

Research Article

A Noise-Immune Boosting Framework for Short-Term Traffic Flow Forecasting

Shiqiang Zheng ¹, Shuangyi Zhang,¹ Youyi Song ², Zhizhe Lin ³, Dazhi Jiang ^{1,4}, and Teng Zhou ^{1,4}

¹Department of Computer Science, Shantou University, Shantou, China

²Center for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

³Shantou Central Hospital, Affiliated Shantou Hospital of Sun Yat-sen University, Shantou, China

⁴Key Laboratory of Intelligent Manufacturing Technology (Shantou University), Ministry of Education, Shantou, China

Correspondence should be addressed to Teng Zhou; zhouteng@stu.edu.cn

Received 6 January 2021; Revised 16 April 2021; Accepted 7 May 2021; Published 30 May 2021

Academic Editor: chuan lin

Copyright © 2021 Shiqiang Zheng 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.

Accurate short-term traffic flow modeling is an essential prerequisite to analyze and control traffic flow. Canonical data-driven methods are a large account of parameters that may be underfitted with limited training samples, yet they cannot adaptively boost their understanding of the spatiotemporal dependencies of the traffic flow. The noisy and unstable traffic flow data also prevent the models from effectively learning the underlying patterns for forecasting future traffic flow. To address these issues, we propose an easy-to-implement yet effective boosting model based on extreme gradient boosting and enhance it by wavelet denoising for short-term traffic flow forecasting. The discrete wavelet denoising is employed to preprocess the noisy traffic flow data. Then, the denoised training datasets are reconstructed to train the extreme gradient boosting model. These two components are integrated seamlessly in a unified framework, and the whole framework can retain the features in the data as much as possible. Our model can precisely capture the hidden spatial dependency in the data. Extensive experiments are conducted on four benchmark datasets compared with frequently used models. The results demonstrate that the proposed model can precisely capture the hidden spatial dependency of the traffic flow data and achieve superior performance.

1. Introduction

Intelligent transportation system (ITS) plays an important role for traffic management and control [1, 2], which significantly benefits traffic safety enhancement, traffic efficiency, traffic congestion alleviation, and so forth. Accurate traffic flow forecasting in a roadway network provides crucial information for the ITS to implement proactive and efficient traffic management decisions. More specifically, traffic flow forecasting estimates traffic state variation tendency by exploiting traffic flow intrinsic patterns via a large amount of historical data [3]. With the fast development of information and electronic technology, the traffic flow data collection changes from original single-source to multiple sources [4], for example, inductive loops, remote microwave,

Bluetooth, video, and float cars with GPS navigation. However, as the external environment of the transportation system is complicated, the unobservable factors may interfere with the raw traffic data collected from detectors [5]. Such interference results in the degrading of reliable and accurate traffic flow forecasting [6].

Traffic flow is a complex dynamic system [7]. The intrinsic periodicity and correlation indicate governing the evolution of the traffic flow. After years of research efforts, traffic flow modeling has achieved considerable results in both theory and practice. The traffic flow forecasting methods are mainly divided into two categories: model-driven methods and data-driven methods. Model-driven methods include Kalman filtering models [8, 9], k -nearest neighbours [10], and time series models [11]. These kinds of

methods are robust and efficient, but they are expertise handcraft. The most representative data-driven methods for traffic flow forecasting are neural networks, such as recurrent neural networks (RNN) [12] and convolution neural networks (CNN) [13]. However, the performance of these neural networks highly depends on the quality and quantity of the training samples [14]. A large number of parameters in the deep networks may be underfitted with limited training samples or noisy samples and thus result in low training efficiency [15]. In particular, it is easy to gradually propagate the errors and prevent the network from achieving high accuracy. It is difficult for a static learning model to reflect the periodicity, nonlinearity, and randomness of the traffic flow. In recent years, the online boosting models receive substantial attention and have been successfully applied in this field. As an important field of machine learning, the boosting models have unique advantages in time series modeling [16]. The boosting models, such as gradient boosting decision tree (GBDT) and adaptive boost strategy (AdaBoost) [17], exhibit their adaptive learning ability for large-scale distribution processing of traffic flow data. They are widely used in complex systems, such as short-term traffic flow forecasting [18], feature recognition of urban road traffic accidents [19], and taxi travel time forecast [20]. The boosting models, such as gradient boosting machine (GBM) [21] and gradient boosting decision tree (GBDT) [22], can approximate periodicity functions well and perform satisfactorily in specific data and applications. Nevertheless, the boosting model has randomness in the selection of weights and thresholds, which affects the convergence speed and results of the network [23, 24]. However, the aforementioned on-shelf boosting models are complicated for the traffic engineers to be integrated into the existing ITS. Exploring an effective and easy-to-implement model for short-term traffic flow forecasting is still essential.

In this paper, we propose a boosting model based on extreme gradient boosting (XGBoost) enhanced by discrete wavelet denoising, which addresses the two shortcomings we have mentioned above. This idea was first present in a conference [25] and has been admired by transportation engineers. XGBoost is a scalable end-to-end tree boosting system [26], improved from GBDT. It learns a set of regression trees (CARTs) in parallel and obtains the result by summing up the score of each CART [26]. However, the noisy and unstable traffic flow makes the XGBoost difficult to identify underlying patterns for predicting future traffic flow [4]. In this regard, we propose to preprocess the traffic flow data by discrete wavelet denoising, which can reduce the impact of noise in traffic flow. Compared with the original GBDT algorithm, one of the special improvements is the regularized objective of the loss function. We further take the spatiotemporal correlation of the traffic flow into consideration. We reconstruct the traffic flow datasets by involving the phase space reconstruction theory. In the end, XGBoost is executed to forecast future traffic flow. The performance of XGBoost for traffic flow forecasting is greatly improved and ensures accuracy and robustness. This work was first accepted as a poster at the 8th International Conference on Digital Home [25]. We have refined our

model, reconducted most of our experiments, and rewritten our paper.

