Assessment of Deep Learning Approaches for the Detection of Cardio-Respiratory Causal Interactions

Andrea Rozo[1], Dries Testelmans[2], Bertien Buyse[2], Carlo Iorio[1], Xiao Hu[3], Carolina Varon[4]

[1] Service Aero-Thermo-Mechanique, Université libre de Bruxelles, Brussels, Belgium
[2] Department of Pneumology, UZ Leuven, Leuven, Belgium
[3] Nell Hodgson Woodruff school of Nursing, Emory University, Atlanta, GA, USA
[4] Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium

Abstract

Granger causality (GC) and transfer entropy (TE) are commonly used methods for studying causality between physiological signals. Recently, neural networks (NN) approaches have been introduced for TE estimation and GC detection. This study compared traditional estimation of TE using adaptive partitioning (DVP), with two NN approaches: neural network GC (nnGC) and neural network TE (nnTE). The comparison was performed based on their ability to detect interactions and their computation times. The study used three simulation models (linear, nonlinear, and linear + nonlinear) and cardio-respiratory data from a polysomnography study of 26 subjects. In the simulations, DVP outperformed the NN-based methods (mean area under the receiving operating curve – AUC: 0.99, computation time: 0.07 s). nnGC performed well (mean AUC: 0.98) but was slower than DVP (167.07 s), while nnTE struggled with the linear and linear + nonlinear models (mean AUC: 0.75 and 0.49, respectively) and was the slowest (943.83 s). In the clinical data, nnGC and nnTE detected interactions on 9 and 10 subjects, respectively, which aligned with the results obtained by DVP (median of 12 subjects). These findings imply that these NN-based methods provide a suitable model-free alternative for studying cardio-respiratory interactions during sleep.

1. Introduction

The study of complex time series obtained from physiological systems can unveil underlying interactions between them, which can be linear, nonlinear, or a combination of both. To study these interactions, different methods have been developed using parametric and nonparametric approaches [1].

Statistics-based methods are a group of parametric approaches typically based on predefined models describing mainly linear interactions. An example of these methods is Granger causality (GC). Its linear version relies on vector autoregressive models and predictability to detect causal and directed relationships between time series. A nonlinear version of GC was introduced in 1992 in the field of economy, but it was not until recently that it was applied in physiology [2].

Information theory-based methods are a group of nonparametric approaches to study causality and information flow. Transfer entropy (TE) is one of the best known information-theoretic methods. TE can identify linear and nonlinear interactions, and their direction, between time series. It is based on Shannon entropies and focuses on the reduction of uncertainty. TE and GC are equivalent for Gaussian variables with linear interactions [1].

Recent developments use neural networks (NN) to detect linear and nonlinear GC [3], estimate entropies, and compute TE [4]. This could provide a model-free solution for the study of physiological interactions, reducing the definition of the parameters needed to estimate GC or TE. The analysis of the cardio-respiratory interactions during sleep is one example of the applications that can benefit from these approaches, as it is hypothesized that the linear and nonlinear dynamics of these time series vary during the sleep cycle.

This study compared two NN-based methods and a traditional method to estimate TE, to understand if the new approaches have advantages over the traditional ones to analyze cardio-respiratory interactions. The selected NN-based methods detected GC (nnGC) using a multilayer perceptron as described in [3], and estimated TE (nnTE) using a fully connected network, with a modified version of the model presented in [4]. The traditional method for TE estimation was adaptive partitioning (DVP), that according to [5] was the best approach to study cardio-respiratory interactions during sleep. The performance was compared on three simulation models (linear, nonlinear, linear + nonlinear). Then, nnGC and nnTE were applied to cardio-respiratory signals of polysomnography recordings during light and deep sleep. These results were compared with the ones obtained by DVP in previous studies [5].

2. Methodology

2.1. Adaptive partitioning

Transfer entropy (TE) is a measure of information transfer between processes, based on Shannon entropy. The

TE between random processes $X = \{x_1, x_2, ..., x_N\}$ and $Y = \{y_1, y_2, ..., y_N\}$ can be understood as the reduction of uncertainty in the future values of Y, when also considering the past values of X rather than only considering the past values of Y. Additionally, it is an asymmetric measure that indicates the direction of the information flow. The TE from X to Y is defined as [5]:

$$TE_{X \to Y} = H(Y|Y^-) - H(Y|X^-, Y^-), \qquad (1)$$

where $H(Y|Y^-)$ is the conditional entropy of Y given its own past, and $H(Y|X^-, Y^-)$ is the conditional entropy of Y given its own past and the past of X. Y^- and X^- are called embedding vectors.

