Measuring information transport

Thomas Schreiber
Max Planck Institut für Physik komplexer Systeme
Nöthnitzer Straße 38, 01187 Dresden
schreibe@mpipks-dresden.mpg.de

Abstract

Many dynamical phenomena in nature as well as in technological applications involve nonlinear behaviour as an essential ingredient. This fact is reflected by recent developments in time series analysis. Specifically, I will report on one example of a novel method that can take nonlinearity into account without assuming determinism. Qualitative as well as quantitative insights can be gained on the structure of a system from an understanding of the information flow between components.

Keywords

Time series, information theory, statistical analysis

1 Introduction

Measuring time series data has become much easier in the last years. Using off-the-shelf hardware including an inexpensive data acquisition card, fluctuations, oscillations, and variations can be recorded of any quantity a sensor exists for. On the one hand, the vast amount of data available makes it necessary to reduce the information to a manageable amount. On the other hand, we can now resolve time scales which make a detailed study of the dynamics of systems possible for which this has not been feasible so far.

Many people see the ultimate goal of analysing such data in establishing a detailed model, say, a set of differential equations, that describes the dynamics of the system and that can be related to known system parameters. In an engineering context, however, this is not the only possible point of view. As an extreme example, consider the dynamics of a fluid flowing through a network of pipes and valves. Despite some mathematical caveats, few people doubt that the Navier-Stokes equation, plus mass conservation, together with a set of boundary conditions, constitute an adequate description of the fluid motion. For many practical situations, however, in particular those involving turbulent flows, this model is completely useless. Therefore, heuristic formulae for specific problems are widely used instead. For most applications in manufacturing, a complete understanding of the dynamics of a device, or a detailed model, will be too much to ask at this stage. Nevertheless, a lot can be gained by asking specific questions and demanding a heuristic but useful answer. Examples for such questions include establishing critical observables for process monitoring, deducing structural information from data, prediction of failure or other transitions, among others.

Concerning recent approaches to the analysis of time series from allegedly nonlinear sources, several general overview articles (see [1] for a recent one), as well as a few text books [2, 3] and software tools [5, 6]. More classical references within the statistics literature are given by Refs. [4]. I will here report on an example for the use of novel time series techniques to extract qualitative, structural information. This example is motivated and applied in the field of clinical medicine, where hopes have been maintained for a longer time to extract clinically relevant information from time series recordings.

2 Some information theory

When looking for a framework that allows for nonlinearity and stochasticity at the same time, information theory is one of the most natural candidates. Consider the decay in time of the knowledge we have on the current state of a system. Stochastic fluctuations and dynamical instability are two mechanisms that produce new information we need to pick up by observing the system continuously, but the concept of informnation is not restricted to these special cases. A mathematical theory for the information content of a source of messages, or a transmission channel was formulated around the middle of the 20 th century by Shannon, Kolmogorov, and a number of other main figures. Original material can be found in Refs. [7, 8], a good text book is Ref. [9]. Shannon's classical text is surprisingly readable and a good paperback edition exists.

Let us briefly recall the most basic concepts. Suppose a discrete variable I is drawn according to a probability distribution $p(i)$. The average number of bits needed to optimally encode each outcome is given by the Shannon entropy [7]

$$H_I = -\sum_i p(i) \log_2 p(i) \,, \tag{1}$$

where the sum extends over all states i the process can assume. The base of the logarithm determines the units, in this case *bits*.

Consider a system that may be approximated by a stationary Markov process of order k, that is, the conditional probability to find I in state i_{n+1} at time $n+1$ depends at most on the k previous states:

$$p(i_{n+1}|i_n, \ldots, i_{n-k+1},\, i_{n-k}) = p(i_{n+1}|i_n, \ldots, i_{n-k+1})\,. \tag{2}$$

The average number of bits needed to encode one *additional* state of the system if all previous states are known is given by the *entropy rate*

$$h_I = -\sum p(i_{n+1},\, i_n, \ldots, i_{n-k+1})\, \log\, p(i_{n+1}|i_n, \ldots, i_{n-k+1})\,. \tag{3}$$

While in the case of a discrete variable (a symbol sequence), very powerful results are available, serious problems arise in the case of continuous variables, which is the case we are usually interested in. Since the amount of information stored in a single real number diverges with the number of digits specified, some coarse graining is indispensable. If I is obtained by coarse graining a continuous system X at resolution r, the entropy $H_X(r)$ and entropy rate $h_X(r)$ will depend on the partitioning and in general diverge like $-\log r$ when $r \to 0$. However, for the special case of a deterministic dynamical system, $\lim_{r\to 0} h_X(r) = h_{KS}$ may exist and is then called the Kolmogorov–Sinai entropy [8]. (For non-Markov systems, also the limit $k \to \infty$ needs to be taken.)

