Neural Process Lett (2016) 43:155–172
DOI 10.1007/s11063-015-9409-6

A Short-term Traffic Flow Forecasting Method Based on the Hybrid PSO-SVR

Wenbin Hu · Liping Yan · Kaizeng Liu · Huan Wang

Published online: 8 February 2015
© Springer Science+Business Media New York 2015

Abstract Accurate short-term flow forecasting is important for the real-time traffic control, but due to its complex nonlinear data pattern, getting a high precision is difficult. The support vector regression model (SVR) has been widely used to solve nonlinear regression and time series predicting problems. To get a higher precision with less learning time, this paper presents a Hybrid PSO-SVR forecasting method, which uses particle swarm optimization (PSO) to search optimal SVR parameters. In order to find a PSO that is more proper for SVR parameters searching, this paper proposes three kinds of strategies to handle the particles flying out of the searching space Through the comparison of three strategies, we find one of the strategies can make PSO get the optimal parameters more quickly. The PSO using this strategy is called fast PSO. Furthermore, aiming at the problem about the decrease of prediction accuracy caused by the noises in the original data, this paper proposes a hybrid PSO-SVR method with historical momentum based on the similarity of historical short-term flow data. The results of extensive comparison experiments indicate that the proposed model can get more accurate forecasting results than other state-of-the-art algorithms, and when the data contain noises, the method with historical momentum still gets accurate forecasting results.

Keywords Traffic flow · Forecasting · SVR · PSO · Short-term

1 Introduction

The short-term traffic forecasting has always been one of the most important problems in real-time traffic control for three decades. The information to be forecasted involves journey

W. Hu (✉) · L. Yan · K. Liu · H. Wang
School of Computer, Wuhan University, Wuhan 430072, Hubei, China
e-mail: hwb@whu.edu.cn

L. Yan
School of Software, East China Jiaotong University, Nanchang 330013, Jiangxi, China

time, vehicle speed, traffic flow density and so on. The forecasting accuracy of traffic flow directly influences the effects of traffic guidance, planning and control.

Various forecasting approaches have been applied to forecast the short-term traffic flow, which can be classified as follows: (1) time series analysis methods, including ARIMA [1], SARIMA [2], Kalman filtering models [3,4] and so on; (2) machine learning methods, including K nearest neighbor(KNN) [5–10], kernel estimator [11], artificial neural network(ANN) [12] and so on. Among the ANN approaches, several improved ones have been proposed, such as back-propagation neural network (BPNN) [13], radial basis function neural network [14], wavelet network [15], fuzzy neural network [16], object-oriented neural network [17] and so on. But these methods are relatively complicated in modeling, or their performances are not good enough in forecasting accuracy [18–22]. Therefore in this paper a hybrid method is proposed, which is less complicated in modeling and can yield accurate real-time flow forecasting.

Unlike ANN, the support vector regression model (SVR) can get the global optimal solution, and can map a nonlinear regression problem into a linear regression problem by applying a kernel function [23]. Particle Swarm Optimization (PSO) is a population-based stochastic approach for solving continuous and discrete optimization problems, and compared with Genetic Algorithm (GA) and other heuristic algorithms, it is easy to be realized and there are not many parameters needing to be adjusted. Through the comprehensive analysis of SVR and PSO, this paper proposes a hybrid short-term traffic flow forecasting method based on PSO and SVR. In addition, to handle the condition when the traffic data contain noises, this paper proposes a PSO-SVR forecasting method with historical momentum based on the historical similarity of the traffic flow data. To sum up, the contribution of this paper can be summarized as follows.

(1) It proposes a short-term traffic flow forecasting method based on hybrid PSO-SVR, which uses PSO to optimize the parameters of SVR.
(2) It proposes three kinds of strategies to handle the particles flying out of the searching space, in order to find an approach to improve the search speed of PSO.
(3) In order to handle the condition when the original traffic data contain noises, it proposes a PSO-SVR forecasting method with historical momentum, which is based on the fact of historical data's similarity.

The rest of this paper is organized as follows. Section 2 introduces the related works on the classical methods for short-term traffic flow forecasting. Section 3 details a hybrid short-term traffic flow forecasting method based on hybrid PSO-SVR. Section 4 details a PSO-SVR method with historical momentum. Section 5 presents the analysis for the results of extensive comparison experiments. Finally, conclusions and future research are made in Sect. 6.

2 Related Work

The related work on the short-term traffic flow forecasting can be presented as follows.

(1) Kalman filtering model and system identification model built the traffic flow prediction as a multidimensional model and involved the relationship between the time and the location. However, this kind of method was difficult to deal with the observation noise in the traffic flow data and the prediction accuracy was limited. Danech-Panjouh and Aron [24] adopted a hierarchical statistical method, which used a mathematical clustering technology to classify the traffic flow data and built a tuned linear regression model for each category respectively.

The modified method improved the prediction accuracy, but it had a high demand for the quality of essential data [25,26].

(2) The ARIMA model developed by Box [20] is one of the most popular methods in traffic flow forecasting. Kamarianakis [27] successfully applied the ARIMA model considering space and time factors to forecast the traffic flow data. However, ARIMA had its own limitation that the tendency to concentrate on the average values of the past data series made it unable to capture the rapid varying process of the traffic flow. In addition, Willianms [2] employed the seasonal ARIMA (SARIMA) model to the traffic flow forecasting, which considered the periodic difference of the peak, non-peak traffic data, and obtained good forecasting results. However, it took much time to detect the needed outliers and to estimate the parameters of SARIMA model.

(3) ANN imitates the human neurological system's information processing, and several kinds of ANN have been widely used in the traffic flow forecasting. Yin [28] applied a fuzzy-neural model (FNM) to predict the traffic flow of urban street network, which was verified by the experiments to be more accurate than ANN. Vlahogria [29] adopted genetic algorithm (GA) and a multilayer structure optimization strategy to determine the appropriate neural network structure. Their experiments indicated that a simple and static neural network with genetic optimization step, momentum and a certain number of hidden units was suitable for modeling the univariate and multivariate traffic data. ANN model is suitable for arbitrary functions especially nonlinear functions, but its disadvantage is that the objective function is difficult to be understood and it is difficult to find the global optimal solution for non-convex problems.

(4) Support Vector Machine (SVM) can effectively overcome the shortcoming of ANN. It can not only use the minimal risk strategy to train, but also use the structure risk minimization strategies to minimize the upper bound of the error. SVM can obtain the global optimal value in theory, while ANN can only get the local optimal value. In addition, through the application of the kernel function, SVM can map a nonlinear problem in the low dimensional input space to a linear problem in the high dimensional feature space. Hong [30,33] applied SVR to short-term traffic flow forecasting, and used simulated annealing algorithm and genetic algorithm to optimize the SVR parameters selection process. However, they failed to consider the condition when the traffic data contain noises.