The main contributions of this work are listed as follows:

- We construct a boosting model for traffic flow forecasting enhanced by discrete wavelet denoising

- We investigate the forecasting performance of different mother wavelets to reveal the best one for traffic flow, and we reconstruct the traffic flow by considering the phase space reconstruction theory

- We evaluate the proposed model on four benchmark datasets

The results demonstrate that the proposed model outperforms frequently used models with lower computation cost.

2. Related Works

2.1. Wavelet Denoising. Denoising algorithm has received considerable attention in various fields [8, 9]. Most of the conventional filtering techniques, such as mean filter, Gaussian filter, and minimum mean squared error filter, cannot always guarantee the acceptable quality of denoised traffic data [27]. In recent decades, the discrete wavelet transform (DWT) has been applied to dispose of the problem of noise reduction, and it has outperformed traditional filters in terms of root mean squared error (RMSE), PSNR, and other evaluations [28].

The wavelet denoising algorithm has been well acknowledged as an essential method. In mathematics, the essence of wavelet denoising is a function approximation problem, in other words, finding the best approximation of the original signal in the wavelet space developed by the scaling and translation of the Wavelet generating function, according to the proposed criteria, to achieve the complete distinction between the original signal and the noise signal. Compared with the noise feature, the larger amplitude in the wavelet domain is the coefficients with important signal characteristics, while the amplitude of noise coefficients is smaller. Therefore, the wavelet coefficients with larger absolute value can be retained or contracted only by setting an appropriate threshold, and the estimated wavelet coefficients (EWC) have been obtained.

For the traffic flow data case, the wavelet denoising algorithm transforms the data to a time-frequency domain under DWT processing. Then we could keep only some large coefficients and throw away the rest using a proper threshold level, too. The result is that a small number of largest coefficients which have crucial information are saved, while most of noise coefficients that are small will be discarded. If we use DWT to decompose the high-frequency noise from the original traffic flow data, the periodical pattern in the traffic flow would be identified efficiently. Furthermore, an example of the application of wavelet analysis in traffic flow denoising is demonstrated in Figure 1. From Figure 1, by comparing the actual data and the denoised data, we can see that the waveform of the denoised data is much smoother than the one of the real data. We expect that the denoised

