IEICE The Institute of Electronics, Information and Communication Engineers

IEICE Proceeding Series

Information-theoretical applications of ordinal patterns

J.M. Amigó, T. Aschenbrenner, W. Bunk, R. Monetti

Vol. 2 pp. 182-185
Publication Date: 2014/03/18
Online ISSN: 2188-5079

2013 International Symposium on Nonlinear Theory and its Applications
NOLTA2013, Santa Fe, USA, September 8-11, 2013

Information-theoretical applications of ordinal patterns

J.M. Amigó†, T. Aschenbrenner‡, W. Bunk‡, and R. Monetti‡

†Centro de Investigacion Operativa, Universidad Miguel Hernandez
Avda. de la Universidad s/n, 03202 Elche (Alicante), Spain
‡Max-Planck Institut für extraterrestrische Physik
Giessenbachstr. 1, 85748 Garching, Germany
Email: jm.amigo@umh.es, whb@mpe.mpg.de, tsa@mpe.mpg.de, monetti@mpe.mpg.de

Abstract—The symbolization of time series by means of ordinal patterns (i.e., permutations) have a number of advantages. One of them is that ordinal patterns are algebraic objects that, therefore, can be further operated with. The transcript of two ordinal patterns, which is the product of one of them by the inverse of the other, is a realization of this possibility. Transcripts have been already used to characterize the synchronization of coupled nonlinear oscillators and, more generally, to quantify the complexity of coupled time series. In this paper we use transcripts to reduce the dimensionality of the permutation conditional mutual information.

1. Introduction

Ordinal time series analysis is a particular sort of symbolic analysis whose "symbols" are ordinal patterns of a given length $L \geq 2$. This concept was introduced by C. Bandt and B. Pompe in their seminal paper [2], in which they also introduced permutation entropy as a complexity measure of time series. Since then, ordinal time series analysis has found a number of interesting applications in biomedical science, physics, engineering, finance, statistics, etc.

One important peculiarity of this new tool in data analysis is the fact that ordinal patterns of length L, which can be identified with permutations of L objects, have a well-known mathematical structure. Indeed, permutations build a (non-commutative) multiplicative group called the symmetric group of order L. That is, the symbols themselves are amenable to non-trivial mathematical operations. This property is exploited by the concept of transcript.

Transcripts were introduced in [4] for characterizing the synchronization of two coupled, chaotic oscillators. In [1] they were used to define two complexity indices for coupled time series. In this communication we present a further application, this time to the dimensional reduction of permutation conditional mutual information.

2. Theoretical setting

2.1. Ordinal patterns

Suppose that $\{x_t\}_{t=t_0}^{\infty}$ is a sequence whose elements (entries, symbols,...) x_t belong to a set (state space, alphabet,...) endowed with a total ordering $\leq$. In practice $\{x_t\}$ is obtained by sampling an analog signal. Let $T \geq 1$ be a *delay time*. We say that a length-L, time delay block (vector, window,...) $\mathbf{v}_{T,L}(x_t) = (x_t, x_{t+T}, ..., x_{t+(L-1)T})$ defines the *ordinal* (L-)*pattern* $\pi = \langle \pi_0, ..., \pi_{L-1} \rangle$ if

$$x_{t+\pi_0 T} < x_{t+\pi_1 T} < ... < x_{t+\pi_{L-1} T}, \tag{1}$$

where in case $x_i = x_j$, we agree to set $x_i < x_j$ if, say, $i < j$. In nonlinear time series analysis, L is called the *embedding dimension.*

Alternatively we also say that the block $\mathbf{v}_{T,L}(x_t)$ is of type π, or that π is realized by $\mathbf{v}_{T,L}(x_t)$, and write $\pi = o(\mathbf{v}_{T,L}(x_t))$. Therefore, an ordinal L-pattern (or ordinal patterns of length L) is nothing else but a permutation of the integer numbers 0, 1,..., $L-1$ showing the ranking (according to their size) of the elements $x_t, x_{t+T}, ..., x_{t+(L-1)T}$, where t is arbitrary and $L \geq 2$. Specifically, $\pi = \langle \pi_0, ..., \pi_{L-1} \rangle$ may be identified with the permutation $i \mapsto \pi_i$, $0 \leq i \leq L-1$, or, in combinatorial notation,

$$\langle \pi_0, \pi_1, ..., \pi_{L-1} \rangle = \begin{pmatrix} 0 & 1 & \cdots & L-1 \\ \pi_0 & \pi_1 & \cdots & \pi_{L-1} \end{pmatrix}. \tag{2}$$

