Article

Short-Term Traffic Flow Forecasting via Multi-Regime Modeling and Ensemble Learning

Zhenbo Lu ¹, Jingxin Xia ^{1,*}, Man Wang ², Qinghui Nie ³ and Jishun Ou ³

¹ Intelligent Transportation System Research Center, Southeast University, No. 2 Southeast University Road, Nanjing 211189, China; 101010360@seu.edu.cn

² Institute of Intelligent Transportation, Zhejiang Scientific Research Institute of Transport, No. 705, Dalong Juwu, Hangzhou 310039, China; zjkitswman@163.com

³ College of Civil Science and Engineering, Yangzhou University, No. 131 Jiangyang Middle Road, Yangzhou 225009, China; nieqinghui@yzu.edu.cn (Q.N.); jishun@yzu.edu.cn (J.O.)

* Correspondence: xiajingxin@seu.edu.cn; Tel.: +86-1381-390-6339

Received: 3 December 2019; Accepted: 30 December 2019; Published: 3 January 2020

Abstract: Short-term traffic flow forecasting is crucial for proactive traffic management and control. One key issue associated with the task is how to properly define and capture the temporal patterns of traffic flow. A feasible solution is to design a multi-regime strategy. In this paper, an effective approach to forecasting short-term traffic flow based on multi-regime modeling and ensemble learning is presented. First, to properly capture the different patterns of traffic flow dynamics, a regime identification model based on probabilistic modeling was developed. Each identified regime represents a specific traffic phase, and was used as the representative feature for the forecasting modeling. Second, a forecasting model built on an ensemble learning strategy was developed, which integrates the forecasts of multiple regression trees. The traffic flow data over 5-min intervals collected from four I-80 freeway segments, in California, USA, was used to evaluate the proposed approach. The experimental results show that the identified regimes are able to well explain the different traffic phases, and play an important role in forecasting. Furthermore, the developed forecasting model outperformed four typical models in terms of root mean square error (RMSE) and mean absolute percentage error (MAPE) on three traffic flow measures.

Keywords: traffic flow forecasting; multi-regime modeling; hidden Markov model; ensemble learning; regression tree

1. Introduction

Traffic congestion brings substantial negative impacts on humanity, such as high travel costs, increased anxiety, and polluted air. To alleviate traffic congestion problems, researchers and authorities all over the world have explored a wide number of feasible solutions. Among them, intelligent transportation system (ITS) has become the most popular and effective one. By effectively and efficiently collecting, processing, and disseminating traffic data, ITS helps traffic researchers and practitioners make reasonable and reliable decisions, and has achieved great success during the past decade [1–5].

Traffic flow describes the traffic conditions over certain time intervals using representative measures, for example, flow rate, speed, and density. As the values of the measures are continuous over time, traffic flow is commonly recorded and described with a time series format. Short-term traffic flow forecasting, the forecasted time scale of which is no more than 30 min (e.g., 5, 10, or 15 min) is a fundamental function in ITS. One key issue in traffic flow forecasting is how to properly define and capture the temporal patterns of traffic flow [6]. To solve this issue, two categories of coping strategies

are commonly adopted. The first is a detrending strategy, which is to build forecasting models by separating the trends of the traffic flow time series from the remaining fluctuations and developing distinct models to describe the trends and fluctuations, respectively. The second is a multi-regime strategy, which assumes that different patterns are contained in traffic flow dynamics. Based on the assumption, different regimes are first identified according to the distinct patterns. After that, the pattern of each regime is characterized and captured by a separate model. A comprehensive comparison study with the two strategies was conducted by Li et al. [6]. They concluded that the models based on the multi-regime strategy are capable of identifying the local trends in the vehicle count time series and describing its fluctuations more effectively. However, the overall forecasting errors of the multi-regime models cannot be notably reduced in their experiments because of the introduced errors by the regime identification model.

As the above analyses show, the forecasting performance of a multi-regime model is closely related to two major aspects. First, a proper regime identification model needs to be built in order to assign the correct regime to each traffic flow observation. Second, an accurate multi-regime forecasting model to capture the patterns of different regimes of traffic flow is required. Inspired by these two facts, this paper presents an effective approach to forecasting short-term traffic flow based on multi-regime modeling and ensemble learning, which includes two key procedures. First, to properly capture the different patterns of traffic flow dynamics, a regime identification model based on a probabilistic approach is developed. Each regime represents a specific and homogeneous traffic condition. The probabilistic approach builds on the concepts of the Markov Process and Markov Chain. Second, a forecasting model using an ensemble learning strategy is developed, which integrates the outputs of multiple individual learners to produce the final forecasts. The ensemble learning strategy introduced can effectively improve the accuracy of the forecasting model.

The organization of this paper is as follows. The next section provides a comprehensive review on the existing studies associated with traffic flow forecasting. Subsequently, the details of the proposed approach are given. After that, the evaluated data sets are described. The following experiments are carried out to evaluate the proposed approach. Lastly, the conclusions and future work are discussed.

2. Literature Review

During the past decades, researchers have explored and developed a vast number of traffic flow forecasting models.

