



2018

A Hybrid Forecasting Framework Based on Support Vector Regression with a Modified Genetic Algorithm and a Random Forest for Traffic Flow Prediction

Lizong Zhang

the School of Computer Science and Engineering, University of Electronic Sciences and Technology of China, Chengdu 611731, China.

Nawaf R. Alharbe

the College of Community, Taibah University, Al-Madinah, Saudi Arabia.

Guangchun Luo

the School of Computer Science and Engineering, University of Electronic Sciences and Technology of China, Chengdu 611731, China.

Zhiyuan Yao

the School of Computer Science and Engineering, University of Electronic Sciences and Technology of China, Chengdu 611731, China.

Ying Li

the School of Computer Science and Engineering, University of Electronic Sciences and Technology of China, Chengdu 611731, China.

Follow this and additional works at: <https://tsinghuauniversitypress.researchcommons.org/tsinghua-science-and-technology>



Part of the [Computer Sciences Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Lizong Zhang, Nawaf R Alharbe, Guangchun Luo et al. A Hybrid Forecasting Framework Based on Support Vector Regression with a Modified Genetic Algorithm and a Random Forest for Traffic Flow Prediction. *Tsinghua Science and Technology* 2018, 23(4): 479-492.

This Research Article is brought to you for free and open access by Tsinghua University Press: Journals Publishing. It has been accepted for inclusion in *Tsinghua Science and Technology* by an authorized editor of Tsinghua University Press: Journals Publishing.

A Hybrid Forecasting Framework Based on Support Vector Regression with a Modified Genetic Algorithm and a Random Forest for Traffic Flow Prediction

Lizong Zhang, Nawaf R Alharbe*, Guangchun Luo, Zhiyuan Yao, and Ying Li

Abstract: The ability to perform short-term traffic flow forecasting is a crucial component of intelligent transportation systems. However, accurate and reliable traffic flow forecasting is still a significant issue due to the complexity and variability of real traffic systems. To improve the accuracy of short-term traffic flow forecasting, this paper presents a novel hybrid prediction framework based on Support Vector Regression (SVR) that uses a Random Forest (RF) to select the most informative feature subset and an enhanced Genetic Algorithm (GA) with chaotic characteristics to identify the optimal forecasting model parameters. The framework is evaluated with real-world traffic data collected from eight sensors located near the I-605 interstate highway in California. Results show that the proposed RF-CGASVR model achieves better performance than other methods.

Key words: traffic flow forecasting; feature selection; parameter optimization; genetic algorithm; machine learning

1 Introduction

As a component of the smart city concept, intelligent transportation systems aim to relieve traffic congestion, reduce fuel consumption, and provide reliable, safe, and green transportation^[1]. An important issue and a crucial component of intelligent transportation systems is the ability to perform real-time forecasting of traffic flow (i.e., short-term traffic flow forecasting), which focuses on predicting traffic flow over a certain road or past a checkpoint for a short future time interval (usually less than 30 min). Traffic flow forecasting can be used for real-time traffic management, including traffic light control, route guidance, and congestion alleviation^[2].

Accurate short-term traffic forecasts are vital for traffic managers and travelers. Given the rapid development of information technology and transportation infrastructure^[3,4], real-time traffic data can be obtained more easily than ever before^[5,6]. Over the past few decades, many techniques have been proposed to solve the short-term traffic flow forecasting problem. However, accurate and reliable traffic flow forecasting is still a significant issue due to the complexity and variability of real traffic systems.

This study proposes a novel prediction framework based on Support Vector Regression (SVR), Genetic Algorithm (GA), and Random Forest (RF) methods for short-term traffic flow forecasting. This framework aims to solve the feature selection and parameter optimization issues of SVR. The framework consists of three major components. The first component is a feature selection method based on RF, but this method performs feature selection during the learning process rather than by applying cross-validation or individual feature ranking. The second component is a parameter optimization method for SVR that uses an enhanced GA with chaotic characteristics (CGA). In particular, a modified

• Lizong Zhang, Guangchun Luo, Zhiyan Yao, and Ying Li are with the School of Computer Science and Engineering, University of Electronic Sciences and Technology of China, Chengdu 611731, China. E-mail: L.zhang@uestc.edu.cn.

• Nawaf R. Alharbe is with the College of Community, Taibah University, Al-Madinah, Saudi Arabia. E-mail: N.paper2018@outlook.com; NAWAF_3130@hotmail.com.

* To whom correspondences should be addressed.

Manuscript received: 2017-11-14; accepted: 2017-12-23

chromosome generation method and a chaotic mutation strategy for GA are proposed to help avoid local optimums. The enhanced GA provides improved solutions to optimization problems, such as SVR parameter selection. The final component of the framework applies the CGA and RF results to optimize the SVR-based forecasting model. The effectiveness of the proposed framework is evaluated using real-world traffic flow data collected from the Performance Measurement System (PeMS) of the California Department of Transportation. The experimental results show that the proposed forecasting model provides more accurate forecasts while requiring fewer features than other methods, such as standard SVR with grid search, Back-Propagation Neural Network (BPNN), and Autoregressive Integrated Moving Average (ARIMA).

The contributions of this study are as follows. First, this work proposes a short-term traffic flow forecasting method based on the hybrid RF, GA, and SVR approaches that uses RF for feature selection and the modified GA to optimize the SVR parameters. Second, this work proposes a feature selection model within the learning process based on RF to select the most informative feature subset. Third, a modified chromosome initial stage and a mutation stage are applied to handle the local convergence problem caused by the random operation of standard GAs.

The rest of this article is organized as follows. Section 2 discusses related work regarding short-term traffic forecast models, parameter optimization, and feature selection. Section 3 describes the proposed hybrid framework in detail. Section 4 evaluates the effectiveness of the framework by using an experiment with real-world data. Finally, Section 5 presents the conclusions of the study and recommendations for future work.

2 Related Work

Accurate short-term traffic flow forecasting is vital for traffic managers and travelers. Various forecasting approaches have been applied to different disciplines. Many studies have employed parametric approaches, treated traffic information as time series data, and employed time series analysis methods to recognize historical data patterns and forecast future trends^[7]. Parametric approaches, including k-nearest neighbor^[8], Kalman filter^[9], and ARIMA^[10,11] methods, provide ideal results when the traffic flow variations are regular and stable. However, the performance of these methods suffers under realistic traffic conditions because the real world is

complex and exhibits irregular variation.

Machine learning approaches, such as Artificial Neural Network (ANN) and Support Vector Machine (SVM), have been proposed to cope with nonlinear patterns of traffic flow data. These methods are used to predict traffic by analyzing the relationships between historic and future data. As a typical machine learning technique, ANN relies on the principle of empirical risk minimization; thus, it may result in low accuracy for small samples and have an overfitting problem with large datasets. Scholars have also proposed several improved ANNs, including BPNN, Radial Basis Function neural network (RBF)^[12], and general regression neural network^[13]. In contrast to ANN, SVM is based on the structural risk minimization principle, which reduces the risk of traditional empiricism and model complexity. SVMs have achieved great success on many real-world problems^[14–16], not only for classification but also for regression problems, in which an extension algorithm, namely, SVR, is applied to introduce the ε -insensitive loss function.

The SVR model can yield a globally optimized result for most nonlinear problems. However, two issues strongly affect the prediction performance of the model. The first issue is feature selection. Training the SVR forecasting model by using the original feature set may reduce the efficiency and effectiveness of the model due to redundant or noise features in the original data. The second issue is parameter optimization. The SVR performance relies on a suitable combination of parameters, and incorrect parameter settings can result in unacceptable performance.