The set of ordinal L-patterns will be denoted by $\mathcal{S}_L$. Needless to say, $\mathcal{S}_L$ can be promoted via the identification (2) to a group (the so-called *symmetric group*) of order $L!$ if equipped with the product of permutations,

$$\begin{aligned} \pi\,\sigma &= \begin{pmatrix} 0 & \cdots & L-1 \\ \pi_0 & \cdots & \pi_{L-1} \end{pmatrix} \begin{pmatrix} 0 & \cdots & L-1 \\ \sigma_0 & \cdots & \sigma_{L-1} \end{pmatrix} \\ &= \begin{pmatrix} 0 & \cdots & L-1 \\ \sigma_{\pi_0} & \cdots & \sigma_{\pi_{L-1}} \end{pmatrix} \\ &= \left\langle \sigma_{\pi_0}, ..., \sigma_{\pi_{L-1}} \right\rangle, \end{aligned} \tag{3}$$

the inverse element being given by

$$\pi^{-1} = o(\pi_0, ..., \pi_{L-1}), \tag{4}$$

and the unity by the identity permutation,

$$id = \langle 0, 1, ..., L-1 \rangle . \quad (5)$$

2.2. Transcripts

One way of exploiting the group-theoretical structure (3)-(5) of the ordinal patterns is the following. Given $\alpha, \beta \in \mathcal{S}_L$ there always exists a unique $\tau = \tau(\alpha, \beta) \in \mathcal{S}_L$, called *transcript* from the *source pattern* α to the *target pattern* β, such that

$$\tau\alpha = \beta, \quad (6)$$

where, according to (3), $\tau\alpha = \langle \alpha_{\tau_0}, \alpha_{\tau_1}, ..., \alpha_{\tau_{L-1}} \rangle$. When the source and target patterns are important for the discussion, we generally write $\tau_{\alpha,\beta}$. It follows from (6) that $\tau_{\beta,\alpha} = (\tau_{\alpha,\beta})^{-1}$.

As the source pattern α and the target pattern β vary over $\mathcal{S}_L$, their transcript varies according to $\tau(\alpha, \beta) = \beta \circ \alpha^{-1}$. Note that different pairs (α, β) can share the same transcription. As an example in $\mathcal{S}_3$, $\tau(\alpha, \beta) = \langle 0, 2, 1 \rangle$ for

$$\begin{aligned} (\alpha, \beta) &= (012, 021), (021, 012), (120, 102), \\ &\quad (102, 120), (201, 210), (210, 201) \end{aligned}$$

(angular parentheses omitted for brevity). More generally, given $\tau \in \mathcal{S}_L$ there exist $L!$ pairs $(\alpha, \beta) \in \mathcal{S}_L \times \mathcal{S}_L$ such that τ is the transcript from α to β.

Consider now two stationary time series $\{x_t\}$, $\{y_t\}$. In turn, they provide two sequences of ordinal patterns of arbitrary length L, $\{\alpha_t\}$ and $\{\beta_t\}$, respectively. The stationarity of the time series guarantees that the statistics of the different $L!$ ordinal patterns α_t, β_t does not depend on the discrete time t. This being the case, we may omit the index t in the sequel. If, moreover, the time series are ergodic, then the statistics of the corresponding ordinal patterns can be obtained from a "typical" series just using sliding windows of size L.

Let $p_L^1(\alpha)$ be the probability for the source L-pattern α to occur in $\{x_t\}$, let $p_L^2(\beta)$ be the probability for the target L-pattern β to occur in $\{y_t\}$, and let $p_L^J(\alpha, \beta)$ be their joint probability. Then, the probability function of the transcripts, $p_L^T(\tau)$, $\tau \in \mathcal{S}_L$, can be written as

$$p_L^T(\tau) = \sum_{(\alpha,\beta):\beta\alpha^{-1}=\tau} p_L^J(\alpha, \beta).$$

Furthermore, let

$$H(p_L^J) = -\sum_{\alpha,\beta\in\mathcal{S}_L} p_L^J(\alpha, \beta) \log p_L^J(\alpha, \beta)$$

be the entropy of the joint probability function p_L^J, and let

$$H(p_L^T) = -\sum_{\tau\in\mathcal{S}_L} p_L^T(\tau) \log p_L^T(\tau)$$

be the entropy of the corresponding transcript probability function p_L^T.

Theorem 1 [1]. The following inequalities hold:

$$0 \leq H(p_L^J) - H(p_L^T) \leq \min\{H(p_L^1), H(p_L^2)\}. \quad (7)$$

The result (7) can be generalized to $N \geq 2$ coupled time series [1]. In this case we write α^1,..., α^N for the ordinal patterns obtained from the time series $\{x_t^1\}$, ..., $\{x_t^N\}$, respectively, and $\tau^{1,2}$, ..., $\tau^{N-1,N}$ for the transcripts τ_{α^1,α^2},..., $\tau_{\alpha^{N-1},\alpha^N}$.

