Partial mutual information for simple model order determination in multivariate EEG signals and its application to transfer entropy.

Jie Zhu, Regine Le Bouquin Jeannes, Chunfeng Yang, Jean-Jacques Bellanger, Huazhong Shu

PARTIAL MUTUAL INFORMATION FOR SIMPLE MODEL ORDER DETERMINATION IN MULTIVARIATE EEG SIGNALS AND ITS APPLICATION TO TRANSFER ENTROPY

J. Zhu[1], R. Le Bouquin Jeannès[2,3], C. Yang[2,3], J.J. Bellanger[2,3], H. Shu[1]

[1]LIST, School of Computer Science and Engineering, Southeast University, Nanjing, China
[2]INSERM, U 1099, Rennes, F-35000, France
[3]Université de Rennes 1, LTSI, F-35000, France
Campus de Beaulieu, Bât 22, Université de Rennes 1, 35042, Rennes Cedex, France

ABSTRACT

The context of this work is the analysis of depth electroencephalographic signals recorded with depth electrodes during seizures in patients with drug-resistant epilepsy. Usually, different phases are observed during the seizure process and we aim to determine how cerebral structures get involved during these phases, in particular whether some structures can "drive" other ones. To this end, we consider a pair of signals and use transfer entropy which needs beforehand to choose efficiently the size of two conditioning vectors built on the past values of these signals. In this contribution, we extend a partial mutual information based technique, first developed for monochannel prediction models, to the case of two channels. Experimental results on signals generated either by a linear autoregressive model or by a physiology-based model of coupled neuronal populations support the relevance of the proposed approach.

Index Terms **—** Effective connectivity, partial mutual information, transfer entropy

1. INTRODUCTION

Over the last decade, some measures have been developed to deal with functional connectivity and effective connectivity between different cortical sites, such as in chaotic systems and multivariate neurobiological signals [1-4], which is an important topic in neuroscience. In [2], a theoretical information measure named Transfer Entropy (TE) was proposed to identify the direction of the information flow and to quantify the strength of coupling between complex systems. This model-free technique can be considered as an extension of the linear Granger causality index, which is based on a parametric linear modeling of observations, to a non parametric index well suited to non linear signals. When defined from signal Y to X, TE depends on two transition (conditional) probabilities measures characterizing stochastic dynamical links. One of them considers transitions from past values of X to the current value, and the other one considers transitions from past values of X and Y to the same current value. Its estimation can be strongly dependent on the sizes (named k and l in the sequel) of the two vectors containing the past values of X and Y respectively. The problem addressed in this contribution is to choose k and l for pairs of signals that could be not well characterized by linear AR models. If we assume AR modeling, the standard Akaike methodology can be used. If we do not assume it, a possible approach is to select nonlinear predictors in a given class. This implies order determination and parameters estimation. For one channel (signal X conditioned only on its own past), a simpler approach based on partial mutual information (PMI) was first proposed by Sharma in [5] to build a robust and effective probabilistic forecast model. The corresponding algorithm requires only computation of mutual information for 2D distributions. Compared to the Mutual Information criterion (MI) [6], PMI takes interdependences between candidate variables into account and is suitable for both linear and nonlinear models. In this paper, transfer entropy is first presented in section 2.1, before giving some details on PMI (section 2.2.1). In section 2.2.2, we extend PMI to select predictors in two-channel signals using two approaches for characterizing models orders and we introduce them in TE estimation algorithm. Experimental results are presented in section 3 before drawing some conclusions.

2. METHODS AND MATERIALS

2.1. Transfer Entropy

In the sequel, $u_n^{(p)}$ denotes $\left[u_n, u_{n-1}..., u_{n-p+1}\right]$. Considering a k-th order Markov process X, we have

$$\forall k' > k : dP_{X_{n+1}/X_n^{(k)}=x_n^{(k)}}\left(x_{n+1}\right) = dP_{X_{n+1}/X_n^{(k')}=x_n^{(k')}}\left(x_{n+1}\right) \quad (1)$$

