ICIC Express Letters ICIC International ©2019 ISSN 1881-803X
Volume **13**, Number **6**, June **2019** pp. **475–482**

SHORT-TERM TRAFFIC FLOW PREDICTION ALGORITHM BY SUPPORT VECTOR REGRESSION BASED ON ARTIFICIAL BEE COLONY OPTIMIZATION

Cheng Li[1,2], Hongtao Zhang[1], Honglie Zhang[1] and Yanju Liu[1]

[1]College of Computer and Control Engineering
Qiqihar University
No. 42, Wenhua Street, Qiqihar 161006, P. R. China
{ lcrb406; zhang_honglie }@163.com; 1916019465@qq.com; liuyanjv@tom.com

[2]College of Computer Science and Technology
Harbin Engineering University
No. 145, Nantong Avenue, Nangang District, Harbin 150001, P. R. China

Received November 2018; accepted February 2019

Abstract. *The prediction of short-term traffic flow is the theoretical basis of intelligent transportation as well as the key technology in traffic flow induction system. The research on short-term traffic flow prediction has shown the considerable social value. At present, support vector regression (SVR) intelligent prediction model has been applied in this domain. Aiming at parameter selection difficulty and prediction accuracy improvement, the artificial bee colony algorithm (ABC) is adopted to optimize the SVR parameters in the paper. Accordingly, the short-term traffic flow prediction algorithm by support vector regression based on artificial bee colony optimization (ABC-SVR) is presented. The simulation experiments are carried out by comparing with other algorithms. The experimental results prove the feasibility and accuracy of the proposed ABC-SVR algorithm and satisfactory effects have been obtained.*
Keywords: Short-time traffic flow prediction, Artificial bee colony algorithm, Support vector regression

1. **Introduction.** The traffic congestion in large and medium-sized cities proceeds to be a severe problem due to the increasing number of car buyers during the process of the continuous acceleration of urbanization and the rapid development of economy. The traffic congestion has not only polluted the environment, but also surged the probability of traffic accidents. Therefore, intelligent transport system (ITS) has emerged in response. Traffic information can be collected and analyzed in real time by ITS so as to induce the traffic control. The key technology of ITS is the short-term (with the sampling interval no more than 15 minutes) traffic flow prediction [1]. In practical applications, how to improve the accuracy of traffic flow prediction is still the great challenge owing to the characteristics of traffic information and the various interference factors in collection process.

The research on short-term traffic flow prediction algorithm has been one of the hot issues at home and abroad. As early as the 1970s, the mature prediction algorithms from the various disciplines ranging from economics to physics were applied to short-term traffic flow prediction by some scholars. Among them, the linear theory and the statistical theory algorithms were mainly adopted, including historical average algorithm [2], time series algorithm [3], Kalman filter algorithm [4], etc. Although the application of these conventional algorithms is relatively mature, the calculation error is pretty larger. In recent years, in order to make up for the defect of these algorithms, various intelligent algorithms have been introduced to the modeling of short-term traffic flow prediction by

DOI: 10.24507/icicel.13.06.475

scholars, and neural network algorithm [5] as well as support vector machine algorithm [6] have emerged correspondingly. In latest years, a multiple of researches have indicated that the short-term traffic flow prediction based on combination prediction algorithm is more effective. For the algorithm, two or more prediction methods are used to predict the short-term traffic flow, and the advantages of various prediction methods are exploited fully, which has improved the prediction accuracy and extensibility. The model covers neural network algorithm and genetic algorithm [7], artificial fish swarm algorithm and support vector machine [8], Kalman filter algorithm and quadratic exponential noise reduction algorithm [9], etc.

Based on summarizing the current research findings, in order to improve the prediction accuracy of short-time traffic flow, an intelligent combination prediction algorithm is proposed in the paper, which is referred to as the short-term traffic flow prediction algorithm by support vector regression based on artificial bee colony optimization. And the nonlinear ε – support vector regression (ε-SVR) is used to perform the modeling of short-term traffic flow prediction.

The remainder of the paper is organized as follows. The support vector regression algorithm is reviewed and analyzed in Section 2. Then, the short-term traffic flow prediction algorithm by support vector regression based on artificial bee colony optimization is presented in Section 3. In Section 4, the experimental results on contrastive data from both the algorithm of the paper and other algorithms are demonstrated. Finally, conclusions are given with the importance and the practical value of the optimal algorithm as well as its future research directions.

