ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

Computational Intelligence approaches to causality detection

Kateřina Hlaváčková-Schindler[1] and Pablo F. Verdes[2]

1- Commission for Scientific Visualization, Austrian Academy of Sciences
Donau-City Str. 1, 1220 Vienna, Austria

2- Institute of Environmental Physics, Heidelberg Academy of Sciences
Im Neuenheimer Feld 229, 69120 Heidelberg, Germany
Current address: Novartis Pharma AG,
Klybeckstrasse 141, 4057 Basel, Switzerland

Abstract. Discovering interdependencies and causal relationships is one of the most relevant challenges raised by the information era. As more and better data become available, there is an urgent need for data-driven techniques with the capability of efficiently detecting hidden interactions. As such, this important issue is receiving increasing attention in the recent literature. The aim of the *Learning Causality* Special Session is to bring together theory-oriented and practitioners of this fascinating discipline. The main streams of causality detection by Computational Intelligence will be covered, namely, the probabilistic, information-theoretic, and Granger approaches.

1 Introduction

Learning is a constant search to relate events, particularly between actions and their consequences, that allows us to understand the world and adapt to it. This pattern, deeply installed in our behavior since childhood, has evolved into the fundamental question in human reasoning and, more generally, in the development of all kinds of science.

The study of causality has a rich history in Philosophy. Particularly influential in this field was the positive account of causality by David Hume ([1]). He asserted that our idea of causation consists of little more than expectation for certain events to result after other events that precede them. In his own words, "We have no other notion of cause and effect, but that of certain objects, which have been always conjoined together, and which in all past instances have been found inseparable. We cannot penetrate into the reason of the conjunction. We only observe the thing itself, and always find that from the constant conjunction the objects acquire a union in the imagination." Accepting, like Hume, our impossibility to understand the real nature of causation, we can nevertheless study the statistical or probabilistical characterization of such "conjunctions", including their assessment.

In this introductory paper we present the main streams to causality assessment. First we give some definitions of causality and some approaches to their detection. These approaches will be discussed in more detail in the papers contributed to our special session: a kernel-basel causal inference methodology by X. Sun *et al.*, an information-theoretic approach by N. Manyakov and M. Van Hulle, and the Granger approach to causality assessment by L. Angelini *et al.* Sections 3 and 4 were inspired by [2].

2 Causality detection and quantification by probability approaches

The following definitions are adopted from [2]. The probabilistic notion of causality described by Suppes [3] is the following: An event X is a cause to event Y if (i) X occurs before Y, (ii) the likelihood of X is non zero, and (iii) the likelihood of occurring Y given X is bigger than the likelihood of Y occurring alone.

Untill 1970 causal modeling was mostly developed within the social sciences. This was primarily due to a pioneering work by Selltiz *et al.* [4] who specified three conditions for the existence of causality:

1. There must be a concomitant covariation between X and Y.
2. There should be a temporal asymmetry or time ordering between the two observed sequences.
3. The covariance between X and Y should not disappear when the effects of any confounding variables (*i.e.*, those variables which are causally prior to both X and Y) are removed.

The first condition implies a correlation between a cause and its effect, though one should explicitly remember that a perfect correlation between two observed variables in no way implies a causal relationship. The second condition is intuitively based on the arrow of time. The third one is rather problematic because it requires to rule out all other possible causal factors. Theoretically, there is potentially an infinite number of unobserved confounding variables, yet the set of measured variables is finite, thus leading to an indeterminacy in the causal modeling approach. In order to avoid this, some structure is imposed on the adopted modeling scheme which should help defining it. The way in which the structure is imposed is crucial in defining as well as quantifying causality.

2.1 Causal inference with kernel-based statistical dependence measures

As described in more detail in the contributions to this Special Session by X. Sun *et al.*, the main idea behind the kernel-basel approach to statistical dependencies detection is to use a kernel to map the input space into an appropriate feature space, where the presence or absence of correlations is then studied. This is possible thanks to the definition of a conditional cross-covariance operator between the kernel spaces of X, Y and Z. The norms of these kernel-based operators are representative of the corresponding coupling intensities. Upon a suitable normalization, the strengths of marginal and conditional dependencies can be compared, and thereby the existence of causality relationships can be assessed.

The detection of causal structures by means of the kernel-based approach is in our Session represented by the paper *Causal inference with kernel-based statistical dependence measures* by X. Sun and D. Janzing.

3 Granger-causality based approaches

N. Wiener gave the first definition of "computationally measurable" causality: "For two simultaneously measured signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one" [5].

