2012 Second International Conference on Business Computing and Global Informatization

Using Multiscale Entropy Method to Analyze the Complexity of Traffic Flow

Zhengtao Xiang[1,2], Yujin Li[2], Li Xiong[2]

[1] School of Electrical and Information Engineering
Hubei University of Automotive Technology
Shiyan, China

[2] School of Management
Shanghai University
Shanghai, China

syxiang2000@163.com; yjli@ecust.edu.cn; xiongli8@163.com

***Abstract*—Multiscale entropy (MSE) analysis is a method of measuring the complexity of time series across multiple time scales. The complexity of traffic flow time series generated by NS model is analyzed with MSE. The results show that both of the randomization probability and vehicle density affect the complexity of time headway. And when we focus on large time scale, we found that with the same randomization probabilities, the complexity is at a stable and high level in higher vehicle density scenarios. However, with the same randomization probabilities, the complexity will decrease as time scale increases in lower vehicle density scenarios.**

Keywords-complexity analysis; traffic flow; cellular automata; multiscale entropy

I. INTRODUCTION

Transportation system is a huge and complex system. Complexity of traffic flow is very important to explore the character of transportation system and then can be used in Intelligent Transportation System (ITS) to manage traffic flow. How to manage traffic flow largely depends on the ability to predict traffic flow.

Time series analysis methods can be used to analyze the complexity of traffic flow. However, there is no consensus definition of complexity [1]. Complexity can be defined as the difficulty in predicting the future behavior of time series and can be measured by entropy-based methods. Traditional entropy measures quantify only the regularity or predictability of time series on a single scale [1], and ignore the temporal structural richness. Costa etc. proposed a new complexity metric called multiscale entropy to analyze the time series across multiple time scales [1].

The traffic time series used in this paper are generated by a traffic flow model called NS model, which is based on Cellular Automata (CA). In traffic flow modeling, CA is very useful for microscopic modeling. The update is decided by the current status and a set of rules. With simple rules, CA can simulate the interaction of cells and produce complex behavior.

The remainder of this paper is organized as follows: the related works are presented in Section II. In Section III, the complexity of traffic flow time series is analyzed and the results are discussed. Finally, conclusion and future works are presented.

II. RELATED WORKS

A. Multiscale Entropy

Multiscale entropy (MSE) analysis [1] is a method of measuring the complexity of time series. Before introducing MSE, we should describe approximate and sample entropies because sample entropy is the basis of MSE.

Approximate entropy can be defined as follows [2]. For a discrete random variable X taking values $\{x_1, \ldots, x_N\}$, we form the vectors $x(t)=(u(t), \ldots, u(t+m-1))$ where m is a positive constant and $m \in N$. Each of the vectors has m consecutive values from the original time series and starts from $u(t)$. The distance $d(x(s), x(t))$ is defined as

$$d(x(s), x(t)) = \max_{0\sim m-1} \left| x(s+k), x(t+k) \right|. \quad (1)$$

Then we count the number of $d(x(s), x(t))<r$ for each s and we get the ratio $C_s^m(r)$ as

$$C_s^m(r) = \frac{1}{N-m+1} num\{d(x(s),x(t)) < r\}, \quad s = 1,2,\ldots,N-m+1. \quad (2)$$

The parameter r specifies the tolerance for two sequences to be considered similar. Then we compute

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{s=1}^{N-m+1} \ln C_s^m(r). \quad (3)$$

The approximate entropy can be defined as

$$ApEn(m,r) = \lim_{N\to\infty} (\Phi^m(r) - \Phi^{m+1}(r)). \quad (4)$$

However, when we consider time series with finite length N, $ApEn(m,r)$ can be approximated as

$$ApEn(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r). \quad (5)$$

For parameter m and r, when $m=2$ and $r=0.1\sim0.25SD$, the value of $ApEn(m,r,N)$ has minimal dependence on N, where SD is the standard deviation of $x(t)$.

978-0-7695-4854-8/12 $26.00 © 2012 IEEE
DOI 10.1109/BCGIN.2012.210

To exclude self-matches in *ApEn(m,r,N)*, Richman and Moorman [3] proposed sample entropy. Sample entropy is defined as

$$SampEn(m,r,N) = -\ln(B^{m+1}(r)/B^{m}(r)), \tag{6}$$

where $B^{m+1}(r)$ and $B^{m}(r)$ are empirical probabilities of *m+1* and *m* matches to the template vector, respectively [4].