In conclusion, short-term traffic flow forecasting has the characteristics of nonlinearity, complexity and real-time performance. However, the present methods are not perfect in constructing adaptive models, ensuring high precision accuracy and providing a real-time solution. In this paper, we combine PSO with SVR and propose a hybrid PSO-SVR short-time traffic flow forecasting method, which can ensure the accurate and real-time prediction of the short-time traffic flow and reduce the influence of the noises in the traffic flow data.

3 A Hybrid PSO-SVR Method for Short-term Traffic Flow Forecasting

3.1 SVR

Support Vector Machines refers to a kind of specific algorithms, which can be used to solve the classification and regression problems. They were invented by Vladimir Vapnik and his colleagues, and were firstly introduced on the computational learning theory (COLT) conference in 1992. Their basic model is a hyper plane with the maximum margin in feature space.

Considering the given train dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, the study target of SVR is to find a function representing the relationship of x and y, and when a new x is given, the function can get the corresponding forecasted value. This function is shown as Eq. (1):

$$f(x) = \sum\nolimits_{i=1}^{n} w\phi(x) + b \tag{1}$$

where w and b are the final study targets of SVR, which decide a linear hyper plane that can fit the training dataset. $\phi(x)$ is the nonlinear mapping about x, which maps x to a new space when the relationship of x and y is nonlinear. In the new space the relationship of $\phi(x)$ and y is linear.

The goal of SVR is to minimize the expected risk, which can be defined as Eq. (2), where L_ϵ is called ϵ-insensitive loss function proposed by Vapnik. L_ϵ is defined as Eq. (3).

$$R_{emp} = \frac{1}{n}\sum\nolimits_{i=1}^{n} L_\epsilon(y_i, f(x_i)) \tag{2}$$

$$L_\epsilon(y, f(x)) = \begin{cases} 0, & \text{if } |y - f(x)| \le \epsilon \\ |y - f(x)| - \epsilon, & \text{otherwise} \end{cases} \tag{3}$$

SVR performs linear regression in the feature space to lower the expected risk using ϵ -insensitive loss and, and at the same time, tries to reduce the complexity of the model by minimizing $\|w^2\|$. This can be realized in Eq. (4), where $\xi_i, \xi_i^*(i = 1, \ldots, n)$ are the non-negative slack variables, representing the deviation between the function $f(x)$ of training dataset and the actual value.

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \quad & \tfrac{1}{2}\|w^2\| + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ s.t. \quad & w\phi(x_i) + b - y_i \le \epsilon + \xi_i, \\ & y_i - w\phi(x_i) - b \le \epsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, n. \end{aligned} \tag{4}$$

This optimization problem can be transformed into the dual problem and its solution is given by Eq. (5), where a_i^*, a_i are the Lagrange multipliers that can be got by solving the dual problem and $K(x_i, x_j)$ is the kernel function that equals the inner product of $\phi(x_i)$ and $\phi(x_j)$. Any function that meets Mercer's condition [31] can be used as the kernel function.

$$\begin{aligned} & f(x) = \sum\nolimits_{i=1}^{n} \left(a_i^* - a_i\right) K(x_i, x) + b \\ & s.t. \quad 0 \le a_i^* \le C, \quad 0 \le a_i \le C \end{aligned} \tag{5}$$

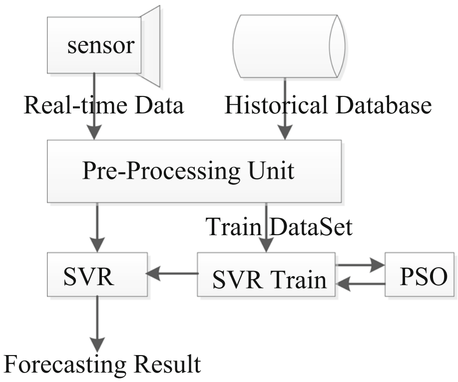
The most frequently used kernel functions are polynomial kernel function, sigmoid kernel function and the radial basis kernel function. This paper uses the radial basis kernel function, which is defined as Eq. (6):

$$K(x, z) = exp\left(\frac{\|x - z\|^2}{2\gamma^2}\right) \tag{6}$$

where γ is the parameter needing to be manually set, as the same as ϵ and C in Eq. (4), all of which have much influence on the forecasting accuracy of SVR.

The hybrid PSO-SVR method proposed in this paper uses SVR to forecast the short-term traffic flow, and uses PSO to optimize the selection procss of the parameters of SVR. The forecasting process is shown in Fig. 1, where the "Pre-Processing Unit" represents a unit that preprocesses the real-time data from the sensor or the history data to obtain the data in the format that SVR needs.

Fig. 1 Short-term traffic flow forecasting based on PSO-SVR algorithm



3.2 Performance Evaluation Index

In this paper *RMSE*(Root Mean Squared Error) and r^2(*Rsquared*) are used to evaluate the forecasting performance of the model, which are respectively defined in Eqs. (7) and (8). Here n is the number of the test samples, $x_i (i = 1, \ldots, n)$ is the instance of a test sample, $f(x_i)$ is the forecasting value of an instance, and y_i $(i = 1, \ldots, n)$ is the true value. The smaller *RMSE* is, the higher the forecasting accuracy is, and the bigger r^2 is, the higher the forecasting accuracy is. r^2 is not more than 1.

$$RMSE = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left(f(x_i) - y_i\right)^2 \tag{7}$$

$$r^2 = \frac{\left(n\sum_{i=1}^{n} f(x_i)y_i - \sum_{i=1}^{n} f(x_i)\sum_{i=1}^{n} y_i\right)^2}{\left(n\sum_{i=1}^{n} f(x_i)^2 - (\sum_{i=1}^{n} f(x_i))^2\right) \times \left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)} \tag{8}$$

3.3 Hybrid PSO-SVR Algorithm

In order to get the optimal parameters of SVR, this paper uses PSO to optimize the selection process of the parameters of SVR, and proposes the PSO-SVR model. In this model, SVR is trained by the method of S-fold cross validation, and *RMSE* is selected to evaluate the performance of SVR. The smaller *RMSE* is, the better the SVR is.

