

Research Article

Short-Term Traffic Flow Local Prediction Based on Combined Kernel Function Relevance Vector Machine Model

Qichun Bing,¹ Bowen Gong,^{1,2,3} Zhaosheng Yang,^{1,2,3} Qiang Shang,¹ and Xiyang Zhou¹

¹College of Transportation, Jilin University, Changchun 130025, China

²State Key Laboratory of Automobile Simulation and Control, Jilin University, Changchun 130025, China

³Jilin Province Key Laboratory of Road Traffic, Jilin University, Changchun 130025, China

Correspondence should be addressed to Bowen Gong; gongbowen@jlu.edu.cn

Received 30 May 2015; Accepted 3 August 2015

Academic Editor: Michael Small

Copyright © 2015 Qichun Bing 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.

Short-term traffic flow prediction is one of the most important issues in the field of adaptive traffic control system and dynamic traffic guidance system. In order to improve the accuracy of short-term traffic flow prediction, a short-term traffic flow local prediction method based on combined kernel function relevance vector machine (CKF-RVM) model is put forward. The C-C method is used to calculate delay time and embedding dimension. The number of neighboring points is determined by use of Hannan-Quinn criteria, and the CKF-RVM model is built based on genetic algorithm. Finally, case validation is carried out using inductive loop data measured from the north-south viaduct in Shanghai. The experimental results demonstrate that the CKF-RVM model is 31.1% and 52.7% higher than GKF-RVM model and GKF-SVM model in the aspect of MAPE. Moreover, it is also superior to the other two models in the aspect of EC.

1. Introduction

Short-term traffic flow prediction is an important basis for intelligent transportation systems (ITS). Real-time and accurate prediction information can be directly applied to the advanced traffic management system (ATMS) and advanced traffic information service system (ATIS). Because of its importance, short-term traffic flow prediction has generated great interest among the scientific community and a large number of relevant methods exist in the literature. These include the spectral analysis model [1, 2], time series model [3, 4], regression model [5, 6], the Kalman filtering model [7, 8], neural network model [9, 10], support vector machine model [11, 12], and wavelet network model [13]. Reader interested in details of models applied in traffic flow prediction field could refer to review papers such as [14–16]. With the development of chaos theory, recent studies such as [17–19] have found that the short-term traffic flow time series data had nonlinear chaotic phenomena. Therefore, short-term traffic flow chaotic predictions have gained special attention. The prediction of chaotic time series could be generally classified into two categories: global prediction and local

prediction. Global prediction methods use all phase points to describe the evolution law and then to predict the future value. A number of researchers have utilized global prediction methods in prediction of chaotic time series. Karunasinghe and Liong [20] investigated the performance of artificial neural network as a global model in chaotic time series predictions compared to local prediction models. Dong et al. [21] adapted the Elman neural network to realize short-time traffic flow prediction based on chaos analysis. Baydaroglu and Kocak [22] used support vector regression model to predict evaporation amounts, and phase space reconstruction is used to prepare input data for SVR. Local prediction methods select K neighboring points to fit the brief evolution trend of phase points and then to obtain the predicted value. Local prediction methods mainly include local average prediction method [23], weighted first-order local prediction method [24], the Lyapunov index prediction method [25], and support vector machine model [26]. Due to the less number of fitting phase points, the local prediction method has the advantage of low computational complexity and high fitting degree. Farmer and Sidorowich [27] had already proved that the performance of local prediction methods

was better than global prediction method under the same embedding dimension. Therefore, local prediction is adopted to achieve short-term traffic flow prediction in this paper.

In order to get the accurate prediction results, we need to find the nonlinear prediction function. However, it is hard to get the accurate function due to the interference of inside and external excitations. But determining the linear function is not hard since detecting linear relations has been focus of much research in statistics and machine learning fields for decades and the resulting algorithms are well understood, well developed, and efficient. So if we could combine both, it will solve the problem. Instead of trying to fit a nonlinear model, we can map the problem from the input space to the feature space by doing a nonlinear transformation using suitably chosen basis functions and then use a linear model in the feature space. The basis function is called kernel function. The linear model in the feature space corresponds to a nonlinear model in the input space. This is the main idea of relevance vector machine (RVM) model. Due to RVM theoretical advantages, it has gained special attention in recent years, such as [28–30]. This paper is motivated to build the short-term traffic flow forecasting model based on RVM because of its ability to deal with the dynamic, nonlinear, and complex traffic flow time series. consequently, it is very suitable for short-term traffic flow prediction.

For these reasons, and with the goal of improving the accuracy of short-term traffic flow prediction, we put forward a short-term traffic flow local prediction method based on combined kernel function relevance vector machine model. The remainder of this paper is structured as follows: Section 2 presents the phase space reconstruction theory. Section 3 gives the process of building combined kernel function relevance vector machine model. Section 4 describes the experiment setup and case study. Section 5 draws some conclusions.

2. Phase Space Reconstruction Theory

Phase space reconstruction theory proposed by Packard et al. [31] is a powerful tool in the study of complicated system. According to the theory of chaos dynamics, the time series contains total useful information and reflects the process of system evolution in a long term. Complex characteristics found in a time series may be the result of temporal evolution on a chaotic attractor, objects of fractal dimension created by means of stretching and folding of space. If we could capture chaotic behavior from the time series signal of traffic flow models, we could enhance our knowledge about the inherent properties of the traffic flow system. Phase space reconstruction theory is used to create topologically equivalent attractors to the original dynamical system using the information from a scalar time series only [32].