As traffic flow is recorded as a time series format, which can be effectively described and modeled by typical time series models, autoregressive integrated moving average (ARIMA) was applied by Ahmed and Cook [7] to establish a short-term traffic flow forecasting model. The model can accurately capture the local trends in traffic flow. Nevertheless, given the cyclical characteristic of traffic flow, the first difference of a traffic flow time series will not produce a stationary series, leading to inaccurate forecasts. In view of this, a seasonal autoregressive integrated moving average (SARIMA) model was developed by Williams et al. [8] to forecast urban freeway traffic flow. The results indicated that the SARIMA models outperform the nonparametric regression (NPR), artificial neural network (ANN), and historical average models. Xia et al. [9] designed a multistep forecaster based on a SARIMA model and an embedded adaptive Kalman filter model. The advantage of the designed model is that it is easy to implement and is computationally inexpensive. Kalman filtering is another efficient model, which uses the state of the previous moment to compute the best estimate of the state at the current moment. Some researchers extended the Kalman filtering to forecast short-term traffic volume [10,11]. The ARIMA, SARIMA, and Kalman filter models are capable of accurately and timely forecasting stable traffic flow, while neither is able to capture the nonlinear patterns in traffic flow.

To effectively capture the nonlinear patterns in traffic flow, a series of more elegant models were proposed. The NPR models [12,13] were presented to avoid the shortcomings of the parametric models (e.g., ARIMA and SARIMA) and to make more accurate forecasts in non-stationary traffic flow. The main advantages of the NPR models are that they do not need to make priori assumptions,

and have a high flexibility and intuitive expression [14]. It should be noted that the NPR models are sensitive to outliers usually underlying in traffic flow data. Recently, SVR has drawn increasing attention, and been successfully applied in short-term traffic forecasting [15,16]. It needs to be indicated that the standard SVR model requires a blind and time-consuming searching procedure to find suitable hyperparameters. To improve the computational efficiency, some more advanced SVRs have been presented [17,18]. Because of the capability to capture the uncertainty and nonlinearity in traffic flow, the ANN models were designed for short-term traffic flow forecasting tasks. Although each neuron in ANN has a very simple function, the network, composed of a large number of neurons, can possess significantly powerful functions [19,20]. In the past decade, deep learning models, in particular deep neural networks (DNN), have achieved great success in various domains. Accordingly, several researchers presented a number of traffic flow forecasting models based on DNN. Lv et al. [21] put forward a deep architecture model using autoencoders as building blocks to represent traffic flow features. Experimental results demonstrated the superiority of the DNN model. Polson and Sokolov [22] showed that deep learning architectures are able to capture the nonlinear spatiotemporal effects resulting from the transitions between free flow, breakdown, recovery, and congestion in traffic flow. More recently, Do et al. [23] introduced the spatial and temporal attentions that are utilized to capture the spatial dependencies between road segments and temporal dependencies between time steps into the DNN model.

While the models mentioned above have been proven to be effective in short-term traffic flow forecasting, an issue that cannot be ignored is that the work mechanisms of the models have not been easily understood by decision-makers. That is, the forecasting results lack a convincing interpretation. Several researchers have noticed this problem, and tried to solve it by adopting some prior feature selection strategies [15,16]. Similarly, we presented a short-term traffic flow forecasting model using a data-driven feature selection strategy and bias-corrected random forests [24], which shows an excellent forecasting performance and good interpretability.

As the well-known “No Free Lunch” theorem [25] states, there is no one model can work best for every problem. Furthermore, the above analyses indicate that each of the analyzed models has its own advantages and limitations. To compare different forecasting models from distinct viewpoints, some researchers conducted comprehensive and deep literature reviews on traffic flow forecasting [1,26–28]. Interested readers can refer to these studies for more comparison details of various models.

A key issue in traffic flow forecasting is how to properly define and capture the temporal patterns of traffic flow. To effectively handle this issue, two categories of coping strategies are commonly adopted. The first is a detrending strategy, which is to build forecasting models by separating the trends of traffic flow time series from the remaining fluctuations, and developing distinct models to describe the trends and fluctuations, respectively. For instance, the SARIMA model assumes that there is a weekly or monthly cyclic trend in the traffic flow [29]. Chen et al. [30] studied different forecasting models built on either the original traffic flow sequence or the remaining time series. They discovered that in the latter case, a significant performance improvement can be achieved. The second is a multi-regime strategy, which assumes that different patterns are contained in traffic flow dynamics. In some relevant studies [31–33], researchers divided the traffic flow data into several regimes according to distinct patterns, and established separate models to characterize the traffic flow dynamics in each regime. The associated results demonstrate the potential of the multi-regime strategy. A comprehensive comparison study with the two strategies was conducted by Li et al. [6]. They indicated that the multi-regime strategy could be utilized to improve the forecasting performance when using an accurate regime identification model. With this in mind, in the subsequent sections, we will develop an effective traffic flow forecasting model based on a multi-regime strategy, which establishes the regime identification model by probabilistic modeling. In addition, an ensemble learning mechanism is introduced to further improve the performance of the developed forecasting model.

3. Methodology

3.1. Overall Framework

The overall framework of the proposed approach is illustrated in Figure 1. As seen, the framework consists of two major procedures. In the first procedure, multiple regimes of traffic flow are identified using a probabilistic approach. Each regime characterizes a pattern that describes a homogeneous traffic condition during the study time period. The identified regimes are then used as the representative features for the forecasting modeling. In the second procedure, the training data set and test data set are separately established according to the constructed features. Subsequently, the forecasting model is built based on the training set and an ensemble learning strategy, and utilized to produce the forecasts of the test data set.

Figure 1. Overall framework of the proposed approach.