Researchers have focused on the feature selection issue. In general, feature selection can be categorized into two main approaches, namely, filter^[17] and wrapper^[18] approaches. The main idea of the filter approach is to rank features according to predefined criteria, such as statistical measures, which are completely independent of the forecasting model^[19]. Huang and Tsai^[20] used a filter-based feature selection approach to choose the most important input features and applied SVR to predict traffic flow. Zhu et al.^[21] proposed a novel adaptive SVR based on the filter approach to efficiently remove the corresponding impulse noise. Wrapper approaches assess the quality of every combination of features through the forecasting model and choose the optimal features based on its performance^[22]. Maldonado and Weber^[23] introduced a wrapper algorithm to decide which features to remove in SVR.

Other studies have focused on identifying the optimal parameters for forecasting training method. Several

algorithms are frequently used to select the optimal parameter values; these algorithms include Particle Swarm Optimization (PSO) algorithm, simulated annealing algorithm, and GA. PSO^[24, 25] is a group-based intelligent optimization algorithm that was originally developed based on the study of the predatory behavior of birds. Several methods have been proposed based on PSO for selection of SVR parameters^[26–28]. The simulated annealing algorithm^[29] is a random search method inspired by physical annealing. Researchers have adopted this approach to optimize SVR parameters^[30, 31]. However, the optimal solution is often abandoned in favor of acceptance probability; consequently, the final solution is not necessarily the optimal solution.

GAs^[32] are based on biological evolution and can be easily combined with other models. GAs have been applied to optimize SVR parameters in many applications^[33–35]. However, the traditional GA has limitations, such as slow search and tendency to fall into the local optima, known as the “pre-maturation” problem.

Performing feature selection and parameter optimization tasks within the same framework, as in the case of traffic flow forecasting, has been rarely reported. The relationship between feature subsets and parameters has not been fully considered. The parameters must “fit” with the feature subsets to achieve the highest accuracy with few features. Thus, these two optimization tasks share the same importance in constructing a forecasting model.

Motivated by these challenges, the present work proposes a novel prediction framework based on SVR, GA, and RF for short-term traffic flow forecasting.

3 Proposed Hybrid Framework

3.1 Overview of the proposed hybrid framework

This study presents a three-stage hybrid method that combines the prediction capability of SVR, the feature-ranking characteristics of RF, and the optimization ability of GA. This framework is illustrated in Fig. 1.

The first stage employs the RF algorithm to roughly evaluate the importance of each feature. This algorithm randomly creates training data and feature subsets from the original data to generate multiple decision trees. An ordered feature list is created by testing the accuracy impact of each feature of each tree. In contrast to other filter-based methods that measure the accuracy of the feature subset with the generated RF model itself^[36, 37], the proposed model simultaneously validates the feature

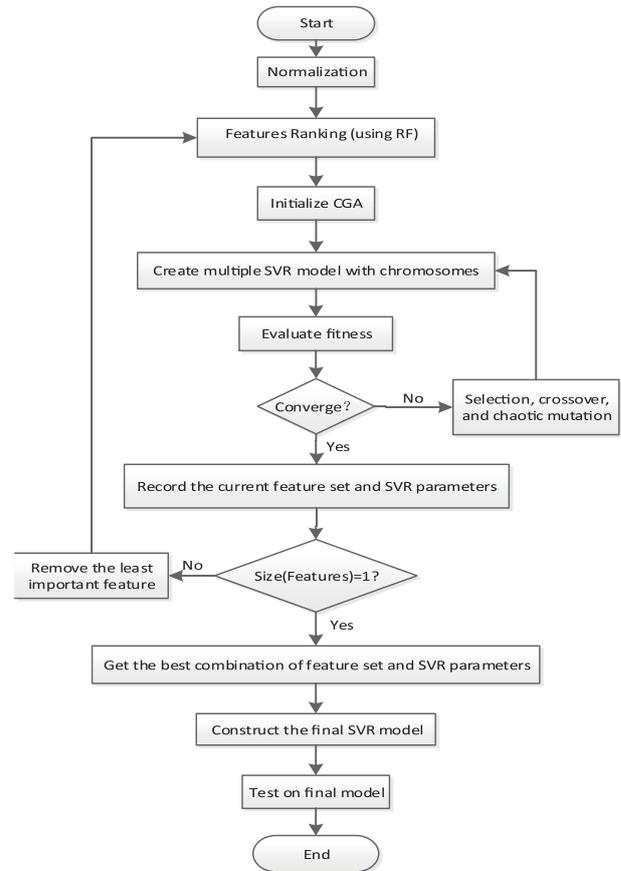


Fig. 1 Proposed RF-CGASVR framework.

subset and the SVR parameters to find the most predictive features and the optimal parameter values. Thus, the goal of the second stage is to find the parameter set given the feature sets obtained from the first stage. CGA, which uses modified chromosome encoding, population initialization, and mutation strategy, is adopted to provide improved convergence by introducing chaotic characteristics that balance randomness and ergodicity^[38]. Chromosomal fitness is assessed by the SVR model using the given feature sets.

The third stage involves determining the best combination of the feature subset and the parameter set to construct a forecasting model. Thus, the procedure described above is executed multiple times with different feature subsets, and features with low effect are removed during each iteration. Finally, the optimized parameters are identified in accordance with the optimized feature subset, and the forecasting model is constructed.

3.2 SVR algorithm

SVR is an extension of SVM for solving regression problems to find a function that represents the relationships in historical data; the identified function can accurately predict future values. The original SVM was introduced

by Cortes and Vapnik at the Computational Learning Theory Conference in 1995^[39]. SVM is designed to solve classification problems by finding an optimal hyperplane for either linear or nonlinear problems by mapping data to a high-dimensional feature space. Using Lagrange multipliers, the problem is transformed into a convex quadratic programming problem^[40] that has a global optimal solution. SVM can be extended to solve regression problems with the ε -insensitive loss function^[41].

Given the training data set $\{(x_1, y_1), \dots, (x_n, y_n)\}$, where each $x_i \in \mathbf{R}^l$ is a l -dimensional input vector that contains *one* feature, in traffic flow forecasting, $l = N \times TP$, where N is the number of sites, TP is the number of data collection time points, and y_i is the corresponding response value (i.e., the forecasted traffic data). The goal of SVR is to find a function that best maps the input x to the output y . The generic SVR function can be represented as follows:

$$f(x) = W\phi(x) + b \quad (1)$$

where W is the weight vector and b is the bias. This function determines a hyperplane that describes the linear relationship between $\phi(x)$ and y in a high feature space, and $\phi(x)$ represents the nonlinear mapping of x . Thus, a complex nonlinear problem is converted into a linear problem. To find the values of W and b , the following objective function is solved^[42]:

$$\min 1/2 \|W\|^2 + C \sum_{i=1}^n L_\varepsilon(f(x_i) - y_i), \quad (2)$$

$$L_\varepsilon(f(x_i) - y_i) = \begin{cases} 0, & \text{if } |y - f(x)| \leq \varepsilon; \\ |y - f(x)| - \varepsilon, & \text{otherwise} \end{cases}$$

where L_ε is the ε -insensitive loss function, and C is a constant parameter that balances the model complexity and training error defined by L_ε . By introducing slack variables, the function can be expressed as follows:

$$\min 1/2 \|W\|^2 + C \left(\sum_{i=1}^n (\xi_i + \xi_i^*) \right),$$

$$\text{s.t.}, \begin{cases} f(x) - y_i \leq \varepsilon + \xi_i, \\ y_i - f(x) \leq \varepsilon + \xi_i^*, \\ \xi_i \geq 0, \xi_i^* \geq 0, \quad i = 1, 2, \dots, n \end{cases} \quad (3)$$

where ξ_i and ξ_i^* are non-negative variables that represent the deviation between the actual data and the edge and are determined by ε and $f(x)$, respectively. This optimization problem can be transformed into a dual problem by using Lagrange multipliers. Under the well-known Karush-Kuhn-Tucker conditions, the prediction function f can be obtained as follows^[43]:

