Research Article

A Multiscale Symbolic Dynamic Entropy Analysis of Traffic Flow

Zhanyou Cui,1 Gaoli Chen,1 Bing Liu,2 and Deguang Li3

1College of Mechanical and Electrical Engineering, Zhengzhou University of Industrial Technology, Zhengzhou 451150, China
2Henan Forestry Vocational College, Luoyang 471002, China
3School of Information Technology, Luoyang Normal University, Luoyang 471934, China

Correspondence should be addressed to Zhanyou Cui; zhanyou_cui@126.com

Received 24 September 2021; Revised 15 November 2021; Accepted 14 March 2022; Published 30 March 2022

Copyright ©2022 Zhanyou Cui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The complexity analysis of traffic flow is important for understanding the property of traffic system. Being good at analyzing the regularity and complexity, multiscale SamEn has attracted much attention and many methods have been proposed for complexity analysis of traffic flow. However, there may exist discontinuity of the calculated entropy value which makes the regularity of the traffic system difficult to understand. Moreover, it is difficult to select an appropriate r for the accurate evaluation of the complexity, which limits the application of multiscale entropy for traffic flow analysis. To solve this problem, a new entropy-based method, multiscale symbolic dynamic entropy, for evaluating the traffic system is proposed here. To verify the effectiveness of the proposed method, traffic data collected from stations in different cities are preprocessed by the proposed method. Both results of two cases show that the weekend patterns and weekday patterns are effectively distinguished using the proposed method, respectively. Specifically, compared with the traditional methods including multiscale SamEn and the multiscale modified SamEn, the complexity of the corresponding traffic system can be better evaluated without considering the selection of r, which demonstrates the effectiveness of the proposed method.

1. Introduction

The research of the traffic system plays great role in reducing the traffic congestion, which increases traveling delays and causes huge economic losses in cities [1]. In recent years, the number of vehicles has increased rapidly and the traffic congestion has become more serious, so the research of the traffic system has become more important and has attracted much attention. Specifically, there are a lot of studies about the research of traffic including the merging behavior [2–4], traffic flow forecasting [5, 6], model of traffic flow [7, 8], traffic theory [9], and traffic resource allocation [10]. For example, Wang et al. [11] proposed a short-term traffic flow prediction approach based on multiple traffic flow basic parameters, in which the chaos theory and support vector regression are utilized. Li et al. [12] proposed a novel hybrid forecasting model by combining three predictors, namely, the autoregressive integrated moving average (ARIMA), backpropagation neural network (BPNN), and support vector regression (SVR). Yao et al. [13] presented a two-level optimization method of scheduling and trajectory planning for connected automated vehicles.

Furthermore, the traffic dynamics and regularity of traffic flow is still a hot topic and a big challenge to transportation engineering [14]. Choi and Lee [15] found that the traffic equation exhibits a variety of behaviors including homogeneous flow, turbulent behavior, and density waves with fluctuations in the appropriate regime, and they investigated the possibility of 1/f fluctuations. Li et al. [16] conducted a comprehensive analysis of the motivations, gap acceptance, duration, and speed adjustment of heavy vehicle lane changes, which gives a better understanding of the lane-changing behaviors of heavy vehicles. Castillo et al. [17] found that the emergent state shows a stochastic resonance-like behavior that the average traffic velocity increases with respect to the system without noise for different initial jammed densities.
With help of these above findings, the dynamics of traffic jams in cities can be understood better. Then, an approach based on algebraic topological methods [18] was used to accurately characterize jamming in dynamical systems with queues. Yin and Shang [19] proposed a multiscale multi-fractal distended cross-correlation analysis method to describe the cross-correlation properties depending on the timescale in which the multifractality is computed. They found that the main distinction between weekday and weekend patterns is the different periodic patterns hidden in them. In addition, these different periodic patterns play an important role in the Hurst surface of cross-correlation investigation. The chaos characteristics of the traffic system were also studied by a nonlinear time series modeling technique [20]. To better evaluate the complexity of the traffic system, the multiscale SamEn (MSE) method is the most used method and has obtained many valuable results. The MSE method was first proposed in 2002 for evaluating the complexity of physiologic time series [21]. Wang et al. [22] introduced the multiscale permutation entropy analysis to investigate the complexities of different traffic time series and found that the complexity of weekend traffic time series is different from that of weekday time series. This finding is helpful for classifying the series when making prediction. Afterwards, to study the correlation degree and complexity between multiple variables, the multivariate multiscale sample entropy [23, 24] was also proposed and the more accurate and helpful knowledge about the complexity of traffic time series was obtained.