Phase space can be reconstructed using delay coordinate method. The basic idea of delay coordinate method is that the evolution of any single variable of a system is determined by the other variables with which it interacts. Information about the relevant variables is thus implicitly contained in the history of any single variable. For a time series $\{x(i),$

$i = 1, 2, \dots, N\}$, the phase space can be reconstructed according to

$$X(i) = \{x(i), x(i + \tau), \dots, x[i + (m - 1)\tau]\}, \quad (1)$$

where τ is delay time and m is embedding dimension.

Embedding dimension and delay time are the key parameters for phase space reconstruction. At present, there are two kinds of views about the selection of these two parameters. One view is that the two parameters are independent and could be determined separately. The methods of calculating delay time include Average Displacement method [33], Mutual Information method [34], and Autocorrelation Function method [35]. The methods of calculating embedding dimension include False Nearest Neighbors method [36], Cao method [37], and G-P method [38]. Another view is that the two parameters are interrelated and should be determined simultaneously, such as C-C method [39]. C-C method can obtain embedding dimension and delay time simultaneously. Compared with other methods, C-C method has the advantage of small amount of calculation and strong anti-interference. Therefore, C-C method is employed to determine delay time τ and embedding window width τ_ω , and then the embedding dimension m is calculated according to $\tau_\omega = (m - 1)\tau$. The principle of C-C method is as follows.

$\{x(i), i = 1, 2, \dots, N\}$ denotes time series data; a new set of vector series denoted by $X = X(i)$ could be obtained through phase space reconstruction. The correlation integral for the embedded time series is the following function:

$$C(m, N, r, \tau) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \theta(r - \|X(i) - X(j)\|_\infty), \quad (2)$$

where r is the neighborhood radius, $X(i)$ is phase point in phase space, τ is delay time, m is embedding dimension, $M = N - (m - 1)\tau$ is the number of embedded points in phase space, N is the length of time series, $\|\cdot\|_\infty$ denotes sup-norm, and $\theta(\cdot)$ is Heaviside unit function; if $x < 0$, $\theta(x) = 0$; if $x \geq 0$, $\theta(x) = 1$. The correlation integral is a cumulative distribution function and denotes the probability that the distance between any two points is less than r . We define the test statistics

$$S(m, N, r, \tau) = C(m, N, r, \tau) - C^m(1, N, r, \tau). \quad (3)$$

The time series $x(i), i = 1, 2, \dots, N$, can be divided into t disjoint time series. The results are as follows:

$$\begin{aligned} &\{x(1), x(t+1), x(2t+1), \dots\} \\ &\{x(2), x(t+2), x(2t+2), \dots\} \\ &\quad \vdots \\ &\{x(t), x(t+t), x(2t+t), \dots\}. \end{aligned} \quad (4)$$

The test statistics is

$$S(m, N, r, t) = \frac{1}{t} \sum_{l=1}^t \left[C_l \left(m, \frac{N}{r}, r, t \right) - C_l^m \left(1, \frac{N}{r}, r, t \right) \right], \quad (5)$$

where C_l denotes the correlation integral of the l th subsequence.

As $N \rightarrow \infty$, we can write

$$S(m, r, t) = \frac{1}{t} \sum_{l=1}^t [C_l(m, r, t) - C_l^m(1, r, t)]. \quad (6)$$

For fixed embedding dimension m and delay time τ , as $N \rightarrow \infty$, $S(m, r, t)$ will be identically equal to 0 for all r if the time series data are independent and identical distribution. However, the actual time series data are finite and correlated, so $S(m, r, t)$ is not equal to 0 generally. Thus, the local optimal times may be either the zero crossings of $S(m, r, t)$ or the times at which $S(m, r, t)$ shows the least variation with r , because this indicates that these points are uniform distribution. Hence, we select the maximum and minimum radius to define quantity.

Consider

$$\Delta S(m, t) = \max \{S(m, r_i, t)\} - \min \{S(m, r_j, t)\}. \quad (7)$$

$\Delta S(m, t)$ measures the maximum deviation of $S(m, r, t) \sim t$ with r . Therefore, the optimal delay time is the first zero crossings of $S(m, r, t) \sim t$ or the first local minimum point of $\Delta S(m, t) \sim t$.

According to the BDS statistic result, we select $m = 2, 3, 4, 5$, $r_j = i\sigma/2$, $i = 1, 2, 3, 4$, to calculate the following variables:

$$\begin{aligned} \bar{S}(t) &= \frac{1}{16} \sum_{m=2}^5 \sum_{j=1}^4 S(m, r_j, t), \\ \Delta \bar{S}(t) &= \frac{1}{4} \sum_{m=2}^5 \Delta S(m, t), \\ S_{\text{cor}}(t) &= \Delta \bar{S}(t) + |\bar{S}(t)|, \end{aligned} \quad (8)$$

where $\bar{S}(t)$ is the mean of $S(m, r, t)$ for all subsequence. The optimal delay time τ is the first local minimum point of $\Delta \bar{S}(t) \sim t$. The delay time window τ_ω is global minimum point of $S_{\text{cor}}(t) \sim t$.

3. Modeling of CKF-RVM Model

3.1. The Principle of RVM Model. The relevance vector machine (RVM) model proposed by Tipping [40] is a sparse probabilistic model based on Bayesian principle. Compared with other intelligent algorithms, RVM owns better performance. For example, the kernel function of RVM model need not be restricted by Mercer's condition. Moreover, it inducts a priori distribution of the weights and then greatly reduces

the complexity of calculation. The principle of RVM model is as follows.

