Physica A 500 (2018) 210–221

Contents lists available at ScienceDirect

Physica A

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

Multidimensional scaling analysis of financial time series based on modified cross-sample entropy methods

Jiayi He, Pengjian Shang *, Hui Xiong

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

HIGHLIGHTS

- We propose MDS-KCSE and MDS-PCSE to analyze multivariate statistical problems.
- We apply MDS-KCSE and MDS-PCSE into the synthetic data and financial time series.
- Our analysis reveals valid classifications, which are consistent with the fact.
- Multidimensional scaling based on Chebyshev distance is employed as a reference.
- The results illustrate that MDS-KCSE and MDS-PCSE behave better than MDSC.

ARTICLE INFO

Article history:
Received 15 September 2017
Received in revised form 29 December 2017
Available online 21 February 2018

Keywords:
Financial time series
Multidimensional scaling
Cross-sample entropy
Kronecker-delta dissimilarity
Permutation cross-sample entropy

ABSTRACT

Stocks, as the concrete manifestation of financial time series with plenty of potential information, are often used in the study of financial time series. In this paper, we utilize the stock data to recognize their patterns through out the dissimilarity matrix based on modified cross-sample entropy, then three-dimensional perceptual maps of the results are provided through multidimensional scaling method. Two modified multidimensional scaling methods are proposed in this paper, that is, multidimensional scaling based on Kronecker-delta cross-sample entropy (MDS-KCSE) and multidimensional scaling based on permutation cross-sample entropy (MDS-PCSE). These two methods use Kronecker-delta based cross-sample entropy and permutation based cross-sample entropy to replace the distance or dissimilarity measurement in classical multidimensional scaling (MDS). Multidimensional scaling based on Chebyshev distance (MDSC) is employed to provide a reference for comparisons. Our analysis reveals a clear clustering both in synthetic data and 18 indices from diverse stock markets. It implies that time series generated by the same model are easier to have similar irregularity than others, and the difference in the stock index, which is caused by the country or region and the different financial policies, can reflect the irregularity in the data. In the synthetic data experiments, not only the time series generated by different models can be distinguished, the one generated under different parameters of the same model can also be detected. In the financial data experiment, the stock indices are clearly divided into five groups. Through analysis, we find that they correspond to five regions, respectively, that is, Europe, North America, South America, Asian-Pacific (with the exception of mainland China), mainland China and Russia. The results also demonstrate that MDS-KCSE and MDS-PCSE provide more effective divisions in experiments than MDSC.

* Corresponding author.
E-mail addresses: 17118443@bjtu.edu.cn (J. He), pjshang@bjtu.edu.cn (P. Shang).

https://doi.org/10.1016/j.physa.2018.02.105
0378-4371/

1. Introduction

During the past several decades, the analysis of financial time series has been attracting increasing attention. Scholars have used many methods to study in this field. Multidimensional scaling (MDS), as a form of non-linear dimensionality reduction, begins to become one of them. It offers an effective way to simplify the variables from high-dimensional space to low-dimensional space with the information contained in a distance matrix. Each object is then assigned coordinates in each of the lower dimensions. Actually, practical considerations are always a combination of many factors. MDS is also quite a good way to fully study the relevant values jointly influenced by multiple factors, in addition, it could offer a perceptual map of values based on the distance or dissimilarity matrix of data. In 1952, Torgerson first gave the mathematical model of multidimensional scaling method for measurement [1]. After the rapidly expanding with times, it has been applied in diverse fields. The paper [2–4] applied this method in signal transactions to establish sensor networks and estimate sensors locations. In sociology, the MDS to the perceived social support(MSPSS) is proposed [5,6] and Chou [7] applied this method to assess Chinese adolescents' social support in 2000. Besides, the regular behaviors, voting and sustainable supply chains are also discussed via MDS [8–11]. In recent years, with the continuous development of the financial markets, Machado et al. analyzed the temperature time-series by MDS in 2014 [12]. In biomedical science, virus diseases [13], mental representation of facial identity [14] and emotional responses to music [15] are all discussed via MDS. Additionally, MDS is utilized for other fields as well [16,17]. MDS is gradually introduced into the financial markets to analyze the relationship between financial systems or recognize the patterns of different financial institutions by Machado, Duarte, Cecilio and so on [18–24]. However, are there any other measurement for distance or dissimilarity in MDS?