Multiscale entropy analysis is defined based on sample entropy. Given time series, we can construct multiple coarse-grained time series by averaging the data points within non-overlapping windows of increasing length τ [1]. τ means the scale factor, and the length of each coarse-grained time series is N/τ.For scale 1, the coarse-grained time series is the original time series. For each coarse-grained time series, we can calculate $SampEn(m,r,N)$, and plot the results as a function of scale factor τ. The time series is more complex if the values of multiscale entropy are higher.

B. Complexity Measure of Traffic Flow Time Series

To describe the complexity of traffic flow, many metrics are used.

Lyapunov exponent [5] and fractal dimension [6] are proposed to analyze the complexity of traffic flow. However, these methods face the problem that the computation of such metrics requires rather large amount of volume of time series and has difficulties in real-time implementation. Thus Lempel-Ziv method, approximate entropy, and statistical complexity method are used to quantify complexity.

In [7], the Lempel-Ziv complexity, statistical complexity, approximate entropy are used to analyze the complexity, as well as the combining the criterion methods of chaos & fractal. Approximate entropy is measured to show the relationship between traffic flow complexity and traffic conflict [8].

For application of statistical complexity method in complexity measurement, Xu etc. analyzed the relationship between statistical complexity and traffic conflict rate on urban road with traffic flow on three typical road sections in Nanjing [9]. Yu etc. analyzed how the chaos of traffic flow simulated on a single lane with two-classification vehicles is influenced by the density of traffic flow and the discrete of following-velocity with approximate entropy and statistical complexity [10].

III. Complexity Analysis of Traffic Flow Time Series

A. NS Model of Traffic Flow and Configuration

NS model is a typical one-dimensional traffic model based on CA, which is proposed by Nagel and Sehreckenberg [11]. In NS model, each site has two status, one for occupied by one vehicle, and the other for empty. The velocity of each vehicle is an integer between zero and $v_{\max}$. The rules are listed below:

1) Acceleration: the velocity v of a vehicle will be advanced as $v_t \to v_t + 1$, if $v_t < v_{\max}$;

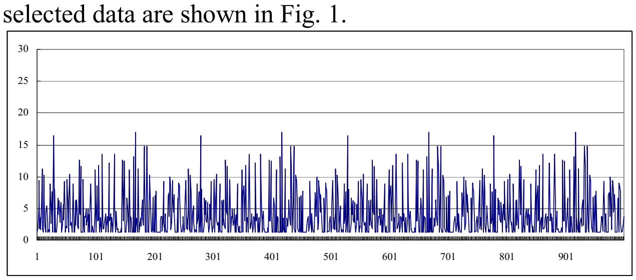
2) Slowing down: the vehicle will reduce its speed v_t to gap, if the vehicle sees the next vehicle ahead and $gap < v_t$, where gap is the distance between them.

3) Randomization: with probability *p*, the vehicle will reduce its speed as $v_t \to \max(v_t - 1, 0)$;

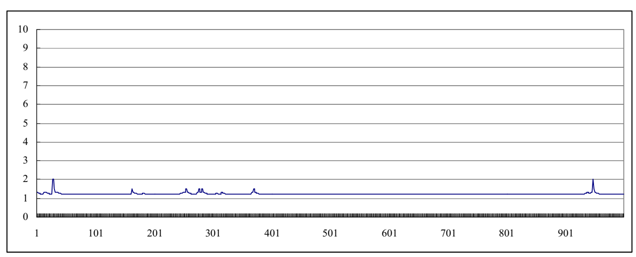
4) Car motion: each vehicle will advance its position as $x_t \to x_t + v_t$.

The parameters of NS model are: the circle lane length *L*, the vehicle density p_{dense}, the maximum of velocity $v_{\max}$, the randomization probability *p*, and the time steps *timeStep*. In our simulation, the configurations are: $L = 5000$; $p_{dense} = \{0.05, 0.1, 0.2, 0.3, 0.4\}$; $v_{\max} = 5$; $p = \{0, 0.1, 0.2, 0.3\}$; *timeStep*=150000. $p=0$ means there is no randomization.