In PSO, assume Φ is the searching space of the particles, which is the value range of the vector (γ, C, ϵ), and n is the swarm size. The ranges of γ, C, ϵ are set as [0, 1000], [1, 10000] and [0, 50] respectively. Each particle has its own position $\vec{x_\iota}$ and velocity $\vec{v_\iota}$, among which $\vec{x_\iota}$ corresponds to (γ, C, ϵ). Whether the particle is excellent depends on its fitness function, which is the *RMSE* of the SVR's train result using this particle. At each iteration, the particles update their position and velocity according to their own historical optimal position $\vec{b_i}$ and the global historical optimal position $\vec{g}$, which are also updated correspondingly after each iteration. The updating formulas of the position and velocity for each particle are shown as Eqs. (9) and (10). The iteration terminates either after the maximum number of iterations or on the condition when *RMSE* is smaller than a preset value. The position $\vec{g}$ is the output that represents the desired parameters combination. The variables discussed in above description are detailed in Table 1.

$$\vec{v}_i^{t+1} = w\vec{v}_i^t + \varphi_1 \vec{U}_1^t(\vec{b}_i - \vec{x}_i^t) + \varphi_2 \vec{U}_2^t(\vec{g} - \vec{x}_i^t) \tag{9}$$

Table 1 Description of PSO variables in PSO-SVR

Variable	Description
Φ	The range of the vector (γ, C, ϵ): $\gamma \in [0, 1000]$, $C \in [1, 10000]$, $\epsilon \in [0, 50]$
n	Swarm size
$\vec{x}_i^t$	Position of the ith particle at the tth iteration
$\vec{v}_i^t$	Velocity of the ith particle at the tth iteration
$\vec{b}_\imath$	Current historical optimal position of the ith particle
$\vec{g}$	Current global historical optimal position of all particles
w	Inertia factor that presents the memory of particles' previous flight direction and can prevent the dramatic change in the direction of flight
φ_1	Cognitive factor that models the tendency of particles to return to previously found best positions
φ_2	Social factor that quantifies the performance of a particle relative to the global optimal particle
$\vec{U}_1$	$n \times n$ diagonal matrices in which the entries on the main diagonal are random numbers uniformly distributed in the interval [0,1]
$\vec{U}_2$	$n \times n$ diagonal matrices in which the entries on the main diagonal are random numbers uniformly distributed in the interval [0,1)

$$\vec{x}_i^{t+1} = \vec{x}_i^t + \vec{v}_i^{t+1} \tag{10}$$

In order to find a proper PSO for selecting the parameters of SVR, this paper proposes three strategies to handle the particles flying out of the searching space, which are described as follows.

The first strategy is the standard strategy. In the traditional PSO, one of the strategies to handle the particles flying out of the searching space is to stop updating the fitness value, and in the latter iterations, the particles will be attracted back to the searching space.

The second strategy is the chaos strategy. Due to the ergodicity and randomness of chaos, the PSO will generate random positions for the particles flying out of the searching space after the chaos.

The third strategy is the convergence strategy. If γ of one particle's position vector is beyond its setting range, it will be set as the γ of $\vec{g}$ and there will be a little disturbance for it. The other vector components are handled equally.

The chaos strategy is proposed for the multiple-peak functions, which slows down the convergence speed of PSO and makes it avoid falling into the local optimal value. This paper adopts the Logistic mapping to do Feigenbaum iteration to generate chaos. The Logistic mapping is a quadratic function defined as Eq. (11), where $\mu \in [0, 4]$, $x \in [0, 1]$. When $\mu > 3.57$, the iteration will generate chaos. The Feigenbaum iteration is defined as Eq. (12), where $\mu \in [3.57, 4]$. The algorithm to generate chaos is called Feigenbaum iteration chaos, which is described in Table 2.

$$f(x) = \mu x (1 - x) \tag{11}$$

$$x_{n+1} = \mu x_n (1 - x_n) \tag{12}$$

The convergence strategy is proposed for finding the optimum value of a single-peak function. As a single-peak function has the unique global optimal solution, the chaos will make the particle further away from the optimal position. The convergence strategy will make the particles closer to the global optimal particle, and the little disturbance could make the particles speed up closer to the global optimal solution.

Table 2 Feigenbaum iteration chaos algorithm

Input:	$x = \left(x^1, x^2, \ldots, x^l\right), k$. $x \in X$. Here, x has l components, x^i is the ith component of x, and the range of x^i is $\left[x^i_{low}, x^i_{high}\right]$. $k = 1, 2, \ldots, l$, denotes the index of the component to be chaos.
Output:	the new x
1	Generates a random number r in the range [0, 1];
2	Use $r = \mu r\,(1 - r)$ to generate 10 iterations, $\mu \in [3.57, 4]$;
3	Make $x^k = x^k_{low} + (x^k_{high} - x^k_{low}) \times r$;
4	Output x.

Table 3 The processing step of the hybrid PSO-SVR algorithm

Input: train dataset, swarm size n, w, φ_1, φ_2, the maximum iteration number T
Output: the optimal combination of (γ, C, ϵ)

Initialization:

Let $t = 0$, do the operation as the following for the n particles.

(1) Initialize $\vec{x}_{\mathrm{i}}^{\mathrm{t}}$ with a value in the searching space Θ;

(2) Initialize $\vec{v}_{\mathrm{i}}^{\mathrm{t}}$ as zero or a little random float number;

(3) Let $\vec{b}_i = \vec{x}_i^t$.

Iteration:

(1) Let $t = 1$;

(2) Update $\vec{g}$;

(3) If $t \leq T$, go to step (4); otherwise terminate the iteration and output $\vec{g}$;

(4) Update the positions and velocities of the n particles, and the historical optimal position of each particle.

(1) Randomly generate $\vec{U}_1^t$, $\vec{U}_2^t$;

(2) Let $\vec{v}_i^{t+1} = w\vec{v}_i^t + \varphi_1 \vec{U}_1^t(\vec{b}_i - \vec{x}_i^t) + \varphi_2 \vec{U}_2^t(\vec{g} - \vec{x}_i^t)$, $\vec{x}_i^{t+1} = \vec{x}_i^t + \vec{v}_i^{t+1}$

(3) When one particle flies out of the searching space, it will be handled by one of the three following strategies.

I standard strategy: Stop updating the particle's fitness value, and go to 5);

II chaos strategy: Use Feigenbaum iteration chaos to generate a new position for the particle;

III convergence strategy: Let the component of $\vec{x}_i^t$ that is out of its range be the value of the same component of $\vec{g}$, and disturb this component slightly.

(5) If $frmse(\vec{x}_i) \leq frmse(\vec{b}_i)$, then $\vec{b}_i = \vec{x}_i^t$;

(6) Let $t = t + 1$.