Consider a data set $\{x_n, t_n\}_{n=1}^N$, where $x_n \in R^n$, $t_n \in R$. The relationship between x_n and t_n is as follows:

$$t_n = y(x_n; w) + \xi_n = \sum_{i=1}^N w_i \varphi_i(x) + w_0 + \xi, \quad (9)$$

where $w = (w_0, w_1, \dots, w_N)$ is weight vector, ξ_n is the independent additive noise term subject to $\xi_n \sim N(0, \sigma^2)$, $\varphi_i(x) = K(x, x_i)$ is the nonlinear basis function, and $K(\cdot)$ is the kernel function. Therefore, $p(t_n | x) = N(t_n | y(x_n), \sigma^2)$ denotes the normal distribution of t_n with mean $y(x_n)$ and variance σ^2 . Assume t_n are independent of each other; the likelihood of the complete data set can be written as

$$p(t | w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2\sigma^2} \|t - \Phi w\|^2 \right\}, \quad (10)$$

where $t = (t_1, t_2, \dots, t_N)^T$ and $\Phi = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_N)]^T$ is the $N \times (N-1)$ kernel function matrix in which $\varphi(x_n) = [1, K(x_n, x_1), K(x_n, x_2), \dots, K(x_n, x_N)]^T$.

Because there are many parameters in the model, the maximum likelihood estimates of ω and σ^2 will lead to severe overfitting. Therefore, the sparse Bayesian theory is adopted and a prior zero-mean Gaussian distribution over ω is as follows:

$$p(\omega | \alpha) = \prod_{i=0}^N N(\omega_i | 0, \alpha_i^{-1}), \quad (11)$$

where $\alpha = \{\alpha_0, \alpha_1, \dots, \alpha_N\}$ is a vector of $N+1$ hyperparameters. Each weight is individually associated with a parameter, which controls the influence of the prior distribution over associated weight.

Because we have defined the prior probability distribution and the likelihood distribution, the posterior probability distribution is as follows according to the Bayesian theory:

$$\begin{aligned} p(w | t, \alpha, \sigma^2) &= (2\pi)^{-(N+1)/2} |\Sigma|^{-1/2} \\ &\cdot \exp \left\{ -\frac{1}{2} (w - \mu)^T \Sigma^{-1} (w - \mu) \right\}. \end{aligned} \quad (12)$$

Posterior covariance matrix and mean value are as follows, respectively:

$$\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1}, \quad (13)$$

$$\mu = \sigma^{-2} \Sigma \Phi^T t, \quad (14)$$

where $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$.

According to the maximum expected hyperparameter estimation, the value of α and σ^2 can be obtained through iterative algorithm. Consider

$$\alpha_i^{\text{new}} = \frac{r_i}{\mu_i^2}, \quad (15)$$

where μ_i is the i th posterior average weight and $r_i = 1 - \alpha_i N_{ii}$, where N_{ii} is the i th diagonal element of the covariance matrix computed by the current α and σ^2 .

The noise variance σ^2 can be obtained through iterative algorithm

$$(\sigma^2)^{\text{new}} = \frac{\|t - \Phi\mu\|^2}{N - \sum_i r_i}. \quad (16)$$

Given a new sample x_* , t_* is the corresponding prediction value. The probability distribution of prediction value follows a normal distribution $N(\mu_*, \sigma_*^2)$ with mean μ_* and variance σ_*^2 . Consider

$$\begin{aligned} \mu_* &= \mu^T \Phi(x_*), \\ \sigma_*^2 &= \sigma_{\text{MP}}^2 + \Phi(x_*)^T \Sigma \Phi(x_*), \end{aligned} \quad (17)$$

where μ_* is the predictive mean on x_* and σ_*^2 is the predictive variance.

3.2. The Construction of Combined Kernel Function. The traditional relevance vector machine model mostly adopts single kernel function to complete the process of feature space mapping, which has achieved good performance in many practical applications. But the single kernel function has great limitations when the sample data contains heterogeneous information. Therefore, this paper integrates the Gaussian kernel function and polynomial kernel function to construct a new combination kernel function. The form of combination kernel function is as follows:

$$\begin{aligned} K(x, x_i) &= \lambda \cdot \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) + (1 - \lambda) \\ &\quad \cdot (xx_i + 1)^d, \end{aligned} \quad (18)$$

where λ is weight coefficient, $0 \leq \lambda \leq 1$, σ is the kernel width of Gaussian kernel function, and d is the order of polynomial kernel function.

Different kernel functions have different advantages; if the weight coefficient of combination kernel function is inappropriate, the performance of combination kernel function may be lower than single kernel function. Therefore, proper weight coefficient is of great importance for the combined kernel function.

3.3. Parameter Optimization Based on Genetic Algorithm. There are three parameters that need to be optimized in the combined kernel function. The commonly used parameter optimization methods mainly include cross validation method [41] and grid search method [42]. But these methods have a large amount of calculation and are often trapped in local optimum. Genetic algorithm (GA) [43] is a heuristic scientific method based on Darwin's biological evolutionism, which has been widely applied to solve high dimensional optimization problem for parameter optimization in engineering and science areas. Genetic algorithm differs from traditional search and optimization methods in four significant points: Genetic algorithms search parallel from a population of points. Therefore, it has the ability to avoid being trapped in

local optimal solution like traditional methods, which search from a single point. Genetic algorithms use probabilistic selection rules, not deterministic ones. Genetic algorithms work on the chromosome, which is encoded version of potential solutions' parameters, rather than the parameters themselves. Genetic algorithms use fitness score, which is obtained from objective functions, without other derivative or auxiliary information.

Therefore, genetic algorithm is used to obtain the optimal parameters of combination kernel function. The specific steps are as follows.

