Proceeding Paper

Approximate Continuous Time Measures of Information Movement in Complex Extended Networks [†]

Paul E. Rapp [1,*], Christopher J. Cellucci [2], Claire E. Gilpin [3] and David M. Darmon [2]

[1] Uniformed Services University
[2] Aquinas, LLC; cellucci@gmail.com (C.J.C.); david.m.darmon@gmail.com (D.M.D.)
[3] Gilpin-Phillips, LLC; cegilpin21@gmail.com (C.E.G.)
[*] Correspondence: paul.rapp@usuhs.edu(P.E.R.)
[†] Presented at the 1st International Online Conference on Mathematics and Applications; Available online: https://iocma2023.sciforum.net/.

Abstract: The differential entropy of a continuous waveform is defined over the time period spanned by the recording. The time-dependent information dynamics of the observed process are not accessible to this measure. We review here the construction and validation of approximate time-dependent measures of information dynamics.

Keywords: local entropy rate, specific entropy rate, differential entropy, network dynamics

1. Introduction

The object of this investigation is to review and extend methods for approximating time-dependent information dynamics and then proceeds to their use in the dynamical characterization of information movement in networks. The investigation is constructed in four steps.

1. Local entropy rate and specific entropy rate: approximate continuous time measures of information generation
2. Specific transfer entropy: an approximate continuous time measure of information movement
3. Time dependent measures of information movement in networks
4. Hierarchical transition chronometries in information networks

2. Local Entropy Rate and Specific Entropy Rate

Consider a time series $\{x_1, \cdots x_N\}$. Time t_i is the time when x_i was observed. Local entropy rate, $H^L(t_i) = H^L(x_i | x_{i-p}^{i-1})$, is a time-dependent, locally determined conditional entropy where the entropy of an observed variable is conditioned on its own past. It quantifies the surprise on observing x_t given the preceding values $x_{t-p}, \cdots x_{t-1}$. $H^L(t_i) = H^L(x_i | x_{i-p}^{i-1})$, can be interpreted as the rate of information generation at time t (Lizier, et al., 2008; Lizier, 2014). It is a single number that is a function of $p + 1$ inputs.

$$H^L(t_i) = H^L(x_i | x_{i-p}^{i-1}) = -\log f_{-i}(x_i | x_{i-p}^{i-1})$$

The essential step in the analysis is to recognize the relationship between a conditional density function, the joint distribution, and the marginal distribution. Specifically,

$$H^L(t_i) = H^L\left(x_i \middle| x_{i-p}^{i-1}\right) = -\log f_{-i}\left(x_i \middle| x_{i-p}^{i-1}\right) = -\log\left\{\frac{f_{-i}(x_i, x_{i-p}^{i-1})}{f_{-i}(x_{i-p}^{i-1})}\right\}$$

The subscript $-i$ is used to indicate that the value x_i is not used when evaluating the local entropy rate at time t_i. Density functions can be approximated by kernel density estimation (Silverman, 1986; Scott, 1992) or by k-th nearest neighbor estimation (Biau and Devroye, 2015). Autoregressive order p can be determined by minimization of a negative log predictive likelihood.

Local entropy rate is exquisitely sensitive to noise. This is consistent with its definition. Noise at time t_i reflects information generation at time t_i and is reflected in the value of local entropy rate. Noise is usually not, however, particularly interesting. This difficulty is addressed with specific entropy rate (Darmon, 2016). Like local entropy rate, the specific entropy rate is a time-dependent locally determined measure. In this case, specific entropy rate gives a quantitative estimate of uncertainty at time t_i. Specific entropy rate differs from local entropy rate in being an average of local entropy rate over a defined set. It is therefore intermediate to local entropy rate where no averaging occurs and total entropy rate which is the average of all available values of local entropy rate.

A summary is given in Gilpin, et al. (2018): "Local and specific entropy rates quantify complementary properties of a dynamical system. If we have a time series representing an observable, the local entropy rate (LER) represents the statistical surprise of seeing an already observed local future given specific past. It also can be thought of as the rate of information generation at a given time point. The specific entropy rate (SER) represents the statistical uncertainty in an as yet unobserved future given a specific past. The LER, in isolation, is a retrospective measure, and the SER is a prospective measure, and each yields distinct information about the behavior of the dynamical system." The notation $H^S(t_i) = H^S(x_{i-p}^{i-1})$ is used to emphasize the fact that specific entropy rate is determined by values in the past. The present, x_i, is not an argument of the specific entropy rate.