The PSOs adopting standard strategy, chaos strategy and convergence strategy are described in Table 3, which are respectively called "standard PSO", "chaos PSO" and "fast PSO".

4 PSO-SVR Algorithm with Historical Momentum

4.1 Weekly Similarity and Holiday Similarity Analysis of Traffic Flow

This paper discovers the weekly similarity and holiday similarity of traffic flow through observing the traffic flow figure of one road. For example, the traffic flow figure of a Sunday

Table 4 Data used to calculate similarity

Weekly	The used data
Sunday	28-04, 05-05, 12-05, 19-05, 29-05
Monday	29-04, 13-05, 20-05
Tuesday	30-04, 07-05, 14-05, 21-05, 28-05
Wednesday	01-05, 08-05, 15-05, 22-05, 29-05
Thursday	02-05, 09-05, 16-05, 23-05, 30-05
Friday	03-05, 10-05, 17-05, 24-05, 31-05
Saturday	04-05, 11-05, 18-05, 25-05, 01-06
Holiday	06-05, 27-05

Table 5 The similarity of traffic flow—AL1770

	Sun.	Mon.	Tue.	Wed.	Thur.	Fri.	Sat.	Holiday
SC	0.99931	0.99281	0.99377	0.99564	0.99466	0.98915	0.99615	1

is very similar with those of other Sundays. To measure the similarity, this paper uses the SC (Similarity Coefficient) as Eq. (13), $SC \leq 1$. The bigger SC is, the more similar the data are. In Eq. (13), n is the number of days needing to be compared, $D_i, D_j (i, j = 1, 2, \ldots, n)$ are the flow data on the ith and jth days of n days. $R\left(D_i, D_j\right)$ is the correlation coefficient's square of D_i, D_j, which is defined in Eq. (14) and is similar with the r^2 defined in Eq. (8). k represents the number of the time periods in one day, and it is set as 96 in this paper.

$$SC = \frac{\sum_{1 \leq i \leq j \leq n} R(D_i, D_j)}{n(n-1)/2} \tag{13}$$

$$R\left(D_i, D_j\right) = \frac{\left(k \sum_{l=1}^{k} D_{il} D_{jl} - \sum_{l=1}^{k} D_{il} \sum_{l=1}^{k} D_{jl}\right)^2}{\left(k \sum_{l=1}^{k} D_{il}^2 - C\left(\sum_{l=1}^{k} D_{il}\right)^2\right)\left(\sum_{l=1}^{k} D_{jl}^2 - C\left(\sum_{l=1}^{k} D_{jl}\right)^2\right)} \tag{14}$$

Tables 5 and 6 display the weekly similarity and holiday similarity of traffic flow on road AL1770 and road AL3179 in April, May and June, 2013. The real data series come from the Highways Agency [32]. Table 4 displays which days are used to calculate the similarity of traffic flow data on some day in one week with the same day in other weeks. The traffic flow data of 06-05 and 27-05 are not used to calculate the similarity of that of Monday, but that of the holiday, because these two days are Bank Holiday in England. From Tables 5 and 6, we can see that the short-term traffic flow has strong weekly and holiday similarity, and especially we find that the SC of the traffic flow data of 06-05 and 27-05 is 1. To verify whether the result is caused by the error of data, this paper also check the SC of the traffic flow data of these two days on road LM1, road LM2, road LM3 and road LM4, and find that all SCs are 1.

4.2 PSO-SVR with Historical Momentum

PSO-SVR needs precise traffic flow data for prediction. The noises in the traffic data reduce the prediction accuracy because they break the original statistical law of the traffic flow data.

Table 6 The similarity of traffic flow—AL3179

	Sun.	Mon.	Tue.	Wed.	Thur.	Fri.	Sat.	Holiday
SC	0.99374	0.97878	0.98556	0.98227	0.98047	0.97327	0.98457	1

Table 7 The forecasting based on PSO-SVR with data containing noise-AL1770

Date	Forecasting	Number of noises			
		5	7	9	11
01-05-2013	*RMSE*	77.12	108.73	201.64	185.51
	R-squared	0.9134718	0.822849	0.443046	0.509819
05-05-2013	*RMSE*	111.48	124.45	191.9	222.83
	R-squared	0.6579193	0.540588	0.232922	0.049421

Table 8 The forecasting based on PSO-SVR with data containing noise-AL3179

Date	Forecasting	Number of noises			
		5	7	9	11
01-05-2013	*RMSE*	47.08	116.91	182.59	220.7
	R-squared	0.618425	0.19527	0.04709	0.03021
05-05-2013	*RMSE*	23.36	119.38	157.53	131.05
	R-squared	0.800312	0.13831	0.07867	0.13766

Table 9 Description of variables in Eq. (15)

Variable	Description
O	The final forecasting result
p	The forecasting result based on PSO-SVR
h	The same day's flow data in the former week to be forecasted. If the day to be forecasted is a holiday, h is the data of the former similar holiday.
α	The adjustment parameter that is used to adjust the formulation to get the best forecast result. In this paper, it is called "historical effect coefficient". The more similar the targeted day and its historical day are, the bigger α is.

Tables 7 and 8 display the prediction result using the traffic flow data with noises manually added, where *RMSE* and *Rsquared* are used to evaluate the forecasting performance of the model. The smaller *RMSE* is, the higher the forecasting accuracy is, and the bigger *Rsquared* is, the higher the forecasting accuracy is. It can be seen that in most cases the prediction result is worse when there are more noises in the traffic data.

To reduce the influence of the noises, this paper presents the "short-term traffic flow forecasting method with historical momentum" on the basis of PSO-SVR, which is based on the similarity of historical traffic flow data. It is defined by Eq. (15), the variables in which are described in Table 9. In Eq. (15), h represents the historical data which will influence the forecasting process, and and p is the forecasting result based on the original PSO-SVR algorithm. When $p < h$, the final forecasting result will increase to h based on the Eq. (15), and when $p > h$, the final forecasting result will decrease to h. It can be seen that due to the effect of the historical traffic flow data in the forecasting process, the new method performs well in the forecasting when the data contain noises, which is verified by the experiments in Sect. 5.

$$O = p + \alpha\,(h - p) \tag{15}$$

5 Experiments and Results Analysis

5.1 Experimental Data

The experimental data series are managed by the Highways Agency [32], known as the Strategic Road Network in England, which contain average journey time, speed and traffic flow information on all motorways and 'A' roads every 15 minutes since April 2009. The data use "Linkref" to identify a road (a junction-to-junction link) in the Highway's Agency managed road network. This paper randomly selects the traffic flow data of two roads in April, May and June, 2013. The Linkref of the two roads is AL1770, AL3179 respectively.