Nowhere have we assumed anything about the nature, deterministic, noisy, or whatever, of the system. For continuous system states, the price is that results will depend on the particular method of coarse graining used. From a fundamental point of view, invariance has been a highly desired property, whence the huge literature on invariant quantities like dimensions and Lyapunov exponents. For practical applications, we have other priorities.

3 Identifying relevant couplings

In this section, we will use information theory to identify relevant couplings within a composed system. Consider the simple setup of Fig. 1 and assume that the two systems, say, tool and workpiece, are physically coupled, for example through mechanical contact. We may ask if both coupling directions are really relevant for the dynamics of the joint system. Can a fluctuation in System I cause a disturbance in System J and vice versa, or will one or both of the perturbations be dampened out? Of course, we have the whole machinery of control theory to answer such

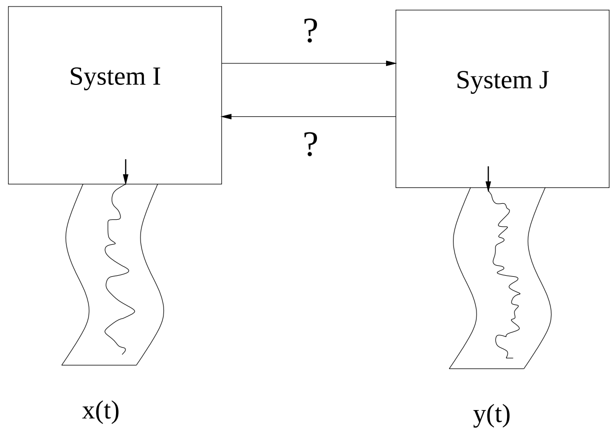

Figure 1: Copuled systems may or may not exchange information.

questions — if we have detailed knowledge of the systems and their coupling. We are here concerned with the realistic case that such knowledge is not available and we want to deduce similar information from time series observations.

Several recent studies [10] of the nonlinear coherence of signals, most notably in physiological systems, address such questions. Here we want to give a solution based on information theory, that is, we ask at what rate information is carried from I to J, and from J to I.

It is often stated incorrectly that the *mutual information* measures information transport. Rather, it quantifies *any* common information, be it due to common influences, the common history, or actual information flow between two coupled systems. If we have two processes I and J with joint probability $p_{IJ}(i,j)$, the deviation from independence of the two processes, i.e. the difference between $p_{IJ}(i,j)$ and $p_I(i)\,p_J(j)$, is given as a special case of the Kullback entropy by the *mutual information*

$$M_{IJ} = \sum p(i,j)\,\log\frac{p(i,j)}{p(i)\;p(j)}\,. \tag{4}$$

It is symmetric and thus lacks a directional sense, which can be introduced in a somewhat ad-hoc way by a time lag in either one of the variables.

Expression (4) reflects only dependencies contained in the static probabilities. The dynamics of the processes and their interrelation is, however, contained in the transition probabilities. Thus, what we have to do is to generalize the entropy rate, rather than Shannon entropy, to more than one system. This can be achieved by measuring the deviation from the generalized Markov property

$$p(i_{n+1}|i_n,\ldots,i_{n-k+1}) = p(i_{n+1}|i_n,\ldots,i_{n-k+1},\,j_n) \tag{5}$$

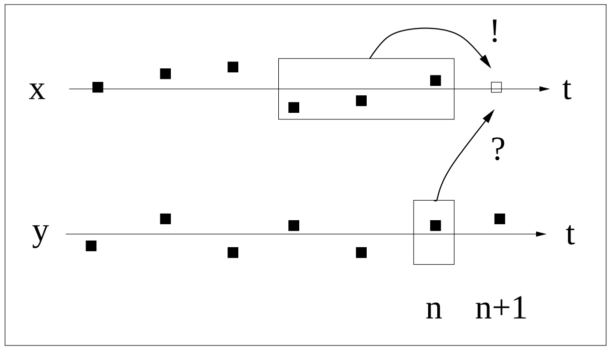

Figure 2: Scheme underlying transfer entropy.