Step 1 (initialize the parameters). The population size and maximal generation count: the population size is 20, and the maximal generation count is 100.

Step 2 (representation). The parameters to be optimized λ , σ , and d are coded in binary to generate the chromosomes.

Step 3 (fitness function definition). The cross validation method is used to prevent overfitting and underfitting. The training data set is randomly divided into K subsets in K -fold cross validation. The RVM model is built using $K - 1$ subset as the training set. The performance of the parameters is checked on the K th subset. In this paper, fivefold cross validation method is used. The fitness function is defined as the mean absolute percentage error of the fivefold validation method on the training data set.

Step 4 (creating new population). Selection, crossover, and mutation are carried out to generate population. The chromosomes with better fitness function values are selected using the roulette wheel method. The crossover probability of creating new chromosomes is set to 0.8. Mutation probability is set to 0.05.

Step 5 (stopping criteria determine). If the generation count reaches its maximum value, the iteration is stopped. Otherwise, the process is repeated from Step 3 to Step 4.

4. Experiment Setup and Case Study

4.1. Data Source. The experimental traffic flow data come from loop detectors located on the north-south viaduct expressway in Shanghai, China. This segment includes 24 mainline detecting sections and 30 ramp detecting sections, equipped with 88 mainline loop detectors and 60 ramp loop detectors, respectively. The experimental data are collected on five consecutive Mondays from September 1, 2008, to September 29, 2008. The original time interval of collected data is 5 min. Figure 1 gives the traffic flow time series data from five consecutive Mondays.

4.2. Phase Space Reconstruction. Phase space reconstruction is the basis of chaotic time series analysis which affects the prediction performance directly. This paper selects C-C method to complete phase space reconstruction. Figure 2 gives the curve graph between $\Delta\hat{S}(t)$ and t . Figure 3 gives the curve graph between $S_{\text{cor}}(t)$ and t .

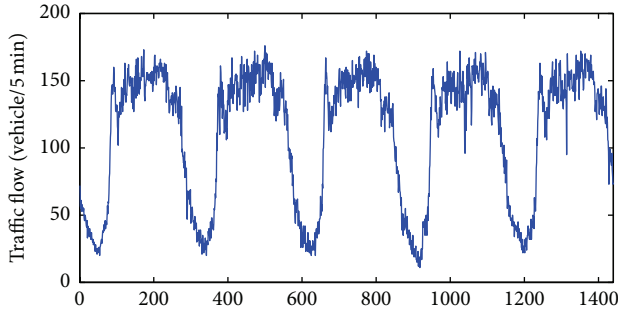


FIGURE 1: The traffic flow time series data from five consecutive Mondays.

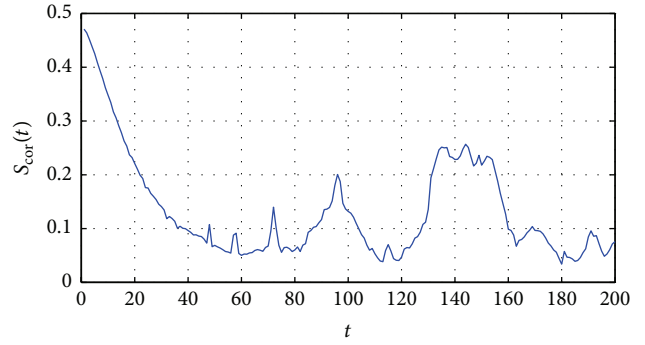


FIGURE 3: The curve graph between $S_{cor}(t)$ and t .

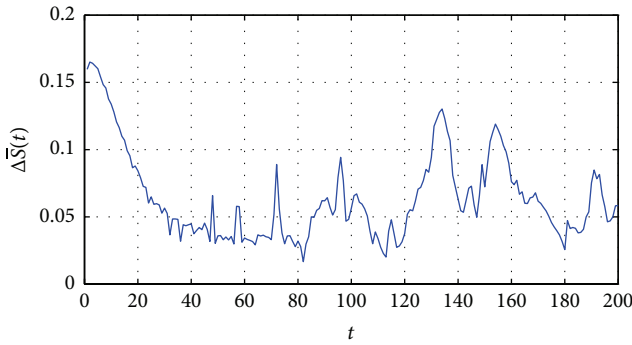


FIGURE 2: The curve graph between $\Delta\bar{S}(t)$ and t .

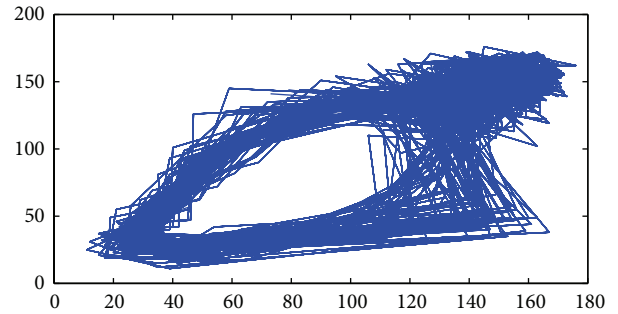


FIGURE 4: The 2D attractor of the reconstructed phase space for traffic flow time series.

From Figure 2, it can be seen that when $\Delta\bar{S}(t)$ get the first minimum, the value of t is 18. Therefore, the value of delay time τ is determined to be 18. From Figure 3, when $S_{cor}(t)$ get the global minimum, the value of t is 113. Therefore, the value of τ_ω is 113, and the embedding dimension m is determined to be 7 according to $\tau_\omega = (m - 1)\tau$.

Figure 4 displays the 2D attractor of the reconstructed phase space for traffic flow time series.

