CHAOS 2009
2nd Chaotic Modeling and Simulation International Conference
June 1 - 5, 2009 Chania Crete Greece www.chaos2009.net

Rendering Statistical Significance of Information Flow Measures

Angeliki Papana, Dimitris Kugiumtzis
Aristotle University of Thessaloniki
agpapana@gen.auth.gr, dkugiu@gen.auth.gr

The estimation of the information flow among variables or sub-systems observed by bivariate or multivariate time series is essential in order to understand the dynamics of the underlying system. It is a subject of active research in many areas ranging from climate processes and electric circuits to finance and neurology. The nature of the inter-dependencies is usually nonlinear and asymmetric deeming linear symmetric measures ineffective and linear causality measures often inadequate. Constraining the problem to bivariate time series, it is essential to assess whether they originate from coupled or decoupled systems, detect the hidden causal dependencies between them and distinguish the driver and responder.

The standard approach for inferring the direction of interactions has been the Granger causality, testing whether one time series is useful in forecasting another. Different linear measures in the time and frequency domain have been developed, such as Directed Transfer Function, Directed Coherence and Partial Directed Coherence. In the recent years Granger causality was extended using approaches based on phase synchronization, nonlinear dynamics evolution and information flow. A number of measures of directed interaction have been proposed but the estimation from time series cautions for a number of complications as indicated in some very recent comparative studies. Aspects influencing the robust estimation of the causal measures are the finite time series length and the presence of noise. The estimation of the measures gets even more complicated when the possibly coupled systems are non-identical or when they have different complexity.

A thorough investigation for the validity and usefulness of the coupling measures should start with a test of significance, i.e. a measure should not identify coupling (or interaction) in any direction when it is not present. The second property of a coupling measure is its power, i.e. how sensitive the measure is in detecting interaction and identifying its direction. To quantify these two statistical properties we use the concept of surrogate data testing and modify the coupling measures in order to attain proper significance. Though this approach can be implemented on a variety of interaction measures we focus here on information-based measures for detecting asymmetric causal interdependencies between two time series. Specifically, we consider the transfer entropy (TE) (Schreiber, 2000) and the symbolic transfer entropy (STE) (Staniek, 2008). Computing TE on the set of the two original bivariate time series and on surrogate sets, where the assumed driving time series is randomly shuffled, one can derive the so-called effective transfer entropy (ETE) (Marschinski, 2002). Thus ETE attempts to remove bias in the estimation of transfer entropy that is not related to the interaction. However, this approach does not fully preserve the dynamics of each sub-system and our simulations have shown that it still lacks significance. We therefore propose a correction to the transfer entropy measure, termed corrected transfer entropy (CTE), using reshuffled points from the driving time series. In the simulation study, we consider also the effective symbolic transfer entropy (ESTE) and the corrected symbolic transfer entropy (CSTE) defined similarly.

To assess the significance and power of the TE, ETE, STE, CTE, ESTE and CSTE measures we compute them on multiple realizations of uncoupled and coupled nonlinear systems (maps and flows), allowing also for a range of the coupling strength. The performance of the measures is examined for variations of the time series length, level of additive noise and embedding dimensions of the two reconstructed state spaces (for driver and responder). Preliminary results have shown that all measures are not crucially affected by the time series length or the addition of noise (we examined noise levels up to 20% of the standard deviation of the time series). However, the selection of the embedding dimensions is important both for the significance and the power of the measures. The proposed CTE and CSTE turn out to attain best significance results and are at the zero level in the case of no causal effect. This is an important statistical feature of CTE and CSTE with regard to applications as it leaves no doubt for actual causal relationship when it is estimated.

References

Schreiber T., 2000. Measuring Information Transfer. *Physical Review Letters* **85** (2), 461-464.

Marschinski R. and Kantz, H., 2002. Analysing the Information flow between financial time series. *European Physical Journal* B **30**, 275–281.

Staniek M. and Lehnertz, K., 2008. Symbolic Transfer Entropy. *Physical Review Letters* **100** (15), 158101.