3.2. Regime Identification Based on Probabilistic Modeling

Traffic behaviors under various traffic conditions can be treated as a stochastic process [34]. In view of this, probabilistic modeling is conducted in this study to properly identify the associated regimes of traffic flow. The hidden Markov model (HMM) is one of the most powerful algorithms in probabilistic modeling [35,36], and is thus used here. An HMM describes two stochastic processes based on the concepts of the Markov Process and Markov Chain. The first stochastic process is hidden, while the second is observable. The hidden stochastic process is established to infer the observations of the observable stochastic process. In the regime identification task, the regimes are described as the hidden states in HMM, and the hidden process depicts how the regimes are transformed into each other. Meanwhile, the values of the traffic flow measures are described as the observations in HMM, and the observable process describes how the collected observations of the measures evolve over time.

3.2.1. Regime Identification

In the study, the regime identification model based on HMM is defined and characterized with the following elements.

- (1) A set of hidden states $S = \{s_1, s_2, \dots, s_N\}$, where each hidden state $s_j (1 \leq j \leq N)$ describes the j th regime of traffic flow. N is the number of hidden states. Specifically, the state at time t is denoted as q_t .

- (2) A set of observation symbols $V = \{v_1, v_2, \dots, v_M\}$, where each observation symbol $v_k (1 \leq k \leq M)$ describes the k th observation of a traffic flow measure. M is the number of observation symbols.
- (3) A state transition probability distribution $A = \{a_{ij}\}_{i=1}^N \{j=1}^N$, where

$$a_{ij} = p[q_{t+1} = s_j | q_t = s_i], \quad a_{ij} \geq 0. \quad (1)$$

A describes the evolving probability distribution of the analyzed traffic flow from one of the defined regimes to another. Specifically, a_{ij} describes the probability of the analyzed traffic flow evolving from regime s_i to regime s_j .

- (4) The observation symbol probability distribution $B = \{b_j(k)\}$ in state s_j , where

$$b_j(k) = p[v_k \text{ at } t | q_t = s_j], \quad (2)$$

B describes the generation probability of the observations of the analyzed traffic flow in each regime. Specifically, $b_j(k)$ describes the generation probability of the k th observation of the analyzed traffic flow in regime s_j .

- (5) The initial state distribution $\pi = \{\pi_i\}$, where

$$\pi_i = p[q_1 = s_i], \quad 1 \leq i \leq N. \quad (3)$$

π describes the probability of the initial regime of the analyzed traffic flow. Specifically, π_i is the probability of the initial regime equaling to s_i .

Given an observation sequence of a traffic flow measure $O = O_1, O_2, \dots, O_T$, the regime identification model aims to identify the corresponding regime sequence $Q = q_1, q_2, \dots, q_T$, based on the determined values of N, M, A, B , and π .

As mentioned above, the hidden state set S defines the regimes of traffic flow. Each regime describes a homogeneous traffic condition. According to the fundamental diagram in traffic flow theory and a series of relevant studies [37–41], five distinct regimes could be defined in practice, as shown in Figure 2. The first two regimes, Regime 1 and Regime 2, describe two kinds of free-flow traffic conditions, while the last two regimes, Regime 4 and Regime 5, represent two categories of congested traffic conditions. The third regime, Regime 3, defines a transition condition from the free-flow stage to the congested stage. Therefore, the number of hidden states in the study, N , is set as five.

Figure 2. Regimes defined in the traffic fundamental diagram.

The observation symbols in the identification model are associated with the values of the traffic flow measures. For example, the speed measure in short-term traffic flow forecasting is usually recorded every 5 min, and its value range determines the representation space of the observation symbols. However, in practice, the value range is usually large, resulting in many observation symbols and hence in a high computational complexity of the model. To effectively tackle the problem, a simple discretization strategy is employed, in which the measure values are discretized into multiple relatively large and equal numerical intervals (e.g., $0 \sim 5$, $5 \sim 10$, $10 \sim 15$). For the other two measures, a similar strategy can be adopted. Based on the strategy, $M = \lceil \frac{o_{max}-o_{min}}{o_{int}} \rceil$, where o_{max} and o_{min} are the maximum and minimum of the measure values, respectively, and o_{int} is the size of the divided numerical interval.

The remaining three parameters, A , B , and π , need to be determined by training the probabilistic model using the traffic flow observations. For simplicity, we denote $\lambda = (A, B, \pi)$. To get the optimal λ according to the given observation sequence $O = O_1, O_2, \dots, O_T$, we defined and solved the training problem by computing the maximum probability $p = (O|\lambda)$ with an iterative procedure. A modified Baum–Welch algorithm was introduced to learn the best λ . The details of the algorithm are skipped here. Interested readers can refer to the work [42] for more details.

The determined λ can be further utilized to uncover the hidden state sequence Q , based on the given observation sequence O . To this end, we implemented the Viterbi algorithm [43] to solve the problem. The algorithm aims to optimize the following objective:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} p[q_1, q_2, \dots, q_t = s_i, O_1, O_2, \dots, O_t | \lambda] \quad (4)$$

where $\delta_t(i)$ is the highest probability along a single path at time t . To obtain the state sequence, we needed to iteratively solve the following optimization problem by considering each t and i :

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij} \right] \cdot b_j(O_{t+1}). \quad (5)$$

The complete computing procedure of the Viterbi algorithm can be seen in the literature [44].

In summary, through the above modeling process, the regimes of traffic flow can be properly identified. The next section is to build the forecasting model by using the regimes to construct the representative features and data sets for the model training.