$$f(x) = \sum_{i=1}^n (a_i^* - a_i) K(x, x_i) + b, \quad (4)$$

$$\text{s.t.}, 0 \leq a_i^* \leq C, \quad 0 \leq a_i \leq C$$

where a_i^* and a_i are the Lagrange multipliers that must be determined by solving the dual problem. K is the kernel function that allows the dot product to be processed in high-dimensional feature space by using known low-dimensional space data without explicit transforming operator ϕ . Any function that meets the Mercer's condition can be used as a kernel function. The present study uses RBF, which is commonly used in regression problems:

$$k(x, x_i) = \exp(-\sigma \|x - x_i\|^2) \quad (5)$$

where σ is a parameter that is normally selected manually (e.g., C and ε). However, an inappropriate parameter setting can result in unacceptable performance. Moreover, large values of σ or C may lead to overfitting because of the overreduction of the training error, whereas small values of σ or C can result in underfitting due to the overreduction of model complexity. Furthermore, all features (dimensions) of x_i are treated equally in the learning process but may either be correlated or irrelevant to the forecasting. Therefore, applying all features without selecting the most predictive features not only increases the computational complexity of the model but also affects its forecasting performance. In addition, the choice of the feature subset can influence the appropriate SVR parameters (and vice versa)^[44].

Considering these two issues, the challenge is to identify the combination of the most informative feature subset and the optimal parameters for the prediction model in the context of traffic flow forecasting.

3.3 Feature selection algorithm

An RF approach is introduced in this study for feature selection. RF can be applied with two main approaches, namely, filter or wrapper methods. Filter methods remove the least effective features by evaluating and ranking them without considering their interrelationships or their effects on the forecasting model. Thus, identifying an optimal feature subset for a specific classification or regression model is difficult using this approach^[45]. By contrast, wrapper methods evaluate features based on the accuracy of the forecasting model and by considering the possible interactions among the features^[46, 47].

This study performs feature ranking by using the RF algorithm, but the features are evaluated and selected by the forecasting model with SVR^[48]. RF is a machine learning algorithm designed for classification and regression problems. This algorithm employs a bootstrap

sampling technique and a Classification And Regression Tree (CART) algorithm to generate multiple unrelated decision trees. RF then combines the results to enhance the generalization performance^[49].

The RF algorithm can be described as follows:

Step 1. Generate n different training sample sets for n CARTs from the original dataset by using a bootstrap sampling technique.

Step 2. For each bootstrap training sample, randomly select p features from all the input features m , where $p < m$.

Step 3. Generate a decision tree from the training sample and selected features by using the CART algorithm.

Step 4. Predict new output by averaging the outputs of n CARTs when new inputs are fed into the RF.

The traditional RF algorithm performs feature ranking using the Out-Of-Bag (OOB) data that are excluded from the bootstrap sampling to calculate feature importance. The procedure is described as follows:

Step 1. For each decision tree in an RF, calculate the OOB error using the corresponding OOB data, denoted as $errOOB1$.

Step 2. For x in all OOB data samples, after randomly adding noise interference, calculate the OOB error again, denoted as $errOOB2$.

Step 3. Calculate the importance of feature x by $\Sigma(errOOB1 - errOOB2)/nTree$, where $nTree$ is the number of trees in the RF.

However, feature selection is important in constructing forecasting models. Using only the ranking information from the RF without considering the forecasting model is insufficient for selecting a small number of crucial features. In addition, the SVR parameters affect the contribution of a feature. Thus, in this study, feature ranking information is used for reference only. Actual feature evaluation and selection are performed by the model itself. The details of this framework are described in Section 3.4.5.

3.4 Parameter optimization

The SVR model contains three key parameters, namely, σ in the kernel function, ε in the loss function, and C in the objective function. These parameters directly affect the generalization performance of the forecasting model. Thus, the values selected for these parameters play a significant role in the traffic forecasting problem.

The forecasting model may perform poorly without suitable parameters. Traditional cross-validation methods

for parameter selection may cause cross error^[50]. To overcome this problem, this study presents an enhanced parameter optimization method using a GA based on tent mapping and chaotic mutation.

3.4.1 Standard GA

GA^[51] is a heuristic search algorithm that has been widely applied in optimization problems. GA was developed from the concepts of genetics, mutation, and natural selection in evolutionary biology. For problems that are difficult to solve with traditional algorithms or contain multiple objectives, GA can help identify the globally optimal solution.

The main procedures of GA are as follows. First, it generates a population of chromosomes. Each individual chromosome in the population represents a candidate solution and consists of multiple genes. The algorithm begins the evolutionary process with a random chromosome population. Second, all chromosomes in the current population are evaluated by a given fitness function that represents the optimization goal. According to the fitness values of the chromosomes, a series of genetic operators, including selection, crossover, and mutation, is applied to generate offspring from the current population and evolve the population toward improved solutions. After many iterations, GA eventually obtains the fittest individual, which represents the optimal solution.

However, slow and premature convergence problems in GA prevent its use in a wide range of applications. These two limitations are directly related to population initialization method and genetic operators. In particular, in the mutation step, the complete random generation of the initial chromosomes cannot guarantee the ergodicity and uniformity of the initial population; moreover, the random selection of gene positions cannot balance the randomness and stable orientation of evolution.

To overcome these problems, this paper presents an enhanced GA that uses tent map method and cloud model for solving the parameter optimization task of SVR for the traffic flow forecasting model. The details of this method are explained as follows.

3.4.2 Chromosome encoding and initialization

The first step of the proposed GA is to represent the solution of the parameter settings as a chromosome and then to generate a set of chromosomes as the initial population. In the case of SVR parameter optimization, a chromosome is a vector that corresponds to the parameters to be tuned. Each element of the vector, called a gene, represents a target parameter. In this study, the

chromosome consists of three elements that represent the three target parameters using real encoding schema. These three parameters are restricted to the range of two pre-defined constants, namely, the minimum and maximum values allowed for the parameter genes.

Tent map method is applied to generate chromosomes and ensure variety in the initial population. This method introduces chaos and thus can balance the ergodicity and uniformity of the initial population. The basic idea of chaos initialization is to generate a chaotic variable sequence according to the population size for each gene and to transform the ergodic range of the chaotic motion into the domain of each parameter. In addition, a random value may be added to the generated chaotic variable to avoid the “fixed periodic points” problem caused by length limitation of a computer “word”. The procedure of chaotic population initialization is as follows:

Step 1. Generate an initial value x_0 that is not in the small periodic point.

Step 2. Generate a chaotic variable by using the following equation:

$$x_{n+1} = \begin{cases} ux_n, & 0 \leq x_n \leq 0.5; \\ u(1-x_n), & 0.5 \leq x_n \leq 1 \end{cases} \quad (6)$$

where n is the number of iterations and u is a control parameter that is usually set to 2.

Step 3. When x_n enters the fixed points or small periodic cycles, i.e., when $x_n = 0, 0.25, 0.5, 0.75$, or $x_n = x(n-k), k = 1, 2, 3, 4$, it is reassigned using the following equation:

$$x_{n+1} = \begin{cases} ux_n + 0.1 \cdot \text{rand}(0, 1), & 0 \leq x_n \leq 0.5; \\ u(1-x_n) + 0.1 \cdot \text{rand}(0, 1), & 0.5 \leq x_n \leq 1 \end{cases} \quad (7)$$

Step 4. Select a chaotic variable after a random number of iterations and repeat three times until each of the three genes have their own chaotic variable.

Step 5. Map the chaotic variable back into the value range of the SVR parameters by using the following equations: $X(i, j) = m + (n-m) \cdot x(i, j), i = 1, 2, \dots, M$ and $j = 1, 2, 3$, where X is the mapped value corresponding to the SVR parameters, M is the population size, j indicates the three genes; and m and n are the minimum and maximum values, respectively, allowed for the gene parameters.