2. **Support Vector Regression Algorithm and Analysis.** For nonlinear SVR, the input data is mapped to the high-dimension data space for linear regression through a nonlinear function, which is expressed as Formula (1).

$$f(x) = w^T \cdot \varphi(x) + b \tag{1}$$

where, w is the weight vector, φ is the nonlinear mapping function, b is the offset vector, and $f(x)$ is the objective function. According to Formula (1), the optimal solution of $(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_n, y_n)$, $x_i, y_i \in R$ is estimated. For the linear regression in high dimensional space, it is necessary to obtain the optimal classification line to maximize the classification interval. The classification linear equation is $w^T \cdot x + b = 0$, and the corresponding interval $2/\|w\|$ between the classification lines is calculated. It is assumed that all the input data pairs are estimated by the objective function f in the range of accuracy ε, and then the solving thought based on the maximization interval is transformed into the convex optimization problem of variables w and b. Then, it is transformed into the convex quadratic programming problem with slack variable. Lagrange multiplier method is adopted for solution, and then the duality transformation is realized.

$$\begin{gathered}\operatorname*{Min}_{a_i, a_i^*} \frac{1}{2} \sum_{i,j=1}^{n} (a_i^* - a_i)(a_j^* - a_j) K(x_i, x_j) - \sum_{i=1}^{n} y_i (a_i^* - a_i) + \varepsilon \sum_{i=1}^{n} (a_i^* + a_i) \\ \text{s.t.} \begin{cases} 0 \le a_i^*, a_i \le C, & i = 1, 2, \ldots, n \\ \sum_{i=1}^{n} y_i (a_i^* - a_i) = 0, & i = 1, 2, \ldots, n \end{cases}\end{gathered} \tag{2}$$

where, a_i and a_i^* are the Lagrange multiplier, and the kernel function $K(x_i, x_j)$ is the inner product of two points mapped to the high-dimension eigenspace.

And a_i and a_i^* are solved, which is calculated as Formula (3).

$$w = \sum_{i=1}^{n} (a_i^* - a_i) \cdot \varphi(x_i) \tag{3}$$

Finally, the decision function of support vector regression is obtained, which is shown in Formula (4).

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

3. SVR Model Optimization by Artificial Bee Colony Algorithm.

3.1. **Artificial bee colony algorithm.** The process of solving the optimization problem by artificial bee colony algorithm simulates that of bee colony's searching for honey source with the highest yield degree (honey content). The algorithm has the advantages of less control parameters, simple realization and favorable global convergence. Compared with other typical heuristic algorithms, it has superior performance in solving the nonlinear optimization function.

(1) The ith initial honey source (feasible solution) is generated randomly.

$$x_{ij} = x_{ij\min} + rand \cdot (x_{ij\max} - x_{ij\min}) \tag{5}$$

where, $j = 1, 2, \ldots, D$, and x_{ij} is the jth dimension parameter of the ith honey source. $x_{ij\max}$ and $x_{ij\min}$ are the upper and lower limits of the jth dimension parameter of the ith honey source respectively, and *rand* is the random number of $[0, 1]$.

(2) The new honey source X_i' is searched near the honey source X_i randomly.

$$x_{ij}' = x_{ij} + R_{ij} \cdot (x_{ij} - x_{kj}) \tag{6}$$

where, $j \in \{1, 2, \ldots, D\}$, $k \in \{1, 2, \ldots, N_e\}$, j and k are generated at random, and $k \neq i$; x_{ij}' is the jth dimension parameter of new honey source X_i'; x_{kj} is the jth dimension parameter of honey source X_k; R_{ij} is the random number of $[-1, 1]$.

(3) The fitness value of the objective function is calculated as follows.

$$fit_i = \frac{1}{1 + f_i} \tag{7}$$

where, f_i is the objective function value of ith honey source; fit_i is the fitness value of the ith honey source.

(4) The probability of the employed bee being selected is calculated by the unemployed foragers based on the observed fitness value, which is expressed as Formula (8).

$$P_i = \frac{fit_i}{\sum_{i=1}^{N_e} fit_i} \tag{8}$$

3.2. **Parameter optimization model.** According to the ABC algorithm optimizing process in Section 3.1, the ABC-SVR prediction model is constructed, which is shown in Figure 1. The specific process is as follows.