3.2.2. Feature Construction

A key step in forecasting modeling is to determine the representative features. For the multi-regime model, the identified regimes are treated as the most important features, and will be used in the modeling. In addition, the temporal correlations of the forecasted traffic flow measure and the interactive correlations of the multiple traffic flow measures are also two significant factors that can be used to improve the forecasting accuracy [14,24,45]. With this in mind, the time-lagged and interactive features of the traffic flow are also added to the representative feature pool. Table 1 lists the constructed features for the forecasting modeling. Note that in the study, the time interval is set as 5-min and the forecasted time is denoted as t .

Table 1. Constructed features for forecasting modeling.

Representative Features	Descriptions
state_5	Regime at time $t - 1$
state_10	Regime at time $t - 2$
state_15	Regime at time $t - 3$
f_5	Flow rate at time $t - 1$
f_10	Flow rate at time $t - 2$
f_15	Flow rate at time $t - 3$
s_5	Speed at time $t - 1$
s_10	Speed at time $t - 2$
s_15	Speed at time $t - 3$
o_5	Occupancy at time $t - 1$
o_10	Occupancy at time $t - 2$
o_15	Occupancy at time $t - 3$

3.3. Forecasting Modeling via Ensemble Learning

Given the historical traffic flow observations and the constructed features, the training data set and the test data set can be separately established. After that, the model to forecast the short-term traffic flow needs to be built. In this study, a forecasting model based on an ensemble learning strategy is developed using the obtained training set. Ensemble learning is a very popular and useful technique in machine learning [46], which can be utilized to improve the generalization performance of the forecasting models, and has achieved great success in various domains. The training and test procedure of the developed forecasting model is as follows:

Step 1: Randomly draw an instance $I_h (1 \leq h \leq n)$ from the training data set D_T and add it to a new training data set D_{NT} . Next, return the instance to D_T . Repeat the sampling process n times and generate the final D_{NT} . That is, $D_{NT} = \{I_1, I_2, \dots, I_n\}$.

Step 2: Use D_{NT} to train an unpruned regression tree, denoted as $T_\alpha (1 \leq \alpha \leq K)$. The training procedure is as follows: for each node of the regression tree, φ , randomly sample β features from the constructed features, and use them to compute the best split that has the maximum mean decrease in impurity. The impurity of the $r (1 \leq r \leq \beta)$ th sampled feature, x_r , associated with node φ is defined as follows:

$$\text{IMP}(\varphi) = \frac{\sum_{x_r \in \varphi} (y_r(\varphi) - \bar{y}(\varphi))^2}{n_\varphi}, \quad (6)$$

where $x_r \in \varphi$ is the value of x_r of the training samples, which falls into the range of the split at node φ ; $y_r(\varphi)$ is the response value of the training samples associated with $x_r \in \varphi$; $\bar{y}(\varphi)$ is the mean of $y_r(\varphi)$; and n_φ is the number of the training samples associated with $x_r \in \varphi$.

In the above training procedure, if φ is a leaf node, $\bar{y}(\varphi)$ is set as the forecast of the node.

Step 3: Repeat Step 1 and Step 2 K times to train a number of K regression trees.

Step 4: Given the instance in the established test data set D_E , denoted as $I_E (I_E \in D_E)$, the forecasts from the K regression trees are first computed. Assume the forecasts are $f_1(I_E), f_2(I_E), \dots, f_K(I_E)$, respectively. The final forecast of the instance is calculated as the median of the forecasts of the K regression trees, as follows:

$$f(I_E) = \text{median}(f_1(I_E), f_2(I_E), \dots, f_K(I_E)). \quad (7)$$

Note that the median rule instead of the average rule is used to combine the outputs of the regression trees, in that the median rule can yield more robust forecasts in the cases where there are some outliers in the leaf nodes [47].

4. Data Description

The traffic flow data collected from four dual-loop detectors on I-80 freeway segments was used to evaluate the proposed approach. The detectors recorded the traffic flow observations of three traffic measures including the flow rate, speed, and occupancy every 30 s. In the study, the lane-by-lane traffic flow observations were aggregated into 5-min intervals at each detector station in order to obtain stable traffic data. The data were acquired through the California Performance Measurement System (PeMS). Figure 3 illustrates the locations of the detector stations. The study period was two months long, from 5 June to 5 August 2007. For the regime identification task, the whole data set was used, because the developed identification model was unsupervised, meaning it did not require label information. For the short-term traffic flow forecasting task, the whole data set was divided into two parts, one for model training and the other for model evaluation. The training data set included the traffic flow observations from 5 June to 20 July, and the evaluation data set included the traffic flow observations from 21 July to 5 August.

Figure 3. Locations of the study detector stations.

5. Experimental Analysis

5.1. Identification of Regimes

5.1.1. Model Calibration

The first step of the developed regime identification models associated with the four detector stations was to determine the model parameters. To this end, the models were implemented and calibrated using the whole study data set. The determined model parameters π and A are listed in Tables 2 and 3, respectively. The statistical means of the traffic measures for each identified regime are depicted in Table 4.

The parameter π describes the probability distribution of the initial state. As seen from Table 2, for each detector station, the values of four of the five states are 0 or close to 0, while the value of the other state is 1.0. This might be because the start time of the traffic flow when the data were collected from each detector station was at 00:00, when the traffic on the road is under a free-flow condition.

Table 2. Determined π in four regime identification models.

