Physica A 506 (2018) 1126–1133

Contents lists available at ScienceDirect

Physica A

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

Diversity analysis based on ordered patterns

Yali Zhang, Pengjian Shang *, Zhenghui Sun

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

HIGHLIGHTS

- Emlen index based on ordered patterns is introduced as a new way is proposed.
- The new model can be used to assess complexity of a complex dynamical system.
- The experiments show this method is more sensitive to the change of system.

ARTICLE INFO

Article history:
Received 27 November 2017
Received in revised form 7 March 2018
Available online 9 May 2018

Keywords:
Emlen index
Weighed Emlen index
Multiscale analysis
Complexity
Traffic system

ABSTRACT

Diversity index based on entropy such as Emlen index (*EI*) is suggested as a highly effective algorithm to detect complexity in nonlinear systems. In this paper, a method considering the amplitude information, called Weighted Emlen index (*WEI*) is proposed to extract the amplitude information and track the signal changes. And then, we expand *EI*/*WEI* methods to multiple time scales. Thus, multiscale Emlen index (*MSEI*) and multiscale Weighted Emlen index (*MSWEI*) will be used to discuss the complexity of traffic time series under different scales. Results show that there are differences between the complexity of weekend traffic time series and workday time series, which gives a better classification of traffic systems for making prediction.

1. Introduction

Complexity in traffic system has been a universal focus of researchers. Especially urban traffic congestion is a gordian knot in course of social reform in modern life [1,2]. Traffic time series often present periodic complex fluctuations which includes the essential information. The data is highly nonlinear and change with time of day [3–5]. In recent years, many researchers have been doing some groundbreaking research about analyzing or forecasting traffic signals from different levels, for instance cellular automata models [6–8], chaos [9–11] and others.

Diversity index based on entropy is considered to be a powerful tool for the analysis of time series, as it allows describing the probability distributions of the possible state of a system, and therefore the information encoded in it [12]. Historically, many indices of diversity have been proposed: Emlen index [13], Gini–Simpson index [14], Renyi entropy [15], Tsallis entropy [16], Hill's diversity number [17], Shannon entropy and Simpson index. This paper offers a new strategical perspective for Emlen index (*EI*) based on ordinal patterns.

Permutation methods assess the appearance of repetitive ordinal patterns in time series to quantify the degree of regularity or orderliness [18–20]. In virtue of the arrangement, these ordinal patterns reflect the comparison of relative adjacent values instead of the actual ones. In this paper, we put forward a modification by extracting amplitude information and adding to traditional permutation [21], and then use Emlen index to calculate the complex information of series that is

* Corresponding author.
E-mail address: 16121596@bjtu.edu.cn (P. Shang).

https://doi.org/10.1016/j.physa.2018.05.026
0378-4371/

called Weighted Emlen index (*WEI*). It can embody volatility information of the regular spiky pattern and it is more sensitive to the abrupt changes in the data.

EI and *WEI* are only on a single scale to quantify the regularity (predictability) of time series. Since we should take into account the multiple temporal scales, more abundant results and property can be got [22–29]. Therefore, *EI* and *WEI* integrated with multiscale analysis called multiscale Emlen index (*MSEI*) and multiscale Weighted Emlen index (*MSWEI*) respectively will be used to study complex system.

The remainder of the paper is organized as follows: Section 2 introduces the procedure of Emlen index (*EI*) and Weighed Emlen index (*WEI*), and provides the background for multiscale analysis. Section 3 shows the analysis and results for both synthetic data and traffic time series. Finally, Section 4 gives a summary.

2. Methodology

2.1. Emlen index and weighted Emlen index

When studying a complex dynamic system, Shannon entropy is often considered as the most basic and the most natural way to reveal the hidden information of the time series. Many methods, such as Shannon and other classical measures, neglect temporal relationships between the values of the time series, which makes the structure and possible temporal patterns present in the process are not accounted for [30]. To better understand the problem, let us consider two sequences $X = \{0, 0, 1, 1\}$ and $Y = \{0, 1, 0, 1\}$. The results reflected by Shannon entropy of these two sequences are the same. However, we can find that the two sequences have different structures. It is well known that a dynamical system can be suitably represented and analyzed by using a symbolic sequence. Recently, permutation was introduced by Bandt and Pompe [31], which is based on symbolic sequences and takes into account the temporal dynamics of time series.

Given a time series $\{x(i), i = 1, 2, \ldots\}$, let us first embed a scalar time series x_t to a n-dimensional space: $X_i = \{x(i), x(i+d), \ldots, x(i+(n-1)d)$, where n is called the embedding dimension and d is the delay time. To this vector, we can get the corresponding ordered patterns, namely the permutation $\pi = \{r_0 r_1 \cdots r_{n-1}\}$ of $(01 \cdots n-1)$ which fulfills

$$x_{i+r_0} \leq x_{i+r_1} \leq x_{i+r_2} \cdots \leq x_{i+r_{n-2}} \leq x_{i+r_{n-1}} \tag{1}$$

In other words, the values of each vector are sorted in an ascending order, and a permutation pattern π is created with the offset of the permuted values. For a n-dimensional reconstructed space, the number of its permutations can be at most $n!$. Assume that the probability of different symbol patterns is $p = p(\pi_i), i = 1, 2, \ldots, n!$. Then we define the Emlen index based on ordinal patterns for the time series with T values $\{x(i), i = 1, 2, \ldots\}$ is:

$$EI = \sum_{i=1}^{n!} p(\pi_i) exp^{-p(\pi_i)} \tag{2}$$

where $p(\pi_i)$ is calculated as

$$p(\pi_i) = \frac{\sharp\{j | j = 1 \cdots\cdots T-n+1;\, x_j \text{ has type } \pi_i\}}{T-n+1} \tag{3}$$

In order to better track abrupt changes in the data signal and assign less complexity to segments that exhibit regularity or are subject to noise and neighboring values effects, we suggest a modification with more weight to amplitude-coded information. The procedure of calculating the *WEI* is briefly summarized here.

First, the weighted vectors x_j relative frequencies are calculated as follows:

$$p_w(\pi_i) = \frac{\sum_{j: x_j \text{ has type } \pi_i} W_j}{\sum_{j \leq T-n+1} W_j} \tag{4}$$

W_j is the weight value of each x_j [32], and the expression records that $\sum_i p_w(\pi_i) = 1$ still holds.

$$W_j = \frac{1}{n} \sum_{k=1}^{n} [x_{j+k-1} - \overline{X}_j]^2 \tag{5}$$

$\overline{X}_j$ is the arithmetic mean of x_j

$$\overline{X}_j = \frac{1}{n} \sum_{k=1}^{n} x_{j+k-1} \tag{6}$$

Then *WEI* is defined as

$$WEI = \sum_{i=1}^{n!} p_w(\pi_i) exp^{-p_w(\pi_i)} \tag{7}$$

By with, *WEI* can be employed to detect sudden change of data in noisy.