Classification-based Causality Detection in Time Series

Abstract. Brain effective connectivity aims to detect causal interactions between distinct brain units and it can be studied through the analysis of megneto/electroencephalography (M/EEG) signals. Methods to evaluate effective connectivity belong to the large body of literature related to detecting causal interactions between multivariate autoregressive (MAR) data, a field of signal processing. Here, we reformulate the problem of causality detection as a classification task and we propose a classification-based approach for it. Our solution takes advantage of the MAR model by generating a labeled data set that contains trials of multivariate signals for each possible configuration of causal interactions. Through the definition of a proper feature space, a classifier is trained to identify the causality structure within each trial. As evidence of the efficacy of the proposed method, we report both the cross-validated results and the details of our submission to a recent causality detection competition, where the method reached the 2nd place.

1 Introduction

A central aspect of neuroscience concerns brain connectivity and aims to investigate the pattern of interactions between distinct units within the brain [10]. The concept of brain units is strongly related to the level of the adopted scale. Thus, brain connectivity can be studied from the microscopic level of single synaptic connections to the macroscopic one of brain regions. Depending on the type of interactions that we focus on, the topic of brain connectivity is divided into structural, functional and effective connectivity. In the first case the connectivity patterns are referred to anatomical links i.e. neural pathways. In the second case to the statistical dependences between brain activity in different units and in the last one to the causal interactions between them [15].

In particular, effective connectivity, that is the one we are interested in, provides information about the direct influence that one unit exerts over another and aims to establish causal interactions among them [7]. To achieve this goal the usefulness of brain signals measured by magneto/electroencephalography (M/EEG) has been largely shown [3]. In fact, they are high temporal resolution signals that directly measure the brain activity. A large body of work was developed about methods to quantify the effective connectivity, mainly in the field of signal processing where it is known as the problem of inferring causality among time series. An overview of the literature is provided below.

A first distinction that can be made in the available causality detection methods, is between linear and nonlinear methods.

Linear approaches are largely used both in time and frequency domain. An example of time domain technique is the Granger Causality index. Granger

Causality is one of the most widespread measure to estimate the direction of causal influence in time series and its basic assumption, that a cause has to precede its effect, has been adopted in many other methods [8]. More precisely, if one or more time series $x_0(t), \ldots, x_k(t)$ are causing the time series $y(t)$, then a future value of $y(t)$ is better predicted by considering also the past values of $x_0(t), \ldots, x_k(t)$ than only those of $y(t)$. Most of the other time domain methods have the property that their multivariate extension is based on the partial auto- and cross-spectra estimation done by frequency-domain methods [16]. Thus, these latter have had a great spread in causality assessment [5]. Examples are the direct transfer function (DTF) [12, 11], the direct coherence (DC) [2] and the partial direct coherence [1].

In situations in which the nonlinear component of the causal interaction is expected to be important, nonlinear multivariate methods are used [14]. A first attempt to deal with nonlinearity was done by the local application of linear multivariate methods in order to perform nonlinear prediction [6]. Further approaches are based on information theory [9], phase synchronization [4] and state space synchronization [13].

The intricate structure of interconnections, the enormous amount of dependence that brain units can exert over each other and, last but not least, the lack of a ground truth, make the assessment of the causal interactions a very complex problem. Thus, generally, new approaches to estimate causal interactions are assessed and validated on a limited set of signals and often by using simulated multivariate autoregressive (MAR) data. This is a common preliminary stage that allows researchers to analyse the performance of their techniques in the fully controlled environment of the MAR model. An example of the interest that has been addressed to causality in multivariate time series is the Biomag2014 Causality Challenge (Causal2014) [1]. Its purpose was to estimate the direct causal interactions in a data set of simulated trials. One trial is meant as three multivariate time series, generated by a known MAR model that is expected to simulate the behaviour of three neuronal populations.