Station ID	State 1	State 2	State 3	State 4	State 5
400976	0	1.0	0	0	0
400081	0	1.0	0	0	0
400329	1.0	0	0	0	0
400825	1.0	0	0	0	0

Another significant model parameter A describes the transition probability distribution from one state to another. The digital numbers in the last five columns of Table 3 are the determined probabilities. Take Station 400976, for example. The transition probability between the same states was much higher than that between the different states. This indicates that the variations of traffic conditions usually occurred after a certain time period. For State 1, its more likely next state is State 2, meaning the traffic evolved from slightly crowded conditions to completely free flow conditions at times. For State 2, the transition probability from it to another state was very small. Occasionally, it transformed into State 1 (slightly crowded) or State 5 (very congested). The traffic goes through from free-flow conditions to very congested conditions, which might be because of the occurrence of traffic crashes. For State 3, its next possible states were State 5 and State 4, implying the traffic evolved from transition conditions to more congested conditions. For State 4, it is possible to transit to the other four states. For State 5, it might change to State 2, State 3, or State 4, meaning that the traffic congestion has been gradually dissipating. For the other stations, similar patterns could be identified. However, there were some slightly differences in the cases where different traffic states were transformed into each other.

Table 3. Determined A in four regime identification models.

Station ID		State 1	State 2	State 3	State 4	State 5
400976	State 1	0.9438	0.0498	0.0000	0.0064	0.0000
	State 2	0.0088	0.9793	0.0000	0.0021	0.0098
	State 3	0.0000	0.0000	0.9839	0.0010	0.0150
	State 4	0.0035	0.0059	0.0018	0.9848	0.0039
	State 5	0.0000	0.0137	0.0132	0.0025	0.9706
400081	State 1	0.9773	0.0000	0.0139	0.0086	0.0002
	State 2	0.0000	0.9022	0.0054	0.0923	0.0000
	State 3	0.0096	0.0006	0.9739	0.0037	0.0123
	State 4	0.0229	0.0143	0.0019	0.9609	0.0000
	State 5	0.0000	0.0000	0.0130	0.0000	0.9870
400329	State 1	0.9829	0.0000	0.0000	0.0171	0.0000
	State 2	0.0000	0.9359	0.0556	0.0045	0.0040
	State 3	0.0000	0.0133	0.9703	0.0000	0.0164
	State 4	0.0195	0.0010	0.0003	0.9574	0.0219
	State 5	0.0003	0.0005	0.0145	0.0129	0.9718
400825	State 1	0.9849	0.0024	0.0000	0.0000	0.0127
	State 2	0.0074	0.9754	0.0059	0.0000	0.0113
	State 3	0.0000	0.0003	0.9794	0.0083	0.0119
	State 4	0.0000	0.0000	0.0314	0.9686	0.0000
	State 5	0.0127	0.0009	0.0159	0.0000	0.9705

5.1.2. Identification Results Analysis

To observe the statistical characteristics of the traffic flow associated with the different regimes, the mean of each measure at each detector station for each regime was calculated, as depicted in Table 4. Meanwhile, the second column of the table provides the corresponding regime in Figure 2, which is associated with each hidden state in the identified model. From the table, we can see that each regime has distinct traffic flow characteristics. Regime 1 has the minimum flow rate mean and occupancy mean, but the maximum speed mean. Compared with Regime 1, Regime 2 shows a higher flow rate mean and an occupancy mean, but a lower speed mean. In Regime 3, the traffic flow shows the maximum flow rate mean and the mild occupancy mean and speed mean. In Regime 4, the flow rate mean begins to drop down while the occupancy mean continues to increase and the speed mean continues to drop down. Finally, in Regime 5, the occupancy mean achieves the maximum and the speed mean comes to the minimum. Meanwhile, the flow rate mean reaches a similar level to that of Regime 2. Figure 4 provides the regime identification results at each detector station. In each subfigure,

the data points with each color represent an identified regime. Comparing Figure 2 with Figure 4, we can see that the proposed approach can properly identify different regimes of traffic flow.

Table 4. Statistical means of traffic measures for each identified regime.

Station ID	States	Regimes	Flow Rate (veh/5-min)	Occupancy (%)	Speed (mi/h)
400976	State 1	Regime 2	378.3687	0.0782	65.2341
	State 2	Regime 1	120.1265	0.0224	69.4739
	State 3	Regime 3	485.8626	0.1272	50.7289
	State 4	Regime 4	475.2812	0.2121	32.3619
	State 5	Regime 5	393.6340	0.3277	19.2915
400081	State 1	Regime 2	391.3933	0.0774	68.2787
	State 2	Regime 1	103.8473	0.0214	69.3800
	State 3	Regime 3	543.7127	0.1219	59.3171
	State 4	Regime 4	504.4221	0.1744	43.5848
	State 5	Regime 5	334.6128	0.4000	17.5781
400329	State 1	Regime 1	113.3446	0.0215	69.4238
	State 2	Regime 2	375.6891	0.0690	67.4715
	State 3	Regime 5	397.1494	0.2806	26.6234
	State 4	Regime 3	506.9161	0.1728	40.1897
	State 5	Regime 4	468.1100	0.2012	34.8338
400825	State 1	Regime 1	104.1068	0.0226	72.2015
	State 2	Regime 2	371.7557	0.0727	69.6464
	State 3	Regime 3	539.3112	0.1592	49.1119
	State 4	Regime 5	397.3773	0.3206	19.9382
	State 5	Regime 4	520.8873	0.2234	32.9221

Figure 4. Regime identification results associated with the four study stations.