From Figure 4, we could see clearly that the 2D attractor for traffic flow time series is well-regulated, which instructs that the C-C method could implement phase space reconstruction of traffic flow time series excellently.

4.3. Identification of Chaos. Among the wide variety of methods available for chaos identification, the most popular one is the largest Lyapunov exponent method. The main methods of calculating largest Lyapunov exponent include Wolf method [44], Jacobian method [45], and small data sets method [46]. Due to the smaller amount of calculation and clear principles, the small data sets method is employed to calculate the largest Lyapunov exponent of traffic flow time series. Figure 5 displays the result of small data sets method. The linear range is from 57 to 98, and the largest Lyapunov exponent corresponding to the slope value is obtained after the least-squares fit for the linear range. The largest Lyapunov exponent is found to be 0.0014, and this positive value implies an exponential divergence of the trajectories and hence a strong signature of chaos.

4.4. The Number of Neighboring Points. The number of neighboring points is one of the most important parameters which affects the prediction accuracy and the amount of calculation. If the number of neighboring points is too little, the nonlinear fitting advantage of relevance vector machine model will not be reflected. However, if the number of neighboring points is too much, the amount of calculation will increase greatly and the overfitting phenomenon will appear. Therefore, the Hannan-Quinn criteria [47] are used to determine the number of neighboring points. Figure 6 shows the results of Hannan-Quinn criteria.

According to Hannan-Quinn criteria, when $C(k)$ gets the minimum value, the corresponding k is the number of neighboring points which we need. From Figure 6, we could see that the number of neighboring points is 26.

4.5. Parameter Optimization. Genetic algorithm is used to optimize λ , σ , and d . The specific parameters of genetic algorithm are as follows: the population size is 20, maximal generation count is 100, the crossover probability is 0.8, and the mutation probability is 0.05. Figure 7 gives the fitness curve.

From Figure 7, we could see that the optimal parameters of combined kernel function are $\lambda = 0.57$, $\sigma = 0.25$, and $d = 3$.

4.6. Performance Evaluation Index. In order to evaluate the performance of the proposed method, two different types of

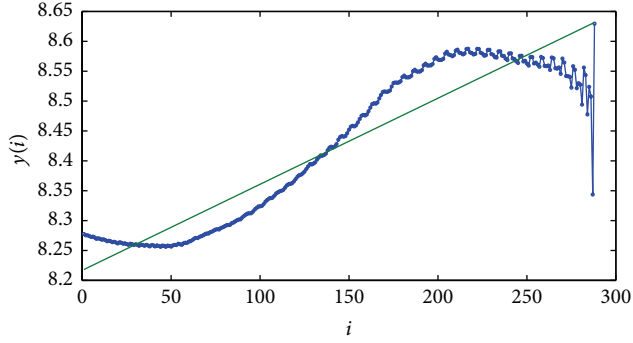


FIGURE 5: The result of small data sets method.

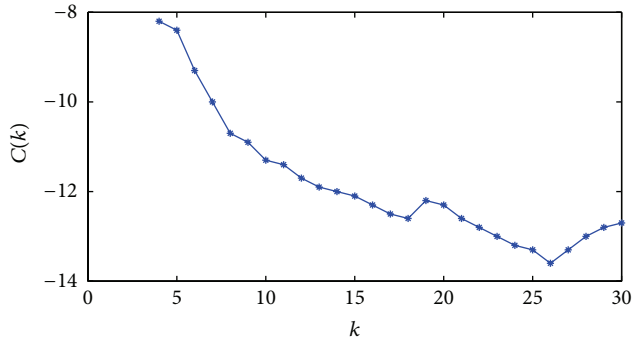


FIGURE 6: The number of neighboring points based on Hannan-Quinn criteria.

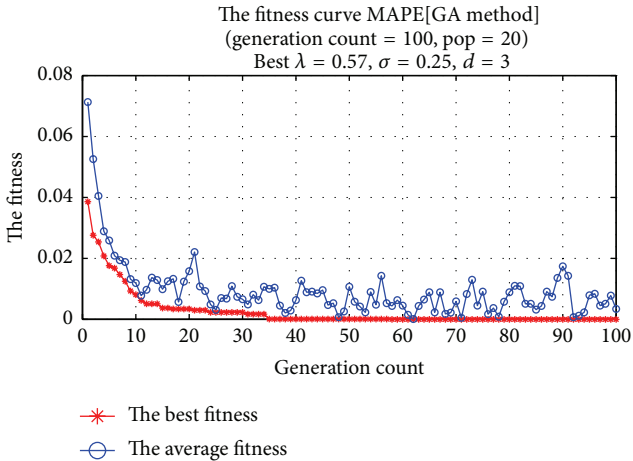


FIGURE 7: The fitness curve of GA.

measurements are introduced: the mean absolute percentage error denoted by MAPE and equal coefficient denoted by EC. The equations for the MAPE and EC are as follows:

$$\begin{aligned} \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \\ \text{EC} &= 1 - \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^n y_i^2 + \sum_{i=1}^n \hat{y}_i^2}}, \end{aligned} \quad (19)$$

TABLE 1: Prediction performance comparison of different methods.

Model	East mainline		West mainline	
	MAPE	EC	MAPE	EC
CKF-RVM	5.0%	0.987	5.4%	0.982
GKF-RVM	7.2%	0.958	7.9%	0.965
GKF-SVM	10.6%	0.940	11.4%	0.935

where y_i denotes the actual value for the i th time interval, \hat{y}_i denotes the predicted value for the i th time interval, and n is the total number of time intervals.