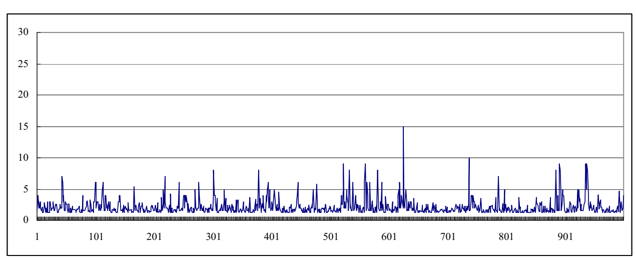
We use the time headway as the time series, which are recorded after the 1000th time step. The *timeStep* is configured to 150000 to guarantee adequate length of the recorded time series for analysis. Thus we can get 20 time series. The time headway is observed at the 100th site. The selected data are shown in Fig. 1.



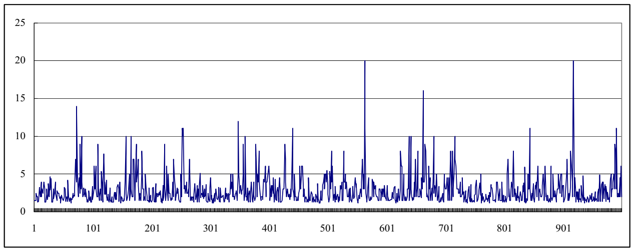
(a) Traffic flow time series with $p_{dense}=0.05$ and $p=0$.



(b) Traffic flow time series with $p_{dense}=0.2$ and $p=0$.



(c) Traffic flow time series with $p_{dense}=0.3$ and $p=0.2$.



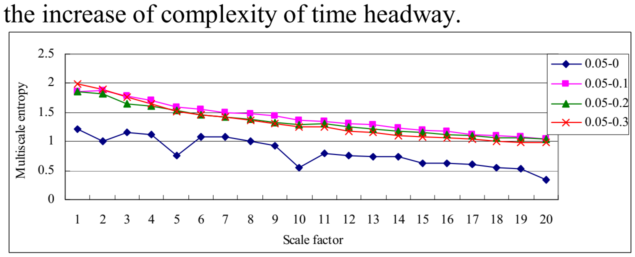
(d) Traffic flow time series with p_{dense}=0.4 and p=0.3.

Figure 1. The selected traffic flow time series.

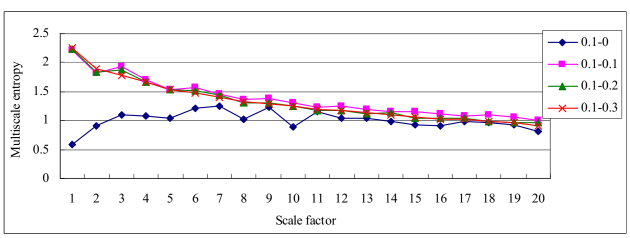
We plot 1000 data of each time series in Fig. 1 because total number of each original time series is too large. And the maximum of y-axis in sub-figure (b) is 10 because the sample values are too low to be plotted clearly if the maximum of y-axis is set to 30.

B. Analysis Results

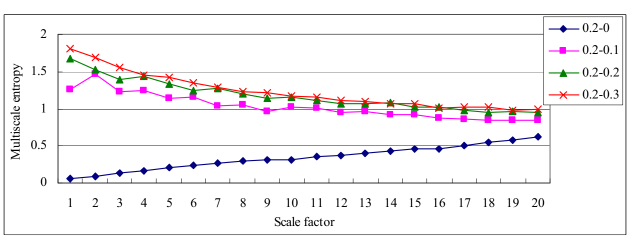
The configurations of MSE are: m=2, r=0.15SD, τ=20. The MSE results for scenarios with same vehicle densities and different randomization probabilities are shown in Fig. 2. With lower vehicle densities, such as 0.05 and 0.1, the deceleration probability due to "Slowing down" is low because the vehicles are in free phase. And the complexity increases with the randomization probability increase, which means the increase of randomization probability will cause the increase of complexity of time headway.



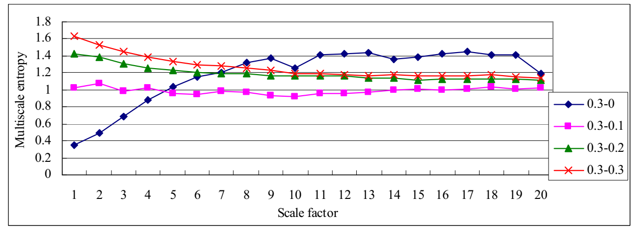
(a) MSE results with p_{dense}=0.05 and p={0,0.1,0.2,0.3}.