5.1.3. Feature Importance Analysis

To check the importance of the constructed features, a measure named the increased node purity [48] (illustrated as IncNodePurity in Figure 5) was calculated for each feature. The measure computes the total decrease in node impurity from splitting on the given feature, and then averages it over all of the component regression trees. The node impurity is quantified by the residual sum of the squares and is calculated only at the node at which that feature is used for that split. Based on the measure, the importance of each feature associated with each station was calculated and ranked. As similar results were obtained for all four stations, for illustration purposes, we only provide the results associated with Station 400081, as shown in Figure 5. From the figure, we can observe that the identified regimes at different times play an important role in forecasting. Accordingly, the regime at time $t - 1$ is more important than the regimes at time $t - 2$ and time $t - 3$. For different forecasting tasks, the most important traffic flow measure is distinct. For example, if the forecasted measure is the flow rate, the most important feature is flow rate at time $t - 1$. For the other two forecasting tasks, the most important feature corresponds to the forecasted traffic flow measure at time $t - 1$. Taken together, it can be seen that it is necessary to use the regimes as the representative features in traffic flow forecasting tasks.

Figure 5. Feature importance associated with Station 400081.

5.2. Analysis of Forecasting Results

5.2.1. Performance Measures

The first step of the developed regime identification models associated with the four detector stations was to determine the model parameters. To this end, the models were implemented and calibrated using the whole study data set. The determined model parameters π and A are listed in Tables 2 and 3. The statistical means of the traffic measures for each identified regime are depicted in Table 4.

To evaluate the developed forecasting models, two performance measures were employed, including root mean square error (RMSE) and mean absolute percentage error (MAPE). The measures are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n_s} \sum_{\gamma=1}^{n_s} (y_{\gamma} - \hat{y}_{\gamma})^2}, \quad (8)$$

$$\text{MAPE} = \frac{1}{n_s} \sum_{\gamma=1}^{n_s} \left| \frac{y_{\gamma} - \hat{y}_{\gamma}}{y_{\gamma}} \right|, \quad (9)$$

where y_{γ} is the true value of the γ th sample of the considered traffic flow measure, $\hat{y}_{\gamma}$ is the forecasted value of the γ th sample of the considered traffic flow measure, and n_s is the number of forecasted samples.

5.2.2. Forecasting Results Analysis

In the study, a one-step forecasting task was carried out. The modeled traffic flow measures include flow rate, occupancy, and speed. Moreover, four forecasting models were implemented and compared, including ARIMA, Regression Tree (RT), Ensemble Regression Trees (ERT), and ensemble regression trees based on multi-regime modeling (ERT-MRM), developed in the study. As a typical kind of time series model, ARIMA has been successfully applied in various domains [7] because of its solid theoretical foundations and good ability to capture local trends in stationary time series. Meanwhile, it is commonly used as a baseline in traffic flow forecasting tasks. There are three parameters to be determined in the ARIMA model (i.e., the number of autoregressive terms, p ; the number of nonseasonal differences needed for stationarity, d ; and the number of lagged forecast errors in the forecasting equation, q). In this paper, the three parameters were determined using the Akaike information criterion (AIC). That is, p , d , and q were chosen when the forecasting model has the lowest AIC. RT is another popular model for forecasting modeling. It has shown to be competitive in many applications while possessing a good interpretability [48]. In this study, it was implemented as a pruned regression tree in order to achieve a good generalization performance. ERT is an ensemble version of the RT model. As mentioned, ensemble learning can improve the accuracy and robustness of the forecasting system. To this end, we added ERT to the comparative list. The number of the considered trees in ERT was set as 100 in order to balance the accuracy and efficiency of the model. The three models mentioned above did not utilize a multi-regime strategy, and hence the regimes were not used as the representative features in modeling. Instead, the ERT-MRM developed in this paper was implemented using a multi-regime strategy, and its component trees were also set as 100 for a fair comparison.

The performances of the implemented models on the traffic flow data sets at the four study stations are illustrated in Figure 6. As the figure shows, the ERT-MRM model achieved the best performance on each data set. The ARIMA and RT models show similar forecasting errors, which are both higher than that of the ERT and ERT-MRM models. This might be because the ensemble strategy introduced in the latter two models can take advantage of the strengths of the component tree forecaster to effectively improve the accuracy and robustness of the forecasting system. Moreover, it can be observed that the ERT-MRM model shows a more competitive performance than the ERT model, meaning the multi-regime strategy is capable of being utilized to improve the accuracy of forecasting models [32]. Taken together, for the ERT-MRM model, the RMSEs associated with the flow rate measure are less than 14 veh/5-min, and the MAPEs are no more than 4%; the RMSEs associated with the occupancy measure are less than 0.01%, and the MAPEs are no more than 4.5%; and the RMSEs associated with the speed measure are less than 2 mi/h, and the MAPEs are no more than 1.5%.

Figure 7 gives the comparisons between the observed and forecasted values of one-day traffic flow randomly selected from the test data set. From the figure, we can see that ERT-MRM can provide reliable traffic flow forecasts at all of the four stations. As a result, the model developed in this study can be perfectly used in proactive freeway management and control.

Figure 6. Performance of the comparative forecasting models associated with different stations.

Figure 7. Cont.

Figure 7. Comparisons between the observed and forecasted values of the three traffic flow measures.