5.2 Experimental Environment

Experiments are conducted under the configuration of Ubuntu Server 12.04, with Intel(R) Core(TM) i3-2100 CPU @ 3.10GHz and 2GB RAM. In this experiment, the input vector length of the SVR and BPNN is set as 4. In order to stop the deviation propagating forward we adopt the static forecasting, which means that the forecasting value of the traffic flow at $t + 1$ will use the real value at t instead of the forecasting value at t.

5.3 Performance Analysis of the Hybrid PSO-SVR Method

5.3.1 Analysis of the Particle Number and the Iteration Number

The aim of this experiment is to find the least particle number and the least iteration number when the PSO converges in parameters searching of SVR. In the following the standard PSO, chaos PSO, fast PSO are respectively called PSO, cPSO, fPSO for short.

The traffic flow data of the road AL1770 on 01-05, 15-05, 31-05 are used as the training datasets. The experiment is done according to the algorithm process described in Table 2, and each training dataset is tested for 10 times. The parameters in this experiment are set as Table 10, among which w, φ_1, φ_2 are described in Eqs. (9) and (10), and γ, C, ϵ are described in Sect. 3.1. The result of the experiment is shown in Tables 11, 12 and 13, where the *E-RMSE* and *E-Rsquared* are the average *RMSE* and *R-squared* values of 10 times training results for each training dataset using SVR, and the $E\text{-}iters$ is the average iteration number of PSO in the 10 times training experiments for each training dataset.

It can be seen from Tables 11, 12 and 13 that:

(1) The fPSO's performance with 5 particles is close to that with 100 particles, so it can converge to a certain result with only 5 particles.
(2) The standard PSO's performance with 5 particles is much different from that with 100 particles, so it cannot converge to a certain result with 5 particles. So is the cPSO.
(3) According to the training result of SVR, we can see that three different PSOs' performances are close to each other when the number of particles is 100.
(4) The iteration number of the cPSO is the least and that of the fast PSO is less.

Table 10 The experimental parameters' settings

Parameter	Value
Particle number	100
The maximum iteration number	200
The range of γ	[0, 1]
The range of C	[1, 10000]
The range of ϵ	[0, 50]
w	0.5
φ_1	0.15
φ_2	0.2

Table 11 The result of SVR parameters selecting by PSO- with data on 01-05-2013

	PSO		fPSO		cPSO	
	5 Particles	100 Particles	5 Particles	100 Particles	5 Particles	100 Particles
E-RMSE	25.75913	24.7248	24.77538	24.7388	39.73128	25.68186
E-R-squared	0.990345	0.990909	0.990762	0.990792	0.97700	0.990246
E-iters	167.4	157.5	155.2	113.7	99.5	103.7

Table 12 The result of SVR parameters selecting by PSO- with data on 15-05-2013

	PSO		fPSO		cPSO	
	5 Particles	100 Particles	5 Particles	100 Particles	5 Particles	100 Particles
E-RMSE	30.50895	27.49711	28.96567	28.17712	40.55561	28.57929
E-R-squared	0.987263	0.989043	0.987642	0.988376	0.975894	0.988152
E-iters	175.4	182.2	156.3	141.1	106.4	119.2

Table 13 The result of SVR parameters selecting by PSO- with data on 31-05-2013

	PSO		fPSO		cPSO	
	5 Particles	100 Particles	5 Particles	100 Particles	5 Particles	100 Particles
E-RMSE	84.4569	27.81281	28.91057	28.07554	41.08362	27.81281
E-Rsquared	0.779964	0.986463	0.985833	0.986194	0.970565	0.986463
E-iters	151.2	139.8	143.3	118.5	92	139.8

5.3.2 Analysis of the Algorithm's Running Time

To find a proper PSO for SVR, we do another experiment to compare the running time of three different PSOs. The parameters and training data are the same with those in the above experiment, except the maximum iteration number of the fPSO, which is set as 150 for the fPSO can converge at about the 150th iteration. Besides, the fPSO's swarm size is 5, and the other two PSOs' swarm sizes are 10 because the standard PSO and the cPSO can converge

Table 14 The comparison of three PSOs

Date	Result	fPSO	PSO	cPSO
01-05	*RMSE*	25.1648	24.5310	27.1197
	R-squared	0.990731	0.990959	0.98936
	Iters	127	149	125
	Time(s)	**8.61**	**15.05**	**13.27**
15-05	*RMSE*	27.9641	28.6820	31.4217
	R-squared	0.988801	0.98833	0.98565
	Iters	147	149	73
	Time(s)	**7.29**	**11.17**	**12.78**
31-05	*RMSE*	27.5041	27.4005	32.8184
	R-squared	0.986999	0.98702	0.982263
	Iters	120	146	86
	Time(s)	**7.01**	**12.05**	**13.51**

to a steady result with 10 particles. One experiment is carried out for each training dataset. The result is displayed in Table 14.

From Table 14 we can see that the fPSO performs closely to the standard PSO, but the former needs less time, and the cPSO performs the worst. To sum up, we can make the conclusion as the following.

(1) The performances of three different PSOs are close to each other with enough particles.
(2) The fPSO needs the least particles to converge to a steady result, which makes the fPSO spend less time to finish the parameter searching of SVR. Therefore, the fPSO is more suitable for the parameter searching of SVR.

5.4 Comparison Experiments

To test the performance of the PSO-SVR for the forecasting of short-term traffic flow, we compare ARIMA and BPNN (Back Propagation Neural Network) with it. This experiment chooses the traffic flow data of road AL1770 in random 7 days of May in 2013 as the static forecasting target. The train dataset and test dataset are the traffic flow data of road AL1770 in the 30 days of April in 2013. The description about each model is detailed as follows.

(1) PSO-SVR: The length of input vector is 4, and the parameter combination (γ, C, ϵ) of SVR, which is selected by PSO, is (0.000214847057443, 833.08646366, 0.00000524635605413).

(2) BPNN: The length of input vector is set as 4. The network has four layers, of which the input layer has 4 nodes, the first hidden layer has 2 nodes, the second hidden layer has 4 nodes, and the output layer has 1 node. It is trained 10,000 times.

(3) ARIMA: It is realized by Eviews, and the model is ARIMA (1,0,1).

The experiment result is displayed in Table 15; from which we can see that ARIMA gets the worst forecasting result, and the PSO-SVR gets the best forecasting result.