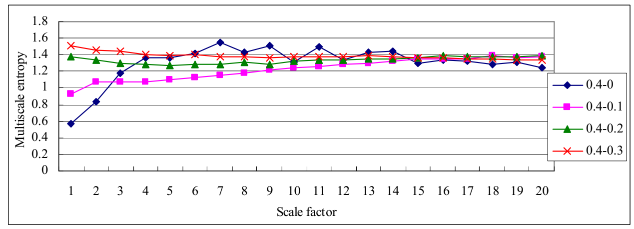
(b) MSE results with p_{dense}=0.1 and p={0,0.1,0.2,0.3}.



(c) MSE results with p_{dense}=0.2 and p={0,0.1,0.2,0.3}.



(d) MSE results with p_{dense}=0.3 and p={0,0.1,0.2,0.3}.



(e) MSE results with p_{dense}=0.4 and p={0,0.1,0.2,0.3}.

Figure 2. MSE results for scenarios with same vehicle densities and different randomization probabilities.

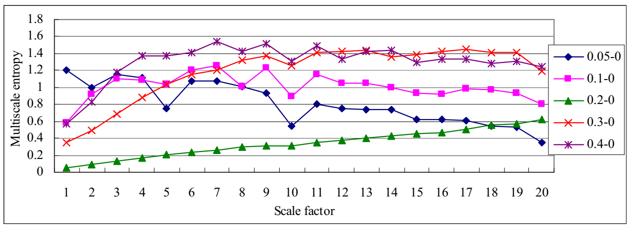
However, with higher vehicle densities, such as 0.3 or 0.4, vehicles are in congestion phase. When the randomization probability is 0.1, the complexity is the minimum as a whole. With higher or lower randomization probabilities, the complexity will increase. If the randomization probability is high, the deceleration probabilities of "Slowing down" of vehicles behind will increase. And if the randomization probability is low, the deceleration probabilities of "Slowing down" of the car itself will increase. Especially, we should note that with the vehicle density of 0.4, the differences of complexities with different randomization probabilities are not distinct, which means the randomization probability is not the major reason of complexity. In general, with high vehicle densities, the major reason of increased complexity of time headway is the operation of "Slowing down".

Also, we compared the results for scenarios with same randomization probabilities and different vehicle densities, as show in Fig. 3. With zero randomization probability, the MSE appears the lowest value when the vehicle density is set to 0.2, which means the complexity is the minimum with such configurations. The result agrees with that in [12]. With other scenarios, the complexity with higher vehicle density is more than that with lower p_{dense}.

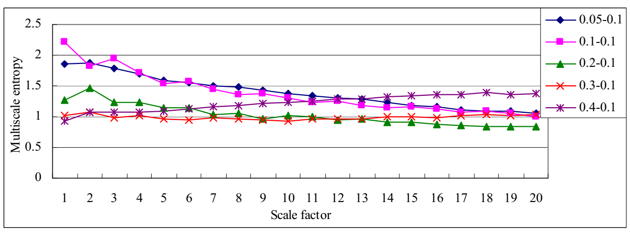
With non-zero randomization probability, the MSE appears stable and high level when the time scale τ increases in higher p_{dense} scenarios, which means the complexity can not be smoothed and the time headway is difficult to be predicted on large time scale. However, the values of MSE decrease with the increase of time scale τ in lower p_{dense} scenarios, which means on large time scale, the time headway is easy to be predicted.

IV. CONCLUSION AND FUTURE WORKS

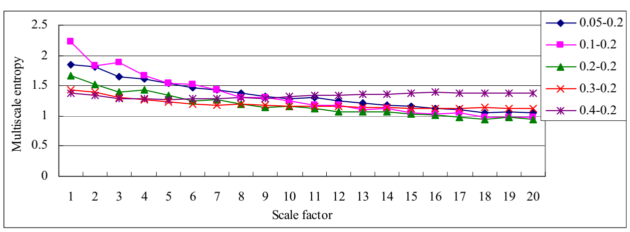
In this paper, the complexity of traffic flow time series generated by NS model is analyzed with multiscale entropy method. The results show that both of the randomization probability and vehicle density affect the complexity of time headway. When focusing on large time scale, we found that with the same randomization probabilities, the complexity is at a stable and high level in higher p_{dense} scenarios. However, with the same randomization probabilities, the complexity will decrease as time scale increases in lower p_{dense} scenarios. The future works include analyzing the impact of "Slowing down" and "Randomization" on the complexity of traffic flow, analyzing the complexity generated by other traffic flow models, and comparing the complexity of realistic and synthetic traffic flows with multiscale entropy method to validate the models.