An operational requirement is specification of the procedure used to calculate specific entropy rate from a subset of local entropy rate values. An unnecessarily simplistic response would be a moving average. As an alternative, note that $(x_{i-p}, \cdots x_{i-1})$ is a point in $\mathbb{R}^p$, a p-dimensional space. Averaging local entropy rate values from points close to $(x_{i-p}, \cdots x_{i-1})$ exploits the possibility, suggested by the Takens embedding theorem (Takens, 1981), that these points are subject to a dynamical process similar to that in operation at point $(x_{i-p}, \cdots x_{i-1})$. Ideally, one is examining the same underlying dynamics at each point in the average and noise will be reduced in the averaging process. We refer to this as dynamical adjacency in contrast with the temporal adjacency of a moving average. The estimator for specific entropy rate is:

$$\hat{H}^S(t_i) = H^S\left(x_{i-p}^{i-1}\right) = \frac{1}{k^*} \sum_{j \in N(x_{i-p}^{i-1})} \hat{H}^L(x_j | x_{j-p}^{j-1})$$

$N\left(x_{i-p}^{i-1}\right)$ is the set of the indices of the k^* points closest to $(x_{i-p}, \cdots x_{i-1}) \in \mathbb{R}^p$. An appropriate value of k^* can be determined by a split-half minimization argument.

3. Specific Transfer Entropy

In the simplest case, consider two sources producing simultaneously measured time series. There is a long history of procedures that can be used to quantify correlation between time series. These correlation measures do not, however, quantify information movement between the two sources. In the "Treatise Concerning Human Knowledge," Bishop Berkeley identified prediction rather than correlation as the essential criterion for causation (Berkeley, 1710). This is consistent with Wiener's operationalization of causation. To paraphrase Wiener (Wiener, 1956), if measuring variable X improves the prediction of variable Y, then Y is, in this very limited operational sense of causation, causally dependent on X. Where we recognize that the relationship is not necessarily unidirectional.

Granger (Granger, 1969) introduced a measure of causal relations based on marginal predictability that has been used extensively in econometrics. It is recognized, however, that Granger causality can fail to detect nonlinear relationships. This motivated the introduction of alternative measures, most notably transfer entropy (Schreiber, 2000). As an example, Amblard and Michel (2012) give an example system that has zero Granger causality in a system that detects information movement with transfer entropy.

Y is said to transfer information to X if the uncertainty in X alone exceeds the uncertainty in X given Y. Schreiber defined transfer entropy $T_{Y \to X}$ as the difference between the marginal conditional entropy and the joint conditional entropy. The corresponding local transfer entropy is defined as

$$\hat{T}^L_{Y \to X}(t_i) = \hat{H}^C\big(x_i \big| x_{i-p}^{i-1}\big) - \hat{H}^C(x_i \big| x_{i-q}^{i-1}, y_{i-r}^{i-1})$$

Conditional entropies can be re-expressed using the Law of Large Numbers

$$\hat{T}^L_{Y \to X}(t_i) = -\log \hat{f}_{-i}\big(x_i \big| x_{i-p}^{i-1}\big) + \log \hat{f}_{-i}(x_i \big| x_{i-q}^{i-1}, y_{i-r}^{i-1})$$

Again, recalling that a conditional density can be expressed as the ratio of a joint density and a marginal density, the estimator of $T^L_{Y \to X}(t_i)$ is

$$\hat{T}^L_{Y \to X}(t_i) = -\log \left\{ \frac{\hat{f}_{-i}\big(x_i, x_{i-p}^{i-1}\big)}{\hat{f}_{-i}\big(x_{i-p}^{i-1}\big)} \right\} + \log \left\{ \frac{\hat{f}_{-i}\big(x_i, x_{i-q}^{i-1}, y_{i-r}^{i-1}\big)}{\hat{f}_{-i}\big(x_{i-q}^{i-1}, y_{i-r}^{i-1}\big)} \right\}$$

The relationship between local transfer entropy and specific transfer entropy follows the logic established by local entropy rate and specific entropy rate. Specific transfer entropy is a local, dynamically informed average of local transfer entropy (Darmon and Rapp, 2017).