In this paper we propose a new approach for the causality detection in time series by tackling the problem from a different prospective. Instead of developing a solution in the context of signal processing, as in the previous literature, we faced the problem from the machine learning point of view. Since modelling causal interactions with a MAR model is a common practice in the literature, we used the competition MAR model to create a set of trials for each possible causal configuration among the time series. Then a classifier was trained on those data in order to discriminate between causal configuration. Finally, it was applied to the competition data set providing a solution that reached the second place of Causal2014.

[1] https://dl.dropboxusercontent.com/u/94877880/causality%20challenge%20biomag%202014/BioMag2014-Causality-Challenge.htm
http://www.biomag2014.org/competition.shtml , see competition 2.

2 Materials

The competition organizers provided the code of the MAR model together with the data set of which to estimate the direct causal interactions. Here, we will describe them both.

The final output of the MAR model is the multivariate time series $\mathbf{X} = \{X(t), t = 0, 1, \ldots, N-1\}, X(t) \in \mathbb{R}^{M\times 1}$ that is defined as the linear combination of two M-dimensional multivariate time series $\mathbf{X_s}$ and $\mathbf{X_s}$

$$\mathbf{X} = (1-\gamma)\mathbf{X_s} + \gamma\mathbf{X_n} \tag{1}$$

$\mathbf{X_s}$ carries the causal information, $\mathbf{X_n}$ represents the noise corruption and $\gamma \in [0, 1]$ tunes the signal-to-noise ratio. Each time point of $\mathbf{X_s}$ and $\mathbf{X_n}$ is computed by following the MAR model

$$\begin{aligned} X_s(t) &= \sum_{\tau=1}^{\min(P,t)} A_s^{(\tau)\top} X_s(t-\tau) + \varepsilon_s(t) \\ X_n(t) &= \sum_{\tau=1}^{\min(P,t)} A_n^{(\tau)\top} X_n(t-\tau) + \varepsilon_n(t) \end{aligned} \tag{2}$$

where P is the order of the MAR model and represents the maximal time lag. $\varepsilon_s(t)$ and $\varepsilon_n(t)$ are realizations from a M-dimensional standard normal distribution. And $A_s^{(\tau)}, A_n^{(\tau)} \in \mathbb{R}^{M\times M}, \tau = 1, 2, \ldots, P$ are the coefficient matrices modelling the influence of the signal values at time $t-\tau$ on the current signal values, i.e. at time t. The coefficient matrices $\{A_s^{(\tau)}\}_\tau$ are involved in the process of causal-informative data generation. They are computed by randomly corrupting the non-zero elements of the $M \times M$ binary matrix A, called configuration matrix. In essence, the configuration matrix A contains the causal structure that leads the MAR model. Specifically $A_{i,j} = 1$ means signal i causes the signal j. On the other hand, coefficient matrices $A_n^{(\tau)}$ lead the noisy part of the signals and they are obtained by randomly generating P diagonal matrices. The diagonality of these latter matrices is needed to avoid noise regressive dependencies across signals. After that, if both sets of matrices $A_s^{(\tau)}$ and $A_n^{(\tau)}$ fulfil the stationarity condition, each time point of $\mathbf{X_s}$ and $\mathbf{X_n}$ can be generated by Equation 2.

In essence, given P, γ and A, it is possible to generate $\mathbf{X}$ following Equation 1 and Equation 2. The goal of the competition is to reconstruct A given $\mathbf{X}$.

The competition data set was built by generating 1000 trials with the following parameter assignments: the number of time series in each trial is $M = 3$, the MAR model order is $P = 10$ and the time series length is $N = 6000$. The trial-specific parameters γ and A were randomly sampled for each trial. For the sake of simplicity, we will refer to the competition data set as $\mathbf{C}$.

3 Methods

The solution that we propose to the causality detection problem is based on a supervised approach. Indeed, this task can be formulated in term of a classification problem. In a general setting, each trial is composed by M time series and

the final goal is to estimate its binary configuration matrix A. Thus, there are $M(M-1)$ free binary parameters and $2^{M(M-1)}$ possible causal configurations [2].