4.7. Model Performance and Analysis. Data collected from September 1 to September 22 are used as training samples, and data collected on September 29 are used as test samples to evaluate the performance of prediction model. In order to illustrate the predictive performance of the proposed method intuitively, Figure 8 presents prediction results based on the proposed method. The black line stands for actual traffic flow data, and the red line stands for the prediction results of CKF-RVM model. Figures 8(a) and 8(b), respectively, show the prediction results and the MAPE for east mainline detector denoted by NBDX16(2). Figures 8(c) and 8(d), respectively, show the prediction results and the MAPE for west mainline detector denoted by NBXX10(1).

As shown in Figure 8, the prediction results are quite close to the actual data, and the MAPE are mostly within 10%. However, the MAPE from 0:00 to 4:00 is high, and this is because the actual traffic flow data during that time period is small. Overall, the CKF-RVM model achieves good prediction performance, which could meet the needs of short-term traffic flow prediction.

To describe the superiority of the proposed method detailedly, comparative analysis is carried out. This paper selects Gaussian kernel function relevance vector machine (GKF-RVM) model and Gaussian kernel function support vector machine (GKF-SVM) model as comparative approaches. For the sake of comparison and analysis in terms of macroscopic and microscopic aspects, Figure 9 gives the microscopic comparative results of different methods. Figure 9(a) shows the prediction results for east mainline detector denoted by NBDX11(1), and Figure 9(b) shows the prediction results for west mainline detector denoted by NBXX15(2). Table 1 gives the macroscopic comparative results of different methods.

As shown in Figure 9, we could see clearly that the prediction results of CKF-RVM model have the best fitting performance comparing to GKF-RVM model and GKF-SVM model. Therefore, the CKF-RVM model could further improve the accuracy of short-term traffic flow prediction.

From Table 1, we could see that the overall improvement of CKF-RVM model is obvious comparing to GKF-RVM model and GKF-SVM model. More precisely, the CKF-RVM model has an extra 31.1% improvement over the GKF-RVM model and an extra 52.7% improvement over the GKF-SVM model in the aspect of MAPE. Meanwhile, the CKF-RVM model is also superior to the other two models in

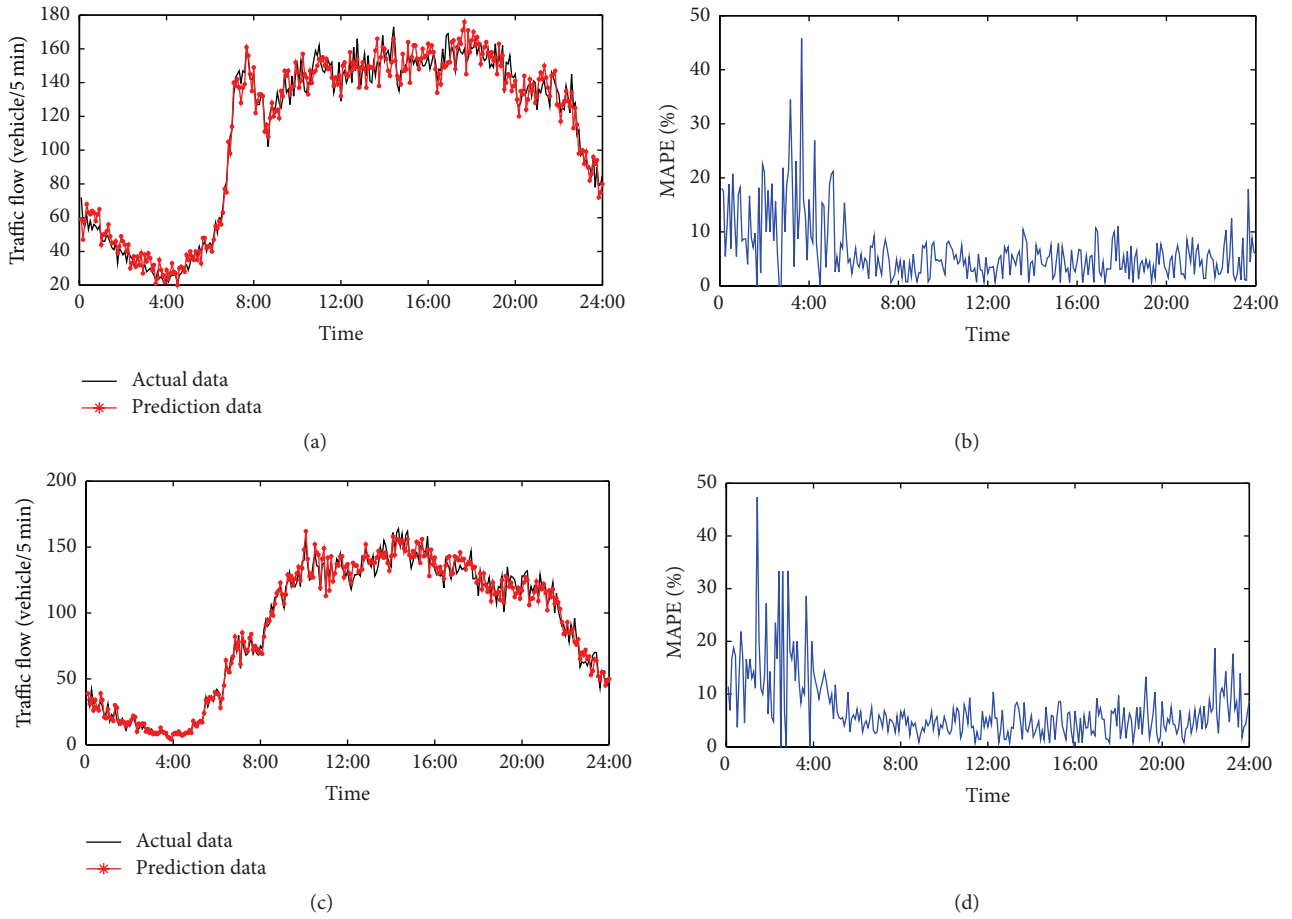


FIGURE 8: The prediction performance based on the proposed method.

