Information Sciences 278 (2014) 298–310

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Dimensional reduction of conditional algebraic multi-information via transcripts

J.M. Amigó [a,*], T. Aschenbrenner [b], W. Bunk [b], R. Monetti [b,c]

[a] Centro de Investigación Operativa, Universidad Miguel Hernández, Avda. de la Universidad s/n, 03202 Elche, Spain
[b] Max-Planck Institut für extraterrestrische Physik, Giessenbachstr. 1, 85748 Garching, Germany
[c] Fundación Escuela Medicina Nuclear (FUESMEN), CNEA-UNC, Garibaldi 405, M5500CJI Mendoza, Argentina

ARTICLE INFO

Article history:
Received 11 July 2013
Received in revised form 10 February 2014
Accepted 17 March 2014
Available online 21 March 2014

Keywords:
Multi-information
Transcript
Transfer entropy
Information direction
Symbolic representation

ABSTRACT

Symbolic representation is a standard and powerful technique in time series analysis. In an ordinal symbolic representation the symbols are the so-called ordinal patterns, which can be identified with permutations. Transcripts exploit the fact that permutations build a group, the transcript of a pair of permutations being the product of the second permutation times the inverse of the first one. This particular setting can be easily generalized to any representation which elements belong to an algebraic group. The dimensional reduction of conditional multi-information via transcripts, proved in this paper, perfectly shows the potential of such algebraic symbolic representations. Specifically, given $N+M$ group-valued random variables, the multi-information of N variables conditioned on the other M variables can also be calculated as a multi-information of N transcripts conditioned on $M-1$ transcripts, under some restrictions. Such a dimensional reduction can be crucial when estimating a conditional multi-information from short time series. Applications include two popular ordinal indicators of the information flow in coupled time series, namely, symbolic transfer entropy and momentary sorting information transfer. As a by-product of the above results, two new information directionality indicators based on ordinal transcripts are proposed, the simplest one being an (unconditioned) mutual information.

1. Introduction

This paper is at the intersection of information theory, symbolic dynamics, and time series analysis. It belongs to information theory because all our results refer one way or the other to entropies of random variables. Most importantly, our main results (Theorems 1 and 2) refer to the dimensional reduction of conditional multi-information, which includes mutual information. Second, the paper belongs to symbolic dynamics because those random variables are symbol-valued, the symbols being elements of an algebraic group. And, finally, it belongs also to time series analysis because one of the applications of our results refers to the information directionality of coupled data. In fact, the latter topic provided much of the inspiration for the present work.

Symbolic representations of real-valued time series arise, for instance, whenever one wants to apply finite-state information-theoretical tools, like the Shannon entropy or the mutual information, to analog signals [12]. In particular, a symbolic

* Corresponding author. Tel.: +34 96 665 8911; fax: +34 96 665 8715.
E-mail address: jm.amigo@umh.es (J.M. Amigó).

http://dx.doi.org/10.1016/j.ins.2014.03.054
0020-0255/

representation by *ordinal patterns* or, for that matter, permutations [8,3] is called an *ordinal symbolic representation*. Among the various theoretical and practical features of this sort of symbolic representation [1,2,4], we highlight here two of them. First, ordinal patterns are not symbols *ad hoc*, but they encapsulate information about the temporal structure of the time series. Further, ordinal patterns, being permutations, are amenable to algebraic operations themselves, like taking the inverse of one pattern or multiplying two of them. The concept of *transcript*, such as it was first formulated in [18], exploits that second feature of ordinal symbolic representations. This original framework will be enlarged below.

Transcripts were introduced in [18] to characterize the synchronization of chaotic oscillators. That work also stimulated the search for further applications. Thus, in [5] the authors proposed two transcript-based quantifiers C_1 and C_2 for the complexity of coupled time series and, furthermore, studied their performance with synthetic and real data. These quantifiers were aptly named *coupling complexity indices*. The basic properties of C_1 is the subject of [19], where the dynamical properties measured by means of ordinal patterns are referred to as *permutation complexity* [3].