From the results of Table 15, the performance of PSO-SVR is better than other typical forecasting methods (SVR, ARIMA and BPNN). In order to verify the effectiveness of the proposed method, the PSO-SVR is compared with other state-of-the-art methods in nowadays, such as GA-SA (Hybrid Genetic Algorithm—Simulated Annealing Algorithm)

Table 15 The comparison of PSO-SVR, ARIMA, SVR and BPNN

Date	Result	ARIMA	SVR	PSO-SVR	BPNN
01-05-2013	*RMSE*	33.91340	24.90132421	16.14574247	23.81314282
	R-squared	0.982927	0.987126	0.996097	0.991807412
02-05-2013	*RMSE*	33.85440	24.90875498	23.34086545	24.49301348
	R-squared	0.983365	0.990541	0.992687	0.991175105
10-05-2013	*RMSE*	32.86041	23.99321421	21.51996747	23.00305686
	R-squared	0.983694	0.991721	0.992917	0.992019036
15-05-2013	*RMSE*	33.89329	23.91856123	21.4498951	23.64542527
	R-squared	0.983195	0.990861	0.993288	0.992239528
20-05-2013	*RMSE*	32.44836	21.12986311	18.7686707	20.78690139
	R-squared	0.984483	0.992134	0.994749	0.993711226

Table 16 The comparison of PSO-SVR with other state-of-the-art methods

Date	Results	PSO-SVR	GA-SA [33]	KNN-LLWNN [8]	SSA [34]
01-05-2013	*RMSE*	16.14574247	16.524879	17.023147	16.741244
	R-squared	0.996097	0.994897	0.963217	0.9799641
02-05-2013	*RMSE*	23.34086545	23.657981	25.147962	24.017621
	R-squared	0.992687	0.991257	0.976854	0.989111
10-05-2013	*RMSE*	21.51996747	21.968794	22.687932	22.0255419
	R-squared	0.992917	0.991574	0.978541	0.989917
15-05-2013	*RMSE*	21.4498951	21.974157	22.674156	22.2875418
	R-squared	0.993288	0.9925874	0.981274	0.9875497
20-05-2013	*RMSE*	18.7686707	19.124713	20.965479	20.0114796
	R-squared	0.994749	0.99401397	0.984102	0.9901876

[33], KNN-LWNN (K Nearest Neighbor based on Linear Wavelet Neural Network) [8] and SSA (Singular Spectrum Analysis) [34]. These three algorithms mentioned above are the most representative of different hybrid intelligent approaches. In the comparison experiments, we implement these algorithms since the corresponding references detail the algorithm steps. These experiments choose the traffic flow data of road AL1770 in random 7 days of May in 2013 as the static forecasting target. The training dataset and test dataset are the traffic flow data of road AL1770 in the 30 days of April in 2013. The forecasting results are shown in Table 16. The proposed PSO-SVR owns better performance than other state-of-the-art methods in forecasting accuracy.

5.5 Performance Analysis of the PSO-SVR Method with Historical Momentum

The aim of this experiment is to test the “short-term traffic flow forecasting method with historical momentum” with the traffic flow data containing noises. The experiment data are the traffic flow data of road AL1770 and road AL3179 on 01-05-2013, 05-05-2013, 27-05-2013, 28-05-2013, 31-05-2013.

The process of this experiment is as follows.

Table 17 The result of PSO-SVR with historical momentum- AL1770

Date	Prediction with momentum			PSO-SVR	
	α	*RMSE*	*R-squared*	*RMSE*	*R-squared*
01-05-2013	0.94	8.702	0.998878872	77.123	0.913471893
05-05-2013	0.98	12.823	0.995429002	111.48	0.65791939
27-05-2013	1	0	1	93.094	0.692948127
28-05-2013	0.9	45.224	0.983386301	129.333	0.741214505
31-05-2013	0.81	38.513	0.986467613	53.594	0.952323133

Table 18 The result of PSO-SVR with historical momentum- AL3179

Date	Prediction with momentum			PSO-SVR	
	α	*RMSE*	*R-squared*	*RMSE*	*R-squared*
01-05-2013	1	6.397	0.99329041	47.077	0.618425185
05-05-2013	0.98	6.596	0.98559425	23.36	0.800312369
27-05-2013	1	0	1	33.453	0.681231342
28-05-2013	0.87	14.73	0.96141761	28.371	0.851862283
31-05-2013	0.81	13.651	0.96527447	23.996	0.892669650

(1) Add 5 noises into each experiment data manually.
(2) Use PSO-SVR to forecast the traffic flow values of road AL1770 and road AL3179 on 01-05-2013, 05-05-2013, 27-05-2013, 28-05-2013, 31-05-2013.
(3) Use Eq. (15) to adjust the result of 2). To get the best α, we change the value of α from 0 to 1 every 0.01 step size, and find the best one according to the last forecasting result. The traffic flow data of 24-04-2013, 28-04-2013, 06-05-2013, 21-05-2013, 24-05-2013 are chosen as the historical data of 01-05-2013, 05-05-2013, 27-05-2013, 28-05-2013, 31-05-2013.

The results of this experiment are displayed in Tables 17, 18 and Fig. 2. In Tables 17 and 18, the "prediction with momentum" is the forecasting result of "short-term traffic flow forecasting with historical momentum", and the "PSO-SVR" is the forecasting result of PSO-SVR model. In Figure 2 the Y-axis represents the traffic flow, and the X-axis represents the 96 time periods of a day.

From the experimental results, we can get the findings as follows.

(1) The PSO-SVR algorithm with historical momentum can improve the static forecasting result when the data contain noises.
(2) The α in this experiment is relatively high, which is caused by the strong historical similarity of the traffic flow data used in this experiment.