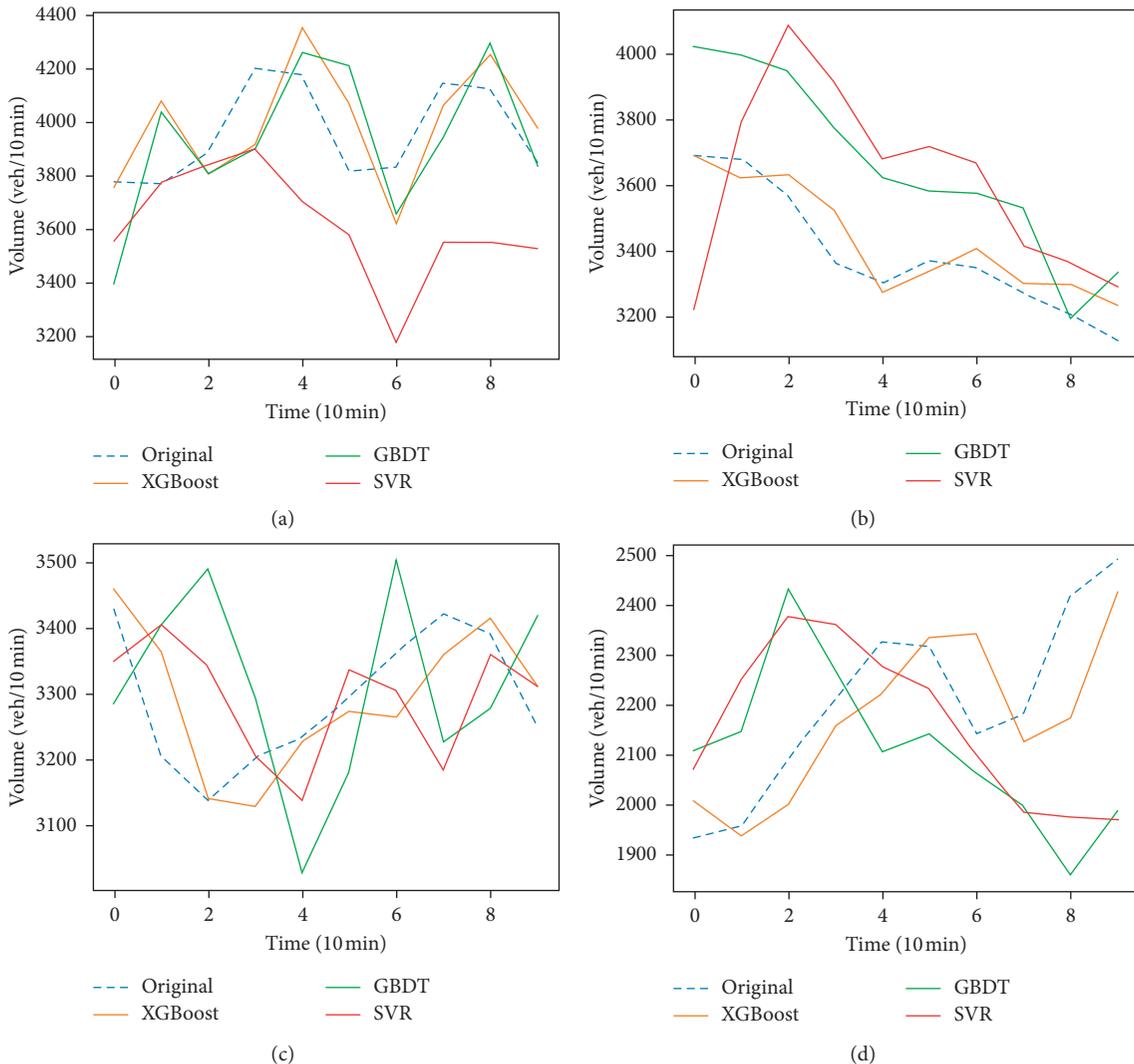


FIGURE 1: The comparison with frequently used models at four challenging scenarios.

data positively affects the following traffic flow analysis and prediction.

2.2. XGBoost. XGBoost is a scalable machine learning system for tree boosting. The system’s impact has been widely recognized in several machine learning and data mining challenges. The system is widely applied in domains such as high energy physics event classification, customer behavior prediction, ad click-through rate prediction, and massive online course dropout rate prediction [29]. The most crucial factor behind the success of XGBoost is its scalability. The system runs more than ten times faster than existing popular solutions on a single machine and scales to billions of examples in distributed or memory-limited settings. The scalability of XGBoost is due to several essential systems and algorithmic optimizations [26]. These innovations include a novel tree learning algorithm for handling sparse data and a theoretically justified weighted quantile sketch procedure, enabling handling instance weights in approximate tree learning. Parallel and distributed computing make learning

faster, which enables quicker model exploration. More importantly, XGBoost exploits out-of-core computation and enables data scientists to process a hundred million of examples on a desktop. Finally, it is even more exciting to combine these techniques to make an end-to-end system that scales to even more extensive data with the least cluster resources.

3. Methodology

In this section, we first give the mathematical definition of the short-term traffic flow forecasting task. Then, we propose to preprocess the raw traffic flow data by wavelet denoising. After that, the adaptive gradient boost algorithm performs more effectively on the denoised traffic flow data Algorithm 1.

3.1. Traffic Flow Formulation. We first formulate the traffic flow at k th measurement point of the road network at t th time interval as $v_t^{(k)}$. Thus, the traffic flow of the road

```

Input:  $V(t)$ , date set of the traffic flow
Output: Split and default directions with max gain;
(1) Step1: Decompose the data  $v(t)$  wavelet into high-frequency information  $cD$  and low-frequency information  $cA$ ;
(2) Step2: Reduce the sampling rate of high-frequency information  $cD$  to half to get new high-frequency information  $l$ ;
(3) Step3: Decompose the low-frequency information  $cA$  and the new high-frequency information  $l$  by inverse wavelet to obtain the
reconstructed data;
(4) Step4: Import the reconstructed data into xgboost for training;
(5)  $gain \leftarrow 0$ 
(6)  $G \leftarrow \sum_{i \in I} g_i, H \leftarrow \sum_{i \in I} h_i$ 
(7) for  $k = 1$  to  $\tau$  do
(8) //enumerate missing value goto right
(9)  $G_l \leftarrow 0, H_l \leftarrow 0$ 
(10) for  $j$  in sorted( $I_k$ , ascent order by  $s_{jk}$ ) do
(11)  $G_l \leftarrow G_l + g_j, H_l \leftarrow H_l + h_j$ 
(12)  $G_r \leftarrow G - G_l, H_r \leftarrow H - H_l$ 
(13)  $score \leftarrow \max(score, (G_l^2/H_l + \lambda) + (G_r^2/H_r + \lambda) - (G^2/H + \lambda))$ 
(14) end for
(15) //enumerate missing value goto right
(16)  $G_r \leftarrow 0, H_r \leftarrow 0$ 
(17) for  $j$  in sorted( $I_k$ , ascent order by  $s_{jk}$ ) do
(18)  $G_r \leftarrow G_r + g_j, H_r \leftarrow H_r + h_j$ 
(19)  $G_l \leftarrow G - G_r, H_l \leftarrow H - H_r$ 
(20)  $score \leftarrow \max(score, (G_l^2/H_l + \lambda) + (G_r^2/H_r + \lambda) - (G^2/H + \lambda))$ 
(21) end for
(22) end for

```

ALGORITHM 1: The proposed framework for short-term traffic flow forecasting.

network at t th time interval can be denoted as $v_t = \{v_t^{(1)}, v_t^{(2)}, \dots, v_t^{(K)}\}$, where K is the number of measurement points on the road network. The traffic flow forecasting task is to train a model \mathcal{F} to predict the traffic flow given a dataset $S = \{s^{(1)}, s^{(2)}, \dots, s^{(T)}\}$, where T is the number of training samples. $s^{(t)} = \{v_{t-\tau+1}, \dots, v_t\}$, and τ is the time lag. In this way, given a query sample $s^{(q)}$, traffic flow prediction at $t + 1$ time interval can be denoted as \hat{v}_{q+1} .

$$\hat{v}_{q+1} = \mathcal{F}(s^{(q)}), \quad (1)$$

3.2. Traffic Flow Preprocessing. As previously mentioned, traffic flow consists of period trend in low frequency and minute-to-minute fluctuations in high frequency [9]. The high-frequency traffic flow fluctuations are often considered as the noises to the periodical traffic flow trend [30]. If these high-frequency noises are learned by a statistical learning models, the models will produce unstable predictions for future traffic flow. Hence, it is important to eliminate the noises in the traffic flow to concentrate the learning-based models to the evolution trend of the traffic flow. In this regard, we propose adopting a wavelet denoising method to eliminate the high-frequency noises in the traffic flow to avoid the learning-based models to learn the minute-to-minute fluctuations.