Firstly, the historical data is preprocessed to form the training sample; then, the swarm is initialized and the parameters are set; SN solution is generated randomly and the fitness value of each solution is calculated; the new solution is generated according to Formula (6) and the fitness value of the new solution is calculated; if the fitness value of the new solution is higher, the old one is updated, food sources are selected according to the calculated P_i, and the new solution is generated again according to Formula (6); if the fitness value of new solution is lower, the old solution remains unchanged, and limit is equal to limit+1; if limit reaches the maximum, the new solution is generated randomly by Formula (5) and optimized continuously; if the maximum iteration number is reached, the optimizing process is ended; finally, the optimal solution is assigned to SVR as parameter for prediction.

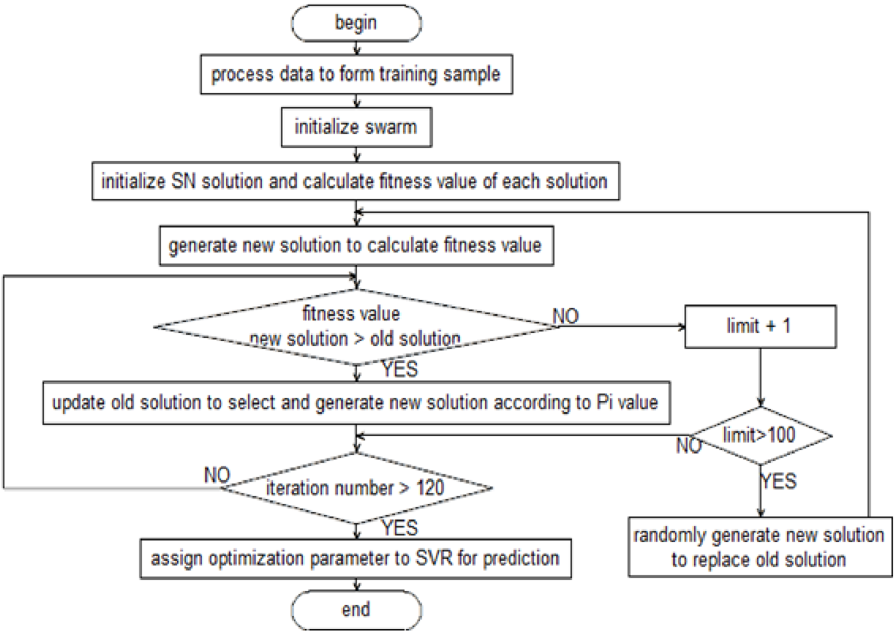


FIGURE 1. ABC-SVR model construction

Parameter optimization is to search the optimal SVR parameter set (C, σ, ε), so as to minimize the error between the predicted value of SVR and the actual value. Therefore, the objective function of the fitness value is calculated, which is shown in Formula (9).

$$\begin{aligned} &\underset{C,\delta,\varepsilon}{\mathrm{Min}}\, f(C,\sigma,\varepsilon) = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^*)^2 \\ &\text{s.t.} \begin{cases} C \in [C_{\min}, C_{\max}] \\ \sigma \in [\sigma_{\min}, \sigma_{\max}] \\ \varepsilon \in [\varepsilon_{\min}, \varepsilon_{\max}] \end{cases} \end{aligned} \tag{9}$$

where, N is the number of the sample, and y_i and y_i^* are the actual traffic flow value and predicted value of the ith sample respectively.

4. Simulation Experiment and Result Analysis.

4.1. Simulation experiment.

4.1.1. *Experiment data.* The data of the experiments is downloaded at certain monitoring points from PEMS system website.

Considering the periodical change of weekly traffic flow, and the different traffic flow situation between the working days and non-working days, the traffic flow data of the three consecutive days from 7th September, 2017 (Thursday), 8th September, 2017 (Friday), and 9th September, 2017 (Saturday) is selected as the research subject. Considering the less traffic flow from 10:00 p.m. to 6:00 a.m. of the second day, and its less reference value on the actual prediction, the data sampling time of the three consecutive days is chosen from 6:00 a.m. to 10:00 p.m. every day to acquire the traffic flow data. Taking account of the importance of the traffic flow data during the traffic peak period, the data from 7:00 a.m. to 8:00 p.m. of the three consecutive days is selected as the simulation experiment sample.

For the three consecutive days, from 7:00 a.m. to 8:00 p.m., with the collection time interval of 5 minutes, 471 data is collected in all. With 157 data per day, the first 130 traffic flow data is used as the training set for support vector regression model, and the remaining 27 data is used as the test set for support vector regression model.