$$\hat{T}^S_{Y \to X}(t_i) = \frac{1}{k^*} \sum_{j \in N^*} \hat{T}^L_{Y \to X}(t_i)$$

The diagram below shows the results of a simulation using a coupled Hénon map (Darmon and Rapp, 2017). Y is independent of X, but X includes a dependence on both Y and X., the top graph shows Y as a function of time, and the second shows X as a function of time. The third graph shows $\hat{T}^L_{Y \to X}(t_i)$ and correctly indicates a nonzero transfer of information from Y to X. The fourth graph shows $\hat{T}^L_{X \to Y}$ and because of simulated noise, it indicates a spurious nonzero transfer of information from X to Y, though the amplitude is less than that observed for $\hat{T}^L_{Y \to X}$. The fifth and sixth graphs show specific transfer entropy. The fifth graph shows the expected nonzero value for $\hat{T}^S_{Y \to X}$. In contrast with $\hat{T}^L_{X \to Y}$, $\hat{T}^S_{X \to Y}$ is effectively zero.
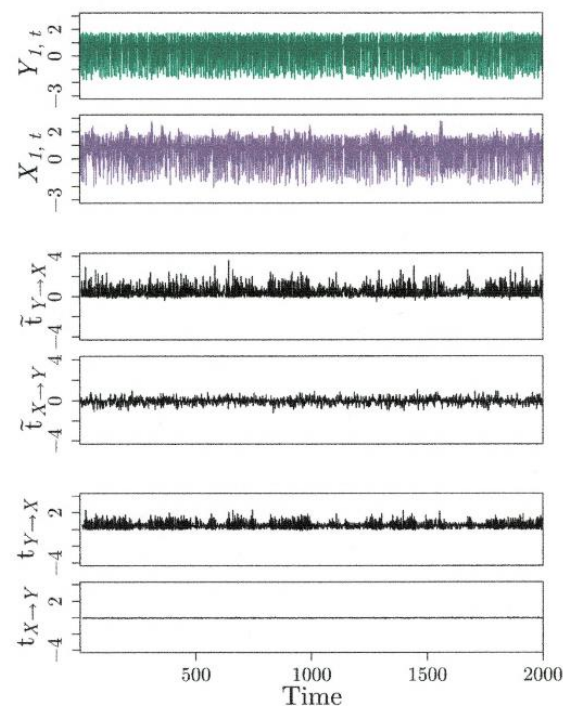
Figure 1. Results for the unidirectionally coupled stochastic Hénon map. (**a**) Y and X time series, (**b**) local transfer entropy, (**c**) specific transfer entropy (from Darmon and Rapp, 2017).

4. Time Dependent Measures of Information Movement in Networks

Networks consist of nodes and edges. Each node is not necessarily connected to all of the other nodes. When applied to the analysis of multichannel EEGs and event related potentials, which is our primary interest area, the electrodes are nodes and edges are quantified by functional connectivity or information transfer. These between node relationships are summarized in the adjacency matrix.

In the simplest case, the relationship between two nodes is quantified by the correlation between the signals measured at the corresponding electrodes. A value of zero is entered in the matrix if the correlation is less than a threshold value, and a value of one is entered if the value is greater than the threshold. The resulting adjacency matrix is binary, symmetric, and static. If the correlation itself is entered, the adjacency matrix is real, symmetric, and static. Alternatively, transfer entropy can be used. As defined, transfer entropy, in contrast with specific transfer entropy, is not time-dependent, but the information movement from Node A to Node B is typically not the same as the information movement from Node B to Node A. Therefore, in this case the adjacency matrix is real, asymmetric, and static. If specific transfer entropy is used, the time-dependent behavior of information movement in a complex network is summarized in a matrix that is real, asymmetric, and time-dependent.

In principle any measure that is applicable to a real, asymmetric, static adjacency matrix can be used iteratively at each time point to produce a time-dependent measure of matrix dynamics which we denote by $\Lambda(t)$. In many applications, particularly in the analysis of biological systems, the detection of transitions in network behavior is of particular interest. This can now be implemented by an examination of the transition behavior of $\Lambda(t)$. This is addressed in the fourth step of the project.