The wavelet decomposition transforms the original traffic data into a couple of oscillatory waveforms in different frequencies, and the structure of each waveform at a specific instant can be determined. In this way, the traffic flow signal can be localized in both the time and frequency domains.

Relying on this property, the wavelet transform is widely adopted for traffic flow denoising.

Given a mother wavelet $\varphi(\omega)$ (i.e., dbN wavelet), the continuous wavelet transform (CWT) of a signal is denoted as

$$c_\varphi = \int_{-\infty}^{+\infty} |\omega| |\varphi(\omega)|^2 d\omega < \infty, \quad (2)$$

where a is the scale or dilation parameter and b is the translation parameter that reflects the position information of the wavelet according to the time information.

Further, since the traffic data is discrete, the wavelet transform is discrete as discrete wavelet transform (DWT). DWT parameter a and the translation parameter b are discretized by the dyadic sequence; for example, $a = a_0^m$, $b = nb_0 a_0^m$, and $m, n \in \mathbb{Z}$.

When $a_0 = 2$ and $b_0 = 1$, the DWT degrades to binary wavelet.

$$\varphi_{m,n}(t) = 2^{-(m/2)} \varphi(2^{-m}t - n), \quad (3)$$

where $X_h(t)$ and $X_l(t)$ are the high-frequency information and low-frequency information of traffic flow, which are calculated by the following equations:

$$X_h(t) = (V * h)(t) = \sum_{k=0}^{N-1} X(k)h(2t - k), \quad (4)$$

$$X_l(t) = (V * l)(t) = \sum_{k=0}^{N-1} X(k)h(2t - k). \quad (5)$$

Then, the high-frequency information is handled by a threshold.

$$X_h(t) = \begin{cases} X_h(t), & \text{if } X_h(t) < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

After the wavelet decomposition process, the high-frequency information and low-frequency information of traffic flow are reconstructed. Reconstruction is the inverse process of decomposition. After an upsampling process of the high-frequency information and low-frequency information, the new training label $X(t)$ is obtained by convolving the inverse transformation of high-pass and low-pass filter with the coefficients.

$$\begin{aligned} X(t) &= (X_l^* L)(t) + (X_h^* H)(t) \\ &= \sum_{k=0}^{N-1} X_l(k)L(t-k) + \sum_{k=0}^{N-1} X_h(k)H(t-k). \end{aligned} \quad (7)$$

By this transformation, the proposed technique removed the high-frequency noise in the traffic flow signal and also can preserve the quality of the original data, fulfilling our purpose.

3.3. Model Training. The extreme gradient boosting (XGBoost) is an efficient tool for large-scale parallel boosted trees, which can be effectively applied to classification and regression tasks [26]. The XGBoost improves the gradient boost decision tree (GBDT) by enhancing parallel computing, approximate tree building, and sparse data processing. It also optimizes the usage of computational cost, making it suitable for multidimensional data feature recognition and classification.

In this paper, we first transform the traffic flow forecasting into a supervised learning task. Different from GBDT, XGBoost adds a regularization term to the objective function to reduce the complexity of the model and avoid overfitting.

$$\text{Obj}(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \quad (8)$$

where \hat{y}_i is the prediction, y_i is the ground truth, $\Omega(f_k)$ is the regular term, f_k is a decision tree, p represents the number of leaf nodes, Ω represents the fraction of leaf nodes, γ controls the number of leaf nodes, and λ controls the fraction of leaf nodes. The objective function constructed by the iteration of the XGBoost is

$$\text{Obj}^p = \sum_{i=1}^n [l(y_i, \hat{y}_i^{(p-1)}) + f_i(x_i)] + \Omega(f_p). \quad (9)$$

By the second-order Taylor process, the convergence speed of the model is accelerated, and the optimal global solution is obtained.

$$\text{Obj}^p = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(p-1)}) + g_t f_p(x_i) + \frac{1}{2} h_t f_p^2(x_i) \right] + \Omega(f_p), \quad (10)$$

where $g_t = \partial_{\hat{y}_t^{(p-1)}} l(y_t, \hat{y}_t^{(p-1)})$ is a first-order derivative and $h_t = \partial_{\hat{y}_t^{(p-1)}}^2 l(y_t, \hat{y}_t^{(p-1)})$ is a second-order derivative. The experiment tries to add partitions to the existing leaf nodes in each step for generating the optimal tree structure. The splitting gain is

$$\text{gain} = \frac{1}{2} \left[\frac{g_l^2}{h_l^2 + \lambda} + \frac{g_r^2}{h_r^2 + \lambda} - \frac{(g_l + g_r)^2}{h_l + h_r + \lambda} \right] - \gamma. \quad (11)$$

When the splitting gain is less than the fixed value or the number of times the division reaches the specified maximum depth, the division stops. We can get the final regression model. Traffic flow prediction is essentially a regression prediction task, so we use regressor as the base learner of XGBoost. We put the wavelet transformed traffic flow data into XGBoost for training. In each training, we minimize each regressor to fit the residual error generated by the last leaf and calculate the split score to determine whether to generate a new leaf. Finally, we can obtain the predicted traffic flow data by adding each leaf. The proposed framework is provided as follows. The computational complexity of the proposed method is slightly higher, and the wavelet denoising and XGBoost are suitable for parallel computing. This means that we can preprocess a proportion of data to train the XGBoost model. As the XGBoost model is training or prediction, we simultaneously preprocess the next batch of data to update the model. The core of XGBoost is approximate calculation, the complexity of which is $O(K d \|x\| \log q)$, where d is the maximum depth of the tree and K is the total number of trees.

