Methods for Reconstructing Causal Networks from Observed Time-Series: Granger-Causality, Transfer Entropy, and Convergent Cross-Mapping

Fermín Moscoso del Prado Martín (fmoscoso@linguistics.ucsb.edu)
Department of Linguistics, South Hall 3521, UCSB
Santa Barbara, CA 93106 USA

Keywords: convergent cross-mapping; development, dynamical systems; Granger-causality; information theory; phase-space reconstruction; time series

Objectives and Scope

A major question that arises in many areas of Cognitive Science is the need to distinguish true causal connections between variables from mere correlations. The most common way of addressing this distinction is the design of well-controlled experiments. However, in many situations, it is extremely difficult –or even outright impossible– to perform such experiments. Researchers are then forced to rely on correlational data in order to make causal inferences. This situation is especially common when one needs to analyze longitudinal data corresponding to historical time-series, symbolic sequences, or developmental data. These inferences are often very problematic. From the correlations alone it is difficult to determine the direction of the causal arrow linking two variables. Worse even, the lack of controls of observational data entail that correlations found between two variables need not reflect any causal connection between them. The possibility always remains that some third variable which the researchers were not able to measure, or were actually unaware of, is the actually driver for both measured variables, giving rise to the mirage of a direct relationship between them.

In recent years, it has been shown that, under particular circumstances, one can use correlational information for making sound causal inferences (cf., Pearl, 2000). In this tutorial I will provide a hands-on introduction to the use of modern causality techniques for the analysis of observational time series. I will cover causality analyses for three types of time-series that are often encountered in Cognitive Science research:

- For numerical time-series of a predominantly stochastic nature I will discuss how to perform *Granger-Causality* (Granger, 1981) analyses used by econometricians, using the methodology introduced by Toda and Yamamoto (1995).
- For symbolic stochastic time-series, I will introduce the *Transfer Entropy* measure developed in Physics (Schreiber, 2000).
- Finally, for numerical series that can be shown to have a predominantly deterministic (even if possibly chaotic) nature, I will discuss *Convergent Cross-Mapping* (Clark et al., 2015; Sugihara et al., 2012), a very powerful technique recently developed in the field of ecology, that relies on the Theory of Dynamical Systems to make causal inferences.

Finally, I will demonstrate how to use each of these techniques for reconstructing networks of causal relations between large sets of variables.

Overview of Causality Methods for Time Series

Granger-Causality

Granger-Causality (cf. Granger, 1981) is a powerful technique developed in Econometrics for assessing whether one time sequence can be said to be the cause of another one (or viceversa). If x and y are stationary time sequences on discrete time (τ), in order to test whether x Granger-causes y, one tests whether the past of x is able to predict the future of y, over and above the predictive power that can be obtained from y's own past. Between just two variables, this is assessed using Autoregressive Models. When more than two variables are involved this is naturally extended by using Vector Autoregressive Models.

This technique is useful for the analysis of numerical time series data that are generated by a process whose nature is predominantly stochastic, which is typical of data resulting from the aggregation of multiple sources. One important requirement of the Granger-causality method is that it is limited to stationary time series. This property is sometimes difficult to guarantee in the types of series that one typically encounters in Cognitive Science, which tend to exhibit a certain degree of co-integration. This limitation can, however, be addressed by using the methodology introduced by Toda and Yamamoto (1995), which I will introduce in the tutorial.

Transfer Entropy

Often, in Cognitive Science, researchers need to analyze sequences of discrete symbols, as could for instance be the sounds uttered by a developing child. Schreiber (2000) extended the main idea of Granger-Causality to symbolic stochastic processes. Instead of analyzing correlations between variables, one moves into using mutual informations, in the sense of Shannon (1948). Even when one's data are actually of a numerical nature, it can be actually beneficial to analyze them in symbolic terms, as the mutual informations are capable of capturing non-linear relations that could be missed by linear correlation-based methods (see, Hlaváčcova-Schindler, Paluš, Vejmelka, & Bhattacharya, 2007, for a review). This later usage will be demonstrated in the tutorial using a dataset from human speech.