From these studies, we can find that the MSE and its variants are very effective for complexity evaluation of the traffic system. However, there are still shortcomings of multiscale entropy which limit its application for traffic flow. As pointed out in [25], the similarity definition of vectors based on Heaviside function may lead to discontinuous and hard boundary and cause some problems in the validity and accuracy of SampEn. Specifically, there are no vectors satisfying the evaluating condition, and the value of Heaviside function is zero. As a result, the denominator of SamEn is zero, which makes the value of SamEn be NaN [26]. This meaningless result makes the complexity of traffic flow difficult to understand. Furthermore, it is difficult to select an optimal parameter \( r \) for the calculation of MSE or multiscale modified SamEn (MMSE) reported in [25]. Therefore, it is difficult to evaluate the complexity of some traffic systems. To solve this problem, inspired the symbolic entropy \( s \) he chaos characteristics of the traffic system [27], a multiscale symbolic dynamic entropy (MSSDE) method is proposed for evaluating the complexity of the traffic system in this paper. In the proposed method, first, the original time series defined is transformed into the coarse-grained time series for representing the system dynamics on different time scales. Second, the coarse-grained time series is transformed into the symbolic time series. Third, we construct the template vector with the length of \( m \) and calculate the frequency of occurrence. Naturally, based on the frequency of occurrence, the average of the frequency can be obtained. Similarly, the average of the frequency with the dimensionality of \( m+1 \) can be obtained. Finally, the MSSDE can be constructed based on the obtained average of the frequency with \( m+1 \) and \( m \). The MSSDE holds both merits of multiscale entropy and symbolic dynamic entropy. The complexity and similarity can be well analyzed with help of the proposed MSSDE. Two cases of traffic data collected from different stations are used to verify the effectiveness of the proposed method. The results show that the proposed method is useful for the complexity analysis of traffic flow.

1.1. Multiscale Symbolic Dynamic Entropy. The MSSDE is proposed based on SamEn for better analysis of the complexity of traffic flow. The MSSDE is described in detail as follows.

(1) First, original time series defined is transformed into the coarse-grained time series for representing the system dynamics on different time scales. The process is achieved by averaging original time series inside consecutive but nonoverlapping windows of length \( \tau \). \( \tau \) is selected from 1 to 6 in the proposed method. The corresponding equation is as follows:

\[
 u^\tau_j = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j} x_i, \quad 1 \leq j \leq N \tau.
\]  

(2) Then, \( \{u^\tau_j, j = 1, 2, \ldots, N/\tau\} \) of the scale \( \tau \) is transformed into the symbolic time series \( s^\tau(s(k), k = 1, 2, \ldots, N) \) by

\[
s^\tau(i) = \begin{cases} \sigma_1, & u^\tau_1 \in C_1 \\ \sigma_2, & u^\tau_2 \in C_2 \\ \vdots & \vdots \\ \sigma_M, & u^\tau_M \in C_M \end{cases},
\]

where \( N \) is the length of the time series and \( M \) \((M < N)\) is the number of symbolic. \( \sigma_i (i = 1, 2, \ldots, N) \) represents different symbolic which is further replaced by Arabic numerals in this paper; specifically, \( \sigma_i = i \). \( C_i (i = 1, 2, \ldots, N) \) are \( N \) disjoint sets obtained by dividing the value space of \( u \). Specifically, \( C_i \) satisfy both equations (3) and (4).

\[
 \bigcup_{i=1}^{N} C_i = \bigcup_{i=1}^{N} u^\tau_i,
\]

\[
 C_j \cap C_j = \emptyset, \forall i \neq j, i, j \in [1, \ldots, N].
\]