In recent years, with the development of measurement methods for complex systems, entropy has a wide application. Approximate entropy (ApEn) was introduced by Pincus [25–27] and used to measure the biologic time series [28,29]. However, ApEn statistics leads to inconsistent results. Richman et al. developed sample entropy (SampEn) [30] and had compared ApEn and SampEn by analyzing sets of random numbers with known probabilistic character. They also evaluated cross-ApEn and cross-SampEn and found that cross-SampEn is a more consistent measure of joint synchrony. Liu's team also reached the same conclusion in the experiments, they found cross-SampEn is superior to describe the correlation between time series by contrast experiment [31]. Yin, shi and Shang used this method in the research of financial time series [32–35]

Inspired by these two approaches, we propose two new MDS methods that based on modified cross-sample entropy to analyze the information from stocks and utilize the MDS based on Chebyshev distance (MDSC) as control. Our analysis reveals a more reliable division of both synthetic data and 18 indices than MDSC. Moreover, the experimental results indicate that the algorithm is more sensitive than MDSC in the clustering process.

The rest of this paper is organized as follows. Section 2 briefly introduces the basic concepts involved in this paper, including Kronecker-delta based cross-sample entropy, permutation based cross-sample entropy and MDS based on the two measurement. We give a description of data set in Section 3. In Section 4, we apply the methods mentioned in this paper to the synthetic data and daily data of 18 indices from diverse stock markets for getting the results. Finally, the conclusion is given.

2. Methodology

Multidimensional scaling (MDS) was first proposed by Torgerson [1]. It is a statistic method for pattern recognition. The main idea is that: Given n data points $x_1, x_2, \dots, x_n$, the vector form is expressed as $X = (x_1, x_2, \dots, x_n)'$. By calculating the distance or dissimilarity between n points through a certain algorithm, an inner product matrix B can be constructed. Then, the coordinates for each point in the lower r dimensional space is obtained. Considering the visualization of the results, in general, $r = 2$ or $r = 3$.

During the whole process of MDS, the selection of distance or dissimilarity matrix is of great concern to the results. Generally, the distance or dissimilarity measurement involved in the MDS is Euclidean distance. However, the Euclidean distance is usually used to deal with the distance between two points in the lower dimensional space. In this paper, we need to deal with the high dimensional data set, so we chose the Chebyshev distance, that is, L_∞, as the dissimilarity measurement in the MDS, which is employed as the control group. The modified cross-sample entropy we newly proposed in this paper are for experiments.

Sample entropy (SampEn) is used for assessing the complexity of time series, proposed by Richman, and it was widely applied in clinical cardiovascular studies at first [30,36]. The steps for calculation are as follows:

Step 1. Given a time series of N points $X = [x(1), x(2), \dots, x(N)]$, through the phase space reconstruction, X transfers into a matrix X_m, where $X_m(i) = [x(i), x(i+1), \dots, x(i+m-1)]$, $i = 1, 2, \dots, N-m+1$. m is the embedding dimension.

Step 2. Define the distance between two vectors by L_∞ measurement as

$$d[X_m(i), X_m(j)] = max(|x(i+k) - x(j+k)|, k = 0, 1, \dots, m-1, i \neq j). \quad (1)$$

Step 3. Let $B_r^m(i)$ be the number of vectors $X_m(j)$ within r of $X_m(i)$ divided by $N-m+1$, $i \leq N-m$, r is a matching tolerance, in general, $r = 0.1 \sim 0.25 * std$. Then define

$$B_r^m = (N-m+1)^{-1} \sum_{n=1}^{N-m+1} B_r^m(i). \quad (2)$$