Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method

Anyu Cheng [a], Xiao Jiang [a], Yongfu Li [a,*], Chao Zhang [b], Hao Zhu [a]

[a] College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[b] Shanghai Key Laboratory of Financial Information Technology, Shanghai University of Finance and Economics, Shanghai 200433, China

HIGHLIGHTS

- A multiple sources and multiple measures based traffic flow prediction algorithm using the chaos theory and support vector regression method is proposed.
- The chaotic characteristics of traffic flow associated with the speed, occupancy, and flow are identified using the maximum Lyapunov exponent.
- The phase space of multiple measures chaotic time series are reconstructed based on the phase space reconstruction theory.
- The support vector regression (SVR) model is designed to predict the traffic flow.
- Results show that the proposed method has better performance in terms of the accuracy and timeliness.

ARTICLE INFO

Article history:
Received 1 June 2016
Received in revised form 26 August 2016
Available online xxxx

Keywords:
Traffic flow
Chaos theory
Bayesian estimation theory
Support vector regression
Phase space reconstruction
Multiple measures

ABSTRACT

This study proposes a multiple sources and multiple measures based traffic flow prediction algorithm using the chaos theory and support vector regression method. In particular, first, the chaotic characteristics of traffic flow associated with the speed, occupancy, and flow are identified using the maximum Lyapunov exponent. Then, the phase space of multiple measures chaotic time series are reconstructed based on the phase space reconstruction theory and fused into a same multi-dimensional phase space using the Bayesian estimation theory. In addition, the support vector regression (SVR) model is designed to predict the traffic flow. Numerical experiments are performed using the data from multiple sources. The results show that, compared with the single measure, the proposed method has better performance for the short-term traffic flow prediction in terms of the accuracy and timeliness.

1. Introduction

Recently, traffic problems, such as traffic congestion, traffic safety and emissions, have received much attention. To address these issues, various traffic flow models have been proposed to capture the characteristics of transportation system [1–8]. In addition, one effective method is to predict the traffic state accurately and timely. It is beneficial to develop

* Corresponding author.
E-mail address: laf1212@163.com (Y. Li).

http://dx.doi.org/10.1016/j.physa.2016.09.041
0378-4371/

new methodology to migrate the traffic congestion. However, on one hand, the traffic state involves multiple measures such as speed, occupancy, and flow; On the other hand, the inherent nonlinearity and complexity of traffic dynamics impede the real-time prediction accuracy of the traffic flow time series [9]. Hence, it is a challenge to deal with the characteristics of multiple measures and nonlinearity of traffic flow for the prediction.

This study is motivated by the need for a more accurate and timely algorithm so that the traffic state evolution can be predicted effectively. One feature of traffic flow associated with the nonlinearity is the chaotic phenomena. It means that small differences in the initial condition would yield large difference in the prediction results. To identify the chaotic characteristic of traffic flow, a common approach is to calculate the maximum Lyapunov exponent of time series [10]. Several methods have been proposed to calculate the maximum Lyapunov exponent. They include the Wolf method [11], the Jacobian method [12], the P norm algorithm [12], the singular value decomposition method [13], and the small-data method [14] as well as the wavelet transform [15]. Regarding the above mentioned methods, small-data based method has been widely used for the reliability and computational complexity. Consequently, this study will identify the chaotic characteristic of traffic flow using the small-data method.

In the literature, traffic flow prediction models can be roughly classified as linear model [16], nonlinear model [17], intelligent model [18], and other prediction model. From the linear model perspective, it can be further classified as historical average prediction model [19], time series prediction model [20], and Kalman filter model [21]. From the nonlinear model perspective, it can be further classified as the wavelet analysis based model [22] and the catastrophe theory based model [23]. From the other prediction model perspective, it can be further classified as the cellular automata model [24], the fuzzy regression model [25] and the nonparametric regression model [26]. The above mentioned traffic flow prediction models are relatively mature. However, these models restrict the application scope in terms of the accuracy as they do not consider the chaotic characteristics of traffic flow. Therefore, the artificial intelligent (AI) theory based intelligent prediction models has been proposed considering the chaotic characteristics of traffic flow. Since these intelligent models do not rely on the precise mathematical expressions between the dependent and independent variables, so they have strong learning ability and robustness. Hence, this study will adopt the intelligent model, i.e., support vector regression (SVR) model, to predict the traffic flow.

The SVR-based prediction model is a typical intelligent model as it has the potential to identify a highly nonlinear system for its ability to approximate complex nonlinear systems. Hence, SVR-based method plays an important role in short-term traffic flow prediction [27]. In this research line, Hu et al. [28] proposed an algorithm by combining the SVR and particle swarm optimization (PSO) for traffic flow prediction. Ahn et al. [29] proposed a real-time traffic flow prediction based on the Bayesian classifier and SVR. The results showed that the performance of the approach using the SVR-based estimation is better than that of the linear-based regression. Liu [30] further verified the effectiveness of the method by combining the phase reconstruction and SVR. Li et al. [31] proposed the method using the SVR model with Gauss loss function, i.e., Gauss-SVR, to forecast urban traffic flow. Wei et al. [32] proposed an adaptive SVR short-term traffic forecasting model. The results showed that the computational efficiency can be improved, compared with the standard SVR model. In summary, SVR model can avoid the local optima and slow convergence rate. Thereby, SVR model has been widely used for short-term traffic flow prediction.

The SVR models mentioned heretofore focus on the time series prediction of a single traffic measure. Considering the nonlinearity and complexity of traffic system, a single traffic measure is difficult to fully capture the characteristics of the system. Hence, it motivates us to characterize the traffic system using multiple measures such as average speed, average occupancy, and average traffic flow [33]. As a result, this study proposes a multiple sources and multiple measures based traffic flow prediction algorithm using chaos theory and support vector regression method. In particular, first, the chaotic characteristics of traffic flow associated with the speed, occupancy, and flow are identified using the small-data based the maximum Lyapunov exponent calculation. Then, the phase space of multiple measures chaotic time series are reconstructed through the selection of the embedding dimension and delay time based on the phase space reconstruction theory, and fused into a same multi-dimensional phase space using the Bayesian estimation theory. In addition, the SVR model is designed to predict the change trend of the time series of traffic states. Numerical experiments are performed using the data from multiple sources. The results show that, compared with the single measure, the proposed method has better performance for the short-term traffic flow prediction in terms of the accuracy and timeliness.

The rest of this paper is organized as follows. Section 2 presents the methodology used in this study, including the framework, chaotic characteristics identification, phase space reconstruction, Bayesian estimation theory based phase point fusion, and SVR based prediction model. Section 3 performs the numerical experiments to evaluate the performance of the proposed method. Section 4concludes this study.

2. Methodology

In this paper, we study the object for the average speed, average occupancy and average traffic flow time series. These three measures can be integrated into a single time series of the fused traffic flow through the multiple sources integration method [33,34], so that the predicted traffic flow can be dynamically updated more accurately due to the correlation between these three measures. Hence, to improve the prediction accuracy, the multiple sources and multiple measures based traffic flow prediction model will be explained hereafter.

Please cite this article in press as: A. Cheng, et al., Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.09.041