By considering the auxiliary random process Y, relation (1) can be extended to formalize the absence of information flow from Y to X (for given k and l values):

$$dP_{X_{n+1}/X_n^{(k)}=x_n^{(k)}}\left(x_{n+1}\right) = dP_{X_{n+1}/X_n^{(k)}=x_n^{(k)},Y_n^{(l)}=y_n^{(l)}}\left(x_{n+1}\right) \quad (2)$$

The deviation from this assumption can be quantified using the Kullback pseudo-metric, which leads to define the transfer entropy from Y to X

$$TE_{Y\to X} = \int_{\mathbb{R}^{k+l+1}} \log_2\left(\left[\frac{dP_{X_{n+1}/X_n^{(k)}=x_n^{(k)},Y_n^{(l)}=y_n^{(l)}}}{dP_{X_{n+1}/X_n^{(k)}=x_n^{(k)}}}\right]\left(x_{n+1}\right)\right) \times dP_{X_{n+1},X_n^{(k)},Y_n^{(l)}}\left(x_{n+1},x_n^{(k)},y_n^{(l)}\right) \quad (3)$$

where the ratio in (3) corresponds to the Radon-Nikodym derivative of the numerator conditional probability measure with respect to the denominator's one [7]. TE is not symmetric. It can be estimated from observations (x_n, y_n), $n=1,...,N$, using a kernel discrete estimation of $\left(X_{n+1},X_n^{(k)},Y_n^{(l)}\right)$ distribution [2]:

$$\hat{P}\left(x_{n+1},x_n^{(k)},y_n^{(l)}\right) \propto \sum_m \Theta\left(\left\|\begin{pmatrix} x_{n+1}-x_{m+1} \\ x_n^{(k)}-x_m^{(k)} \\ y_n^{(l)}-y_m^{(l)}\end{pmatrix}\right\| - r\right) \quad (4)$$
$$\triangleq C_{n,r}, \qquad n=k+1,...,N-1$$

which depends on a neighborhood size (radius r). Then, it can be used to compute the estimation

$$\widehat{TE}_{Y\to X} = \sum_n \hat{P}\left(x_{n+1},x_n^{(k)},y_n^{(l)}\right)\log_2\frac{\hat{P}\left(x_{n+1}\mid x_n^{(k)},y_n^{(l)}\right)}{\hat{P}\left(x_{n+1}\mid x_n^{(k)}\right)} \quad (5)$$

where conditional probabilities in (5) are obtained from estimated joint probabilities in (4). $\Theta(.)$ is defined by $\Theta(x>0)=1$, $\Theta(x\le 0)=0$; $|\bullet|$ is the sup-norm and summation is performed on $[k+1,...,N-1]$. The value of r is chosen in the linear region of the curve $\ln r \to \ln C(r)$ where $C(r)$ is the average of the $C_{n,r}$ values. In [4], the value of k is determined heuristically from the auto-correlation function of X and the parameter l is set to 1.

2.2. Determination of *k* and *l* using PMI

2.2.1. Order determination for one signal

PMI measures the amount of information shared by X and Y while discounting the possibility that a third variable Z drives both X and Y. It was used in [5] to select efficient regressors (predictors) in past values of a process to predict (linearly or non-linearly) its current value. It takes the interdependences among candidate variables into account, and hence can be more relevant than the Mutual Information (MI) criterion [6]. PMI of (X,Y) given Z is defined as

$$\mathrm{PMI}(X,Y) = \mathrm{MI}\left(X',Y'\right) = E\left(\ln\left[\frac{dP_{X',Y'}}{d\left(P_{X'}\otimes P_{Y'}\right)}\left(X',Y'\right)\right]\right) \quad (6)$$

where $X' = X - E[X \mid Z]$ and $Y' = Y - E[Y \mid Z]$.