C. W. J. Granger introduced Wiener's concept of causality into the analysis of time series. He identified two components of the statement about causality: 1) the cause occurs before the effect, and 2) the cause contains information about the effect that is unique, and is present in no other variable. It follows that the causal variable can help to forecast the effect variable after other data had been used [6]. This restricted sense of causality, referred to as *Granger causality*, characterizes the extent to which a process X_t is leading another process Y_t, and builds upon the notion of incremental predictability. It is said that the process X_t *Granger causes* another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t. The standard test developed by Granger [7] is based on a linear regression model

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \xi_t, \tag{1}$$

where ξ_t is an uncorrelated random variable with zero mean and variance σ^2, L is the specified number of time lags, and $t = L+1, \ldots, N$. The null hypothesis that X_t does not Granger cause Y_t is supported when $b_{2k} = 0$ for $k = 1, \ldots, L$, reducing Eq. (1) to

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \tilde{\xi}_t. \tag{2}$$

This model leads to two well-known alternative test statistics, the Granger-Sargent and the Granger-Wald test [8].

This linear framework for measuring and testing causality has been widely applied in economy and finance (see Geweke [9] for a comprehensive survey of the literature), and in natural sciences. Nevertheless, the limitation of the present concept to linear relations required further generalizations.

Baek and Brock [10] and Hiemstra and Jones [11] proposed a nonlinear extension of the Granger causality concept. Their non-parametric dependence estimator is based on the so-called correlation integral, which is a probability distribution and entropy estimator developed by physicists Grassberger and Procaccia in the field of nonlinear dynamics and deterministic chaos as a characterization tool for chaotic attractors [12]. Following Hiemstra and Jones [11], Aparicio and Escribano [13] suggested an information-theoretic definition of causality that includes both linear and nonlinear dependencies.

In Physics in general, and in nonlinear dynamics in particular, a considerable interest has recently emerged in studying cooperative behavior of coupled complex systems [14]. Synchronization and related phenomena were observed not only in physical, but also in many biological systems. In such systems it is not only important to detect synchronized states, but also to identify drive-response relationships and thus causality in the evolution of the interacting (sub)systems.

Arnold *et al.* [15] proposed asymmetric dependence measures based on averaged relative distances of the (mutual) nearest neighbors. As pointed out by Quian Quiroga *et al.* [16] and Schmitz [17], these measures, however, might be influenced by the different dynamics of the individual signals, and different dimensionality of the underlying processes, rather than by coupling asymmetry.

Another nonlinear extension of the Granger causality approach was proposed by Chen *et al.* [18] using local linear predictors. An important class of nonlinear predictors are based on radial basis functions, which were used for nonlinear parametric extensions of the Granger causality concept [19, 20].

The detection of causal communities by means of the Granger causality approach is in our Session represented by the paper *Causality and communities in neural networks* by L. Angelini *et al.*

4 Causality detection by information theory

A non-parametric method for measuring causal information transfer between systems was proposed by T. Schreiber [21]. His *transfer entropy* is designed as a Kullback-Leibler distance of transition probabilities. This measure is in fact an information-theoretic functional of probability distribution functions.

Paluš *et al.* [22] studied synchronization phenomena in experimental time series also by means of information theory, concretely by mutual information (M.I.), which is a measure of general statistical dependence. M.I. and conditional M.I. can also be used for inferring causal relationships. It was shown in [22] that Schreiber's transfer entropy is equivalent to the conditional mutual information. Diks and DeGoede [23] applied a nonparametric approach to nonlinear Granger causality using the concept of correlation integrals [12] and pointed out the connection between them and information theory. Diks and Panchenko [24] critically discussed previous tests by Hiemstra and Jones [11]. As the most recent development in economics, Baghli [25] proposed information-theoretic statistics for a model-free characterization of causality, based on an evaluation of conditional entropy.

Nonlinear extensions of Granger causality based on an information-theoretic formulation have had many applications. For example, Schreiber's transfer entropy has been applied in Climatology [26, 27], Physiology [27, 28], and Neurophysiology [29]. Paluš *et al.* [22, 30] applied their conditional mutual information based measures in analyses of electroencephalograms of patients suffering from epilepsy. Paluš and Stefanovska demonstrated the suitability of the conditional mutual information approach to assess causality in the cardio-respiratory interaction [22, 31].

4.1 Coarse-grained information rates

Paluš *et al.* [22, 30] introduced several coarse-grained information rates (CIR) to describe the interaction between two given systems. All of them are based on mutual information. Since for causality detection we are interested in assessing the direction of the coupling between two systems, we will mention here only two coarse-grained information rates (for other CIR definitions, see e.g. [2]).