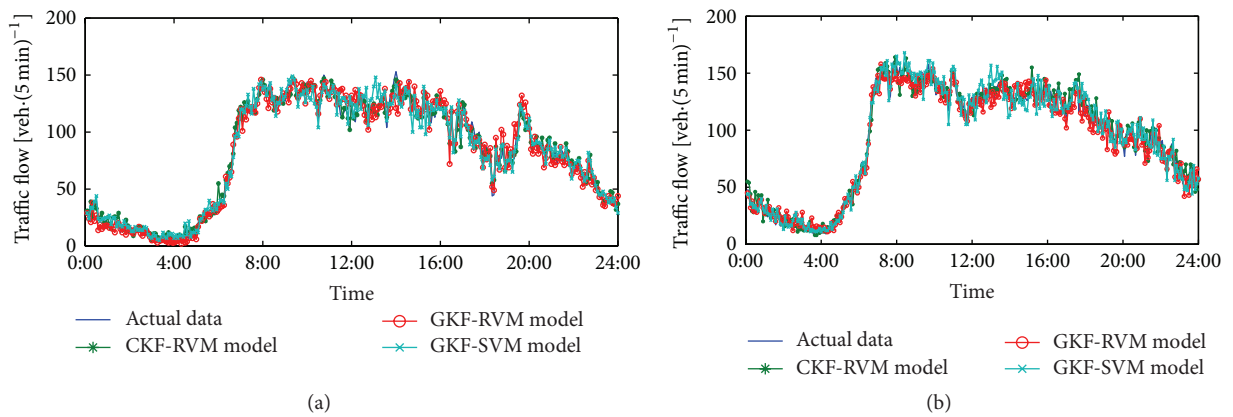


FIGURE 9: The microscopic comparative results of different methods.

the aspect of EC. Furthermore, the experimental results also demonstrate that the CKF-RVM model achieves good prediction performance for both east mainline data and west mainline data, which proves that CKF-RVM model has strong generalization ability. Overall, the CKF-RVM model is an effective and accurate method for short-time traffic flow prediction, which can provide satisfactory prediction results.

5. Discussion and Conclusions

This paper proposes a new short-term traffic flow local prediction method based on combined kernel function relevance vector machine model. The proposed method is more in line with the short-term traffic flow characteristic, which are nonlinear, chaotic, and nonstationary. The main

contribution of this paper is not the specific techniques but rather the demonstration that the forecasting model should take the dynamic characteristics of short-term traffic flow into consideration. The most important contribution is that this paper provides the new idea and methodology to the relevance vector machine model on how to construct the combined kernel function for the short-term traffic flow forecasting model and how to optimize and identify the model structure parameters efficiently and effectively.

Traffic flow data collected from expressway are employed to evaluate the prediction performance of the proposed method, and the results are encouraging. The theoretical advantage and better performance from our studies indicate that the CKF-RVM model has good potential to be developed and is feasible in applying for short-term traffic flow prediction. In order to have more general and robust conclusions, traffic data from different roadways require further exploration. And future studies need to apply the model to other traffic variable data sets (such as traffic speed, travel time, and average occupancy; this study chooses the traffic flow as the demonstration). Moreover, it will be interesting to test traffic data set in different time intervals in the model.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors express their sincere appreciation to the Chinese National High Technology Research and Development Program Committee for the financial support provided under Grant no. 2014BAG03B03, National Natural Science Foundation of China (no. 51408257 and no. 51308248), and Jilin Province Science and Technology Development Plan of Youth Research Fund Project no. 20140520134JH.