Step 6. Repeat Steps 2 to 6 for M times until all the genes have values and all the chromosomes have been

constructed. At this point, the population is initialized.

3.4.3 Selection, crossover, and mutation

According to the “principle of the survival of the fittest”, the fittest solutions (chromosomes) should survive, whereas the less fit solutions should be removed from the current population. The surviving chromosomes from the current population are used as parents to produce new offspring. In this study, the selection operator follows “rank selection” method and keeps half of the current population.

The crossover operation is then conducted to generate child chromosomes from the selected chromosomes with a possibility of cp . A traditional linear recombination method is applied, and genes in child chromosomes are determined using the following equation:

$$\begin{cases} G_1^c(i) = aG_1^p(i) + (1-a)G_2^p(i); \\ G_2^c(i) = (1-a)G_1^p(i) + aG_2^p(i), \\ i = 1, 2, 3, \quad a \in U[-0.25, 1.25], \quad \text{if rand} > cp \end{cases} \quad (8)$$

where $G_1^p(i)$ and $G_2^p(i)$ are the genes in the parent chromosomes; $G_1^c(i)$ and $G_2^c(i)$ are the genes in the child chromosomes; and a is a uniformly distributed random number.

The next step of GA is a mutation operation. Tent map method is employed again to generate a new value based on the original gene. First, the gene values are mapped back to an interval of $[0, 1]$ by using the following equation:

$$G^*(i) = (G(i) - m) / (n - m) \quad (9)$$

A mutation gene is then generated to replace the original gene by using tent map method:

$$G_{\text{mut}}(i) = f(x_n) = \begin{cases} uG^*(i)(n-m) + m, & 0 \leq G^*(i) < 0.5; \\ u(1-G^*(i)), & 0.5 \leq G^*(i) \leq 1 \end{cases} \quad (10)$$

where $G_{\text{mut}}(i)$ represents the new mutated genes; m and n are the minimum and maximum allowed values, respectively, for parameter genes; and u is a control parameter.

3.4.4 Fitness function and performance evaluation

The fitness function is an important component in GA used to estimate the quality of each chromosome. In this study, an SVR-based fitness evaluation is introduced with widely used measurement criteria, such as Root Mean Square Error (RMSE). RMSE is used to evaluate the performance of traffic flow forecasting models and

serves as the fitness function in GA. In particular, the parameter settings of each chromosome are applied to the SVR forecasting model to perform the forecasting task. RMSE is calculated and used as the fitness value for the corresponding chromosome.

In addition, the Mean Absolute Percentage Error (MAPE) is introduced for performance evaluation. MAPE reflects the error between the predicted and actual values and is typically suitable for measuring datasets with large outliers. By contrast, RMSE is mainly used to evaluate error distribution. For the given validation data set $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, RMSE and MAPE are used for performance evaluation and calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2} \quad (11)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{f(x_i) - y_i}{y_i} \right| \quad (12)$$

where n is the number of test samples, $f(x_i)$ is the forecasting value, and y_i is the actual value for the i -th sample. Thus, the smaller the RMSE or MAPE is, the higher the forecasting accuracy will be, and the fitter the chromosome will be. In this study, RMSE is selected as the fitness function for the proposed CGA.

3.4.5 Framework

A flowchart of the proposed RF-CGASVR parameter optimization is shown in Fig. 1, and the details are described as follows:

Step 1. Collect traffic flow data from a real-world road network and perform data normalization.

Step 2. Rank the features by using the RF algorithm with the training data.

Step 3. Initialize the GA population by generating genes encoded for the three SVR parameters through tent map method.

Step 4. Create multiple SVR models based on the training data with the given feature subsets from the RF model and the chromosome population from GA.

Step 5. Evaluate the fitness of each model on the validation data.

Step 6. If CGA does not converge, then go to Step 7; otherwise, go to Step 8.

Step 7. Perform the selection, crossover, and chaotic mutation operations to generate a new population and then go to Step 4.

Step 8. Record the current feature subset and SVR parameters; if the number of features is 1, then go to step

10; otherwise, go to Step 9.

Step 9. Remove the least important feature from the feature set, and then go to Step 2.

Step 10. Select the best combination of feature subsets and SVR parameters.

Step 11. Construct the final SVR forecasting model.

Step 12. Evaluate the SVR model on the testing data to obtain the forecasting results.

4 Experiments

To evaluate the effectiveness of the proposed RF-CGASVR method for short-term traffic flow forecasting, this study conducted experiments with two typical road layouts, namely, straight road and crossroad. Experiments in the straight road layout were designed to evaluate the forecasting performance in a simple scenario, where all the observation sites were located along the same road. Experiments in the crossroad layout focused on an intersection and used more observation sites located along different roads. The experiments were also designed to assess the capability of the proposed RF-CGASVR method to reveal the spatiotemporal relationships between other road regions and the target sites from the selected feature subset.

4.1 Experiment data

In this study, traffic flow data obtained from eight observation sites located along the I-605 interstate highway in California were used. The data were obtained from the Caltrans PeMS^[52] and can be downloaded from <http://pems.dot.ca.gov>. The eight sites are located from the intersection of Artesia Fwy and I-605 to the intersection of Del Amo Blvd and I-605 (Figs. 2 and 3).

In the straight road layout (Fig. 2), the model predicted the traffic flow of a specific site by using data collected from sensors located along the same road. Three nearby observation sites at I-605 were selected. Sites 2 and 3 were selected because they are located at two nearby intersections, and site 1 was selected because it is the nearest site to the two intersections. The traffic at these three sites flows from north to south. The data from sites 1 and 2 were used to predict traffic at site 3.

In the crossroad layout (Fig. 3), the model adopted data from multiple sensors located along different roads but around an intersection to perform the forecasting task. To evaluate the performance and reduce the effect from distant road regions, this study selected five nearby observation sites located around an intersection that have obverse traffic links with one another. The traffic at sites 2, 5, and

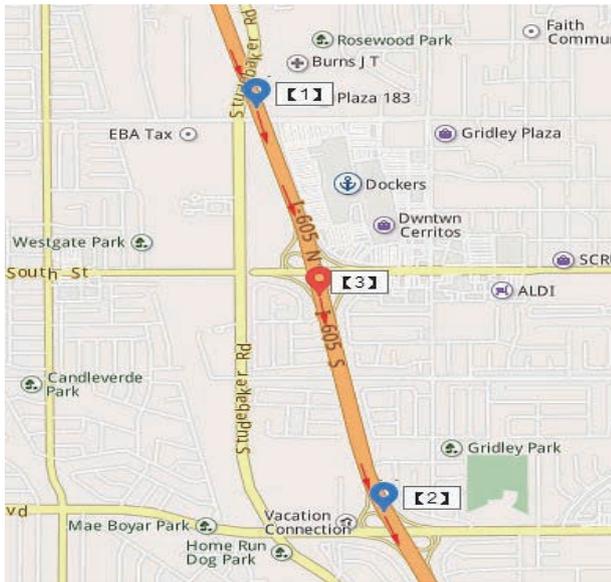


Fig. 2 Selected sensor sites for the straight road layout.

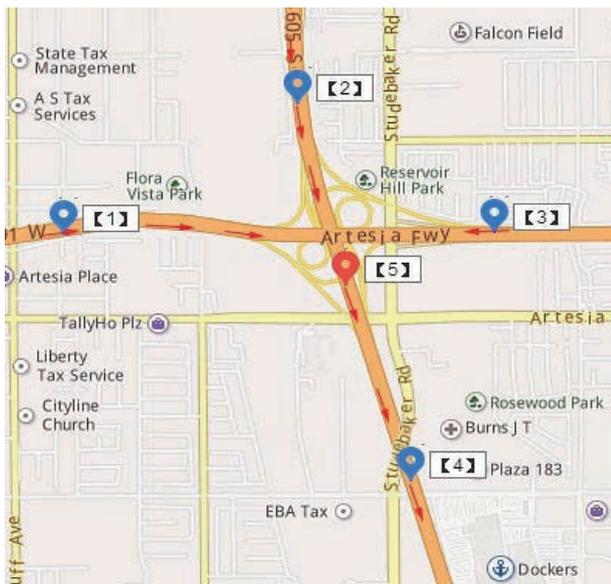


Fig. 3 Selected sensor sites for the crossroad layout.