$P_{X',Y'}$, $P_{X'}$ and $P_{Y'}$ are the joint and marginal probability measures for the pair (X',Y'). As in [5] the joint density probability of the pair (X,Z) is estimated as a kernel estimation from realizations (x_i,z_i), $i=1,..,n$. With a Gaussian kernel whose covariance is chosen to be equal to the sample joint covariance of (X,Z), then $E[X\mid Z]$ can be written as

$$E[X\mid Z] = \sum_{i=1}^{n} w_i\left(x_i + (Z-z_i)^T S_{ZZ}^{-1} S_{XZ}\right) \quad (7)$$

where S_{XZ} is the sample cross-covariance between X and Z, S_{ZZ} is the sample covariance of Z, and w_i is given by:

$$w_i = GK(Z-z_i) / \sum_{l=1,..,n} GK(Z-z_l) \quad (8)$$

with $GK(u) = \left((2\pi)^K \left|S_{ZZ}\right| \lambda^2\right)^{-1/2} \exp\left(-\frac{u^T S_{ZZ}^{-1} u}{2\lambda^2}\right)$.

λ is a smoothing parameter chosen as in [5]. Let us consider a univariate stationary process X_n and suppose we want to choose an efficient subset Z of regressors in the set $Z_c = \{X_{n-i}, i=1,..,n_{\max}\}$ to predict X_n (*i.e.* $Y = X_n$ in Figure 1), where $n_{\max}$ is the maximal predictor lag value. The algorithm proposed by Sharma to build Z is described in Figure 1.

2.2.2. Extension of PMI to 2 signals

We proposed two approaches (Method 1 and Method 2) to extend the original PMI to a bivariate time series to determine the orders k and l needed to compute $\widehat{TE}_{Y\to X}$ as in section 2.1.

Method 1: At the first step, the candidate set $Z_c = Z_{cx} \cup Z_{cy}$ initially contains lag versions of both X (Z_{cx}) and Y (Z_{cy}), and the forecasting target is X_n. Running PMI algorithm (see Figure 1), we get $Z = Z_{xx} \cup Z_{xy}$ where Z_{xx} and Z_{xy} are subsets of selected predictors extracted from past values of X and Y respectively. The orders k and l are finally equal to the maximal lags in Z_{xx} and Z_{xy}.

Method 2: We run PMI algorithm twice. Firstly, we do not use Y. From the initial lag versions of X, Z_{cx}, we obtain a new set Z_{xx}. In a second step, we keep this set Z_{xx} and the algorithm searches in the set $Z_{xx} \cup Z_{cy}$ to get $Z_{xx} \cup Z_{xy}$. Finally, the orders k and l are obtained as in the first method.