4. Experiments

4.1. Data Description. We employ four benchmark datasets to evaluate the performance of the proposed method. The traffic flow data were collected from four sites on the highways ending on the ring road in Amsterdam, Netherlands, a short distance before the merged points. The data on the four sites (i.e., A1, A2, A4, and A8) were collected from May 20, 2010, to June 24, 2010. Highway A1 connects Amsterdam and the German border. It is the first high-capacity road, and its flow pattern is difficult to find. The A2 highway is one of the busiest highways in Netherlands, connecting Amsterdam and the Belgian border. In the experiment, we used the data before widening. It can test whether our model can predict congestion well. The A4 motorway is part of Rijksweg 4, starting from Amsterdam to the Belgian border. It is representative of a mature highway, which can prove the universality of the model. The A8 highway starts from the A10 road at the Coenplein interchange and is less than 10 kilometers from Zaandijk. Because the road has more connections with other highways, the vehicles' speed on the road is constantly changing. By predicting the road as mentioned above traffic flow, we can study the impact of expressway capacity, travel time changes, and model prediction accidents. The 1-minute average traffic data over five weeks were collected by MONICA sensors (velocity-flow measurement points). The datasets are split in

chronological order with 70% for training, 10% for validation, and 20% for testing. Missing values are excluded from both training and testing.

4.2. *Baselines.* We compare XGBoost with the following frequently used models in intelligent transportation systems:

Decision tree (DT) is a decision support tool that uses a tree-like model of decisions and their possible consequences

Artificial neural networks (ANN) are computer programs inspired by biological design to simulate how the human brain processes information

ANNs gather their knowledge by detecting the patterns and relationships in data and learn (or are trained) through experience, not from programming.

Support vector regression (SVR) is a version of support vector machines (SVM) for regression

Gradient boosting decision tree (GBDT) is an iterative decision tree algorithm

The algorithm consists of multiple decision trees, and the conclusions of all trees are accumulated to make the final answer.

4.3. *Experimental Setup.* We use two criteria, root mean squared error (RMSE) and mean absolute percentage error (MAPE), to evaluate the performance of the proposed model, as defined in the two following equations, respectively.

$$\text{RMSE} = \sqrt{\sum_{t=1}^n \frac{(y_t - \hat{y}_t)^2}{n}}, \quad (12)$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\%. \quad (13)$$

Our experiments are conducted under a computational environment of Intel Core i7@3.60 GHz with 8 GB RAM. To determine the number of optimal lags in the model, we set the forecasting lags of $n = \{6, 8, 10, 12, 15, 20\}$ for the model and use the MAPE to evaluate the optimal lags. Figure 2 shows that when $n = 12$, the value of MAPE is the lowest, so we set the lags for the forecasting at 12. We train our model using an optimizer with an initial learning rate of 0.01, the number of decision trees of 100, and the random sampling ratio of each tree of 0.5.

4.4. *Performance Evaluation.* Table 1 lists the performances of denoised XGBoost model and baseline models for 10-minute ahead prediction on four benchmark datasets. XGBoost archives superior results on all datasets. It outperforms frequently used models, including DT, ANN, SVR, and GBDT. Compared to other prediction models, XGBoost outperforms the DT, ANN, and SVR significantly.

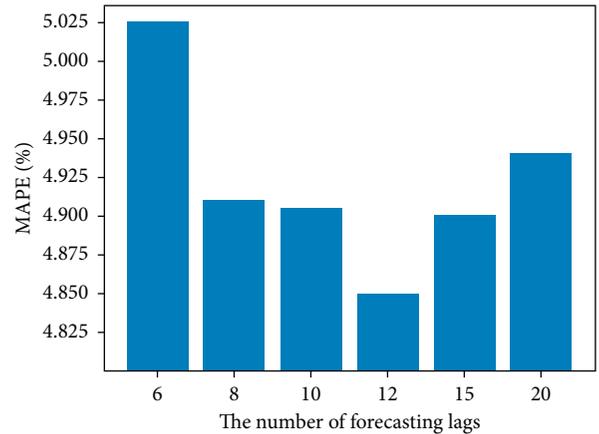


FIGURE 2: The MAPE of the model with regard to the forecasting lags.

TABLE 1: Comparison of prediction performances of various models.