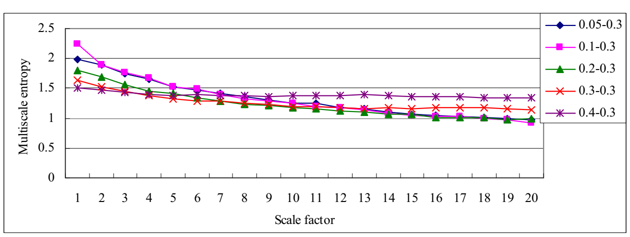
(a) MSE results with p_{dense}=*{0.05,0.1,0.2,0.3,0.4}* and *p=0*.



(b) MSE results with p_{dense}=*{0.05,0.1,0.2,0.3,0.4}* and *p=0.1*.



(c) MSE results with p_{dense}=*{0.05,0.1,0.2,0.3,0.4}* and *p=0.2*.



(d) MSE results with p_{dense}=*{0.05,0.1,0.2,0.3,0.4}* and *p=0.3*.

Figure 3. MSE results for scenarios with same randomization probabilities and different vehicle densities.

ACKNOWLEDGMENT

This research was supported by Shanghai Philosophy and Social Science Planning Project (Grant No. 2011BTQ001), Ph.D. Programs Foundation of Ministry of Education of China (Grant No. 20093108110019), Hubei Key Laboratory of Automotive Power Train and Electronic Control (Hubei University of Automotive Technology) (Grant No. ZDK201002), and Scientific Research Foundation for Doctors of Hubei University of Automotive Technology (Grant No. BK200921).

REFERENCES

[1] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," Phys Rev E, vol. 71, Feb. 2005, pp. 021906.

[2] S. M. Pincus, "Approximate entropy as a measure of system complexity," Proc. the National Academy of Sciences of the United States of America, Mar. 1991, pp. 2297-2301

[3] J. Richman and J. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," American Journal of Physiology-Heart and Circulatory Physiology, vol. 278, Jun. 2000, pp. 2039-2049.

[4] J. Riihijarvi M. Wellens, and P. Mahonen, "Measuring complexity and predictability in networks with multiscale entropy analysis," Proc. IEEE INFOCOM, Apr. 2009, pp. 1107-1115, doi:0.1109/INFCOM. 2009.5062023.

[5] L. W. Lan, J. B. Sheu, and Y. S. Huang, "Investigation of temporal freeway traffic patterns in reconstructed state spaces," Transportation Research Part C, vol. 16, Feb. 2008, pp. 116-13.

[6] L. A. Safonov, E. Tomer, V. V. Strygin et al, "Multifractal chaotic attractors in a system of delay differential equations modeling road traffic," Chaos, vol. 12, Dec. 2002, pp. 1006-1014.

[7] F. T. Liu and Y. Tang, "Criterion and measurement of traffic flow system complexity: Chaos, fractal and complexity," Proc. International Conference on Transportation Engineering , American Society of Civil Engineers, Jul. 2007, pp. 1129-1134.

[8] H. T. Wu, and G. Z. Ma, "Estimating the relationship between traffic flow complexity and traffic conflict," Proc. the 8th International Conference of Chinese Logistics and Transportation Professionals - Logistics: The Emerging Frontiers of Transportation and Development, Jul. 2008, pp. 4625-4630.

[9] Z. H. Xu, X. W. Chen, and M. K. Zhang, "Study on the relationship of traffic flow complexity and traffic conflict on urban road," Proc. the 9th International Conference of Chinese Transportation Professionals, vol. 358, Aug. 2009, pp. 555-561.

[10] D. Yu, X. M. Yin, and J. X. Xie, "The influence of discrete character of following-velocity on chaotic character of traffic flow in different density," Proc. 2009 International Conference on Measuring Technology and Mechatronics Automation, Apr. 2009, pp. 617-621

[11] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," Phys I France, vol. 2, Dec. 1992, pp. 2221-2229, doi:10.1051/jp1:1992277.

[12] X. J. Kong, "Research on modelling and charaeteristics analysis of traffic flow based on cellular automaton," PH.D. thesis, 2007, Beijing Jiaotong University.