In this work, an available implementation of the adaptive partitioning (DVP) estimator combined with a uniform embedding technique was used. The embedding vectors were generated as $X^- = \{x_{t-k}, x_{t-2k}, ..., x_{t-mk}\}$ and $Y^- = \{y_{t-k}, y_{t-2k}, ..., y_{t-nk}\}$, with k the embedding delay and m and n the embedding dimensions, and $m = n$. The DVP estimator combined the Darbellay-Vajda algorithm with ordinal sampling. For this, the time series resulting from sampling X and Y were replaced by their ordinal representations U and V, respectively, and then the space defined by V, U^- and V^- was iteratively partitioned into cubes of different sizes, until all partitions had an even distribution of points [6].

This method was selected because in [5] it was identified as the best option to study interactions in sleep data.

A detailed account of the definition of TE is presented in [1,5]. Additional information for embedding techniques and other entropy estimators can be found in [1, 5–7].

2.2. Neural network transfer entropy

Previous studies [1, 5–7] have shown that the traditional approaches to estimate (1) can introduce significant bias and, in some cases (e.g., linear estimator), a model for the time series dynamics is assumed. To overcome these difficulties, an approach based on neural networks (NN) to estimate conditional entropies was proposed in [4]. The network consisted of one input layer, two fully connected layers with 50 nodes, a rectified linear unit (ReLU) activation function and an output layer with softmax activation function. The training was based on the cross-entropy loss minimization. A quantization was performed on the time series, as this model is not defined for continuous data.

This work proposed a modified version of the model of [4]. The probability estimation is modelled as a regression problem and not as a classification problem, which led to the definition of a new loss function, composed of two terms as:

$$L = -\frac{1}{N} \sum_n \ln(G_\theta(x_n)) + \lambda \left(1 - \sum_c G_\theta(x_c)\right)^2 \qquad (2)$$

The first one is the empirical estimator for the cross-entropy as presented in [4], where x_n is the n-th sample of X, N is the total number of samples, and $G_\theta(x_n)$ is the output of the NN for x_n. The second one is a regularization term to ensure that the sum of the probabilities of all possible values was equal to 1, where λ is a regularization parameter and x_c is the c-th element of the finite set of values that x_n can take. Additionally, the activation function of the last layer was changed to a sigmoid, so that the estimated probabilities were always between 0 and 1. For this model, the number of quantization levels was tuned and defined as 5. The output corresponds to the conditional probabilities needed to estimate the conditional entropies in (1) and compute the TE.

2.3. Neural network Granger causality

Granger causality (GC) is a statistical measure of causality between stochastic processes. Assuming an autoregressive linear model for Y, it can be said that Y is explained by its present and past values, and that the part that is not explained by the model corresponds to a random white noise. Now, if X Granger-causes Y, a model that better explains Y is defined as:

$$Y_t = \sum_{j=1}^{N} (A_j Y_{t-j}) + \sum_{j=1}^{N} (B_j X_{t-j}) + \epsilon_t, \qquad (3)$$

where there is at least one non-zero B_j coefficient [1]. However, this definition does not consider nonlinear interactions. For this reason, nonlinear GC approaches have been proposed [1], but in practice their applicability has been limited due to over-fitting leading to the detection of many false positives (i.e., non existing interactions) [2]. Lately, new ways to apply these approaches using NN have been proposed. In [3], the authors presented a multilayer perceptron model with a hierarchical group lasso loss function and a proximal gradient descent optimization, to identify the presence of GC between time series. The network consisted of one input layer, a fully connected layer and a ReLU activation function. Thanks to the combination of the loss function and optimization method, the weights of the first layer were forced to be zero if there is no GC between the time series. If the norm of these weights was different from zero, X Granger-caused Y. This value is a proxy of the (nonlinear) GC and is only an indicative of the presence of the interactions between the sampled processes.

In this work, the number of units in the hidden layer and the regularization term for the loss function were tuned based on [3], and defined as 5 and 0.002, respectively.

2.4. Simulation data

The three simulation models presented in [5] were used to test if the NN-based approaches could identify the predefined interactions. In all cases, the interactions were from X to Y, but not in the opposite direction. Fifty trials (pairs of time series) of each model were simulated. For each trial a different seed for the random components was used. The simulated time series had N = 200, 500, 1000

and 2000 samples, to study the effect of the time series length on the detection of the interactions.

The first model was a linear model defined as $x_n = -0.5x_{n-2} + \varepsilon_{x_n}$, $y_n = -0.5y_{n-2} + ax_{n-\tau} + \varepsilon_{y_n}$, where ε_{x_n} and ε_{y_n} were gaussian noises with zero mean (μ) and unit variance (σ^2). The interaction was modulated by $a = 0.5$. The interaction occurred at $\tau = 1$.