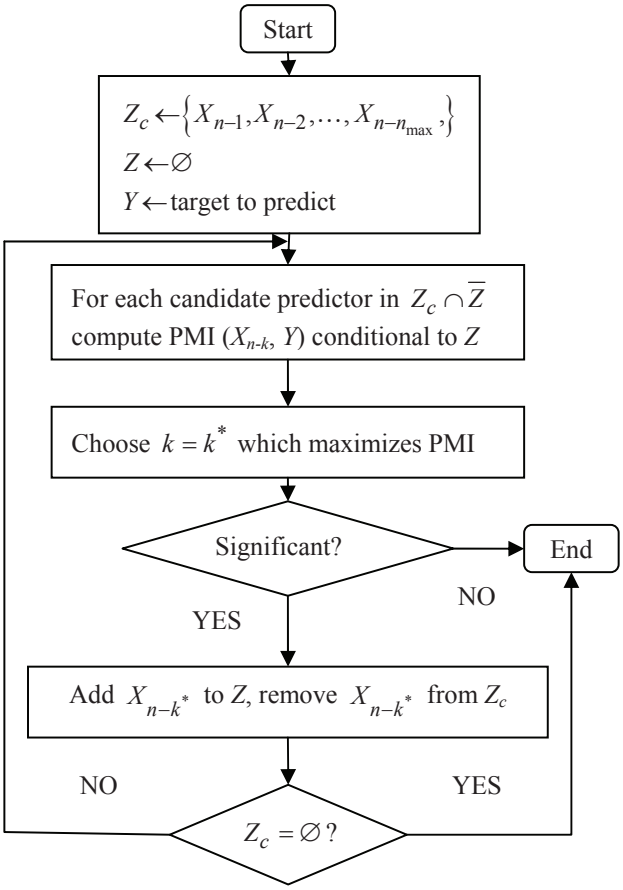


Fig. 1. Algorithm flow chart of PMI

3. EXPERIMENTAL RESULTS

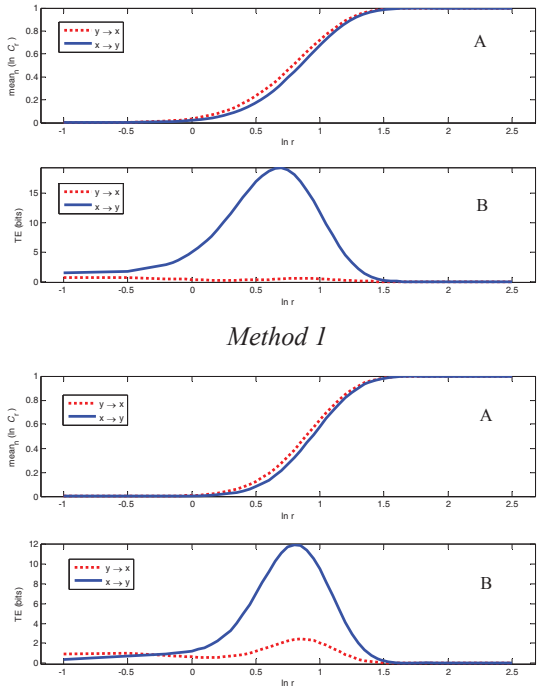
In this section, we used Method 1 and Method 2 to compute $TE_{X\to Y}$ and $TE_{Y\to X}$ on two types of signals. The first one was a toy linear AR model and the other one was a realistic EEG model. Each experiment was repeated 100 times and the following results correspond to averaged values.

3.1. Linear Model

For the linear stochastic system we considered, the following two signals were generated:

$$\begin{cases} X_n = 0.6X_{n-1} - 0.4X_{n-4} + W_n \\ Y_n = 0.6X_{n-1} + W_n' \end{cases} \quad (9)$$

where W_n and W_n' were independent white Gaussian noises with zero means and unit variances. Figure 2 displays $C(r)$ (in ln space) between signals X and Y (Figure 2.A) and the corresponding TE values vs. ln r (Figure 2.B). According to Figure 2, transfer entropy is well estimated using both methods. The flow propagation from signal X to signal Y is correctly established, the highest dynamics being obtained using Method 1. It is also clear that there was no influence of signal Y to X.



Method 1

Method 2

Fig. 2. Results of Transfer Entropy between X and Y in AR4 model (Eq. (9)).
Top: Results for method 1. Bottom: Results for method 2.
A. Plot of the mean of $C(r)$ (in ln scale) over the trials.
B. Plot of TE (bits) vs. radius r (in ln scale), for direction of flow $X \to Y$ (blue solid line) and $Y \to X$ (red dotted line).

3.2. Physiology-based Model

We used a physiology-based time continuous SDE (Stochastic Differential Equation) model to represent the electrical activity of two distant — and possibly coupled — neuronal populations Pop_X and Pop_Y. This model produces two outputs X and Y which can be interpreted as two iEEG (intracranial EEG) signals similar to those recorded with proximal electrodes [8]. In the model, each population contains three subpopulations of neurons that mutually interact via excitatory or inhibitory feedback linking main pyramidal cells with two other types of local interneurons. Since pyramidal cells are excitatory neurons that project their axons to other areas of the brain, the model accounts for this organization by using the average pulse rate of action potentials from the main cells of one population, Pop_X, as an excitatory input to main cells inputs of another population, Pop_Y. The connection from the first population to the second one is represented by a parameter K_{XY} which is proportional to the number of corresponding axonal links. An appropriate setting of this

parameter allows adjusting effective connectivity. Other introduced parameters are intra-population parameters whose tuning allows modifying the type of activity (normal/epileptic). This model was used to simulate 100 blocks of 4-second length signals. Sampling rate was equal to 256 Hz. Model parameters were such that a fast quasi-sinusoidal (25 Hz) activity (similar to that observed at seizure onset) was generated by the two populations which were unidirectionally coupled as already mentioned. Referencing to [8], we used parameters values for (A , B , G) equal respectively to (5, 3, 20) and (3.5, 3.5, 84) in Pop_X and Pop_Y with $K_{XY} = 1500$ (unidirectional flow). Experimental results are shown in Figure 3. Even if the dynamics of transfer entropy is reduced compared to the linear model, the second method appears relevant and outperforms the first one in detecting the flow propagation from X to Y .