References

- [1] H. Nicholson and C. D. Swann, "The prediction of traffic flow volumes based on spectral analysis," *Transportation Research Part C*, vol. 8, no. 6, pp. 533–538, 1974.
- [2] Y. R. Zhang, Y. L. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 65–78, 2014.
- [3] S. Ishak and H. Al-Deek, "Performance evaluation of short-term time-series traffic prediction model," *Journal of Transportation Engineering*, vol. 128, no. 6, pp. 490–498, 2002.
- [4] J. Yeon, L. Eleftheriadou, and S. Lawphongpanich, "Travel time estimation on a freeway using Discrete Time Markov Chains," *Transportation Research Part B: Methodological*, vol. 42, no. 4, pp. 325–338, 2008.
- [5] S. Clark, "Traffic prediction using multivariate nonparametric regression," *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 161–168, 2003.
- [6] Y. Kamarianakis, H. Oliver Gao, and P. Prastacos, "Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 821–840, 2010.
- [7] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B*, vol. 18, no. 1, pp. 1–11, 1984.
- [8] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141–167, 2005.
- [9] Z. S. Yang, Q. C. Bing, C. Y. Lin, N. Y. Yang, and D. Mei, "Research on short-term traffic flow prediction method based on similarity search of time series," *Mathematical Problems in Engineering*, vol. 2014, Article ID 184632, 8 pages, 2014.
- [10] J. Z. Zhu, J. X. Cao, and Y. Zhu, "Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections," *Transportation Research Part C: Emerging Technologies*, vol. 47, no. 2, pp. 139–154, 2014.
- [11] G. Fu, G. Q. Han, and F. Lu, "Short-term traffic flow forecasting model based on support vector machine regression," *Journal of South China University of Technology (Natural Science Edition)*, vol. 41, no. 9, pp. 71–76, 2013.
- [12] Y. L. Zhang and Y. C. Xie, "Forecasting of short-term freeway volume with v-support vector machines," *Transportation Research Record*, vol. 2024, no. 2007, pp. 92–99, 2007.
- [13] Y. Xie and Y. Zhang, "A wavelet network model for short-term traffic volume forecasting," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 10, no. 3, pp. 141–150, 2006.
- [14] H. Adeli, "Neural networks in civil engineering: 1989–2000," *Computer-Aided Civil and Infrastructure Engineering*, vol. 16, no. 2, pp. 126–142, 2001.
- [15] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: overview of objectives and methods," *Transport Reviews*, vol. 24, no. 5, pp. 533–557, 2004.
- [16] J. W. C. Van Lint and C. P. I. J. Van Hinsbergen, "Short-term traffic and travel time prediction models," *Artificial Intelligence Applications to Critical Transportation Issues*, vol. 22, pp. 22–41, 2012.
- [17] D. S. Dendrinos, "Traffic-flow dynamics: a search for chaos," *Chaos, Solitons and Fractals*, vol. 4, no. 4, pp. 605–617, 1994.
- [18] L. W. Lan, F. Y. Lin, and A. Y. Kuo, "Testing and prediction of traffic flow dynamics with chaos," *Journal of Eastern Asia Society for Transportation Studies*, vol. 5, pp. 1975–1990, 2003.
- [19] C. Frazier and K. M. Kocklman, "Chaos theory and transportation systems: instructive example," *Journal of Transportation Research Board*, vol. 1897, pp. 9–17, 2007.
- [20] D. S. K. Karunasinghe and S.-Y. Liang, "Chaotic time series prediction with a global model: artificial neural network," *Journal of Hydrology*, vol. 323, no. 1–4, pp. 92–105, 2006.
- [21] C. J. Dong, C. F. Shao, J. Li et al., "Short-term traffic flow prediction of road network based on chaos theory," *Journal of System Engineering*, vol. 26, no. 3, pp. 340–345, 2011.
- [22] O. Baydaroglu and K. Kocak, "SVR-based prediction of evaporation combined with chaotic approach," *Journal of Hydrology*, vol. 508, pp. 356–363, 2014.
- [23] G. Sugihara and R. M. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature*, vol. 344, no. 6268, pp. 734–741, 1990.

- [24] L.-L. Zang, L. Jia, L.-C. Yang, and T. Liu, "Chaotic time series model of real-time prediction of traffic flow," *China Journal of Highway and Transport*, vol. 20, no. 6, pp. 95–99, 2007.
- [25] Y. Zhang and W. Guan, "Prediction of multivariable chaotic time series based on maximal Lyapunov exponent," *Acta Physica Sinica*, vol. 58, no. 2, pp. 756–763, 2009.
- [26] J.-S. Zhang, J.-L. Dang, and H.-C. Li, "Local support vector machine prediction of spatiotemporal chaotic time series," *Acta Physica Sinica*, vol. 56, no. 1, pp. 67–77, 2007.
- [27] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Physical Review Letters*, vol. 59, no. 8, pp. 845–848, 1987.
- [28] J. Yan, Y. Liu, S. Han, and M. Qiu, "Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 613–621, 2013.
- [29] Y. Bai, P. Wang, C. Li, J. Xie, and Y. Wang, "A multi-scale relevance vector regression approach for daily urban water demand forecasting," *Journal of Hydrology*, vol. 517, pp. 236–245, 2014.
- [30] S. W. Fei and Y. H. He, "Wind speed prediction using the hybrid model of wavelet decomposition and artificial bee colony algorithm-based relevance vector machine," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 625–631, 2015.
- [31] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.
- [32] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, D. A. Rand and L. S. Young, Eds., vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381, Springer, Berlin, Germany, 1980.
- [33] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica D: Nonlinear Phenomena*, vol. 73, no. 1-2, pp. 82–98, 1994.
- [34] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.
- [35] J. Holzfuss and G. Mayer-Kress, "An approach to error-estimation in the application of dimension algorithms," *Dimensions and Entropies in Chaotic Systems*, vol. 32, pp. 114–122, 1986.
- [36] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 6, pp. 3403–3411, 1992.
- [37] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43–50, 1997.
- [38] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, no. 5, pp. 346–349, 1983.
- [39] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics, delay times, and embedding windows," *Physica D: Nonlinear Phenomena*, vol. 127, no. 1-2, pp. 48–60, 1999.
- [40] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
- [41] Y. L. Zhang and Y. H. Yang, "Cross-validation for selecting a model selection procedure," *Journal of Econometrics*, vol. 187, no. 1, pp. 95–112, 2015.
- [42] X. L. Liu, D. X. Jia, and H. Li, "Research on kernel parameter optimization of support machine in speaker recognition," *Science Technology and Engineering*, vol. 10, no. 7, pp. 1669–1673, 2010.
- [43] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1976.
- [44] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D: Nonlinear Phenomena*, vol. 16, no. 3, pp. 285–317, 1985.
- [45] M. Sano and Y. Sawada, "Measurement of the Lyapunov spectrum from a chaotic time series," *Physical Review Letters*, vol. 55, no. 10, pp. 1082–1085, 1985.
- [46] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica D: Nonlinear Phenomena*, vol. 65, no. 1-2, pp. 117–134, 1993.
- [47] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 41, no. 2, pp. 190–195, 1979.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

