Commun Nonlinear Sci Numer Simulat 56 (2018) 49–61

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

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

Research paper

Multiscale sample entropy and cross-sample entropy based on symbolic representation and similarity of stock markets

Yue Wu[a,*], Pengjian Shang[a], Yilong Li[b]

[a] Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, PR China
[b] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, PR China

ARTICLE INFO

Article history:
Received 21 December 2016
Revised 2 July 2017
Accepted 22 July 2017
Available online 28 July 2017

Keywords:
Modified multiscale entropy
Symbolic representation
Similarity
Stock markets

ABSTRACT

A modified multiscale sample entropy measure based on symbolic representation and similarity (MSEBSS) is proposed in this paper to research the complexity of stock markets. The modified algorithm reduces the probability of inducing undefined entropies and is confirmed to be robust to strong noise. Considering the validity and accuracy, MSEBSS is more reliable than Multiscale entropy (MSE) for time series mingled with much noise like financial time series. We apply MSEBSS to financial markets and results show American stock markets have the lowest complexity compared with European and Asian markets. There are exceptions to the regularity that stock markets show a decreasing complexity over the time scale, indicating a periodicity at certain scales. Based on MSEBSS, we introduce the modified multiscale cross-sample entropy measure based on symbolic representation and similarity (MCSEBSS) to consider the degree of the asynchrony between distinct time series. Stock markets from the same area have higher synchrony than those from different areas. And for stock markets having relative high synchrony, the entropy values will decrease with the increasing scale factor. While for stock markets having high asynchrony, the entropy values will not decrease with the increasing scale factor sometimes they tend to increase. So both MSEBSS and MCSEBSS are able to distinguish stock markets of different areas, and they are more helpful if used together for studying other features of financial time series.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Entropy is the rate of information production as it has a strong relation with nonlinear time series and dynamical systems. As a result, entropy is widely applied in various kinds of fields as a measure of quantifying the degree of uncertainty to detect the complexity in time series. To study time series from different angles, many concepts derived from entropy have been proposed, including approximate entropy [1,2], transfer entropy [3,4], sample entropy [5], cross-sample entropy [6–8], permutation entropy [9], etc. There have been successful attempts by using a range of entropy-based approaches [10–17]. Pincus [1] introduced approximate entropy (ApEn), which is well suited to the analysis of clinical cardiovascular and other time series and quantify the concept of changing complexity. But there are shortcomings for ApEn, for instance, it may induce inconsistent results. Furthermore, sample entropy (SampEn) was developed by Richman et al. [5], which agrees with theory much more closely than ApEn over a broad range of conditions and is more useful in clinical cardiovascular and

* Corresponding author.
E-mail addresses: 13271014@bjtu.edu.cn, 614035603@qq.com (Y. Wu).

http://dx.doi.org/10.1016/j.cnsns.2017.07.021
1007-5704/© 2017 Elsevier B.V. All rights reserved.

other biological studies. Similarly, cross-sample entropy (Cross-SampEn) was introduced for measuring the similarity of two distinct time series. However, contradictory findings hold in real-world datasets obtained in health and disease states when the SampEn algorithm is applied [18]. The reason for these findings may be due to the scale factor, so then Costa et al. [11] introduced the multiscale entropy (MSE) to take into account the time scales which can present the complexity more comprehensively. Based above, multiscale cross-sample entropy (MCSE) was also proposed to detect financial time series [19,20]. Nowadays, stock markets have attracted much attention [21–25] since they are hard to predict because of their high degree of complex and randomness.

In the MSE and MCSE method, however, there exists disadvantages such as inducing the probability of undefined entropies and obtaining inaccurate entropy values which challenge the two algorithms' application. Considering these limitation, Wu et al. [26] proposed the composite multiscale entropy (CMSE) algorithm to address the accuracy concern of the MSE algorithm. But the method increases the probability of inducing undefined entropies, so they developed a refined composite multiscale entropy (RCMSE) [27] algorithm successfully solving the two problems mentioned above. Inspired by this treatment, Yin et al. [28] introduced similar methods for calculating cross-sample entropy.

