Applied Mathematics and Computation 289 (2016) 98–110

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Modified multiscale cross-sample entropy for complex time series

Yi Yin, Pengjian Shang*, Guochen Feng

Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, PR China

ARTICLE INFO

Keywords:
Refined composite multiscale cross-sample entropy (RCMCSE)
Composite multiscale cross-sample entropy (CMCSE)
Multiscale cross-sample entropy (MCSE)
Artificial time series
Stock indices

ABSTRACT

In this paper, we introduce the composite multiscale cross-sample entropy (CMCSE) which may induce undefined entropies and then further propose the refined composite multiscale cross-sample entropy (RCMCSE) which modifies CMCSE. First, we apply multiscale cross-sample entropy (MCSE), CMCSE and RCMCSE methods to three types of artificial time series in order to test the validity and accuracy of these methods. Results show that RCMCSE reduces not only standard deviation, but also the probability of inducing undefined entropy effectively, which can provide better robustness and more accurate entropies. Then, these three methods are employed to investigate financial time series including US and Chinese stock indices. For the study between stock indices in the same region, some conclusions which are consistent with previous study are drawn by the RCMCSE results. Meanwhile, it can be found that undefined entropies are induced and the numbers of inducing undefined entropy by three methods for investigation between three US stock indices and two Chinese mainland stock indices are given. Compared with MCSE and CMCSE, RCMCSE method is capable of reducing the number of undefined entropy and providing more accurate entropies. Moreover, the differences on inducing undefined entropy between results for US stock indices & two Chinese mainland stock indices and results for US stock indices & HSI demonstrate a much closer relation between US stock markets and HSI than between US stock markets and two Chinese mainland stock markets. Hence, it can be concluded that RCMCSE is more applicable for the study between US and Chinese stock markets.

1. Introduction

Recently, numerous techniques have focused on the analysis of nonlinear time series [1–3]. Meanwhile, many approaches have been proposed and applied to measure the complexity of time series in the complex systems [4–7]. As a measure of degree of uncertainty to detect the system complexity from time series, entropy has a wide application. Approximate entropy (ApEn) was introduced by Pincus to quantify the concept of changing complexity [8–11] and had been used to measure the biologic time series [11,12]. Furthermore, the shortcomings of the ApEn method were analyzed and sample entropy (SampEn) was developed by Richman et al., which agreed with theory results much more closely than ApEn over a broad range of conditions and had a wide applications in clinical cardiovascular studies [13,14]. Cross-sample entropy (Cross-SampEn) was introduced to measure the similarity of two distinct time series. Compared with correlation coefficient, Cross-SampEn is superior to describe the correlation between time series [15]. Generally, the larger the entropy is, the more random and

* Corresponding author. Tel.: +86 10 51684410.
E-mail address: pjshangdu@163.com (P. Shang).

http://dx.doi.org/10.1016/j.amc.2016.05.013
0096-3003/

complex a system is. However, an increase in the entropy may not always be associated with an increase in dynamical complexity. For example, when entropy-based algorithms are applied to real-world datasets obtained in health and disease states, we obtain contradictory findings [16]. These results may be attributed to the factor of scale. Thus, Costa et al. introduced the multiscale entropy (MSE) to calculate SampEn over a range of scales which can represent the complexity more comprehensively. MSE can resolve the contradiction and has been successfully applied to measure the complexity of time series generated from various dynamical systems including physiological signals [17–19] and vibrational signals [20,21]. Similarly, multiscale cross-sample entropy (MCSE) is proposed to show Cross-SampEn at different time scales and applied in the financial time series [22,23].

However, in the MCSE method, the coarse-graining procedure construct coarse-grained time series whose length is equal to N/τ, from the original N-points time series with the scale factor τ, which may result in an inaccurate entropy value for the coarse-grained time series at large scales may not be adequately long. Besides, the Cross-SampEn is undefined in some cases because no template vectors are matched to one another. The reliability of MCSE reduces due to the inaccurate or undefined Cross-SampEn. These issues in accuracy and validity challenge the application of MCSE method. As a result, inspired by [24] we propose the composite multiscale cross sample entropy (CMCSE) to address the accuracy concern of the MCSE method. CMCSE algorithm calculates the Cross-SampEns of all coarse-grained time series at a scale factor τ and then defines the CMCSE value as the means of τ Cross-SampEn values. CMCSE method provides more accurate entropy values but increases the probability of inducing undefined entropy. Hence, we modify the CMCSE algorithm and further propose the refined composite multiscale cross sample entropy (RCMCSE).

The remainder of this paper is organized as follows. Section 2 introduces the MCSE, CMCSE and RCMCSE methods briefly. In Section 3, three types of artificial time series are used to evaluate the effectiveness of these methods. Section 4 presents the application to the financial time series. A conclusion is drawn in Section 5.

2. Methodologies

2.1. Composite multiscale cross-sample entropy

MSE is based on the application of SampEn, which is proposed by Costa et al. [16,25] used MSE to analyze the biological time series, and succeed in separating healthy and pathologic groups. It has been indicated the difficulty in distinguishing the inter-beat interval time series of different diseased and healthy states if only a single-scale SampEn is used [16,25]. As a result, MSE is proposed. Similarly, MCSE is based on MSE and defined as using Cross-SampEn to solve the same problem in the study of two time series and analyze two time series over scale factor. The MCSE method is also based on Cross-SampEn. Thus, the MCSE algorithm consists of two procedures: 1) a coarse-graining procedure, which can be used to obtain the representations of the original time series on different time scales, and 2) the Cross-SampEn, which is capable of measuring the degree of the asynchrony of two time series. First, we review Cross-SampEn procedure briefly. Given two time series of N points: $u:\{u(j):j=1,\ldots,N\}$ and $v:\{v(j):j=1,\ldots,N\}$. From vector length m sequences

$$x_m(i)=(u(i),u(i+1),\ldots,u(i+m-1)),\{i:1\le i\le N-m+1\}, \tag{1}$$

$$y_m(j)=(v(j),v(j+1),\ldots,v(j+m-1)),\{j:1\le j\le N-m+1\}, \tag{2}$$

from u and v respectively. Let $n_i^{(m)}$ be the number of vectors $y_m(j)$ whose distance of $x_m(i)$

$$d(x_m(i),y_m(j))=max\{|u(i+k)-v(j+k)|:0\le k\le m-1\} \tag{3}$$

is within the tolerance r. Similarly, $n_i^{(m+1)}$ is the number of matches of length $m+1$. Finally, Cross-SampEn is calculated with the equation:

$$Cross-SampEn(u,v,m,r)=-ln\left(\sum_{i=1}^{N-m}n_i^{(m+1)}\Big/\sum_{i=1}^{N-m}n_i^{(m)}\right) \tag{4}$$

The essential feature of Cross-SampEn is to measure the degree of the asynchrony of two time series. The value of Cross-SampEn is higher, if the pair of series is more asynchronous [16,26,27].

Meanwhile, the coarse-graining procedure need construct coarse-grained time series from the original series u and v with the scale factor τ, respectively. Then we get $\{x^{(\tau)}\}$ and $\{y^{(\tau)}\}$. Each point of the coarse-grained time series is defined as

$$x_j^{(\tau)}=\frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau}u_i,\ 1\le j\le N/\tau \tag{5}$$

$$y_j^{(\tau)}=\frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau}v_i,\ 1\le j\le N/\tau \tag{6}$$