Our supervised approach aims to train a classifier on a new simulated data set generated by the MAR model described in Section 2. The new data set, $\mathbf{L}$, contains multiple trials for each of the possible $2^{M(M-1)}$ causal configurations. And the causal configuration represents the class label of the trial. After the definition of a proper feature space, a classifier f is trained on $\mathbf{L}$. In order to evaluate it, its discriminative power is estimated through cross-validation and, finally, f is applied to the competition data set $\mathbf{C}$ to predict the configuration matrix of each trial.

The feature space, that we built, is strongly based on the concept of Granger causality. Indeed, it is a collection of measures that quantifies the ability to predict the value at a given time point of a certain time series (effect) from the past values of each possible subset of the M time series in the trial (causes). The pair, made by causes and effect, is called causality scenario and, for M time series, there are $\sum_{i=1}^{M} \binom{M}{i} M$ scenarios. In the case of the competition, where $M = 3$, the possible causality scenarios are 21 and they are summarized in Table 1, when $x_i(t), i = 0, 1, 2$, denotes each of the time series that defines a trial.

Causes	Effect
$x_0(t)$	$x_i(t)$
$x_1(t)$	$x_i(t)$
$x_2(t)$	$x_i(t)$
$x_0(t), x_1(t)$	$x_i(t)$
$x_0(t), x_2(t)$	$x_i(t)$
$x_1(t), x_2(t)$	$x_i(t)$
$x_0(t), x_1(t), x_2(t)$	$x_i(t)$

Table 1. The possible causality scenarios for three time series $x_i(t), i = 0, 1, 2$.

For each causality scenario, a plain linear regression problem was built by selecting as dependent variable a set of time points from the signal in the effect column. Each of these dependent variables has a P-dimensional vector of regressors composed by the P previous time points selected from the signals in the causes column. Table 2 shows how the regression problems are defined when $M = 3$, by specifying from which time series and time points, regressors and dependent variables are extracted. In the following, in order to simplify the notation, we will use x_i^t instead of $x_i(t)$, $i = 0, 1, 2$ and $t \in \mathbf{T}, \mathbf{T} \subseteq \{P, P+1, \ldots, N-1\}$.

The regression problem of each causality scenario was cross-validated and its performance was quantified through multiple regression metrics, e.g. mean square error. The ensemble of the regression metrics of each causal scenario defined the initial feature vector of the trial. We then applied standard feature engineering techniques to enrich the feature space. See Section 4 for details.

[2] The diagonal is not relevant since by definition the time series are autoregressive.

Regressors (causes)	Dependent variable (effect)
$[x_0^{t-1}, \ldots, x_0^{t-10}]$	x_i^t
$[x_1^{t-1}, \ldots, x_1^{t-10}]$	x_i^t
$[x_2^{t-1}, \ldots, x_2^{t-10}]$	x_i^t
$[x_0^{t-1}, \ldots, x_0^{t-10}, x_1^{t-1}, \ldots, x_1^{t-10}]$	x_i^t
$[x_0^{t-1}, \ldots, x_0^{t-10}, x_2^{t-1}, \ldots, x_2^{t-10}]$	x_i^t
$[x_1^{t-1}, \ldots, x_1^{t-10}, x_2^{t-1}, \ldots, x_2^{t-10}]$	x_i^t
$[x_0^{t-1}, \ldots, x_0^{t-10}, x_1^{t-1}, \ldots, x_1^{t-10}, x_2^{t-1}, \ldots, x_2^{t-10}]$	x_i^t

Table 2. Description of how the 21 linear regression problems are defined for each trial. x_i^t, $i = 0, 1, 2$ and $t \in \mathbf{T}, \mathbf{T} \subseteq \{10, 11, \ldots, N-1\}$, are the three time series of a trial.

4 Experiments

In this section we present the technical details and results of the experiments that were conducted to evaluate the method described in Section 3. In particular, we show two different types of results. The first one is an estimate of the discriminative power of a classifier trained on the **L** data set and it provides a quantification of how well the defined feature space is able to express the causal structure behind a trial. The second result is the competition score obtained by our submission and it gives us some insights into how our approach works compared to the ones adopted by the other participants.