5. Hierarchical Transition Chronometries in Information Networks

There is a large well developed mathematical technology for identifying transitions in a time series, but care must be exercised when using these methods. As previously argued, "The typically stated challenge, 'find transition times in the time series' is

unacceptably naïve. Transitions are timescale dependent. A stock market change that is significant to a day trader is not necessarily important to a pension fund manager. The challenge to find a time series' transitions must be revised: 'find the transitions in time series X that are important on a timescale of Y' (Rapp, et al., 2018).

One possible approach to meeting this requirement uses quantification of recurrence diagrams. Recurrence diagrams were introduced by Eckmann, Kamphorst and Ruelle (1987) with the specific objective of identifying subtle time-dependent transitions in dynamical systems. Recurrence diagrams provide a graphical representation of geometrical relationships between points in an embedded set. Quadrant scans (Rapp, et al., 2013) provide a quantification of this visually presented structure. A change in the dynamics of the system generating the time series produces a local maximum in the quadrant scan. A change in the embedding dimension changes the time scale of the transitions identified in the recurrence diagram, and there is a corresponding change in the time scale of the quadrant scan. The utility of this method in identifying hierarchical transition chronometries is largely unexplored.

6. Conclusions

We have presented a four step analysis trajectory that begins with the measurement of a multichannel time series, proceeds with the construction of a time-dependent network, and concludes with the identification of transitions in the behavior of that network. Applications in the investigation of central nervous system psychopathology are now underway.

Conflicts of Interest: The opinions and assertions contained herein do not necessarily reflect the official policy or position of the Uniformed Services University, the Department of Defense, or the Henry M Jackson Foundation for the Advancement of Military Medicine. No Conflicts of Interest are reported.

References

1. Amblard, P.-O.; Michel, O.J.J. The relation between Granger causality and directed information flow: A review. *Entropy* **2013**, *15*, 113–143.
2. Berkeley, G. *A Treatise on Principles of Human Knowledge*; Wilkins, D.R., Ed.; Trinity College Dublin: Dublin, Ireland, 1710.
3. Biau, G.; Devroye, L *Lectures on the Nearest Neighbor Method*; Springer Series in the Data Sciences; Springer: New York, NY, USA, 2015.
4. Darmon, D. Specific differential entropy rate estimation for continuous-valued time series. *Entropy* **2016**, *18*, 190.
5. Darmon, D.; Rapp, P.E. Specific transfer entropy and other state-dependent transfer entropies for continuous-state input-output systems. *Phys. Rev. E* **2017**, *96*, 022121.
6. Eckmann, J.-P.; Kamphorst, S.O.; Ruelle, D. Recurrence plots of dynamical systems. *Europhys. Lett.* **1987**, *4*, 973–977.
7. Gilpin, C.; Darmon, D.; Ziwy, Z.; Martens, C. Information dynamics of a nonlinear stochastic nanopore system. *Entropy* **2018**, *20*, 221.
8. Granger, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438.
9. Lizier, J.T.; Prokopenko, M.; Zomaya, A.Y. Local information transfer as spatiotemporal filter for complex systems. *Phys. Rev. E* **2008**, *77*, 026110.
10. Lizier, J.T. Measuring the dynamics of information processing on a local scale in time and space. In *Directed Information Measures in Neuroscience*; Wibral, M., Vicente, R., Lizier, J.T., Eds.; Springer: New York, NY, USA, 2014; pp. 161–193.
11. Rapp, P.E.; Darmon, D.M.; Cellucci, C.J. Hierarchical transition chronometries in the human central nervous system. In Proceedings of the International Conference on Nonlinear Theory and Applications. Santa Fe, NM, USA, 8–11 September 2013.
12. Rapp, P.E.; Darmon, D.; Cellucci, C.J.; Keyser, D.O. The physiological basis of consciousness: A clinical ambition and the insufficiency of current philosophical proposals. *J. Conscious. Stud.* **2018**, *25*, 191–205
13. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.
14. Scott, D.W. *Multivariate Density Estimation: Theory, Practice and Visualization*; John Wiley: New York, NY, USA, 1992.
15. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: New York, NY, USA, 1986.
16. Takens, F. *Detecting Strange Attractors in Turbulence*; Lecture Notes in Mathematics; Rand, D.A., Young, L.S., Eds.; Springer: New York, NY, USA, 1981; Volume 898, pp. 365–381.

17. Wiener, N. The theory of prediction. In *Modern Mathematics for Engineers*; Beckenbach, E.F., Ed.; McGraw Hill: New York, NY, USA, 1956; pp. 165–190.