Let $\{x(t)\}$ be a time series considered as a realization of a stationary and ergodic stochastic process $\{X(t)\}$, $t = 1, 2, 3, \dots$ In the following we denote $x(t)$ as x and $x(t+\tau)$ as x_τ for simplicity. To define the simplest form of CIR, we compute the mutual information $I(x; x_\tau)$ for all analyzed datasets and find τ_{max} such that, for $\tau' \geq \tau_{max}$, $I(x; x_{\tau'}) \approx 0$ for all the datasets. Then we define a **norm of the mutual information**

$$||I(x; x_\tau)|| = \frac{\Delta\tau}{\tau_{max} - \tau_{min} + \Delta\tau} \sum_{\tau=\tau_{min}}^{\tau_{max}} I(x; x_\tau) \tag{3}$$

with $\tau_{min} = \Delta\tau = 1$ sample as a usual choice. The CIR h^1 is then defined as

$$h^1 = I(x, x_{\tau_0}) - ||I(x; x_\tau)||. \tag{4}$$

Since usually $\tau_0 = 0$ and $I(x; x) = H(X)$, which is given by the marginal probability distribution $p(x)$, the sole quantitative descriptor of the underlying dynamics is the M.I. norm (3). Paluš et al. [22] called this descriptor the **coarse-grained information rate** of the process $\{X(t)\}$ and denoted it by $i(X)$. Now, consider two time series $\{x(t)\}$ and $\{y(t)\}$ regarded as realizations of two processes $\{X(t)\}$ and $\{Y(t)\}$ which represent two possibly linked (sub) systems. These two systems can be characterized by their respective CIR's $i(X)$ and $i(Y)$. In order to characterize the interaction between the two systems, in analogy to the above CIR, Paluš et al. [22] defined their **mutual coarse-grained information rate** (MCIR) by

$$i(X, Y) = \frac{1}{2\tau_{max}} \sum_{\tau=-\tau_{max}}^{\tau_{max};\tau\neq 0} I(x; y_\tau). \tag{5}$$

Due to the symmetry properties of $I(x; y_\tau)$, the mutual CIR $i(X, Y)$ is symmetric, i.e., $i(X, Y) = i(Y, X)$.

Assessing the direction of coupling between the two systems, i.e., causality in their evolution, we ask the question of how the dynamics of one of the processes, say $\{X\}$, is influenced by the other process, $\{Y\}$. For a quantitative answer to this question, Paluš et al. [22] proposed to evaluate the **conditional coarse-grained information rate** CCIR $i_0(X|Y)$ of $\{X\}$ given $\{Y\}$:

$$i_0(X|Y) = \frac{1}{\tau_{max}} \sum_{\tau=1}^{\tau_{max}} I(x; x_\tau | y), \tag{6}$$

considering the usual choice $\tau_{min} = \Delta\tau = 1$ sample.

Paluš and Stefanovska [31] adapted the conditional mutual information approach [22] to the analysis of instantaneous phases of interacting oscillators and demonstrated the suitability of this approach for analyzing causality in the cardio-respiratory interaction [22]. The paper from N. Manyakov and M. Van Hulle (*Causality analysis of LFPs in micro-electrode arrays based on mutual information*) represents an application of the CCIR in our Session.

5 Conclusions

In this manuscript we have given an overview of the main lines of research in causality assessment, assuming that we observe dynamical systems by recording time-series. Given the information explosion of the recent years, the problem of causality detection bears a growing importance. We have first given a very brief account of some causality definitions drawn from Philosophy, followed by a presentation of the main streams of development within the probabilistic, statistical, Granger, and information-theoretic approaches to interdependencies detection. These methodologies are discussed in more detail in the papers contributed to the *Learning Causality* Special Session.

Acknowledgements

K. H.-S. was supported by grant of Austrian Research Fonds FWF-H-226 (2005) under Charlotte Bühler Program and partly by ASCR 1ET 100 750 401, Project Baddyr. P. F. V. acknowledges support from the Alexander von Humboldt Foundation.

References

[1] D. Hume (1740), *A Treatise of Human Nature*, Oxford University Press, Oxford, 1967.

[2] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka and J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis, *Phys. Rep.*, in press, electronically available at http://dx.doi.org/10.1016/j.physrep.2006.12.004

[3] P. Suppes, *A Probabilistic Theory of Causality*, Amsterdam: North-Holland, 1970.

[4] C. Selltiz, L.S. Wrightsman, and S.W. Cook, *Research Methods in Social Relations*, Holt, Rinehart and Winston, New York, 1959.

[5] N. Wiener, The theory of prediction. In E.F. Beckenbach, editor, *Modern Mathematics for Engineers*, McGraw-Hill, New York, 1956.