6. Conclusions

Traffic flow forecasting has been a significant and hot research topic during the past decades because of its key role in proactive traffic management and control. In this paper, a short-term traffic flow forecasting approach based on multi-regime modeling and ensemble learning is presented. The approach consists of two procedures. In the first procedure, multiple regimes of traffic flow were properly identified using a probabilistic modeling method, and further used to construct the representative features for the forecasting modeling. In the second procedure, the constructed features

were utilized to establish the training and test data sets that are employed to train and evaluate the forecasting model. To improve the generalization performance of the model, an ensemble learning strategy from the machine learning domain was introduced. To evaluate the proposed approach, 5-min traffic flow data collected from four dual-loop detectors on I-80 freeway segments were used. The experimental results show that the identified regimes are able to explain the different phases of traffic flow well, and play an important role in forecasting. Furthermore, the developed forecasting model outperformed four comparative models in terms of RMSE and MAPE on three traffic flow measures.

For a forecasting model, its accuracy is closely associated with two aspects, that is, data and algorithm. To ensure the quality of the used data, the representative features need to be properly determined. In the study, we developed a multi-regime modeling strategy to enable the model to have good input features. On the other hand, an elegant algorithm needs to be designed to make the forecasting model well-fitting with the training data, while possessing a good generalization ability. To achieve this goal, a typical ensemble learning strategy was employed in the study. The above two strategies ensure the good performance of the proposed approach.

In the future, more traffic flow data collected from different road types will be used to evaluate the proposed approach. As the number of identified regimes could affect the performance of the forecasting model, it is necessary to check the sensitivity of the parameter and to explore how to optimally determine the parameter. In addition, more forecasting models will be implemented and compared. Finally, more multi-regime modeling and ensemble learning strategies will be developed and integrated into the framework of the proposed approach.

Author Contributions: The paper was written by Z.L. in collaboration with all co-authors. The model was established and tested by J.O. and M.W. The data was collected by Q.N. The research and key elements of the models were reviewed by J.X. The major revision of this paper was completed by Z.L. and J.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by National Natural Science Foundation of China (grant number 71871055).

Conflicts of Interest: The authors declare that there is no conflict of interest.

References

1. Vlahogianni, E.I.; Karlaftis, M.G.; Golias, J.C. Short-term traffic forecasting: Where we are and where we're going. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 3–19. [\[CrossRef\]](#)
2. Ou, J.; Yang, S.; Wu, Y.-J.; An, C.; Xia, J. Systematic clustering method to identify and characterise spatiotemporal congestion on freeway corridors. *IET Intell. Transp. Syst.* **2018**, *12*, 826–837. [\[CrossRef\]](#)
3. Wang, C.; Xu, C.; Dai, Y. A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accid. Anal. Prev.* **2019**, *123*, 365–373. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wang, C.; Xu, C.; Xia, J.; Qian, Z.; Lu, L. A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation. *Transp. Res. Part C Emerg. Technol.* **2018**, *90*, 281–291. [\[CrossRef\]](#)
5. Rao, W.; Wu, Y.-J.; Xia, J.; Ou, J.; Kluger, R. Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data. *Transp. Res. Part C Emerg. Technol.* **2018**, *95*, 29–46. [\[CrossRef\]](#)
6. Li, Z.; Li, Y.; Li, L. A comparison of detrending models and multi-regime models for traffic flow prediction. *IEEE Intell. Transp. Syst. Mag.* **2014**, *6*, 34–44. [\[CrossRef\]](#)
7. Ahmed, M.S.; Cook, A.R. Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques. *Transp. Res. Rec.* **1979**, *722*, 1–9.
8. Williams, B.; Durvasula, P.; Brown, D. Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transp. Res. Rec. J. Transp. Res. Board* **1998**, *1644*, 132–141. [\[CrossRef\]](#)
9. Xia, J.; Chen, M.; Huang, W. A multistep corridor travel-time prediction method using presence-type vehicle detector data. *J. Intell. Transp. Syst.* **2011**, *15*, 104–113. [\[CrossRef\]](#)
10. Okutani, I.; Stephanedes, Y.J. Dynamic prediction of traffic volume through Kalman filtering theory. *Transp. Res. Part B Methodol.* **1984**, *18*, 1–11. [\[CrossRef\]](#)