For financial time series in stock markets, volatility and sheer complexity are most common natures which have been serious obstacles for investigators to study the regularities. We study many aspects of financial time series, but most of the time, we prefer to know the trend, so we think of symbolic representation for financial time series. The distance in MSE is the infinite norm between two vectors constructed by original series. When symbolic representation is applied, the finite norm is inappropriate accordingly. Since the distance can be regarded as similarity to some extent, we develop another way to quantify the degree of similarity. So we propose a modified multiscale sample entropy measure based on symbolic representation and similarity (MSEBSS) to investigate stock markets. MSEBSS is confirmed to be useful for time series mingled with much noise like financial time series, then we modify the MSEBSS and further propose the modified multiscale cross-sample entropy measure based on symbolic representation and similarity (MCSEBSS) to research the asynchrony among stock markets from different areas.

The remainder of this paper is organized as follows. Section 2 introduces the SampEn, MSE, MSEBSS and MCSEBSS algorithms briefly. In Section 3, two types of artificial time series are used to evaluate the effectiveness of MSEBSS and MCSEBSS. Sections 4 and 5 illustrate our points through the application to stock markets by the two algorithms. Finally, we make a conclusion in Section 6.

2. Methodology

2.1. Sample entropy (SampEn)

At first, we take a brief review at SampEn. Let N be the length of the time series, m be the length of sequences to be compared, and r be the tolerance for accepting matches. For a time series $\boldsymbol{x} = \{x_1, x_2, \cdots, x_N\}$ of N points, the SampEn algorithm can be summarized as follows [5]:

Step 1. Construct template vectors $x_m(i)$, the sets of points x from i to $i+m-1$ with dimension m by using Eq. (1):

$$x_m(i) = \{\, x_{i+k} : 0 \le k \le m-1 \,\}, \quad 1 \le i \le N-m. \tag{1}$$

Step 2. For each $x_m(i)$, the distance between two such vectors $x_m(i)$ and $x_m(j)$ is calculated by using the infinite norm (the maximum difference of their corresponding scalar components):

$$d(x_m(i), x_m(j)) = \|x_m(i), x_m(j)\| = \max\left\{\left|x_{i+k} - x_{j+k}\right| : 0 \le k \le m-1\right\}, \quad 1 \le i, j \le N-m, j \ne i. \tag{2}$$

Step 3. An instance where a vector $x_m(j)$ is within r of $x_m(i)$ is called a m-dimensional template match. To exclude self-matches, we must have $j \ne i$. For $x_m(1)$, we count the number of template matches $n_1^{(m)}$, then from $x_m(2)$ to $x_m(N-m)$, we get the number of template matches for themselves in turn. The sum of $n_i^{(m)} (1 \le i \le N-m)$ is assigned to $n^{(m)}$. Finally, let $n^{(m)}$ represent the total number of m-dimensional template matches. Next, repeat the above process for $m+1$, and $n^{(m+1)}$ is obtained to represent the total number of $m+1$-dimensional template matches.

Step 4. Sample entropy is calculated with the equation:

$$SampEn(\boldsymbol{x}, m, r, N) = -ln\frac{n^{(m+1)}}{n^{(m)}}. \tag{3}$$

For a short time series, the SampEn algorithm may cause the problem that it yields inaccurate values and induce an undefined SampEn. In order to make the SampEn algorithm reasonable, the length of the time series is suggested to be in the range of 10^m to 30^m [29]. And in the condition where $m = 2$, the length of the time series should be longer than 750 [10].

From another perspective, the SampEn algorithm can be precisely seen as the negative natural logarithm of the conditional probability that two sequences similar for m points remain similar at the next point, where self-matches are not included in calculating the probability. Hence, it indicates that a lower value of SampEn means more self-similarity in the time series.

Download English Version:

https://daneshyari.com/en/article/5011313

Download Persian Version:

https://daneshyari.com/article/5011313

Daneshyari.com