		A1	A2	A4	A8
DT	MAPE	14.76	11.60	12.89	12.62
	RMSE	246.58	204.06	236.02	203.22
ANN	MAPE	10.67	10.89	10.73	11.34
	RMSE	287.60	249.35	233.91	150.91
SVR	MAPE	8.44	10.20	8.44	8.67
	RMSE	215.07	247.05	170.35	130.48
GBDT	MAPE	7.87	6.87	7.94	7.76
	RMSE	182.67	127.42	138.94	111.07
SVRGSA [23]	MAPE	11.15	9.42	10.65	11.81
	RMSE	284.97	192.68	213.69	161.07
SVRPSO [24]	MAPE	11.63	10.08	10.99	12.20
	RMSE	300.97	205.94	224.63	163.95
OiKF [4]	MAPE	8.57	7.93	10.61	11.56
	RMSE	203.34	154.98	184.96	132.41
XGBoost	MAPE	5.64	4.85	6.87	6.70
	RMSE	157.11	111.01	132.30	100.38

Regarding the second-best model GBDT in Table 1, XGBoost achieves more accurate predictions over GBDT on all the datasets. This is because our model eliminates the noises in the traffic flow, which prevent the learning-based models from learning the temporal dependencies. Therefore, XGBoost can discover implicit relationships within data.

We also compare the proposed framework with the conventional models in Table 2. From Table 2, we can observe that XGBoost outperforms the conventional ones.

4.5. *Ablation Study.* The observed traffic flow data quality is crucial for traffic flow prediction accuracy, and thus data quality control is essential to smooth the noisy traffic flow data. To comprehensively compare varied denoising framework performance, we employ the wavelet denoising model with different wavelet bases to preprocess the raw data. The RMSE and MAPE statistics help us analyze varied smoothing methods quantitatively. Overall, there is no significant difference between varied smoothing models at the same time span data samples. Taking traffic flow

TABLE 2: Comparison with the conventional models.

		A1	A2	A4	A8
Historical average	MAPE	7.13	5.97	7.13	7.49
	RMSE	182.31	116.48	137.75	103.45
Kalman filter	MAPE	165.17	189.80	145.95	137.80
	RMSE	1267.49	1291.23	1100.64	702.49
XGBoost	MAPE	5.64	4.85	6.87	6.70
	RMSE	157.11	111.01	132.30	100.38

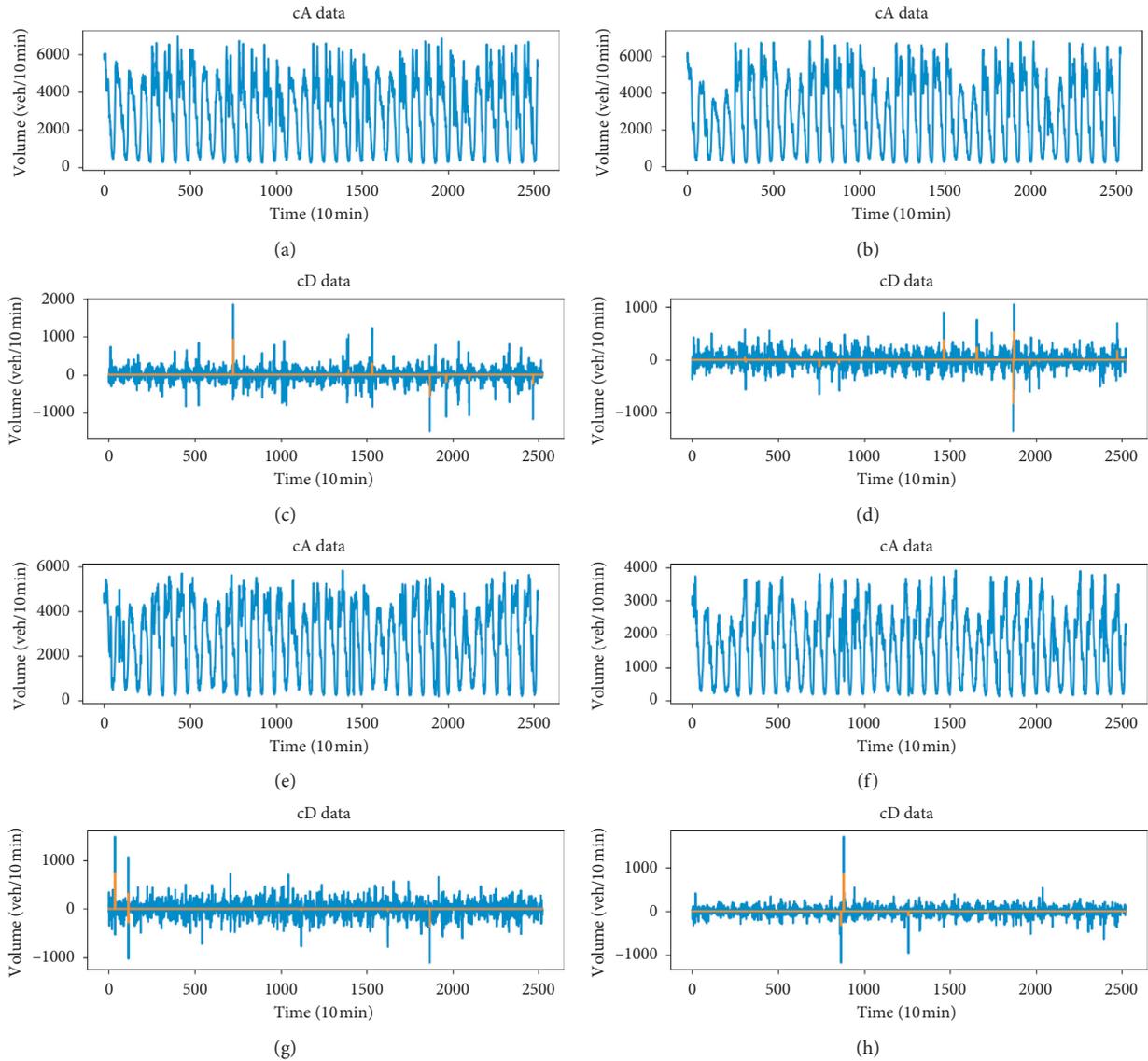


FIGURE 3: Denoising effect of traffic flow data with sensor A1 under 10 min scale.