11. Stathopoulos, A.; Karlaftis, M.G. A multivariate state space approach for urban traffic flow modeling and prediction. *Transp. Res. Part C Emerg. Technol.* **2003**, *11*, 121–135. [\[CrossRef\]](#)
12. Zheng, Z.D.; Su, D.C. Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 143–157. [\[CrossRef\]](#)
13. Dell'Acqua, P.; Bellotti, F.; Berta, R.; De Gloria, A. time-aware multivariate nearest neighbor regression methods for traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 3393–3402. [\[CrossRef\]](#)
14. Clark, S. Traffic prediction using multivariate nonparametric regression. *Transp. Eng.* **2003**, *129*, 161–168. [\[CrossRef\]](#)
15. Wei, D.L.; Liu, H.C. An adaptive margin support vector regression for short term traffic flow forecast. *J. Intell. Transp. Syst.* **2013**, *17*, 317–327. [\[CrossRef\]](#)
16. Jeong, Y.S.; Byon, Y.J.; Castro-Neto, M.M.; Easa, S.M. Supervised weighting-online learning algorithm for short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1700–1707. [\[CrossRef\]](#)
17. Zhang, Y.; Xie, Y. Forecasting of short-term freeway volume with v-support vector machines. *Transp. Res. Rec. J. Transp. Res. Board* **2008**, *2024*, 92–99. [\[CrossRef\]](#)
18. Castro-Neto, M.; Jeong, Y.S.; Jeong, M.K.; Han, L.D. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Syst. Appl.* **2009**, *36*, 6164–6173. [\[CrossRef\]](#)
19. Kumar, K.; Parida, M.; Katiyar, V.K. Short term traffic flow prediction in heterogeneous condition using artificial neural network. *Transport* **2013**, *30*, 1–9. [\[CrossRef\]](#)
20. Zhu, J.Z.; Cao, J.X.; Zhu, Y. Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections. *Transp. Res. Part C Emerg. Technol.* **2014**, *47*, 139–154. [\[CrossRef\]](#)
21. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F.Y. Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 865–873. [\[CrossRef\]](#)
22. Polson, N.G.; Sokolov, V.O. Deep learning for short-term traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 1–17. [\[CrossRef\]](#)
23. Do, L.N.; Vu, H.L.; Vo, B.Q.; Liu, Z.; Phung, D. An effective spatial-temporal attention based neural network for traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* **2019**, *108*, 12–28. [\[CrossRef\]](#)
24. Ou, J.; Xia, J.; Wu, Y.-J.; Rao, W. Short-term traffic flow forecasting for urban roads using data-driven feature selection strategy and Bias-corrected random forests. *Transp. Res. Rec.* **2017**, *2645*, 157–167. [\[CrossRef\]](#)
25. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [\[CrossRef\]](#)
26. Vlahogianni, E.I.; Golias, J.C.; Karlaftis, M.G. Short-term traffic forecasting: Overview of objectives and methods. *Transp. Rev.* **2004**, *24*, 533–557. [\[CrossRef\]](#)
27. Oh, S.; Byon, Y.J.; Jang, K.; Yeo, H. Short-term travel-time prediction on highway: A review of the data-driven approach. *Transp. Rev.* **2015**, *35*, 4–32. [\[CrossRef\]](#)
28. Ermagun, A.; Levinson, D. Spatiotemporal traffic forecasting: Review and proposed directions. *Transp. Rev.* **2018**, *38*, 786–814. [\[CrossRef\]](#)
29. Williams, B.M.; Hoel, L.A. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *J. Transp. Eng.* **2003**, *129*, 664–672. [\[CrossRef\]](#)
30. Chen, C.; Wang, Y.; Li, L.; Hu, J.; Zhang, Z. The retrieval of intra-day trend and its influence on traffic prediction. *Transp. Res. Part C Emerg. Technol.* **2012**, *22*, 103–118. [\[CrossRef\]](#)
31. Cetin, M.; Comert, G. Short-term traffic flow prediction with regime-switching models. *Transp. Res. Rec.* **2006**, *1965*, 23–31. [\[CrossRef\]](#)
32. Kamarianakis, Y.; Shen, W.; Wynter, L. Real-time road traffic forecasting using regime-switching space-time models and adaptive lasso. *Appl. Stoch. Models Bus. Ind.* **2012**, *28*, 297–315. [\[CrossRef\]](#)
33. Comert, G.; Bezuglov, A. An online change-point-based model for traffic parameter prediction. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1360–1369. [\[CrossRef\]](#)
34. Qi, Y.; Ishak, S. A Hidden Markov Model for short term prediction of traffic conditions on freeways. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 95–111. [\[CrossRef\]](#)
35. Rabiner, L.R.; Juang, B.H. An introduction to hidden Markov models. *IEEE ASSP Mag.* **1986**, *3*, 4–16. [\[CrossRef\]](#)

36. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286. [[CrossRef](#)]
37. Xia, J.; Chen, M. Defining traffic flow phases using intelligent transportation systems-generated data. *J. Intell. Transp. Syst.* **2007**, *11*, 15–24. [[CrossRef](#)]
38. Xia, J.; Chen, M. A nested clustering technique for freeway operating condition classification. *Comput. Aided Civ. Infrastruct. Eng.* **2007**, *22*, 430–437. [[CrossRef](#)]
39. Xia, J.; Huang, W.; Guo, J. A clustering approach to online freeway traffic state identification using ITS data. *KSCE J. Civ. Eng.* **2012**, *16*, 426–432. [[CrossRef](#)]
40. Antoniou, C.; Koutsopoulos, H.N.; Yannis, G. Dynamic data-driven local traffic state estimation and prediction. *Transp. Res. Part C Emerg. Technol.* **2013**, *34*, 89–107. [[CrossRef](#)]
41. Rao, W.; Xia, J.; Lyu, W.; Lu, Z. An interval data-based k-means clustering method for traffic state identification at urban intersections. *IET Intell. Transp. Syst.* **2019**, *13*, 1106–1115. [[CrossRef](#)]
42. Baggenstoss, P.M. A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces. *IEEE Trans. Speech Audio Process.* **2001**, *9*, 411–416. [[CrossRef](#)]
43. Forney, G.D. The Viterbi algorithm. *Proc. IEEE* **1973**, *61*, 268–278. [[CrossRef](#)]
44. Lou, H.L. Implementing the Viterbi algorithm. *IEEE Signal Process. Mag.* **1995**, *12*, 42–52. [[CrossRef](#)]
45. Ma, T.; Zhou, Z.; Abdulhai, B. Nonlinear multivariate time-space threshold vector error correction model for short term traffic state prediction. *Transp. Res. Part B Methodol.* **2015**, *76*, 27–47. [[CrossRef](#)]
46. Dietterich, T.G. Ensemble learning. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 2002; pp. 110–125. [[CrossRef](#)]
47. Nguyen, T.-T.; Huang, J.Z.; Nguyen, T.T. Two-Level Quantile Regression Forests for Bias Correction in Range Prediction. *Mach. Learn.* **2015**, *101*, 325–343. [[CrossRef](#)]
48. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).