[6] C.W.J. Granger, Time series analysis, cointegration, and applications. Nobel Lecture, December 8, 2003. In Tore Frängsmyr, editor, *Les Prix Nobel.*

The Nobel Prizes 2003, [Nobel Foundation] (Stockholm), pages 360-366, 2004.

[7] C.W.J. Granger, Investigating causal relations by econometric and cross-spectral methods, *Econometrica*, 37:424-438, 1969.

[8] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, 1972, pages 946-949.

[9] J. Geweke, Inference and causality in economic time series models. In Z. Griliches and M.D. Intriligator, editors, *Handbook of Econometrics*, vol. 2, pages 1101-1144, North-Holland, 1984.

[10] E.G. Baek and W.A. Brock, A general test for nonlinear Granger causality: Bivariate model, *Working paper*, Iowa State University and University of Wisconsin, Madison, 1992.

[11] C. Hiemstra and J.D. Jones, Testing for linear and nonlinear Granger causality in the stock pricevolume relation, *Journal of Finance*, 49:1639-1664, 1994.

[12] P. Grassberger and I. Procaccia, Measuring of strangeness of strange attractors, *Physica D*, 9:189-208, 1983.

[13] F.M. Aparicio and A. Escribano, Information-theoretic analysis of serial dependence and cointegration, *Studies in Nonlinear Dynamics and Econometrics*, 3:119-140, 1998.

[14] S. Boccaletti, J. Kurths, G. Osipov, D.L. Valladares and C.S. Zhou, The synchronization of chaotic systems, *Phys. Rep.*, 366:1-101, 2002.

[15] B.C. Arnold, *Pareto Distributions*, International Co-Operative Publishing House, Burtonsvile, MD, 1985.

[16] R. Quian Quiroga, J. Arnhold and P. Grassberger, Learning driver-response relationships from synchronization patterns, *Phys. Rev. E*, 61:5142-5148, 2000.

[17] A. Schmitz, Measuring statistical dependence and coupling of subsystems, *Phys. Rev. E*, 62:7508-7511, 2000.

[18] Y. Chen, G. Rangarajan, J. Feng and M. Ding, Analyzing multiple nonlinear time series with extended Granger causality, *Phys. Lett. A*, 324:26-35, 2004.

[19] N. Ancona, D. Marinazzo and S. Stramaglia, Radial basis function approach to nonlinear Granger causality of time series, *Phys. Rev. E*, 70:056221, 2004.

[20] D. Marinazzo, M. Pellicoro and S. Stramaglia, Nonlinear parametric model for Granger causality of time series, *Phys. Rev. E*, 73:066216, 2006.

[21] T. Schreiber, Measuring information transfer, *Phys. Rev. Lett.*, 85:461-464, 2000.

[22] M. Paluš, V. Komárek, Z. Hrnčíř and K. Štěrbová, Synchronization as adjustment of information rates: Detection from bivariate time series, *Phys. Rev. E*, 63:046211 (2001).

[23] C. Diks and J. DeGoede, A general nonparametric bootstrap test for Granger causality. In Broer, Krauskopf and Vegter, editors, *Global Analysis of Dynamical Systems*, Chapter 16, pages 391-403, 2001.

[24] C. Diks and V. Panchenko, A note on the Hiemstra-Jones test for Granger non-causality, *Studies in Nonlinear Dynamics and Econometrics*, in press, 2006.

[25] M. Baghli, A model-free characterization of causality, *Econ. Lett.*, 91:380-388, 2006.

[26] I.I. Mokhov and D.A. Smirnov, El Nino-Southern Oscillation drives North Atlantic Oscillation as revealed with nonlinear techniques from climatic indices, *Geophys. Res. Lett.*, 33:L03708, 2006.

[27] P. F. Verdes, Assessing causality from multivariate time series, *Phys. Rev. E*, 72:026222, 2005.

[28] T. Katura, N. Tanaka, A. Obata, H. Sato and A. Maki, Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics, *Neuroimage*, 31:1592-1600, 2006.

[29] M. Chávez, J. Martinerie and M. Le Van Quyen, Statistical assessment of nonlinear causality: Application to epileptic EEG signals, *J. of Neurosci. Meth.*, 124:113-128, 2003.

[30] M. Paluš, V. Komárek, T. Procházka, Z. Hrnčíř and K. Štěrbová, Synchronization and information flow in EEG of epileptic patients, *IEEE Engineering in Medicine and Biology Magazine*, 20:65-71, 2001.

[31] M. Paluš, A. Stefanovska, Direction of coupling from phases of interacting oscillators: An information-theoretic approach, *Phys. Rev. E*, 67:055201(R), 2003.