4 flows from north to south; at site 1, the traffic flows from west to east; and at site 3, the traffic flows from east to west. The experiment used data from sites 1–4 located in each of the four directions to predict traffic flow over the next 5 minutes at site 5, which is located at the center of the intersection.

The traffic flow data used in the experiments covered 10 weekdays (6 March 2017–19 March 2017) for each site. Weekend data were removed due to differences in traffic patterns. In addition, the experiments used only traffic data from morning (6:00–10:00) and evening (16:00–20:00) rush hours because few vehicles are in transit during the rest of the day. The data were collected at 5-minute

aggregated intervals; thus, the unit is vehicles per 5 min (veh/5 min), i.e., the number of vehicles over the past five minutes. Therefore, each site had 96 sample points per day.

The traffic flow data were split into three groups to construct the forecasting model and evaluate its performance. The first eight weekdays (6–9 and 13–16 March 2017) were used as the training dataset. The data for 10 July 2017 were used as the validation dataset. The data from 17 March 2017 were used as the test dataset. Finally, for the two different road layouts, we obtained four experimental datasets, namely, straight road morning rush hour (straight-M), straight road evening rush hour (straight-E), crossroad morning rush hour (cross-M), and crossroad evening rush hour (cross-E) datasets.

In this study, the goal of the forecasting task is to predict traffic flow for the next 5 minutes by using the short-term historic data of related sites, including the target site itself. The traffic flow data collected at times t , $t-i$, $t-2i$, and $t-3i$ from all sites were used to predict traffic flow at the target site at time $t+i$, where i is a 5-min sampling interval. The collected data were used as input features for the forecasting model; thus, the straight road layout experiment had 12 original input features (three sites with four features each), and the crossroad layout experiment had 20 features (five sites with four features each).

4.2 Configuration

The experiments in the two road layouts were conducted using the real-world traffic dataset described in the previous section. We compared the proposed model with related methods, including ARIMA^[53], BPNN^[54], SVR with grid optimization (GRIDSVR), and SVR with the proposed chaos GA optimization (CGASVR).

ARIMA and BPNN are two algorithms widely used for regression problems. These methods were compared with the proposed algorithm to evaluate forecasting ability. GRIDSVR^[55] and CGASVR are SVR-based prediction methods that can optimize the SVR parameters by using different algorithms, similar to the proposed model; however, these methods lack a feature selection process. GRIDSVR employs the GRID algorithm to search for the optimal parameters, and CGASVR uses the enhanced GA to evaluate the performance of the selected features. The proposed RF-CGASVR method is also based on the enhanced GA, but it executes an RF algorithm to perform feature selection. The related parameters for these methods are shown in Tables 1 and 2.

To compare the prediction performances of these forecasting methods, we adopted RMSE and MAPE as

Table 1 CGA settings.

Parameter	Value
Maximum evolution generation	50
Population size	100
SVM cross-validation number	3
Crossover probability	0.4
Mutation probability	0.01
Value range of C	[0.1, 100]
Value range of σ	[0.01, 100]
Value range of ε	[0.01, 1]

Table 2 RF settings.

Parameter	Value
Number of decision trees	200
Number of predictors sampled for splitting at each node	$\max\left(\text{floor}\left(\frac{\text{size}(\text{Feature}, 2)}{3}\right), 1\right)$

criteria. RMSE was used as a GA fitness function as in Eq. (11), and MAPE was calculated using Eq. (12).

4.3 Analysis of results

All the experiments were conducted using the four datasets described in Section 4.1. The performances of the different methods in terms of RMSE and MAPE were analyzed to determine prediction errors. The prediction errors of the different methods with the different datasets are listed in Table 3. A histogram of the comparison results is presented in Fig. 4 to show the differences among the methods.

For traffic flow forecasting, Table 3 and Fig. 4 show that SVR-based methods are superior to BPNN and ARIMA methods on all four datasets. This result could be due to the highly nonlinear pattern of traffic flow and the difficulty in considering many potential factors affecting real-world road networks in forecasting. ARIMA is a simple time series forecasting model that focuses more on historical average values than on patterns; as such this method is suitable when the relationship between the input and the output is approximately linear. In the present work, ARIMA performed the worst. BPNN

requires a relatively large training dataset and has complex network structures that create features that are difficult to realize and determine in a real-world scenario for short-term traffic flow forecasting. However, SVR-based methods with an RBF kernel provide nonlinear modeling capability and thus achieved better performances and smaller prediction error than the other methods.

Comparison of SVR-based methods GRIDSVR, CGASVR, and the proposed RF-CGASVR shows that the proposed model is better than CGASVR and considerably outperforms GRID-SVR on all the experimental datasets. These three methods employ different parameter optimization algorithms: RF-CGASVR and CGASVR use the enhanced GA to find the optimal parameters, and GRID-SVR uses the GRID search algorithm. GA is a heuristic algorithm that can search a broad solution space than the GRID algorithm, which searches only in a given space with a given step size; as such, finding the optimal solution using the GRID algorithm is more difficult than when using GA. In addition, given the ergodic characteristics of tent map method for population initialization and mutation operations, GA search can be focused in the range of the optimal solutions to offer a better fitness value. Thus, CGA is efficient at finding the optimal solution and is suitable for SVR parameter optimization.

From the viewpoint of feature selection, the proposed method, which adopts the RF algorithm for feature selection, significantly improves the forecasting accuracy. This result is reasonable because competing methods use all available features to construct a forecasting model; thus, they are unable to remove redundant or noise data. In addition, the obtained feature subset of the model implies the contribution of each site to the target sites via the traffic network. The feature subsets of the four experimental datasets selected by the proposed method are shown in Table 4. The feature selection process reduced the number of features from 12 to 4 in the straight-M and 8 in the straight-E datasets, and from 20 to 8 in the cross-M

Table 3 Comparison of RMSE and MAPE for different datasets and algorithms.

	Straight-M		Straight-E		Cross-M		Cross-E	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ARIMA	73.7405	0.1192	62.6106	0.0974	40.5423	0.0777	35.5719	0.0703
BPNN	62.3429	0.0932	36.6302	0.0563	42.8166	0.0842	36.8955	0.0712
GA-BPNN	60.2726	0.0873	35.0750	0.0511	40.0641	0.0803	34.6850	0.0710
GRIDSVR	52.6397	0.0796	31.7426	0.0478	38.9752	0.0727	32.8228	0.0645
CGASVR	52.1030	0.0783	31.8273	0.0482	37.6650	0.0703	32.2336	0.0638
RF-CGASVR	49.8129	0.0726	31.1140	0.0469	33.0953	0.0675	31.3078	0.0636