6 Conclusions

In order to ensure the accurate and real-time prediction of the short-time traffic flow, this paper proposes a hybrid PSO-SVR forecasting method. This method uses PSO to optimize the parameter setting of SVR and puts forward different strategies to handle the particles flying

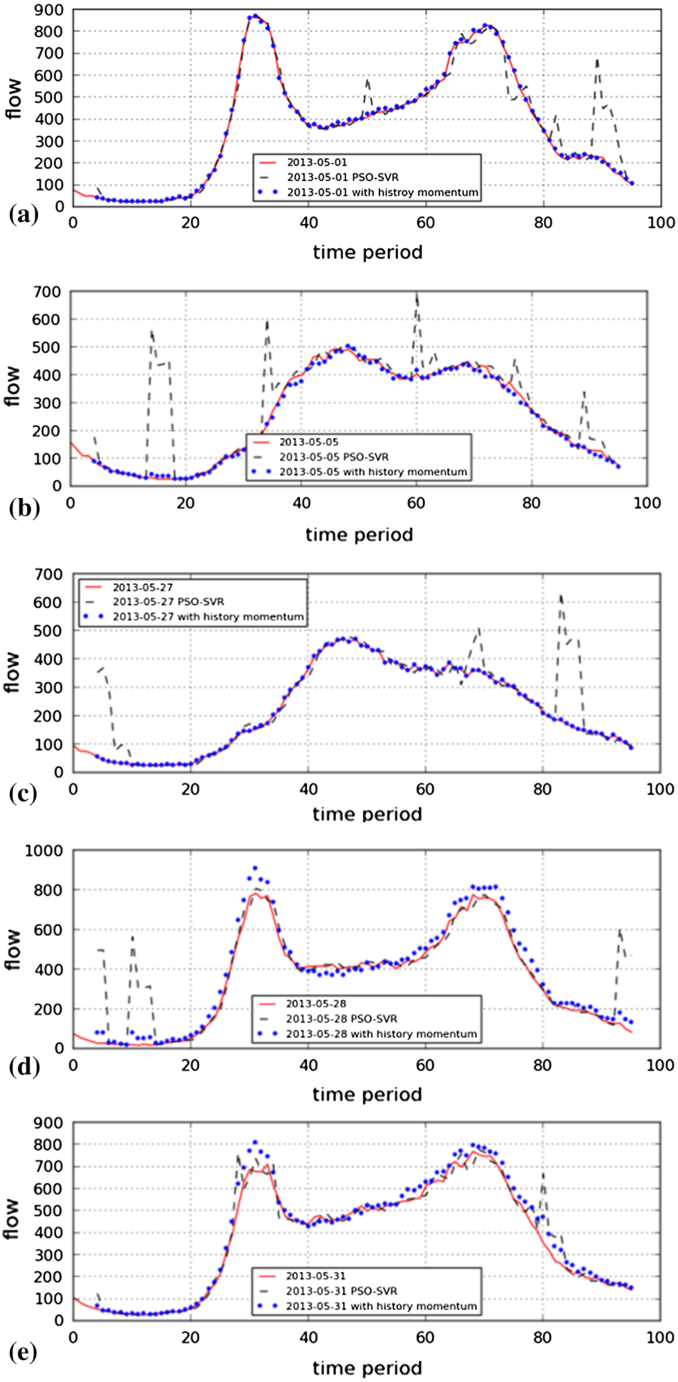


Fig. 2 The result of PSO-SVR with historical momentum: **a** AL1770(01-05-2013); **b** AL1770(05-05-2013); **c** AL1770(27-05-2013); **d** AL1770(28-05-2013); **e** AL1770(31-05-2013); **f** AL3179(01-05-2013); **g** AL3179 (05-05-2013); **h** AL3179(27-05-2013); **i** AL3179(28-05-2013); **j** AL3179(31-05-2013). The Y-axis unit is vehicles, and the X-axis unit is a quarter

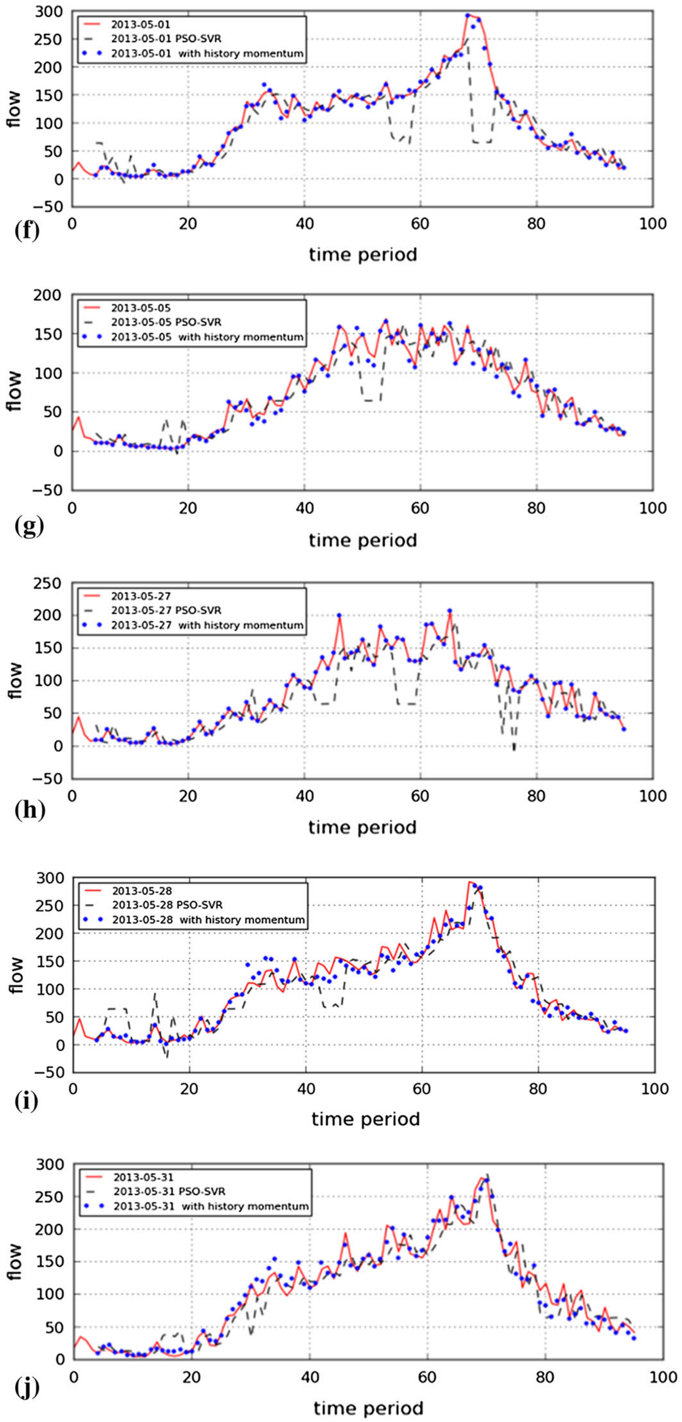


Fig. 2 continued

out of the searching space to adapt to the difference of the forecasting traffic data distribution. Besides, a PSO-SVR forecasting method with histrorical momentum is proposed to reduce the influence of the noises in the traffic flow data. In order to verify the validity of the proposed method, we carry out a large number of performance experiments and comparative experiments. To sum up we can make the conclusions as follows.