The space of \( C \) is divided into \( M \) partitions based on the normal distribution [28]. Through this step, the time series \( u^\tau_i \) with length \( N/\tau \) at various \( \tau \) are synchronized into the symbolic time series with length \( M \).
(3) Next, the template vector is constructed by

$$S^m_i = [s(i), s(i+1), \ldots, s(i + m - 1)], i = 1, 2, \ldots, \frac{N}{\tau - m + 1}$$  \hspace{1cm} (5)

where \( m \) is the length of sequences to be compared. We can find that the number of possible values of \( S^m_i \) is \( M_m \). In addition, \( N/\tau - m + 1 \) template vectors can be obtained.

(4) Afterwards, for a given \( S^m_i \), calculate the number of vectors \( Q^m_i \) satisfying \( S^m_i = S^m_j \), where \( 1 \leq j \leq N/\tau - m + 1 \).

(5) Calculate the frequency of occurrence

$$C^m_i = 1/N - m + 1Q^m_i, j = 1, 2, \ldots, N/\tau - m + 1.$$  \hspace{1cm} (6)

(6) The average of the frequency is defined by

$$C^m = 1/N - m + 1 \sum_{i=1}^{N/\tau - m + 1} C^m_i.$$  \hspace{1cm} (7)

(7) Extend the dimensionality from \( m \) to \( m+1 \), and then \( S^{m+1}_i \) are calculated as follows:

$$S^{m+1}_i = [s(i), s(i+1), \ldots, s(i + m + 1)], i = 1, 2, \ldots, \frac{N}{\tau - m}.$$  \hspace{1cm} (8)

(8) Similarly, the frequency of occurrence is calculated according to

$$C^{m+1}_i = \frac{1}{N - m}Q^{m+1}_i, (j = 1, 2, \ldots, \frac{N}{\tau - m}).$$  \hspace{1cm} (9)

where \( Q^{m+1}_i \) represent the number of vectors satisfying \( S^{m+1}_i = S^{m+1}_j \). The corresponding average of \( C^{m+1}_i \) is

$$C^{m+1} = \frac{1}{N - m} \sum_{i=1}^{N/\tau - m + 1} C^{m+1}_i.$$  \hspace{1cm} (10)

(9) Finally, the MSSDE can be obtained according to

$$\text{MSSDE}(m, \tau, M) = -\text{ln} \left( \frac{C^{m+1}}{C^m} \right).$$  \hspace{1cm} (11)

The computational complexity of the proposed method is more expensive than the traditional methods including symbolic dynamic entropy and sample entropy but is almost the same as the other traditional multiscale entropy-based methods. In other words, the computational complexity depends greatly on the scales of entropy. However, the computation complexity is still small and the proposed method can be quickly computed with help of a personal computer.

2. Results and Analysis

2.1. Case 1: Complexity Analysis of Traffic Flow from PeMS.

To verify the effectiveness of the proposed method, the test dataset from PeMS open-access traffic flow database is used. The database was recorded every 5 min in Sacramento County. Therefore, 288 data elements were collected for a daily flow series of one loop detector, and the data of 7 days from March 31 to April 6, 2014, as shown in Figure 1, were analyzed by the proposed MSSDE to explore the complexities of the traffic time series [29, 30].

We first apply the traditional MSE to process the data, and the results are shown below. The parameter \( r \) is set to be 0.2 * std (std stands for the standard deviation of the data to be analyzed), empirically. The corresponding MSE of various scales from Monday to Sunday is shown in Figure 2(a). MSE of many days at some scales has no value, and the existing MSE values on the traffic data of Saturday and Sunday show that the traffic data of the two days belong to the same pattern with the weekday pattern. The MSE with no value refers to the meaningless value, which is obtained because some values are divided by zero. This condition is commonly seen because the template vectors are coarsely segmented just based on the original time series without preprocessing. As a result, the different patterns of traffic data cannot be well extracted and the difference between different patterns is so small that the frequency of occurrence is zero. Therefore, the complexity of the traffic data cannot be evaluated by the MSE when \( r = 0.2 \). To investigate the influence of \( r \) on MSE, \( r \) varying from 0.01 to 0.6 at a step of 0.05 is used to calculate the MSE at different scales from 1 to 6 for the traffic data of March 31 as shown in Figure 2(b). It can be seen that \( r \) has a great influence on the MSE and the MSE gave no entropy values when \( r \) is very small at some scales. Moreover, the MSE at the same \( r \) of various scale changes randomly, which is inconsistent with the fact of the traffic flow. Because these analysis results are conducted under one pattern of traffic flow, they should have the same pattern even under different \( r \). Undoubtedly, it is a hard work to select a suitable \( r \) for the traditional MSE.

