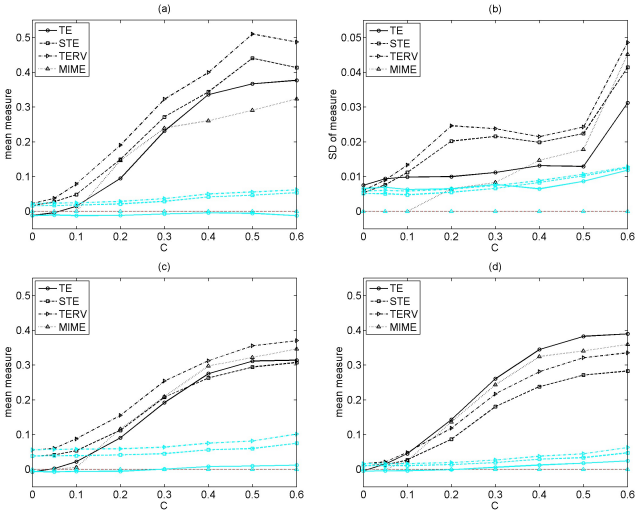

Figure 1: (a) The mean of the measures, as shown in the legend, from 100 realizations of the unidirectionally coupled Henon map for the correct direction $X \rightarrow Y$ (black line) and the opposite direction (gray line, cyan online). The time series length is $n = 1024$ and the data are noise-free. (b) The standard deviation (SD) of the measures in (a). (c) As in (a) but when adding normal white noise to both time series with SD being 20% of the data SD. (d) As in (c) but for $n = 4096$.

The same results on MIME are found from the simulations on the next system, the coupled Rössler–Lorenz system given by

$$\begin{aligned}
\dot{x}_1(t) &= 6(-y_1(t) - z(t)_1) \\
\dot{y}_1(t) &= 6(x_1(t) + 0.2y_1(t)) \\
\dot{z}_1(t) &= 6(0.2 + x_1(t)z_1(t) - 5.7z_1(t)) \\
\dot{x}_2(t) &= 10(y_2(t) - x_2(t)) \\
\dot{y}_2(t) &= 28x_2(t) - y_2(t) - x_2(t)z_2(t) + Cy_1(t)^2 \\
\dot{z}_2(t) &= -8/3z_2(t) + x_2(t)y_2(t),
\end{aligned}$$

where the driving time series regards y_1 and the response y_2, for C being $0, 0.5, 1, 1.5, 2, 3, 4$. We compute the measures for $T = 3$ to account for driving effects over several steps ahead. This complicates the computation of TE, STE and TERV because the entropy terms take as arguments larger vectors, introducing more bias in the estimation of the entropies and subsequently the estimation of the measures. The bias is particularly large for small time series and increases with the addition of noise as shown in Fig. 2 for $m_x = m_y = 3$ and $n = 1024$. It also increases with the embedding dimensions, particularly for STE and TERV. In the computation of MIME we let $L_x = L_y = 15$ to account for all significant delays. Again $\text{MIME}_{Y\rightarrow X}$ lies at the zero level and increases slightly only for noisy data and large C (see Fig. 2b). This zero level of $\text{MIME}_{Y\rightarrow X}$ combined with the significantly positive $\text{MIME}_{X\rightarrow Y}$ for all $C > 0$ suggests a reliable detection of the direction of the causal effect and estimation of its strength. The other measures fail to provide reliable estimation due to the presence of biased positive measure values for the direction $Y \rightarrow X$. In particular, STE tends to give larger values for the wrong direction and this is corrected by TERV. The variance of TE, STE and TERV is again at the same (significant) level in both directions, blurring the observed difference of the mean values of TE and TERV in the two directions in Fig. 2, whereas $\text{MIME}_{Y\rightarrow X}$ is again zero for all 100 realizations except for the noisy data and large C, where it has also large variance.

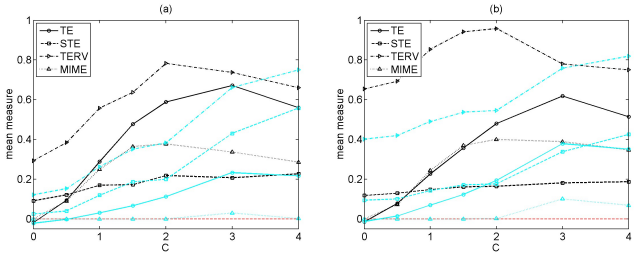

Figure 2: (a) Results as in Fig. 1a but for $T = 3$ and for the unidirectionally coupled Rössler–Lorenz system. (b) As in (a) but for 20% additive normal white noise.