by a suitable Kullback entropy, the *transfer entropy* [11]

$$T_{J\to I} = \sum p(i_{n+1},\, i_n, \ldots, i_{n-k+1},\, j_n) \,\log \frac{p(i_{n+1}|i_n, \ldots, i_{n-k+1},\, j_n)}{p(i_{n+1}|i_n, \ldots, i_{n-k+1})}\,. \quad (6)$$

The heuristic meaning of the two preceeding equations is the following. In the absence of information flow from J to I, the state of J has no influence on the transition probabilities on system I. In that case, the argument of the $\log$ cancels and $T_{J\to I} = 0$. If there is any kind of information flow, Eq. (5) is violated and $T_{J\to I} > 0$. This concept is illustrated in Fig. 2. $T_{J\to I}$ is now explicitly non-symmetric since it measures the degree of dependence of I on J and not vice versa.

For coarse grained states of continuous systems (X, Y), the limit of fine partitions $\lim_{r\to 0} T_{Y\to X}(r)$ exists and is independent of the partition, except for the case of deterministic coupling, when $T_{Y\to X}(r)$ diverges as $r \to 0$. In this respect, transfer entropy behaves like mutual information. For practical applications, the limit $r \to 0$ is not obtainable and has to be replaced appropriately. Either one can study transfer entropy as a function of the resolution, or one can fix a resolution for the scope of a study. Furthermore, there are several methods of coarse graining. A partition consisting of a fixed mesh of boxes is only suitable when data can be produced with little effort. For time series applications, an alternative implementation using generalized correlation integrals [12] is preferable. Since these technical issues are the same as in many nonlinear time series methods, for example estimates of the correlation dimension, the reader is referred to the discussion in the literature [3].

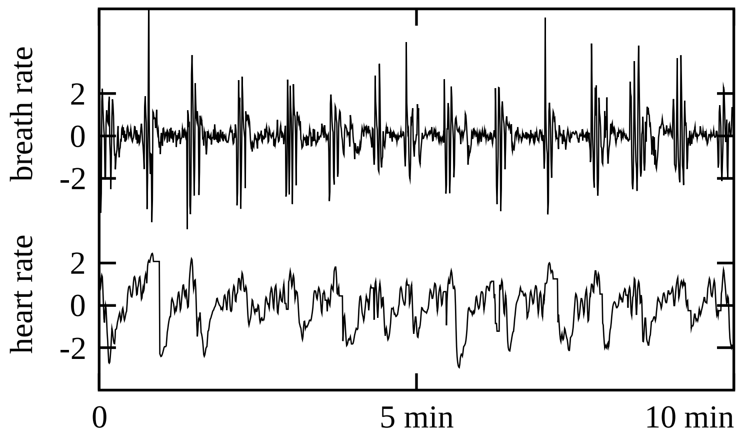

Figure 3: Bi-variate time series of the breath rate (upper) and instantaneous heart rate (lower) of a sleeping human suffering from sleep apnoea (here: periodic breathing) . The data is sampled at 2 Hz. Both traces have been normalized to zero mean and unit variance.

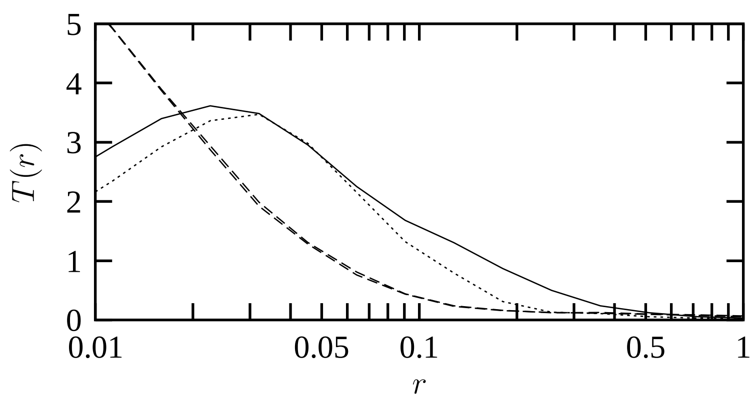

Figure 4: Depending on the resolution r, both directions of the transfer entropy, $T(\text{heart} \to \text{breath})$ (solid line), $T(\text{breath} \to \text{heart})$ (dotted line), and the time delayed mutual information $M(\tau = 0.5 \text{ s})$ (directions indistinguishable, dashed lines) are shown for the patho-physiological time series shown in Fig. 3.