The second model was a nonlinear model defined as $x_n = s_{x_n} + \xi_{x_n}$, $y_n = (b\, x_{n-\tau})^2 + \xi_{y_n}$, where s_{x_n} was a random gaussian signal with $\mu = 10$ and $\sigma^2 = 1$, ξ_{x_n} and ξ_{y_n} were laplacian noises with $\mu = 0$ and $\sigma^2 = 1$, and $b = 0.4$ was the coupling factor. The interaction occurred at $\tau = 2$.

The last model included linear and nonlinear interactions, and it was defined as $x_n = 0.3x_{n-1} + \varepsilon_{x_n}$, $y_n = 0.3y_{n-1} + cx_{n-\tau_l} + d\left(\frac{2.4-0.9(x_{n-\tau_{nl}})}{1+\exp(-4(x_{n-\tau_{nl}}))}\right) + \varepsilon_{y_n}$, where ε_{x_n} and ε_{y_n} were gaussian noises with $\mu = 0$ and $\sigma^2 = 1$, and $c = 0.4$ and $d = 0.6$ controlled the linear and nonlinear dynamics, respectively. The linear and nonlinear interactions occurred at $\tau_l = 2$ and $\tau_{nl} = 4$, respectively.

2.5. Clinical data

Polysomnography (PSG) recordings of 26 subjects (median (25th;75th) age and BMI: 37 (34; 47) years, 24.98 (23; 32.19) kg/m^2), referred to the sleep laboratory of the UZ Leuven were used. The study was approved by the ethical committee of UZ Leuven (S53746, S60319) and all subjects signed an informed consent. One-minute segments from the ECG and respiratory (nasal airflow) signals (fs = 500 Hz) were extracted, for light (NREM1) and deep (NREM3) sleep. In total, 1891 apnea-free segments were used.

The ECG signals (lead II) were filtered with a Butterworth bandpass filter (0.03 - 150 Hz), and then with a notch filter (50 Hz). The locations of the R-peaks were identified and the missed, false and ectopic beats were corrected using the integral pulse frequency modulation model. The corrected locations were used to compute the heart rate variability (HRV) time series, which was then resampled at 4Hz. A final Butterworth bandpass filter (0.03 - 1 Hz) was applied.

The respiratory (RESP) signal was bandpass filtered with a Butterworth filter (0.03 - 1 Hz), and then resampled at 4Hz.

For the detailed description of the data and the preprocessing, see [5].

2.6. Performance evaluation

2.6.1. Simulation models

Since nnGC is only able to detect the presence of interactions, the comparison of the methods was done by a binary estimation of presence or absence of interactions between time series. For the simulation models, the estimates of each of the approaches was binarized by defining a decision threshold. For the nnGC, the threshold was defined for the weights of the first layer of the network. For the nnTE and DVP the threshold was defined for the estimated TE. There were interactions detected if the norm of the weights or the estimation of TE was higher than the threshold, for nnGC, and for nnTE and DVP, respectively.

A receiver operating characteristic (ROC) curve was calculated varying the decision threshold of the weights of nnGC and the estimated TE of DVP and nnTE.

For this analysis, a maximum lag to observe the interactions was defined. This lag was fixed at 2 for the linear and nonlinear models, and at 4 for the linear + nonlinear model.

2.6.2. Clinical data

The interactions during sleep were analyzed from RESP to HRV, and the maximum lag was defined as 5 seconds (20 samples). In this case, there was no ground truth available. The results obtained by nnGC and nnTE were assessed by selecting an appropriate threshold, and comparing the norm of the weights or the estimation of the TE with this threshold.

The threshold for the nnGC was defined using the simulation ROC curves. The operating point with a false positive rate of 5% was selected, and The threshold corresponding to this point was used.

For nnTE, the threshold was defined based on the results obtained in [5]. The threshold was selected as the lower TE value of the significant segments for the $TE_{RESP \to HRV}$ when studying the nonlinear interactions. This value was selected because, as mentioned in [5], if the TE was significant for nonlinear interactions there was a possibility that a linear interaction was also present.

The results, in terms of for how many patients the interactions were detected, were compared against the results obtained in [5].

3. Results and Discussion

Table 1 presents the area under the ROC curve (AUCs) for each method and for each simulation model, for the four time series lengths.

DVP and nnGC are highly specific, obtaining values very close to zero when there is no interaction between the time series, and values almost perfectly separable from zero when the interaction exists. Nevertheless, DVP outperforms the NN-based approaches. This can be because DVP is based on a pre-defined method of probability estimation, compared to the behavior of a NN that only relies on the loss function optimization to estimate the desired probabilities. Additionally, NN-based methods are data driven, which renders them more sensitive to the samples used during training.