4. Application of the Causality Measures to epileptic EEG

The application regards human scalp electroencephalographic (EEG) recordings from several hours before epileptic seizure onset to many minutes after the seizure end. We consider anti-hemispheric channels in pairs from the left and right frontal (F3 and F4), central (C3 and C4), temporal (T7 and T8) and parietal lobe (P3 and P4). We used 6 records of generalized tonic-clonic seizures of different patients. Each EEG record was split in segments of 30 sec (sampling time 0.01 sec) and the measures TE, STE, TERV and MIME were computed on each EEG segment of channel pairs ($T = 1$, $L_x = L_y = 20$). For TE, STE, TERV, $m_x = m_y$ were set to 3, 6 and 10, giving varying results: the STE and TERV profiles over the whole recording increased a lot with the embedding dimension, while the TE profile was always at the same level but varied in shape (see Fig. 3a and b for the TE profiles from one EEG record). The STE and TERV measures produced the same profile with small changes for $m_x = m_y = 3$ and being almost identical for larger embedding dimensions. Therefore only the TERV profile is shown for the same episode in Fig. 3c. The profiles of TE, STE and TERV are positive for all segments and channel pairs, and for both directions, so that one would conclude that information flows constantly from left to right and vice versa at all brain areas and regardless of the epileptic state (preictal, ictal, postictal). Given the presence of positive bias in all these measures, as observed in the simulations, this conclusion may not be correct and then one should look at differences in the level of these measures in the two directions in order

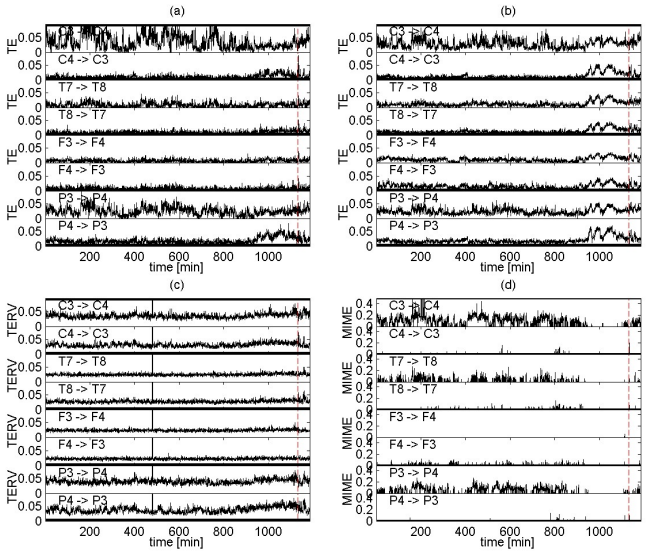

Figure 3: (a) The profile of TE for $m_x = m_y = 3$ for one epileptic EEG record for the 4 channel pairs and the two directions for each pair, as denoted at each panel. (b) Same as (a) but for $m_x = m_y = 10$. (c) Same as (a) but for TERV. (d) Same as (a) but for MIME. The vertical dashed line denotes the seizure onset.

to find specific patterns of information flows. For the specific profiles in Fig. 3, it seems that there is larger causal effect (information flow) from left to right central, parietal and temporal lobes than in the opposite direction. This is exactly what we observe clearly with the MIME measure as MIME is positive in the left to right direction at these lobes and zero in the opposite direction in almost all segments. The same characteristics were observed in the other 5 epileptic records. For the last preictal period in Fig. 3, MIME stays at zero for all directions and channel pairs, whereas the TE measure gets large as the embedding dimension increases in both directions and all channel pairs (and STE and TERV become even larger, not shown here), suggesting that there are other effects than driving, causing this large increase of TE, STE and TERV but not affecting MIME. This pattern at the late preictal period was absent in the other 5 epileptic records. No significant differences were observed after the seizure onset with TE, STE and TERV, whereas MIME often turned to zero.

5. Discussion

It has been already pointed in the literature that the causality measures, including the measures TE, STE and TERV, contain positive bias that can be attributed to effects other than the driving effect, mainly the dynamics of the individual systems, the state space reconstruction, the time series length and noise. We observed the positive bias of the three measures both in the simulations with the coupled Henon maps and the Rössler–Lorenz system, as well as in the application to epileptic EEG records. On the other hand, the causality measure MIME, which is a normalized conditional mutual information derived from a non-uniform mixed embedding, turns out to be a less biased estimate of causal effect. MIME has the nice property of being exactly zero when no causal effect is found, i.e. no components of the driving time series are present in the vector of mixed embedding. This property is particularly useful in real-world applications, as it detects only significant driving effects, whereas TE, STE and TERV are always positive at a varying level due to bias, so that it cannot be concluded whether the driving effect is true. Therefore opposite to MIME, these measures, as any other causality measure, cannot be applied without including a bias correction or a significance hypothesis test.

Acknowledgments

I thank Pål G. Larsson for providing the EEG data.

References

[1] B. Schelter, M. Winterhalder, and J. Timmer, editors. *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. Wiley-VCH, Berlin, 2006.

[2] A. Papana and D. Kugiumtzis. Detection of directionality of information transfer in nonlinear dynamical systems. In C. H. Skiadas, I. Dimotikalis, and C. Skiadas, editors, *Topics on Chaotic Systems, Selected Papers from CHAOS 2008 International Conference*, pages 251 – 266. World Scientific, 2009.

[3] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461 – 464, 2000.

[4] M. Staniek and K. Lehnertz. Symbolic transfer entropy. *Physical Review Letters*, 100(15):158101, 2008.

[5] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1):1 – 46, 2007.

[6] D. Kugiumtzis. Transfer entropy on rank vectors. accepted in Journal of Nonlinear Systems and Applications, 2010.

[7] I. Vlachos and D. Kugiumtzis. Non-uniform state space reconstruction and coupling detection. accepted in Physical Review E, 2010.

[8] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.