denoising results on the 10 min data from sensor A1 as an example, the db4 model obtains optimal noise removal performance compared to other wavelet based denoising results. To further examine the denoising effects of different models, we looked at smoothing details of how each of the models addresses the outliers in the original traffic flow data. It is observed that the wavelet denoising can successfully

smooth the anomaly oscillations without discarding data details. Taking the denoising effect on data samples with sensor A1 at a 10 min scale as an example, the variation tendency was successfully shown in the denoised traffic flow data in Figure 3. Table 3 shows a similar smoothing result for the traffic flow data at sensor A1 under 10 min. In sum, various wavelet-based models showed similar results on

TABLE 3: The performances of different mother wavelets.

	Nonwavlet	Haar	db4	sym4
MAPE (%)	11.71	6.81	5.64	5.88
RMSE (vehs/h)	287.93	209.06	157.11	176.05

TABLE 4: The performance of the model in different step size.

		10	15	20	30
A1	MAPE (%)	5.64	5.63	5.66	5.56
	RMSE (vehs/h)	157.11	158.13	159.23	151.31
A2	MAPE (%)	4.85	4.90	4.94	4.95
	RMSE (vehs/h)	111.01	111.32	109.74	111.39
A4	MAPE (%)	6.87	6.86	6.70	6.54
	RMSE (vehs/h)	132.30	132.04	130.21	129.20
A8	MAPE (%)	6.70	6.71	6.62	6.63
	RMSE (vehs/h)	100.38	100.52	100.70	99.07

TABLE 5: Significant test result.

	Df	sum_sq	mean_sq	F	PR
C (method)	4.0	119.250250	29.812562	35.890103	0.000001
C (road)	3.0	1.122375	0.374125	0.450394	0.721672
Residual	12.0	9.967950	0.830662	NaN	NaN

suppressing the data outliers, and db4 obtained slightly better performance compared to other smoothing methods. We also compare the performance of the model when the window is 20 and 30 in Table 4.

The following significant test for the experiment results is illustrated in Table 5. From Table 5, the *PR* value of the model is 0.000001, which is much smaller than 0.01. The *PR* value of the highway is 0.72, which is much larger than 0.01. In this regard, we conclude that the choice of model has a significant influence, while the choice of highway has no influence.

4.6. Cases Study. Figure 1 shows the result for comparing the observed traffic flow data and the prediction value of the A1 detector in the test sample. XGBoost model has a more accurate and smoother result of short-term traffic flow. The integrated model is more suitable for periodic data than canonical methods.

5. Conclusion

In this paper, we propose an easy-to-implement and effective boosting model for accurate short-term traffic flow forecasting. The noisy and unstable traffic flow data are firstly preprocessed by discrete wavelet denoising. We conclude that the db4 mother wavelet is more suitable for traffic flow data denoising by ablation study. Then, the extreme gradient boosting model is trained by the denoised dataset reconstructed by considering the phase space reconstruction theory. Extensive experiments on four benchmark datasets reveal that the proposed method can better learn the periodicity, nonlinearity, and randomness of the traffic flow. The

result demonstrates the outperformance by comparing the proposed model with the frequently used models. In the future, we plan to extend the method to other forecasting applications, such as load forecasting and taxi demand forecasting.

Data Availability

The data and source code that support the findings of this study are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the NSFC (no. 61902232), the Natural Science Foundation of Guangdong Province (no. 2018A030313291), the Education Science Planning Project of Guangdong Province (2018GXJK048), the STU Scientific Research Foundation for Talents (NTF18006), and the 2020 Li Ka Shing Foundation Cross-Disciplinary Research Grant (no. 2020LKSG05D).

References

- [1] H. Lu, D. Huang, Y. Song, D. Jiang, T. Zhou, and J. Qin, "St-trafficnet: a spatial-temporal deep learning network for traffic forecasting," *Electronics*, vol. 9, no. 9, pp. 1474–1517, 2020.
- [2] Z. Ge, Y. Li, C. Liang et al., "Adaptive cross-scale network with feature maps refusion for vehicle density detection," in

- Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2021)*, Shenzhen, China, July 2019.
- [3] H. Lu, Z. Ge, Y. Song, D. Jiang, T. Zhou, and J. Qin, "A temporal-aware lstm enhanced by loss-switch mechanism for traffic flow forecasting," *Neurocomputing*, vol. 427, pp. 169–178, 2021.
 - [4] S. Zhang, Y. Song, D. Jiang, T. Zhou, and J. Qin, "Noise-identified kalman filter for short-term traffic flow forecasting," in *Proceedings of 2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pp. 462–466, IEEE, Shenzhen, China, December 2019.
 - [5] Y. Lin, L. Li, H. Jing, B. Ran, and D. Sun, "Automated traffic incident detection with a smaller dataset based on generative adversarial networks," *Accident Analysis & Prevention*, vol. 144, Article ID 105628, 2020.
 - [6] L. Li, B. Du, Y. Wang, L. Qin, and H. Tan, "Estimation of missing values in heterogeneous traffic data: application of multimodal deep learning model," *Knowledge-Based Systems*, vol. 194, Article ID 105592, 2020.
 - [7] L. Li, L. Qin, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowledge-Based Systems*, vol. 172, pp. 1–14, 2019.
 - [8] L. Cai, Z. Zhang, J. Yang, Y. Yu, T. Zhou, and J. Qin, "A noise-immune kalman filter for short-term traffic flow forecasting," *Physica A: Statistical Mechanics and Its Applications*, vol. 536, Article ID 122601, 2019.
 - [9] T. Zhou, D. Jiang, Z. Lin, G. Han, X. Xu, and J. Qin, "Hybrid dual Kalman filtering model for short-term traffic flow forecasting," *IET Intelligent Transport Systems*, vol. 13, no. 6, pp. 1023–1032, 2019.
 - [10] L. Cai, Y. Yu, S. Zhang, Y. Song, Z. Xiong, and T. Zhou, "A sample-rebalanced outlier-rejected k -nearest neighbor regression model for short-term traffic flow forecasting," *IEEE Access*, vol. 8, pp. 22686–22696, 2020.
 - [11] S. Shahriari, M. Ghasri, S. A. Sisson, and T. Rashidi, "Ensemble of arima: combining parametric and bootstrapping technique for traffic flow prediction," *Transportmetrica A: Transport Science*, vol. 16, no. 3, pp. 1552–1573, 2020.
 - [12] Z. Lv, J. Xu, K. Zheng, H. Yin, P. Zhao, and X. Zhou, "Lc-rnn: a deep learning model for traffic speed prediction," in *Proceedings of International Joint Conferences on Artificial Intelligence IJCAI*, pp. 3470–3476, Virtual, Japan, August 2018.
 - [13] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and cnn deep learning," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1688–1711, 2019.
 - [14] Y. Li, Z. Ge, Z. Zhang et al., "Broad learning enhanced 1h-mrs for early diagnosis of neuropsychiatric systemic lupus erythematosus," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 8874521, 13 pages, 2020.
 - [15] Y. Song, Z. Yu, T. Zhou et al., "Cnn in ct image segmentation: beyond loss function for exploiting ground truth images," in *Proceedings of 2020 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 325–328, IEEE, Iowa City, IA, USA, May 2020.
 - [16] T. Zhou, G. Han, X. Xu, C. Han, Y. Huang, and J. Qin, "A learning-based multimodel integrated framework for dynamic traffic flow forecasting," *Neural Processing Letters*, vol. 49, no. 1, pp. 407–430, 2019.
 - [17] T. Zhou, G. Han, X. Xu et al., " δ -agree AdaBoost stacked autoencoder for short-term traffic flow forecasting," *Neurocomputing*, vol. 247, no. 4, pp. 31–38, 2017.
 - [18] L. Cai, M. Lei, S. Zhang, Y. Yu, T. Zhou, and J. Qin, "A noise-immune lstm network for short-term traffic flow forecasting," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 2, Article ID 023135, 2020.
 - [19] Y. Qu, Z. Lin, H. Li, and X. Zhang, "Feature recognition of urban road traffic accidents based on ga-xgboost in the context of big data," *IEEE Access*, vol. 7, pp. 170106–170115, 2019.
 - [20] K. D. Kankanamge, Y. R. Witharanage, C. S. Withanage, M. Hansini, D. Lakmal, and U. Thayasivam, "Taxi trip travel time prediction with isolated xgboost regression," in *Proceedings of 2019 Moratuwa Engineering Research Conference (MERCon)*, pp. 54–59, IEEE, Moratuwa, Sri Lanka, July 2019.
 - [21] C. Ding, X. Wu, G. Yu, and Y. Wang, "A gradient boosting logit model to investigate driver's stop-or-run behavior at signalized intersections using high-resolution traffic data," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 225–238, 2016.
 - [22] S. Yang, J. Wu, Y. Du, Y. He, and X. Chen, "Ensemble learning for short-term traffic prediction based on gradient boosting machine," *Journal of Sensors*, vol. 2017, Article ID 7074143, 15 pages, 2017.
 - [23] L. Cai, Q. Chen, W. Cai, X. Xu, T. Zhou, and J. Qin, "SVRGSA: a hybrid learning based model for short-term traffic flow forecasting," *IET Intelligent Transport Systems*, vol. 13, no. 9, pp. 1348–1355, 2019.
 - [24] W. Cai, J. Yang, Y. Yu et al., "PSO-ELM: a hybrid learning model for short-term traffic flow forecasting," *IEEE Access*, vol. 8, pp. 6505–6514, 2020.
 - [25] S. Zheng, S. Zhang, Y. Song, Z. Lin, F. Wang, and T. Zhou, "A noise-eliminated gradient boosting model for short-term traffic flow forecasting," in *Proceedings of The 8th International Conference on Digital Home*, New York, NY, USA, May 2020.
 - [26] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, San Francisco, CA, USA, August 2016.
 - [27] L. Zhu, Y. Zhu, H. Mao, and M. Gu, "A new method for sparse signal denoising based on compressed sensing," in *Proceedings of 2009 Second International Symposium on Knowledge Acquisition and Modeling*, pp. 35–38, IEEE, Wuhan, China, December 2009.
 - [28] A. Tavakoli and A. Pourmohammad, "Image denoising based on compressed sensing," *International Journal of Computer Theory and Engineering*, vol. 4, no. 2, pp. 266–269, 2012.
 - [29] S. Tyree, K. Q. Weinberger, K. Agrawal, and J. Paykin, "Parallel boosted regression trees for web search ranking," in *Proceedings of the 20th international conference on World wide web*, pp. 387–396, Hyderabad, India, March 2011.
 - [30] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using kalman filter with discrete wavelet decomposition," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 326–334, 2007.