As shown in Figure 2(c), the MSE of scale 1 versus different \( r \) is plotted. It can be seen that the MSE gave no values when \( r \) is smaller than 0.1 and the MSE of various \( r \) is greatly different. Therefore, it is difficult to select an appropriate \( r \) for the effective evaluation of traffic data using MSE. A MMSE method is proposed in [25] to solve the problem of discontinuousness boundary caused by the Heaviside function, and the MMSE is constructed by replacing the nonlinear function, sigmoid function, in equation (10) with the Heaviside function.

$$f(d, c) = \frac{1}{1 + \exp[(d - 0.5)/c]}$$  \hspace{1cm} (10)

where \( c \) and \( d \) are parameters for the slope or the steepness of the function. Although the discontinuity problem is well solved in MMSE, it is doubtful that the MMSE is effective to the evaluate the complexity of traffic data. In addition, it is also difficult to select an optimal \( r \) in equation (11) of Castillo et al. [17] for MMSE to accurately evaluate the nonlinearity of the traffic system.

$$D^m_{ij} = f(d^m_{ij}, r) = \frac{1}{1 + \exp[(d^m_{ij} - 0.5)/r]}.$$  \hspace{1cm} (11)

where \( d^m_{ij} \) denotes the similarity degree between two different vectors in MMSE. To illustrate this point, the MMSE is also used to analyze the traffic flow for PeMS. The corresponding result when \( r = 0.1 \) in MMSE is shown in Figure 3(a). As
Figure 1: The traffic data of week model from Monday to Sunday.

Figure 2: Continued.
shown in Figure 3(a), different day patterns including weekday pattern and weekend data pattern cannot be found, which indicates that the MMSE method fails to analyze the complexity of these traffic data. To further explain this, the MMSE values with various $r$ from 0.01 to 0.6 with a step of 0.05 are also calculated for the traffic data on Monday. The result is shown in Figure 3(b), and we can find that the MMSE values are different for various $r$, which shows that $r$ influences MMSE greatly. Therefore, the MMSE may not be a proper method for evaluating the complexity of the traffic system. The problem of meaningless values is well solved by the MMSE method because the method replaces the division with a similarity distance. Consequently, even when two patterns are very similar, the MMSE can also be successfully computed with help of the similarity degree. But we can find that this method also suffered from the problem of coarse evaluation, leading to the weakness of the proposed method.

To evaluate the complexity of these traffic data, the proposed method MSSDE is used to process these traffic data. Because the MSSDE is calculated by counting the number of the same patterns of symbolic vectors rather than counting the number of two vectors smaller than $r$, we can complete the calculation of MSSDE without considering the selection of $r$. Therefore, the problem of traditional MSE about how to select an optimal $r$ is well solved. The traffic data of 7 days are analyzed by the MSSDE, and the result is shown in Figure 4. A
monotonic increase in the MSSDE values versus the scale is found, and the gap between the MSSDE value of Saturday and Sunday is small, while the gap between weekdays and the weekend days is large. This result indicates that Saturday and Sunday fall into the same pattern named weekend pattern and have different patterns with the weekdays, which is consistent with the fact. In particular, the difference between the weekday pattern and the weekend pattern becomes larger with the increase of the scale. As we all know, at workday, people have to go to work or school, while a lot of people prefer to stay at home or nearby areas regularly at weekends. As a result, the traffic data of weekdays are less complex than the traffic data of weekends. Obviously, the traffic data of Saturday and Sunday belong to the different pattern from the other days. It can be inferred that the weekend pattern can be effectively distinguished with the help of MSSDE.