3. Coupling complexity index

Based on the concept of transcript, we proposed in [1] two different indices to measure the complexity of two interacting (deterministic or random) dynamical systems, as observed at two data series $\{x_t\}$ and $\{y_t\}$ output by them. For the information-theoretical application envisaged in this communication, only one of them, called $C_1(L)$ in [1] and $C(\alpha^1, ..., \alpha^N)$ here, will be needed. The following are equivalent definitions [5]:

$$\begin{aligned} & C(\alpha^1, ..., \alpha^N) \\ = & \min_{1\leq n\leq N} H(\alpha^n) - \big(H(\alpha^1, ..., \alpha^N) \\ & -H(\tau^{1,2}, ..., \tau^{N-1,N})\big) \\ = & \min_{1\leq n\leq N} I(\alpha^n; \tau^{1,2}, ..., \tau^{N-1,N}) \\ = & H(\tau^{1,2}, ..., \tau^{N-1,N}) \\ & - \max_{1\leq n\leq N} H\left(\alpha^1, ..., \hat{\alpha}^n, ..., \alpha^N \middle| \, \alpha^n\right), \end{aligned}$$

where $I(\alpha^n; \tau^{1,2}, ..., \tau^{N-1,N})$ is the mutual information between the univariate random ordinal pattern α^n and the multivariate random ordinal pattern $\tau^{1,2}, ..., \tau^{N-1,N}$, and $H\left(\alpha^1, ..., \hat{\alpha}^n, ..., \alpha^N \middle| \, \alpha^n\right)$ is the conditional entropy of the multivariate random ordinal pattern $\alpha^1, ..., \hat{\alpha}^n, ..., \alpha^N$ ($\hat{\alpha}^n$ means that the pattern α^n has been omitted) conditioned on the univariate random ordinal pattern α^n. It can be proved that $C_L(\alpha^1, ..., \alpha^N)$ is invariant under permutations of its arguments [5]. Note for further reference that $C(\alpha^1, ..., \alpha^N) = 0$ implies

$$\begin{aligned} & H(\alpha^1, ..., \alpha^N) \\ = & \min_{1\leq n\leq N} H(\alpha^n) + H(\tau^{1,2}, ..., \tau^{N-1,N}). \end{aligned} \quad (8)$$

The complexity index $C_L(\alpha^1, ..., \alpha^N)$ has a number of very interesting properties [5]. Here we state only two properties that will be needed in the next section.

Property 1 [5]. Let $N \geq 3$.

(i) If $\alpha^{n_{\min}}$ is such that $\min_{1\le n\le N} H(\alpha^n) = H(\alpha^{n_{\min}})$, then

$$C(\alpha^1, ..., \alpha^N) \ge C(\alpha^1, ..., \hat{\alpha}^k, ..., \alpha^N) \quad (9)$$

for all $k \ne n_{\min}$.

(ii) If there are at least two random variables with minimum entropy, then (9) holds for all $k = 1, ..., N$.

The following result is a corollary of Property 1 (ii).

Property 2. If $C(\alpha^1, ..., \alpha^N) = 0$ and $H(\alpha^1) = ... = H(\alpha^N)$, then all complexity indices $C(\alpha^{n_1}, ..., \alpha^{n_k})$ with $1 \le k \le N-1$ (i.e., with less than N variables) vanish as well.

4. Dimensional reduction of the permutation conditional mutual information

Given two random variables X^1, X^2, the *mutual information* between X^1 and X^2, $I(X^1, X^2)$, is defined as [3]

$$I(X^1, X^2) = H(X^1) + H(X^2) - H(X^1, X^2). \quad (10)$$

From the point of view of information theory, $I(X^1, X^2)$ represents the amount of information shared by the variables X^1 and X^2.

Conditional mutual information is analogously defined by adding the condition to each term in (10). We refer to the number of variables in the argument of I (including the conditioning ones) as the *dimension* of I.

We are going to show that the conditional mutual information of random ordinal L-patterns, i.e., the permutation conditional mutual information (PCMI) can be calculated via transcripts with one conditioning variable less, under some restrictions.