4. CONCLUSIONS

In this paper, we focused on information propagation between two observations using TE and employed the extended PMI method in the determination of models orders. The proposed approach appeared efficient to reveal the information flow direction between two signals in unidirectional case. In a future work, we will test it on multivariate systems including bidirectional flows. We also plan (i) to conduct statistical analysis on the estimators we proposed to compare them with other order determination methods, and (ii) to take into account supplementary EEG observations for conditioning surrounding information.

5. ACKNOWLEDGMENTS

The authors thank F. Wendling for providing the database on the physiological model and A. Sharma for fruitful discussions on PMI.

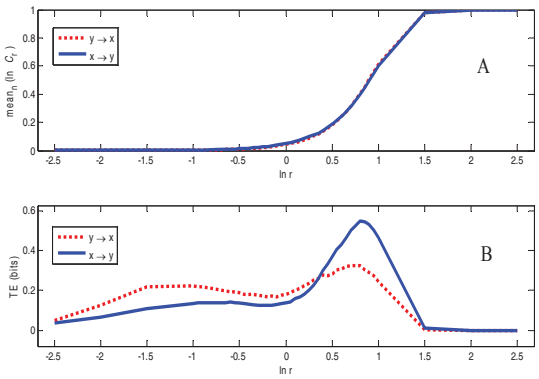
6. REFERENCES

[1] J.J. Engel, P.C. Van Ness, T.B. Rasmussen, and L.M. Ojemann, Outcome with respect to epileptic seizures. New York: Raven Press, 1993.

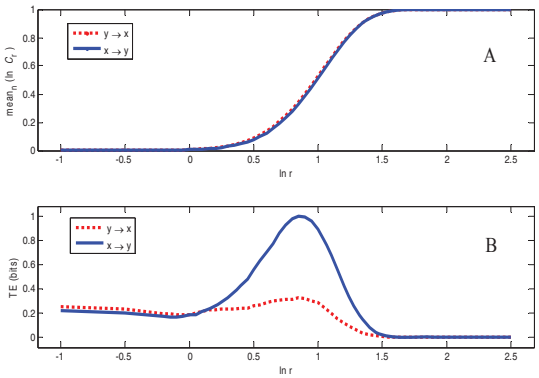
[2] T. Schreiber, "Measuring information transfer", *Physical review letters*, vol. 85, pp. 461-464, 2000.

[3] S. Sabesan, K. Narayanan, A. Prasad, L.D. Iasemidis, A. Spanias, and K. Tsakalis, *Information flow in couple nonlinear system: Application to the epileptic human brain*. New York: Springer, 2007.

[4] S. Sabesan, L. Good, K. Tsakalis, A. Spanias, D. Treiman, and L. Iasemidis, "Information flow and application to epileptogenic focus localization from intracranial EEG", *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 17, pp. 244-253, 2009.



Method 1



Method 2

Fig. 3. Simulation results on physiological model.
Top: Results for method 1. Bottom: Results for method 2.
A. Plot of the mean of $C(r)$ (in ln scale) over the trials.
B. Plot of TE (bits) vs. radius r (in ln scale), for direction of flow $X \to Y$ (blue solid line) and $Y \to X$ (red dotted line).

[5] A. Sharma, "Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1-A strategy for system predictor identification", *Journal of Hydrology* 239(1-4): 232-239, 2000.

[6] A.M. Fraser and H.L. Swinney, "Independent coordinates for strange attractors from mutual information", *Physical Review A* 33(2): 1134-1140, 1986.

[7] R. M. Gray, "Entropy and information theory", New Work, USA, Springer Science+business Media, 2011.

[8] F. Wendling, A. Hernandez, J.J. Bellanger, P. Chauvel, and F. Bartolomei, "Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG", *Journal of Clinical Neurophysiology*, vol. 22, pp. 343-356, 2005.