DVP detected the interactions for all three simulation methods with an AUC higher than 0.97. nnGC had a higher performance for the linear and linear + nonlinear models (mean AUC 0.99) and a slightly lower performance for the nonlinear model (mean AUC 0.96). These results can

Table 1. Area under the curve (AUC) for each of the approaches for the simulation models

Model	Samples	DVP	nnGC	nnTE
Linear	200	0.973	0.986	0.472
	500	1	1	0.665
	1000	1	0.974	0.858
	2000	1	0.990	0.993
Nonlinear	200	1	1	0.870
	500	1	0.980	0.973
	1000	1	0.940	0.984
	2000	1	0.910	1
Linear + Nonlinear	200	0.975	1	0.462
	500	1	1	0.490
	1000	1	1	0.505
	2000	1	0.960	0.507

be explained considering that the proposed solution is still based on an autoregressive model, and the nonlinear simulation model does not follow this pattern. The nnTE struggled to identify the linear and linear + nonlinear interactions, with a worse performance in the latter, which can be related to the fact that by considering a higher maximum lag the dimensionality of the problem increases making it necessary to use a huge amount of data for the estimation of the probabilities. To address this issue it could be useful to test different architectures of the network, like an encoder-decoder to reduce the dimensionality of the input and then compute the probabilities with the extracted features.

Additionally, the methods differed in terms of computation times expressed as the average time to process a pair of time series in both directions of interactions, with DVP being the fastest (0.07 seconds), nnGC following (167.01 seconds) and nnTE the slowest (943.83 seconds). This factor could affect the usability of nnGC and nnTE.

For the clinical data, the threshold for nnGC was 0.097 and the threshold for nnTE was 0.089. With these thresholds, the number of patients for which interactions from RESP to HRV were detected for both NREM1 and NREM3 were 9 with nnGC and 10 with nnTE. Separately, for NREM1 and NREM3, nnGC detected interactions for 13 and 12 patients, respectively. In the case of nnTE, for NREM1 it detected interactions for 14 patients and for NREM3 for 19 patients.

These results are consistent with the ones obtained in [5], in which a median of 12 patients were observed to have significant $TE_{RESP \rightarrow HRV}$ at 4 Hz, for both linear and nonlinear interactions and over all the lags (1 to 5 seconds). However, these results could benefit from a surrogate analysis similar to the one performed in [5]. This analysis should be designed for each of the methods separately, as they do not focus on the same type of measure (i.e., GC vs. TE).

4. Conclusion

This study compared two NN-based methods (nnGC and nnTE) and a traditional TE estimation approach (DVP) for the detection of interactions in simulation models and in clinical data. The results suggest that nnGC is preferred over nnTE given that it is more specific when detecting the interactions and presents a lower computation time. However, as it only detects the interactions based on a proxy of GC, more research is suggested to interpret its outputs for the quantification of the interaction. The performance of nnTE was affected by the selection of the maximum lag. To overcome this, tests with other architectures could be done, for example, with an encoder-decoder network to handle the changes in dimensionality. Regarding their application to clinical data, both methods show potential for the analysis of cardio-respiratory interactions during sleep, but additional developments like surrogates analysis are needed to confirm these results.

Acknowledgements

A.R., C.I. and C.V. acknowledge the financial support of the European Space Agency (ESA), BELSPO - NEPTUNE.

References

[1] Hlaváčková-Schindler K, et al. Causality detection based on information-theoretic approaches in time series analysis. Phys Rep 2007;441:1–46.

[2] Marinazzo D, et al. Kernel method for nonlinear granger causality. Phys Rev Lett 4 2008;100:144103.

[3] Tank A, et al. Neural granger causality. IEEE Trans Pattern Anal Mach Intell 2 2021;1–1.

[4] Shalev Y, et al. Neural joint entropy estimation. IEEE Trans Neural Netw Learn Syst 2022;1–13.

[5] Rozo A, et al. Benchmarking transfer entropy methods for the study of linear and nonlinear cardio-respiratory interactions. Entropy 2021 Vol 23 Page 939 7 2021;23:939.

[6] Lee J, et al. Transfer entropy estimation and directional coupling change detection in biomedical time series. Biomed Eng Online 2012;11:19.

[7] Montalto A, et al. Mute: A matlab toolbox to compare established and novel estimators of the multivariate transfer entropy. PLoS ONE 2014;9.

Address for correspondence:

Andrea Rozo
CREST/ATM/Université libre de Bruxelles
Av. Paul Heger, 1050 Bruxelles, CP 165/43
andrea.rozo@ulb.be