Theorem 1. If (i)

$$C(\alpha^1, \alpha^2, \beta^1, ..., \beta^M) = 0, \quad (11)$$

and (ii)

$$\min\{H(\alpha^1), H(\alpha^2)\} \ge \min_{1\le m\le M} H(\beta^m), \quad (12)$$

then in case $M = 1$,

$$I(\alpha^1, \alpha^2 \,|\beta^1) = I(\tau_{\alpha^1,\beta^1}, \tau_{\alpha^2,\beta^1}), \quad (13)$$

and in case $M \ge 2$,

$$\begin{aligned} & I(\alpha^1, \alpha^2 \,|\beta^1, ..., \beta^M) \quad (14) \\ = \; & I\left(\tau_{\alpha^1,\beta^1}, \tau_{\alpha^2,\beta^1} \,\middle|\tau_{\beta^1,\beta^2}, ..., \tau_{\beta^{M-1},\beta^M}\right). \end{aligned}$$

Proof. First of all, from the assumption (11) and the monotonicity property (9), it follows that, along with $C(\alpha^1, \alpha^2, \beta^1, ..., \beta^M)$, the complexity indices $C(\alpha^1, \beta^1, ..., \beta^M)$, $C(\alpha^2, \beta^1, ..., \beta^M)$, and $C(\beta^1, ..., \beta^M)$ vanish too.

Secondly, by the definition of mutual information, Eqn. (10) with X^1, X^2 replaced by α^1, α^2, and the chain rule of the joint entropy [3, Eqn. 2.14],

$$\begin{aligned} & I(\alpha^1, \alpha^2 \,|\beta^1, ..., \beta^M) \quad (15) \\ & = \sum_{n=1}^{2} H(\alpha^n \,|\beta^1, ..., \beta^M) - H(\alpha^1, \alpha^2 \,|\beta^1, ..., \beta^M) \\ & = \sum_{n=1}^{2} H(\alpha^n, \beta^1, ..., \beta^M) - H(\alpha^1, \alpha^2, \beta^1, ..., \beta^M) \\ & - H(\beta^1, ..., \beta^M). \end{aligned}$$

Set $H_\beta = \min_{1\le m\le M} H(\beta^m)$, and apply (8) to the last three terms in (15) to transform them as follows. First,

$$\begin{aligned} & H(\alpha^n, \beta^1, ..., \beta^M) \quad (16) \\ = \; & H_\beta + H(\tau_{\alpha^n,\beta^1}, \tau_{\beta^1,\beta^2}, ..., \tau_{\beta^{M-1},\beta^M}), \end{aligned}$$

since $\min\{H(\alpha^n), H(\beta^1), ..., H(\beta^M)\} = H_\beta$ for every $1 \le n \le N$ by the assumption (12).

By the same token,

$$\begin{aligned} & H(\alpha^1, \alpha^2, \beta^1, ..., \beta^M) \\ & = H_\beta + H(\tau_{\alpha^1,\alpha^2}, \tau_{\alpha^2,\beta^1}, \tau_{\beta^1,\beta^2}, ..., \tau_{\beta^{M-1},\beta^M}) \\ & = H_\beta + H(\tau_{\alpha^1,\beta^1}, \tau_{\alpha^2,\beta^1}, \tau_{\beta^1,\beta^2}, ..., \tau_{\beta^{M-1},\beta^M}), \quad (17) \end{aligned}$$

since $\min\{H(\alpha^1), ..., H(\alpha^N), H(\beta^1), ..., H(\beta^M)\} = H_\beta$ by the assumption (12). The second equality (17) follows from the fact that each variable in $\{\tau_{\alpha^1,\alpha^2}, \tau_{\alpha^2,\beta^1}\}$ can be determined from variables in $\{\tau_{\alpha^1,\beta^1}, \tau_{\alpha^2,\beta^1}\}$ and the other way around, so the corresponding entropies are the same.

Finally,

$$H(\beta^1, ..., \beta^M) = H_\beta + H(\tau_{\beta^1,\beta^2}, ..., \tau_{\beta^{M-1},\beta^M}) \quad (18)$$

if $M \ge 2$, or

$$H(\beta^1) = H_\beta \quad (19)$$

if $M = 1$.

The dimensional reductions (13) and (14) follow now upon replacing (16)-(19) in (15) . □

The dimension reduction by one unit of the PCMI stated in Theorem 1 is very important in time series analysis because the number of joint symbols grows exponentially with M, while the length of real-world time series is finite. Therefore, the expressions (13)-(14) can avoid undersampling in some cases and, in any case, they improve the statistical significance of the estimations.

We are going to see in the next section that the condition (11) can be usually met, at least approximately, in time series analysis by adjusting the delay time T.