(1) Through the comparision with ARIMA and BPNN, the proposed hybrid PSO-SVR method is confirmed to have good performance on the short-term traffic flow forecasting.
(2) To find a proper PSO for SVR, this paper proposes three different strategies to handle the particles flying out of the searching space. One of these strategies is proved to make PSO converge with fewer particles and reduce the time used in the parameters searching of SVR, which is called fast PSO.
(3) To solve the problem that the noises in traffic flow data reduce the prediction accuracy of PSO-SVR, this paper proposes the "short-term traffic flow forecasting with historical momentum" based on the similarity of the historical traffic flow data. The experimental results indicate that this method can get a good forecasting result when the traffic flow data contain noises.

In spite of the work discussed above, further research is still needed to continue in the following two aspects.

(1) The similarity of the historical traffic flow data are not only used to select the training dataset, but also to quantify the historical influence factor α.
(2) From the forecasting research, it is found that the traffic flow data of different roads in different time periods have obvious characteristics. For example, the traffic flow data of the same road are different in weekends, workdays and holidays, even in the same time period the traffic flow data of different roads are not the same. Therefore, combined with different factors of roads, the characteristic of the traffic flow data should be further focused on to study the influence of the road structure, geographical positions and the daily life of people on the traffic flow data.

References

1. Ahmed MS, Cook AR (1979) Analysis of freeway traffic time-series data by using Box-Jenkins techniques. Transport Res Rec 722:1–9
2. Williams BM, Hoel LA (1999) Modeling and forecasting vehicular traffic flow as a seasonal stochastic time series process (No. LTVA/29242/CE99/103)
3. Guo J, Huang W, Williams BM (2014) Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. Transport Res Part C 43:50–64
4. Okutani I, Stephanedes YJ (1984) Dynamic prediction of traffic volume through Kalman filtering theory. Transport Res Part B 18(1):1–11
5. Hu WB, Du B, Wu Y, Liang HL, Peng C, Hu Q (2013) A Hybrid column generation algorithm based on metaheuristic optimization. Transport, publish online
6. Dong C, Shao C, Zhao D, Liu Y (2013) Short-term traffic flow forecasting based on periodicity similarity characteristics. Recent Prog Data Eng Internet Technol 1:15–21
7. Hu WB, Wang H, Du B, Yan LP (2014) A Multi-intersection model and signal timing plan algorithm for urban traffic signal control. Transport, publish online
8. Lin L, Li Y, Sadek A (2013) A k nearest neighbor based local linear wavelet neural network model for on-line short-term traffic volume prediction. Procedia-Soc Behav Sci 96:2066–2077
9. Xiaoyu H, Yisheng W, Siyu H (2013) Short-term traffic flow forecasting based on two-tier k-nearest neighbor algorithm. Procedia-Soc Behav Sci 96:2529–2536
10. Zheng Z, Su D (2014) Short-term traffic volume forecasting: a k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. Transport Res Part C 43:143–157

11. Hu WB, Liang HL, Peng C, Du B, Hu Q (2013) A hybrid chaos-particle swarm optimization algorithm for the vehicle routing problem with time window. Entropy 15:1247–1270
12. Xue J, Shi Z (2008) Short-time traffic flow prediction based on chaos time series theory. J Transport Syst Eng Inf Technol 8(5):68–72
13. Smith BL, Demetsky MJ (1994) Short-term traffic flow prediction: neural network approach. Transport Res Rec 1453:98–104
14. Park B, Messer CJ, Urbanik T II (1998) Short-term freeway traffic volume forecasting using radial basis function neural network. Transport Res Rec 1651(1):39–47
15. Xie Y, Zhang Y (2006) A wavelet network model for short-term traffic volume forecasting. J Intell Transport Syst 10(3):141–150
16. Abdi J, Moshiri B, Abdulhai B, Sedigh AK (2012) Forecasting of short-term traffic-flow based on improved neurofuzzy models via emotional temporal difference learning algorithm. Eng Appl Artif Intell 25(5):1022–1042
17. Dia H (2001) An object-oriented neural network approach to short-term traffic forecasting. Eur J Oper Res 131(2):253–261
18. Abdi J, Moshiri B, Abdulhai B, Sedigh AK (2013) Short-term traffic flow forecasting: parametric and nonparametric approaches via emotional temporal difference learning. Neural Comput Appl 23(1): 141–159
19. Abdi J, Moshiri B (2013) Application of temporal difference learning rules in short-term traffic flow prediction. Expert Systems, publish online
20. Box GE, Jenkins GM, Reinsel GC (2013) Time series analysis: forecasting and control. Wiley, New York
21. Ghosh B, Basu B, O'Mahony M (2007) Bayesian time-series model for short-term traffic flow forecasting. J Transport Eng 133(3):180–189
22. Zhou H, Hirasawa K (2014) Traffic conduction analysis model with time series rule mining. Expert Syst Appl 41:6524–6535
23. Castro-Neto M, Jeong YS, Jeong MK, Han LD (2009) Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. Expert Syst Appl 36(3):6164–6173
24. Danech-Pajouh M, Aron M (1991) ATHENA: a method for short-term inter-urban motorway traffic forecasting. Rech Transp Sécurité 6:6–11
25. Smith BL, Demetsky MJ (1997) Traffic flow forecasting: comparison of modeling approaches. J Transport Eng 123(4):261–266
26. Vapnik V (2000) The nature of statistical learning theory. Springer, Berlin
27. Kamarianakis Y, Prastacos P (2005) Space-time modeling of traffic flow. Comput Geosci 31(2):119–133
28. Yin H, Wong S, Xu J, Wong CK (2002) Urban traffic flow prediction using a fuzzy-neural approach. Transport Res Part C 10(2):85–98
29. Vlahogianni EI, Karlaftis MG, Golias JC (2005) Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transport Res Part C 13(3):211–234
30. Hong WC (2012) Application of seasonal SVR with chaotic immune algorithm in traffic flow forecasting. Neural Comput Appl 21(3):583–593
31. Smith BL, Williams BM (2002) Comparison of parametric and nonparametric models for traffic flow forecasting. Transport Res Part C 10(4):303–321
32. Highways Agency(UK). Highways Agency network journey time and traffic flow data. http://data.gov.uk/dataset/dft-eng-srn-routes-journey-times [Access in March, 2014]
33. Hong WC, Dong Y, Zheng F, Wei SY (2011) Hybrid evolutionary algorithms in a SVR traffic flow forecasting model. Appl Math Comput 217(15):6733–6747
34. Guo F, Krishnan R, Polak J (2013) A computationally efficient two-stage method for short-term traffic prediction on urban roads. Transport Plan Technol 36(1):62–75