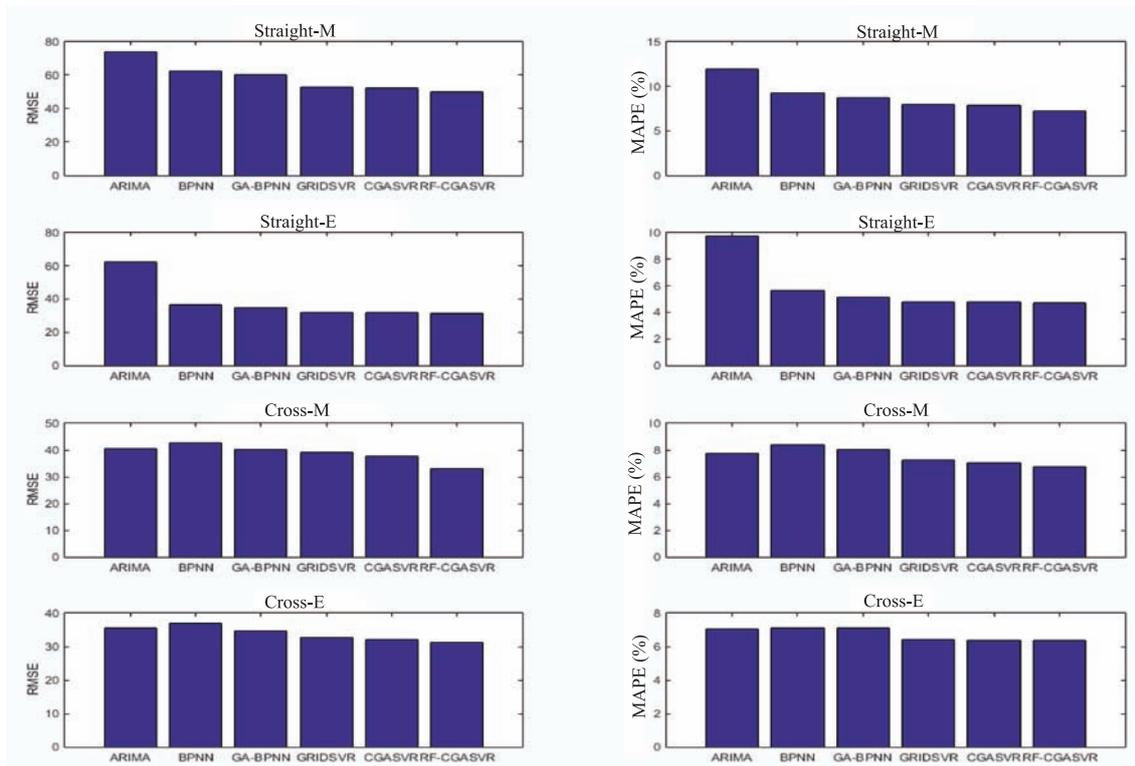


Fig. 4 Histogram of RMSE and MAPE for different datasets and algorithms.

Table 4 Features selected by RF-CGASVR.

Dataset	Selected features*
Straight-M	2-1, 3-1, 3-2, 3-3
Straight-E	1-1, 1-2, 1-3, 2-1, 3-1, 3-2, 3-3, 3-4
Cross-M	2-1, 2-3, 4-1, 4-2, 4-3, 5-1, 5-2, 5-4
Cross-E	1-1, 2-1, 2-2, 2-3, 3-2, 4-1, 4-2, 4-3, 5-1, 5-2

Note: *, The feature ID is coded with two numbers: the first number is the site ID, and the second number indicates the time point. For example, 2-1 means the traffic flow data from site 2 at the current time, and 3-4 is the traffic flow data of site 3 collected 15 minutes ago.

and 10 in cross-E datasets. These results indicate that the traffic pattern during evening rush hours is more complex than that during morning rush hours; thus, the RF algorithm selected more features from the evening dataset than from the morning dataset. One explanation might be that morning traffic has numerous fixed destinations (i.e., most people need to arrive at work at specific times). By contrast, the times and destinations where they leave work are variable. The results also reflect the spatiotemporal relationships of the sites: that is, features that occur immediately prior to the prediction time should be selected, and features from the target sites provide great contributions.

The results provide useful insights about the features and reveal the relationship between the target site and traffic flow. For example, from the cross-M and cross-E

experiments, traffic flow from site 1 has more impact on site 5 during the evening rush hours than that during the morning rush hours. We also observed that more features were selected from sites 2 and 4, indicating that these two sites have a greater effect on the target site traffic flow than the other sites.

To illustrate the performances of different methods, we compared the real traffic flow data and the forecasting results of the various modeling methods (Fig. 5). The results of the proposed RF-CGASVR method have minimal differences with the real traffic flow, and its residual value is more stable than that of the other methods in the erratic and stable parts of the dataset. Thus, the proposed method achieves a traffic flow prediction curve that is most similar to the observed data. The figure clearly indicates that the proposed RF-CGASVR method achieves smaller prediction errors than the other methods in the four experiments while employing fewer (but critical) features. Overall, the proposed method is more suitable for short-term traffic prediction than the other tested methods.

5 Conclusion

Accurate forecasting of short-term traffic flow can effectively save travel time, reduce traffic jams, and provide route guidance. This paper proposes a novel

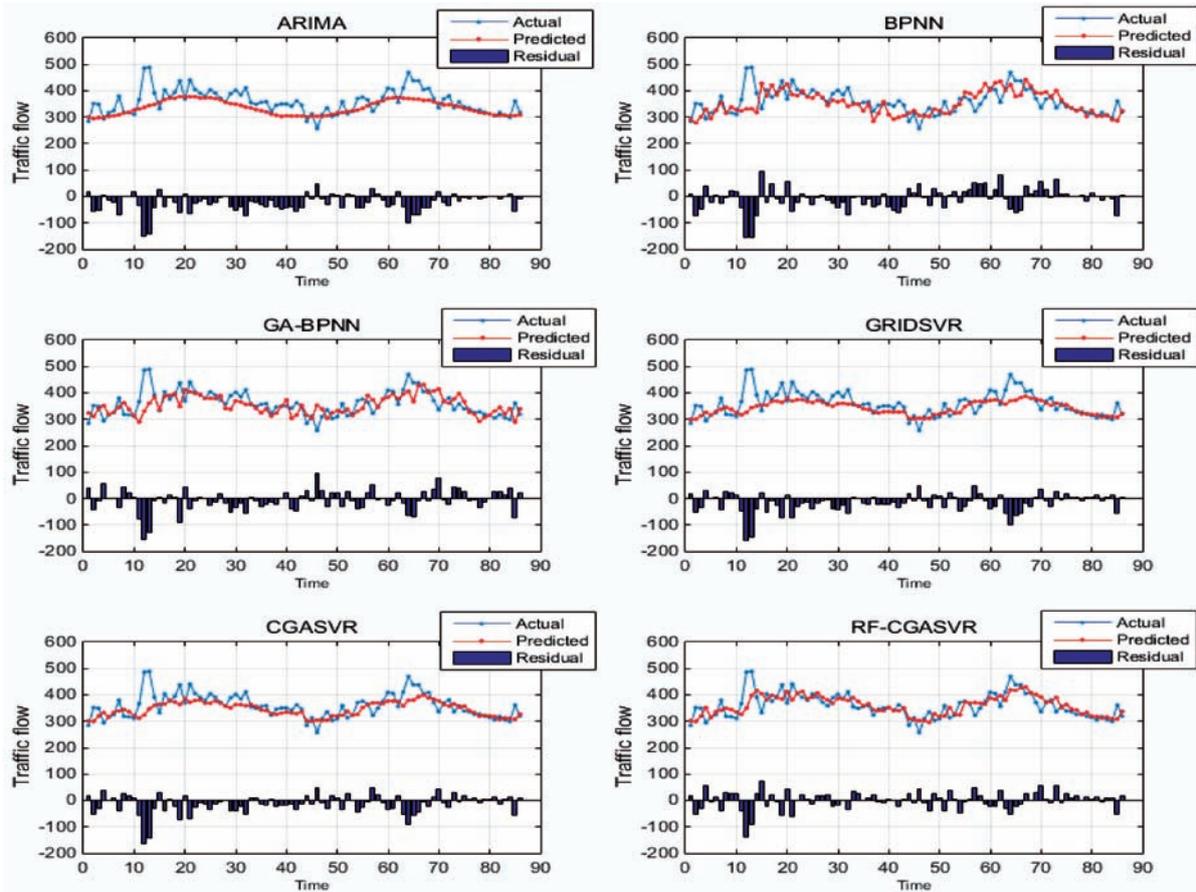


Fig. 5 Comparison of prediction results for the cross-M dataset. x -axis is sampling times with 5-min sampling intervals, y -axis is the number of vehicles.

method for short-term traffic flow forecasting. The method is based on the combination of RF, GA, and SVR for feature selection, parameter optimization, and prediction, respectively, in an integrated framework. A modified chromosome initialization stage and a mutation stage are applied to handle the local convergence problem caused by the random operation of the standard GA. The selected features and parameter values are directly related to the final forecasting performance; thus, the proposed method reveals the relationships between the target site and other road regions and exhibits better forecasting performance than its competitors. The experiments conducted in this study confirm the performance of the proposed method with real-world traffic flow data obtained from the Caltrans PeMS in the USA. The proposed RF-CGASVR method provides better forecasting performance than the other tested methods and constitutes a valid approach for short-term traffic flow forecasting.