The new simulated labeled data set **L** was generated by keeping the same parameter initialization of **C**, except for the number of trials that was increased to 64000 in order to have 1000 trials for each class. Indeed, since $M = 3$ the amount of causal configurations is $2^6 = 64$. The regression metrics used to build the feature space are the mean square error and the coefficient of determination r^2. Both were included since we noticed a significant improvement in the cross-validated results, although, intuitively, they could seem redundant. We also added an estimate of the Granger causality coefficients [3] to the feature space.

As a final step we increased the number of features through standard feature engineering techniques. This consisted in extracting the 2nd power, 3rd power and square root of the previously defined features, together with the pairwise product of all features. Adding extracted features was motivated both by the need to overcome the limitation of the adopted linear classifier and because they proved to be effective in increasing the cross-validated performance.

Both of the data sets **L** and **C** were mapped to the proposed feature space. Then the performance of the logistic regression classifier [4], with ℓ_2 regularisation, was evaluated on **L** by the 5-folds cross-validation. In this way we quantified the discriminative capability of the proposed method.

We present the results both in terms of confusion matrices and competition score. The competition score was defined in the following way. For each entry $\hat{a}_{ij}$, $i \neq j$, of each predicted $\hat{A}$, if $\hat{a}_{ij}$ was 1 and correct, then $+1$ point was given. If $\hat{a}_{ij}$ was 1 but incorrect, then -3 points were given. If $\hat{a}_{ij}$ was 0, then 0 points were given. In practice, false discoveries were punished three times more than what true discoveries were rewarded.

[3] http://nipy.org/nitime

[4] http://scikit-learn.org

In order to take into account the strong false positive penalisation, we added a cost model to our predictions, by combining the probability of each of the 64 causality scenario with the cost of predicting one scenario instead of another. Given S_{ij} the cost of predicting i when the true class/scenario was j, the optimal way to assign the class l to a trial is

$$l = \underset{i=0,1,\ldots,63}{\operatorname{argmax}} \sum_{j=0}^{63} S_{i,j} p_j \tag{3}$$

where p_j is the probability of class j for the trial, as estimated by logistic regression.

Table 3 and Table 4 show the classification results cross-validated in **L** by means of confusion matrices. In particular, Table 3 is related to the percentage of causal interactions predicted by assigning to each test trial the most probable class, i.e. $l = \operatorname{argmax} p_i$ and its accuracy is 81%. In Table 4 the assignments are done by Equation 3 according to the cost matrix, i.e. by penalizing the false positives, and the related accuracy is 77.5%. Through their comparison, the effect of S is evident since in Table 4, false positives are strongly decreased, due to the score penalization, but to the detriment of some true positives.

		Predicted	
		1	0
True	1	79%	21%
	0	17%	83%

Table 3. Confusion matrix computed by assigning to each test trial the most probable class.

		Predicted	
		1	0
True	1	56%	44%
	0	1%	99%

Table 4. Confusion matrix in which the test trial class labels are computed by Equation 3.

Finally logistic regression was trained on **L** and tested on **C** to predict the configuration matrices of the competition. According to the number of trials in **C** and the assumptions of the generative process, the expected range of the score is $[-9000, 3000]$. The score of our submission was 1571, which reached the 2nd place in the final ranking of the competition.

5 Discussion, Conclusion and Future Works

In this paper, we proposed a new approach to detect causal interactions in multivariate time series. Specifically, we developed a classification-based causality detection method by defining a feature space based on the Granger causality concept and by exploiting the MAR model as data generator.

The proposed method was assessed by cross-validating the generated labeled data set providing promising results, as shown in Table 3 and Table 4 by means of confusion matrices. Then, the submitted solution to the Causal2014 competition was computed by a classifier trained on the generated labeled data set used the cross-validation. The achieved results, both in terms of cross-validation and competition ranking, are evidence that classification-based techniques are a feasible alternative to the signal processing methods for inferring causality between time series. And furthermore, that the defined feature space is able to well capture the causal structures among signals.

As an improvement of our approach, we are working on a tractable extension to the case of detecting causality in more than three time series.