4 An example from physiology

Let us illustrate the use of transfer entropy with a bi-variate time series (see Fig. 3) of the breath rate and instantaneous heart rate of a sleeping human suffering from sleep apnoea (samples 2350–3550 of data set B of the Santa Fe Institute time series contest held in 1991 [13]). Figure 4 shows that time delayed mutual information is almost symmetric between both series. The transfer entropy also finds information transport in both directions but indicates a stronger flow of information from the heart rate to the breath rate than vice versa over a significant range of length scales r. Note that for small r, the curves deflect down to zero due to the finite sample size. In this particular pathological state of periodic breathing, the dominant direction of information flow from the heart to the breath signal is consistent with the observation that the patient breathes in bursts which seem to occur whenever the heart rate crosses some threshold.

It is worth noting that the two channels studied here differ in their individual information contents. In such a situation, unless we find zero information transfer in one of the directions, too rash conclusions about the nature of the interaction have to be avoided. Different rates of information production and transport between length scales will naturally cause some asymmetry in the rate of information transfer, as measured by T. Reducing the analysis to the identification of a "drive" and a "response" may not be useful and could even be misleading. Also, note that the findings could also be explained with a coupling of both signals to a common external trigger.

In conclusion, we may use information theory to obtain qualitative insights into the structure and dynamics of an observed system. As an example, the *transfer entropy* is able to detect the directed exchange of information between two systems. Analysing the flow of communication between parts of a system can be an important first step in understanding and modelling the dynamics of the system.

References

1. T. Schreiber. Interdisciplinary application of nonlinear time series methods. *Phys. Rep.*, 308:2, 1999.

2. H. D. I. Abarbanel. *Analysis of observed chaotic data.* Springer, New York, 1996; C. Diks. *Nonlinear time series analysis.* World Scientific, River Edge, NJ, 1999.

3. H. Kantz and T. Schreiber. *Nonlinear time series analysis.* Cambridge University Press, Cambridge, 1997.

4. H. Tong. *Non-linear time series analysis.* Oxford University Press, Oxford, 1990; M. B. Priestley. *Non-linear and non-stationary time series analysis.* Academic Press, London, 1988.

5. C. Merkwirth, U. Parlitz, and W. Lauterborn. TSTOOL – A software package for nonlinear time series analysis. In J. A. Suykens and J. Vandewalle, editors, *Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling*, 1998. (`http://www.physik3.gwdg.de/tstool/index.html`)

6. R. Hegger, H. Kantz, and T. Schreiber. Practical implementation of nonlinear time series methods: The TISEAN package. *CHAOS*, 9:413, 1999. (`http://www.mpipks-dresden.mpg.de/~tisean`)

7. C. E. Shannon and W. Weaver. *The Mathematical Theory of Information.* University of Illinois Press, Urbana, IL, 1949.

8. A. N. Kolmogorov. *Information theory and the theory of algorithms.* Selected works, Vol. 3, Kluwer, Dordrecht, 1993.

9. P. Billingsley. *Ergodic Theory and Information.* Wiley, New York, 1965.

10. J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger. A Robust Method for Detecting Interdependences: Application to Intracranially Recorded EEG. *Physica D*, 134:419, 1999.

11. T. Schreiber. Measuring Information Transfer. *Phys. Rev. Lett.*, 85:461, 2000.

12. D. Prichard and J. Theiler. Generalized redundancies for time series analysis. *Physica D*, 84:476, 1995.

13. D. R. Rigney, A. L. Goldberger, W. Ocasio, Y. Ichimaru, G. B. Moody, and R. Mark. Multi-channel physiological data: Description and analysis. in A. S. Weigend and N. A. Gershenfeld, editors. *Time series prediction: Forecasting the future and understanding the past.* Santa Fe Institute Studies in the Science of Complexity, Proc. Vol. XV, Addison-Wesley, Reading, MA, 1993.