The traffic flow data of different roads for different time periods exhibit clearly different patterns. This study

focuses primarily on the spatial relationships among the sites. However, this work does not consider long-term scale traffic flow patterns, such as weekly similarities and holiday similarities. Future studies should include traffic type, road conditions, and other driving restrictions, such as truck routes, across-traffic turns, school zones, traffic light times, and traffic densities, which influence traffic patterns.

Acknowledgment

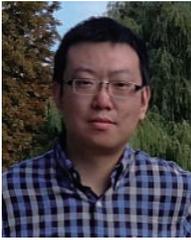
This work was supported by the Science and Technology Department of Sichuan Province of China (Nos. 2017JY0007, 2016JY0073, and 2016JZ0031), the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, and the Fundamental Research Funds for the Central Universities (No. ZYGX2015J063).

References

- [1] Y. Huang, X. Guan, Z. P. Cai, and T. Ohtsuki, Multicast capacity analysis for social-proximity urban bus-assisted

- vanets, in *Proc. 2013 IEEE Int. Conf. Communications (ICC)*, Budapest, Hungary, 2013, pp. 6138–6142.
- [2] X. Zheng, Z. P. Cai, J. Z. Li, and H. Gao, An application-aware scheduling policy for real-time traffic, in *Proc. 2015 IEEE 35th Int. Conf. Distributed Computing Systems*, Columbus, OH, USA, 2015, pp. 421–430.
- [3] X. Guan, Y. Huang, Z. P. Cai, and T. Ohtsuki, Intersection-based forwarding protocol for vehicular ad hoc networks, *Telecommun. Syst.*, vol. 62, no. 1, pp. 67–76, 2016.
- [4] Y. Huang, M. Chen, Z. P. Cai, X. Guan, T. Ohtsuki, and Y. Zhang, Graph theory based capacity analysis for vehicular ad hoc networks, in *Proc. 2015 IEEE Global Communications Conf. (GLOBECOM)*, San Diego, CA, USA, 2015, pp. 1–5.
- [5] L. A. Tang, J. W. Han, and G. F. Jiang, Mining sensor data in cyber-physical systems, *Tsinghua Sci. Technol.*, vol. 19, no. 3, pp. 225–234, 2014.
- [6] X. J. Wang, L. J. Guo, C. Y. Ai, J. B. Li, and Z. P. Cai, An urban area-oriented traffic information query strategy in VANETs, in *Wireless Algorithms, Systems, and Applications*, K. Ren, X. Liu, W. F. Liang, M. Xu, X. H. Jia, and K. Xing, eds. Springer, 2013, pp. 313–324.
- [7] M. van der Voort, M. Dougherty, and S. Watson, Combining kohonen maps with ARIMA time series models to forecast traffic flow, *Transport. Res. C Emer. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.
- [8] H. K. Hong, W. H. Huang, X. B. Zhou, S. Z. Du, K. G. Bian, and K. Q. Xie, Short-term traffic flow forecasting: Multi-metric KNN with related station discovery, in *Proc. 2015 12th Int. Conf. Fuzzy Systems and Knowledge Discovery (FSKD)*, Zhangjiajie, China, 2015, pp. 1670–1675.
- [9] J. H. Guo, W. Huang, and B. M. Williams, Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification, *Transport. Res. C Emer. Technol.*, vol. 43, pp. 50–64, 2014.
- [10] G. Q. Yu and C. S. Zhang, Switching ARIMA model based forecasting for traffic flow, in *Proc. 2004 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Montreal, Canada, vol. 2, pp. ii-429–ii-432, 2004.
- [11] K. I. Wong and Y. C. Hsieh, Short-term traffic flow forecasting for urban roads using space-time ARIMA, in *Transportation and Urban Sustainability*, Hong Kong, China, 2010.
- [12] X. B. Li, RBF neural network optimized by particle swarm optimization for forecasting urban traffic flow, in *Proc. 2009 Third Int. Symp. Intelligent Information Technology Application*, Shanghai, China, 2009, pp. 124–127.
- [13] X. Y. Kuang, L. H. Xu, Y. G. Huang, and F. L. Liu, Real-time forecasting for short-term traffic flow based on general regression neural network, in *Proc. 2010 8th World Congress on Intelligent Control and Automation*, Ji'nan, China, 2010, pp. 2776–2780.
- [14] G. L. Duan, P. Liu, P. Chen, Q. Jiang, and N. Li, Short-term traffic flow prediction based on rough set and support vector machine, in *Proc. 2011 Eighth Int. Conf. Fuzzy Systems and Knowledge Discovery (FSKD)*, Shanghai, China, 2011, pp. 1526–1530.
- [15] H. Y. Wang and Y. Z. Wang, Vessel traffic flow forecasting with the combined model based on support vector machine, in *Proc. 2015 Int. Conf. Transportation Information and Safety (ICTIS)*, Wuhan, China, 2015, pp. 695–698.
- [16] M. G. Arenas, N. Rico, V. M. Rivas, P. Castillo, A. Fernandez-Ares, P. Garcia-Fernandez, P. Garcia-Sanchez, A. M. Mora, J. J. Asensio, G. Romero, et al., Traffic flow forecasting in real world, in *Int. Work-Conf. Time Series (ITISE 2014)*, 2014, pp. 1436–1445.
- [17] N. Sánchez-Marroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán, Filter methods for feature selection—A comparative study, in *Proc. 8th Int. Conf. Intelligent Data Engineering and Automated Learning—IDEAL 2007*, Birmingham, UK, 2007, pp. 178–187.
- [18] R. Kohavi and G. H. John, Wrappers for feature subset selection, *Artif. Intell.*, vol. 97, nos. 1&2, pp. 273–324, 1997.
- [19] C. Yao, Y. F. Liu, B. Jiang, J. G. Han, and J. W. Han, LLE score: A new filter-based unsupervised feature selection method based on nonlinear manifold embedding and its application to image recognition, *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5257–5269, 2017.
- [20] C. L. Huang and C. Y. Tsai, A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting, *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1529–1539, 2009.
- [21] J. G. Zhu, S. T. Wang, X. S. Wu, and F. L. Chung, A novel adaptive SVR based filter ASBF for image restoration, *Soft Comput.*, vol. 10, no. 8, pp. 665–672, 2006.
- [22] M. Aladeemy, S. Tutun, and M. T. Khasawneh, A new hybrid approach for feature selection and support vector machine model selection based on self-adaptive cohort intelligence, *Expert Syst. Appl.*, vol. 88, pp. 118–131, 2017.
- [23] S. Maldonado and R. Weber, A wrapper method for feature selection using support vector machines, *Inf. Sci.*, vol. 179, no. 13, pp. 2208–2217, 2009.
- [24] J. Kennedy and R. Eberhart, Particle swarm optimization, in *Proc. 1995 IEEE Int. Conf. Neural Networks*, Perth, Australia, vol. 4, pp. 1942–1948, 1995.
- [25] X. Q. Wang and J. Gao, Application of particle swarm optimization for tuning the SVR parameters, in *Proc. 2015 IEEE Int. Conf. Advanced Intelligent Mechatronics (AIM)*, Busan, South Korea, 2015, pp. 1173–1177.
- [26] P. Duan, K. G. Xie, T. T. Guo, and X. G. Huang, Short-term load forecasting for electric power systems using the PSO-SVR and FCM clustering techniques, *Energies*, vol. 4, no. 1, pp. 173–184, 2011.
- [27] H. H. Tsai, Y. J. Jhuang, and Y. S. Lai, An SVD-based image watermarking in wavelet domain using SVR and