As shown in [20], transcripts can also be used to determine the information direction between coupled time series. The means for this are provided by the ordinal versions of the transfer entropy [22] and momentary transfer information [21], which go by the names of *symbolic transfer entropy* (STE) and *momentary sorting information transfer* (MSIT), respectively. We found the remarkable result that, under some provisos involving the complexity index C_1, transcripts allow to calculate the STE and the MSIT with one variable less in each case. To be specific, STE is a trivariate conditional mutual information which can be calculated as a bivariate one using transcripts. Similarly, MSIT can be viewed as a 4-variate conditional mutual information which can be calculated as a 3-variate one also by means of transcripts. Moreover, it turns out that those provisos can be usually satisfied in practice. We will briefly discuss the practical implications of those two dimensional reductions in Sections 5.1 and 5.2. This being the case, why not to use the transcript-based dimensional reductions of STE and MSIT themselves to measure the information direction in general? We ponder over this question in Section 5.3.

The present paper generalizes the results of [20] on the above mentioned dimensional reduction of STE and MSIT in two directions.

(i) We go beyond ordinal symbolic representations of time series by allowing for any symbolic representation via the elements of an algebraic group. In this more general framework we talk of *algebraic symbolic representations*.
(ii) We also go beyond conditional mutual information by using the conditional multi-information (CMI).

The multi-information (function) is a tool for measuring the stochastic dependence among multiple random variables. In particular, the bivariate multi-information is the well-known mutual information. Furthermore, transfer entropy and momentary information transfer are instances of conditional multi-information functions. As we will see below, the reduction scheme of the STE and the MSIT generalizes to any number of variables once the mutual information is replaced by the multi-information. Specifically, the multi-information of N group-valued random variables conditioned on M such variables can be calculated as a multi-information of N group-valued random variables conditioned on $M-1$ such variables. The extra leverage brought in by the algebraic structure of the symbols is instrumental in this dimensional reduction.

This paper is divided in two parts. The first part (Sections 2–4) deals with group-valued random variables, and the mathematical underpinnings of the corresponding transcripts, complexity coefficient C_1 (denoted hereafter by C), and conditional multi-information, the latter being called *conditional algebraic multi-information* (CAMI) in this context. The second part (Section 5) is devoted to the applications of the theoretical results proved in the first part to *ordinal* time series analysis. The choice of this particular symbolic representation responds to our experience [18,5,7,11,19] and previous results in [20].

Thus, in Section 2 we introduce the basics of transcripts without reference to ordinal patterns. To pave the way to the core results, we prove in Section 3 some properties of the complexity index C in the new, group-theoretical setting. Section 4 contains the main theoretical results (Theorems 1 and 2) of the paper, which refer to the dimensional reduction of the CAMI, made possible precisely by the additional algebraic structure of the state space. Some applications of Theorem 2 to ordinal time series analysis are discussed in Section 5. These applications are further substantiated with the help of the STE (Section 5.1) and MSIT (Section 5.2). Last but not least, we propose in Section 5.3 the transcript-based, dimensional reductions of the STE and MSIT as information directionality indicators. The paper concludes with a summary of results in Section 6.

2. Transcripts

The concept of transcript was introduced in [18] for permutations. In this section we generalize it to elements of any (Abelian or non-Abelian) algebraic group.

Let $\mathcal{G}$ be a group. Its elements will be denoted by low case Greek letters throughout. Since $\mathcal{G}$ may be non-Abelian, we prefer to use the multiplicative notation. Thus, the composition or product of two elements $\alpha, \beta \in \mathcal{G}$ will be written $\alpha \circ \beta$ or just $\alpha\beta$. The *transcript from the element α to the element β* is defined as the element

$$\tau = \beta\alpha^{-1}. \qquad (1)$$

Equivalently, $\tau\alpha = \beta$, and $\alpha = \tau^{-1}\beta$. Sometimes we write $\tau_{\alpha,\beta}$ to explicitly indicate that τ is the transcript from α to β.

From a formal point of view, the transformation $(\alpha, \beta) \mapsto \tau_{\alpha,\beta}$ defines a projection from $\mathcal{G} \times \mathcal{G}$ onto $\mathcal{G}$ sending $|\mathcal{G}|$ pairs of group elements to each $\tau \in \mathcal{G}$. As usual, $|\mathcal{G}|$ denotes the cardinality of $\mathcal{G}$ (possibly infinite). Indeed, for any $\tau \in \mathcal{G}$ the transcript