5. Applications to transfer entropy

Let $\xi^1, ..., \xi^K$ random ordinal patterns obtained with sliding windows of size L from K coupled time series $\{x_t^1\}$, ..., $\{x_t^K\}$. Then it happens often [6] that $C(\xi^1, ..., \xi^N) \to 0$ when the delay time T grows or, at least, $C(\xi^1, ..., \xi^K)$ becomes very small as compared to its maximum value. This fact allows to satisfy in practice the condition (11) with a high degree of accuracy.

Let us see next how the above observation applies to one of the simplest instances of PCMI, namely, the symbolic transfer entropy. Thus let ξ^1, ξ^2 be the random ordinal L-patterns corresponding to the time series $\{x_t^1\}, \{x_t^2\}$, respectively, and likewise let ξ_Λ^1 be the random ordinal L-pattern obtained as above from the time series $\{x_{t+\Lambda}^1\}$, $\Lambda \geq 1$ being a time shift. Therefore, $H(\xi_\Lambda^1) = H(\xi^1)$.

The *symbolic transfer entropy from system 2 to system 1* is defined as [7]

$$T_{2\to1}^S := I(\xi_\Lambda^1, \xi^2 \left| \xi^1 \right.). \qquad (20)$$

Recall that if system 2 drives system 1, then $T_{2\to1}^S > 0$ for some Λ. Theorem 1 with $M = 1$, $\alpha^1 = \xi_\Lambda^1$, $\alpha^2 = \xi^2$, $\beta^1 = \xi^1$ yields the following result.

Corollary 1. [6] If

$$C(\xi_\Lambda^1, \xi^2, \xi^1) = 0 \qquad (21)$$

and $H(\xi^2) \geq H(\xi^1)$, then

$$T_{2\to1}^S = I(\tau_{\xi_\Lambda^1, \xi^1}, \tau_{\xi^2, \xi^1}). \qquad (22)$$

Therefore, if the delay time T can be so chosen that $C(\xi_\Lambda^1, \xi^2, \xi^1) \simeq 0$ and $H(\xi^1) = H(\xi^2)$ (because, for instance, the time series $\{x_t^1\}, \{x_t^2\}$ are different read-outs of the same dynamics), then both assumptions of Corollary 1 will be fulfilled.

The case $T_{1\to2}^S := I(\xi_\Lambda^2, \xi^1 \left| \xi^2 \right.)$ is dealt with analogously: If

$$C(\xi_\Lambda^2, \xi^1, \xi^2) = 0 \qquad (23)$$

and $H(\xi^1) \geq H(\xi^2)$, then

$$T_{1\to2}^S = I(\tau_{\xi_\Lambda^2, \xi^2}, \tau_{\xi^1, \xi^2}). \qquad (24)$$

6. Information direction using transcripts

In the last section we have illustrated with the transfer entropy how transcripts, together with the conditions (21) or (23), can make possible in practice the reduction of dimensionality of PCMI proved in Theorem 1, thereby enhancing the statistical quality of the estimation.

Another by-product of Theorem 1 is the justification for using the transcript mutual informations (22) and (24) as information direction indicators between two coupled time series, independently of whether the conditions (21) and (23) are fulfilled. In fact, the analysis of numerical and real time series supports this approach [6].

We conclude that one further advantage of transcripts as compared to ordinal patterns when studying coupled dynamical system is the possibility of using *mutual information* to ascertain the direction of the coupling. The ultimate reason for this magic is that transcripts themselves carry information about the coupling.

Acknowledgments

J.M.A. was financially supported by the Spanish *Ministerio de Economía y Competitividad*, grant MTM2012-31698.

References

[1] J.M. Amigó, R. Monetti, T. Aschenbrenner, and W. Bunk, "Transcripts: An algebraic approach to coupled time series," *Chaos* 22, 013105, 2012.

[2] C. Bandt, and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters* 88, 174102, 2002.

[3] T.M. Cover, J.A. Thomas, Elements of Information Theory (2nd edition), John Wiley, Hoboken, 2006.

[4] R. Monetti, W. Bunk, T. Aschenbrenner, and F. Jamitzky, "Characterizing synchronization in time series using information measures extracted from symbolic representations," *Physical Review E* 79, 046207, 2009.

[5] R. Monetti, J.M. Amigó, Aschenbrenner, and W. Bunk, "Permutation complexity of interacting dynamical systems, European Physical Journal Special Topics 222 (to appear)

[6] R. Monetti, W. Bunk, T. Aschenbrenner, S. Springer, and J.M. Amigó, "Information directionality in coupled time series using transcripts", `arXiv` 1304.7613v1.

[7] M. Staniek, K. Lehnertz, Symbolic transfer entropy, Phys. Rev. Lett. 100 (2008) 158101.