- PSO, *Appl. Soft Comput.*, vol. 12, no. 8, pp. 2442–2453, 2012.
- [28] M. W. Li, W. C. Hong, and H. G. Kang, Urban traffic flow forecasting using Gauss-SVR with cat mapping, cloud model and PSO hybrid algorithm, *Neurocomputing*, vol. 99, pp. 230–240, 2013.
- [29] D. Bertsimas and J. Tsitsiklis, Simulated annealing, *Statist. Sci.*, vol. 8, no. 1, pp. 10–15, 1993.
- [30] Y. F. Tan, Z. H. Jia, X. Z. Qin, C. Chang, and H. Wang, Application of SA-SVR in mobile communication traffic forecasting, (in Chinese), *Comput. Eng.*, vol. 36, no. 22, pp. 195–196, 2010.
- [31] W. C. Hong, Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm, *Neurocomputing*, vol. 74, nos. 12&13, pp. 2096–2107, 2011.
- [32] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Boston, MA, USA: Addison-Wesley, 1989.
- [33] M. Mia and N. R. Dhar, Prediction and optimization by using SVR, RSM and GA in hard turning of tempered AISI 1060 steel under effective cooling condition, *Neural Comput. Appl.*, doi: 10.1007/s00521-017-3192-4.
- [34] L. Chato, S. Tayeb, and S. Latifi, A genetic algorithm to optimize the adaptive Support Vector Regression model for forecasting the reliability of diesel engine systems, in *Proc. 2017 IEEE 7th Annu. Computing and Communication Workshop and Conf. (CCWC)*, Las Vegas, NV, USA, 2017, pp. 1–7.
- [35] H. Shafizadeh-Moghadam, A. Tayyebi, M. Ahmadlou, M. R. Delavar, and M. Hasanlou, Integration of genetic algorithm and multiple kernel support vector regression for modeling urban growth, *Comput. Environ. Urban Syst.*, vol. 65, pp. 28–40, 2017.
- [36] C. Mandriota, M. Nitti, N. Ancona, E. Stella, and A. Distanto, Filter-based feature selection for rail defect detection, *Mach. Vis. Appl.*, vol. 15, no. 4, pp. 179–185, 2004.
- [37] Z. B. Cao, Y. Wang, Y. Sun, W. Du, and Y. C. Liang, Effective and stable feature selection method based on filter for gene signature identification in paired microarray data, in *Proc. 2013 IEEE Int. Conf. Bioinformatics and Biomedicine*, Shanghai, China, 2013, pp. 189–192.
- [38] K. Y. Leong, A. Sitiol, and K. S. Anbananthen, Enhance neural networks training using GA with chaos theory, in *Proc. 6th Int. Symp. Neural Networks: Advances in Neural Networks - Part II*, Wuhan, China, 2009, pp. 520–529.
- [39] C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [40] M. S. Emami and K. Omar, A low-cost method for reliable ownership identification of medical images using SVM and Lagrange duality, *Expert Syst. Appl.*, vol. 40, no. 18, pp. 7579–7587, 2013.
- [41] H. Z. Yang, X. G. Shao, and F. Ding, A new smooth support vector regression based on ϵ -insensitive logistic loss function, in *Proc. 1st Int. Conf. Advances in Natural Computation*, Changsha, China, 2005, pp. 25–32.
- [42] O. L. Mangasarian and D. R. Musicant, Robust linear and support vector regression, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 950–955, 2000.
- [43] J. Jahn, Karush–Kuhn–Tucker conditions in set optimization, *J. Optim. Theory Appl.*, vol. 172, no. 3, pp. 707–725, 2017.
- [44] H. Frohlich, O. Chapelle, and B. Scholkopf, Feature selection for support vector machines by means of genetic algorithm, in *Proc. 15th IEEE Int. Conf. Tools with Artificial Intelligence*, Sacramento, CA, USA, 2003, pp. 142–148.
- [45] D. Agnihotri, K. Verma, and P. Tripathi, Variable global feature selection scheme for automatic classification of text documents, *Expert Syst. Appl.*, vol. 81, pp. 268–281, 2017.
- [46] D. Antonucci, U. F. Oberegger, W. Pasut, and A. Gasparella, Building performance evaluation through a novel feature selection algorithm for automated ARX model identification procedures, *Energy and Buildings*, vol. 150, pp. 432–446, 2017.
- [47] L. M. Abualigah, A. T. Khader, M. A. Al-Betar, and O. A. Alomari, Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering, *Expert Syst. Appl.*, vol. 84, pp. 24–36, 2017.
- [48] L. Breiman, Random forests, *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [49] L. Lin, F. Wang, X. L. Xie, and S. S. Zhong, Random forests-based extreme learning machine ensemble for multi-regime time series prediction, *Expert Syst. Appl.*, vol. 83, pp. 164–176, 2017.
- [50] M. L. Huang, Intersection traffic flow forecasting based on ν -GSVR with a new hybrid evolutionary algorithm, *Neurocomputing*, vol. 147, pp. 343–349, 2015.
- [51] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1996.
- [52] California Department of Transportation, Caltrans Performance Measurement System (PeMS), <http://pems.dot.ca.gov>, 2017.
- [53] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA, USA: Holden-Day, 1976.
- [54] J. Wen, J. L. Zhao, S. W. Luo, and Z. Han, The improvements of BP neural network learning algorithm, in *Proc. 5th Int. Conf. Signal Processing, 16th World Computer Congress 2000*, Beijing, China, vol. 3, pp. 1647–1649, 2000.
- [55] M. L. Guo, X. Z. Qin, L. Chen, and Z. H. Jia, The prediction of holidays' traffic based on improved grid search SVR, in *Proc. 2012 4th Int. C. Environmental Science and Information Application Technology, Ser. Advances in Biomedical Engineering*, Newark, NJ, USA, 2013, pp. 688–693.



Lizong Zhang is a lecturer in computing science at the School of Computer Science and Engineering, University of Electronic Sciences and Technology of China, China. He received the Msc degree in computer science in 2008 and the PhD degree in 2013 both from Staffordshire University, UK. His main interests include machine

learning, knowledge management, and decision support systems. His other interests include big data, cloud computing, and artificial intelligence systems.



Guangchun Luo is a professor of the School of Computer Science and Engineering, University of Electronic Sciences and Technology of China, China. He has published over 60 papers. He received the Mse degree in computer science in 1999 and the PhD degree in 2002 from the University of Electronic

Science and Technology of China. His publications include 20 SCI publications, 40 EI articles, and 6 published computer monographs. His main interests include machine learning, big data, and cloud computing.



Nawaf R. Alharbe is an assistant professor with the Community College at Taibah University in Saudi Arabia. He got an MSc degree in advanced computing science and a PhD degree in health informatics from Staffordshire University, UK. His main interests include knowledge management, RFID, and sensor

technology in healthcare environment.



Zhiyuan Yao received the BS degree in information security in 2014 and the MS in computer application technology from the University of Electronic Science and Technology of China in 2017. He is currently a cloud computing engineer in Huawei Technologies. His research interests include machine learning and cloud computing.



Ying Li received the BS degree in computer science and technology in 2016 from Changsha University of Science and Technology and is now a master student in computer application technology at the University of Electronic Science and Technology of China. Her research interests include machine learning and big data.