Convergent Cross-Mapping

The Granger-Causality and Transfer-Entropy approaches outlined above are suitable only for stochastic systems. In some

Cognitive Science domains, especially those dealing with longitudinal developmental data, one also encounters numerical data that can be argued to originate from a predominantly deterministic Dynamical System. Such cases can be modelled explicitly using systems of coupled differential equations. However, in many cases, only a few of the variables that relevant for the system are available to the researchers (the others being unmeasurable or plainly unknown). However, Takens' Theorem (Takens, 1981) states that the crucial properties of a dynamical system's attractor can be successfully recovered using a single one of its variables, in what is known as *Phase-State Reconstruction*. Using this fact, Sugihara et al. (2012) developed the Convergent-Cross Mapping (CCM) technique, which enables recovering the direction of causality between any two time sequences generated by the same dynamical system. Importantly, and in contrast with the methods discussed above, CCM is also capable of distinguishing the case when two correlated variables are not actually causally related, but rather they are both driven by a third unstudied variable. A limitation of CCM is that it requires relatively long time series, which are often unavailable in many actual research problems. Clark et al. (2015) extended CCM to allow combining multiple short time series originating from similar processes (i.e., as if considering random effects in a regression model), introducing the "multispatial" variant of the CCM method. I will demonstrate how the multispatial CCM analyses can be performed.

Format and Organization

This tutorial is designed to cover half a day (three hours) broken into two sections of 1.5 hours each. The first session in the tutorial will discuss the theoretical basis, conditions of applicability, advantages, drawbacks of each of the three causal analysis methods. The second session will be hands-on, guiding attendants on how to perform each of these analyses, together with the necessary diagnostics, using the R statistical software. For this, I will make use of previously published datasets, covering three different time-scales: historical (Moscoso del Prado Martín, 2014), developmental (Irvin, Spokoyny, & Moscoso del Prado Martín, 2016), and the time-scale of a typical behavioral experiment (Moscoso del Prado Martín, 2011).

Target Audience

The tutorial is aimed at advanced graduate students, postdocs, and senior researchers wishing to use this type of causal analyses in this research. A familiarity with basic statistics and with programming (preferably using R) will be necessary to be able to follow the theoretical arguments, and to be able to perform the analyses.

Tutor Information

Fermín Moscoso del Prado Martín is assistant professor of Linguistics at the University of California, Santa Barbara. Previously he has held positions at the Max Planck Institute for Psycholinguistics, the Medical Research Council – Cognition and Brain Sciences Unit, and at the Cognitive Psychology Laboratory of the French National Research Center. He holds an MEng in Computer Science by the Technical University of Madrid, and a PhD in Linguistics by the University of Nijmegen, where he was a student of Prof. R. H. Baayen. Over the last decade he has published multiple papers combining information-theoretical methods, computational modeling, corpus analyses, and psycholinguistic experiments.

References

Clark, A. T., Ye, H., Isbell, F., Deyle, E. R., Cowles, J., Tilman, G. D., & Sugihara, G. (2015). Spatial convergent cross mapping to detect causal relationships from short time series. *Ecolology*, *96*, 1174–1181.

Granger, C. W. J. (1981). Testing for causality. *Journal of Economic Dynamics and Control*, *2*, 329–352.

Hlaváčcova-Schindler, K., Paluš, M., Vejmelka, M., & Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, *441*, 1–46.

Irvin, J., Spokoyny, D., & Moscoso del Prado Martín, F. (2016). Dynamical systems modeling of the child–mother dyad: Causality between child-directed language complexity and language development. In A. Papafragou, J. Trueswell, D. Grodner, & D. Mirman (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Moscoso del Prado Martín, F. (2014). Grammatical change begins within the word: Causal modeling of the coevolution of Icelandic morphology and syntax. In P. Bello, M. Guarini, M. McShane, & B. Scasselatti (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2657–2662). Austin, TX: Cognitive Science Society.

Moscoso del Prado Martín, F. (2011). Causality, criticality, and reading words: distinct sources of fractal scaling in behavioral sequences. *Cognitive Science*, *35*, 785–837.

Pearl, J. (2000). *Causality: models, Reasoning and Inference*. Cambridge, England: Cambridge University Press.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, *85*, 461–464.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, *27*, 379–423, 623–656.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E. R., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, *338*, 496–500.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. A. Rand & L.-S. Young (Eds.), *Dynamical Systems and Turbulence* (pp. 366–381). Berlin, Germany: Springer Verlag.

Toda, H., & Yamamoto. (1995). Statistical inference in Vector Autoregressions with possible integrated sources. *Journal of Econometrics*, *66*, 225-250.