References

1. Baccalá, L.A., Sameshima, K.: Partial directed coherence: a new concept in neural structure determination. Biological cybernetics 84(6), 463–474 (Jun 2001), http://view.ncbi.nlm.nih.gov/pubmed/11417058
2. Baccalà, L.A., Sameshima, K., Ballester, G., Do Valle, A.C., Timo-Iaria, C.: Studying the Interaction Between Brain Structures via Directed Coherence and Granger Causality. Applied Signal Processing 5, 40–48 (1998), http://www.lcs.poli.usp.br/~baccala/pdc/papers/asp.pdf
3. Brookes, M.J., Woolrich, M.W., Barnes, G.R.: Measuring functional connectivity in MEG: a multivariate approach insensitive to linear source leakage. NeuroImage 63(2), 910–920 (Nov 2012), http://view.ncbi.nlm.nih.gov/pubmed/22484306
4. Butler, S.R., Glass, A.: Asymmetries in the electroencephalogram associated with cerebral dominance. Electroencephalography and clinical neurophysiology 36(5), 481–491 (May 1974), http://view.ncbi.nlm.nih.gov/pubmed/4135345
5. Faes, L., Erla, S., Nollo, G.: Measuring Connectivity in Linear Multivariate Processes: Definitions, Interpretation, and Practical Analysis. Computational and Mathematical Methods in Medicine 2012, 1–18 (2012), http://dx.doi.org/10.1155/2012/140513
6. Freiwald, W.A., Valdes, P., Bosch, J., Biscay, R., Jimenez, J.C., Rodriguez, L.M., Rodriguez, V., Kreiter, A.K., Singer, W.: Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. Journal of neuroscience methods 94(1), 105–119 (Dec 1999), http://view.ncbi.nlm.nih.gov/pubmed/10638819
7. Friston, K.J.: Functional and effective connectivity: a review. Brain Connectivity 1(1), 13–36 (2011), http://dx.doi.org/10.1089/brain.2011.0008
8. Granger, C.W.J.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37(3), 424–438 (Aug 1969), http://dx.doi.org/10.2307/1912791
9. Hlavackovaschindler, K., Palus, M., Vejmelka, M., Bhattacharya, J.: Causality detection based on information-theoretic approaches in time series analysis. Physics Reports 441(1), 1–46 (Mar 2007), http://dx.doi.org/10.1016/j.physrep.2006.12.004
10. Horwitz, B.: The elusive concept of brain connectivity. NeuroImage 19(2 Pt 1), 466–470 (Jun 2003), http://view.ncbi.nlm.nih.gov/pubmed/12814595
11. Kamiński, M., Ding, M., Truccolo, W.A., Bressler, S.L.: Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. Biological cybernetics 85(2), 145–157 (Aug 2001), http://view.ncbi.nlm.nih.gov/pubmed/11508777
12. Kaminski, M.J., Blinowska, K.J.: A new method of the description of the information flow in the brain structures. Biological Cybernetics 65(3), 203–210 (Jul 1991), http://dx.doi.org/10.1007/bf00198091
13. Papana, A., Kugiumtzis, D., Larsson, P.G.: Reducing the bias of causality measures. Physical Review E 83(3) (Mar 2011), http://dx.doi.org/10.1103/physreve.83.036207
14. Pereda, E., Quiroga, R.Q.Q., Bhattacharya, J.: Nonlinear multivariate analysis of neurophysiological signals. Progress in neurobiology 77(1-2), 1–37 (Sep 2005), http://dx.doi.org/10.1016/j.pneurobio.2005.10.003

15. Sakkalis, V.: Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. Computers in Biology and Medicine 41(12), 1110–1117 (Dec 2011), http://dx.doi.org/10.1016/j.compbiomed.2011.06.020
16. Winterhalder, M., Schelter, B., Hesse, W., Schwab, K., Leistritz, L., Klan, D., Bauer, R., Timmer, J., Witte, H.: Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. Signal Processing 85(11), 2137–2160 (Nov 2005), http://dx.doi.org/10.1016/j.sigpro.2005.07.011