2.2. Case 2: Complexity Analysis of Traffic Flow from the WisTransPortal System. To further demonstrate the effectiveness of the proposed method, traffic flow data obtained from the WisTransPortal system [31] are used. The traffic data collection station is shown in Figure 5, which is marked by a blue triangle. Moreover, it can be found that there exist many companies and hospitals around the station, which indicates that the complexity of these traffic data is similar to the traffic data of Case 1. Specifically, these traffic data probably contain two patterns including the weekday pattern and the weekend pattern. Detector 5561 is used to collect the traffic flow of entrance ramp, while there also exist detectors 5562, 5566, and 5570 to collect the traffic flow of other lanes.

The traffic flow data collected by detector 5561 versus various times are shown in Figure 6(a). To evaluate and learn the complexity of this traffic day, we first apply the proposed MSSDE method to deal with this day, and the corresponding result is shown in Figure 6(b). It can be seen that the difference between the weekday traffic flow and weekend traffic flow is large, which indicates that there also exist a weekday pattern and weekend pattern in the traffic data. This finding is in agreement with the condition of the data collection station, which verifies the performance of our proposed method. The tractional MSE and the MMSE methods are also used, and the results are shown in Figures 6(c) and 6(d), respectively. Figure 6(c) shows that discontinuities occur in the MSE result, and the complexity of the traffic data cannot be illustrated from this result. Moreover, as shown in Figure 6(d), though the discontinuity problem is solved, the weekday and weekend patterns cannot be found from the MMSE result. Compared with our proposed method, both
MSE and MMSE methods fail to analyze the traffic flow. In addition, more traffic data collected by 5562, 5566, and 5570 are also used for verifying the proposed MSSDE. The traffic data versus different times of detectors 5562, 5566, and 5570 are shown in Figures 7(a), 8(a), and 9(a), respectively. The proposed method is applied to these data, and the corresponding results are shown in Figures 7(b), 8(b), and 9(b). It can be found that the MSSDE of Saturday and Sunday is larger than that of other days. The MSSDE of Saturday approximately equals MSSDE of Sunday, which indicates that Saturday and Sunday have the same pattern, weekend pattern. Specifically, two patterns, including weekday pattern and weekend pattern, occur in the traffic system, which is similar to the result of the analysis of traffic data obtained from PeMS. In other words, different patterns of both weekday and weekend can also be clearly distinguished with the help of the proposed method. Therefore, the proposed MSSDE yields excellent performance and can be used for evaluating the complexity of traffic flow.
Figure 7: The result of traffic data collected by detector 5562. (a) The traffic data collected of week model from Monday to Sunday. (b) The result of analysis on the traffic day using the proposed MSSDE.

Figure 8: The result of traffic data collected by detector 5566. (a) The traffic data collected of week model from Monday to Sunday. (b) The result of analysis on the traffic day using the proposed MSSDE.
In this paper, the MSSDE is proposed for accurately analyzing the complexity of traffic data. With help of the proposed method, the problem of discontinuity phenomenon caused by the Heaviside function in traditional multiscale entropy is well solved by the proposed method. Furthermore, the traffic flow can be effectively evaluated without considering the selection of $r$. In addition, the result of the analysis on the traffic data from PeMS and WisTransPortal systems shows that weekday and weekend patterns of the traffic system can be distinguished using the MSSDE, which demonstrates the effectiveness of the proposed method. Therefore, the proposed MSSDE method is able to evaluate the regularity of time series effectively and thus can be more convenient and powerful for traffic system analysis. In the future, we will consider the application of the MSSDE method to the multivariate traffic data analysis. Furthermore, some inspired methods [32–34] will also be considered and introduced for traffic analysis.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Authors’ Contributions

ZC and DL were responsible for investigation and original draft preparation. GC and BL were responsible for methodology and review and editing. ZC was responsible for project administration.

### Acknowledgments

This study was supported in part by the Science and Technology Key Project of Henan Province under grant no. 202102210370.

### References