The data is preprocessed, including missing data processing, abnormal data processing and normalization processing.

The input data and output data of SVR are illustrated as follows. Set $x(t)$ to be the traffic flow of time t, $x(t-1)$ to be the traffic flow of time $t-1$, $x(t-2)$ to be the traffic flow of time $t-2$, $x(t-3)$ to be the traffic flow of time $t-3$, $x(t-4)$ to be the traffic flow of time $t-4$, and $x(t-5)$ to be the traffic flow of time $t-5$. Because the traffic has the change rule in time series trend, the traffic flow of the first 5 moments is adopted to predict the traffic flow of the current time. Therefore, the input data is $[x(t-5); x(t-4); x(t-3); x(t-2); x(t-1)]$, and the output data is $x(t)$.

4.1.2. *Experiment design.* For verifying the accuracy of the ABC-SVR of the paper, the particle swarm optimization SVR (PSO-SVR) and the genetic algorithm optimization SVR (GA-SVR) are compared with the ABC-SVR. The three algorithms have the same optimization subject, and the optimization subject is the parameter of SVR.

4.1.3. *Experiment parameter.* The parameters of SVR are C, ε and σ, which satisfy $C \in [0, 100]$, $\varepsilon \in [0.01, 0.5]$, and $\sigma \in [0.1, 10]$. The specific parameter settings of each model in the simulation experiments are shown in Table 1.

Table 1. Initial parameter settings of each model

Optimization model	Parameter	Initial value
PSO	weight ω	1.0
	c_1, c_2	1.7, 2.0
	elasticity coefficient	1.0
GA	crossover probability	0.85
	mutation probability	0.02
ABC	quantity of food source	20
	employed bee, onlooker bee	10, 10
	limit	100
	maximum number of iteration	120

4.1.4. *Experiment tool.* The simulation experiments are performed using Libsvm-3.21 toolkit, which is installed on MATLAB R2010b.

The sketch procedure of the core code for a flow prediction value is presented below.

```
tic;
x = [x1; x2; x3; x4; x5]; y = [x2; x3; x4; x5; x6];
model = svmtrain(y, x, '−s 3 −t 2 −c 48.22 −h 1 −g 0.56 −p 0.23');
[py, mse, ∼] = svmpredict(y, x, model);
testx = [x6]
display('real data')
Testy = [x7]
[ptesty, tmse, ∼] = svmpredict(testy, testx, model);
display('predicted data');
ptesty
toc
```

where, $-s$ 3 is ε-SVR; $-t$ 2 is the radical basis function (RBF); $-c$ 48.22 is the value of penalty coefficient C; $-h$ 1 denotes that the heuristic method is adopted; $-g$ 0.56 is the value of RBF kernel function parameter σ; $-p$ 0.23 is the value of non-sensitivity loss coefficient ε.

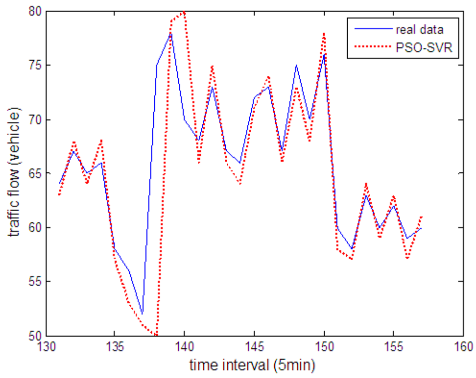
4.2. **Experimental results and analysis.**

4.2.1. *Experimental results.* According to the experimental setup, the three parameters of SVR are optimized, which is shown in Table 2.

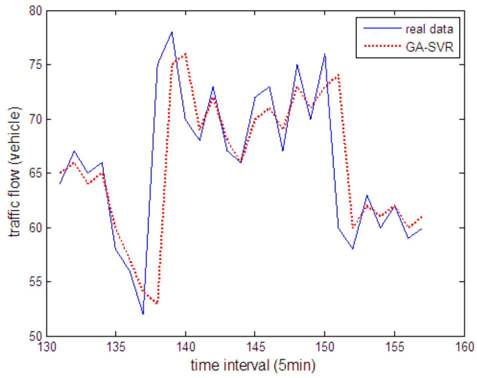
The horizontal contrast experimental results among PSO-SVR, GA-SVR, ABC-SVR and the real value are shown in Figures 2-4.

TABLE 2. Parameter optimization result of each optimized algorithm

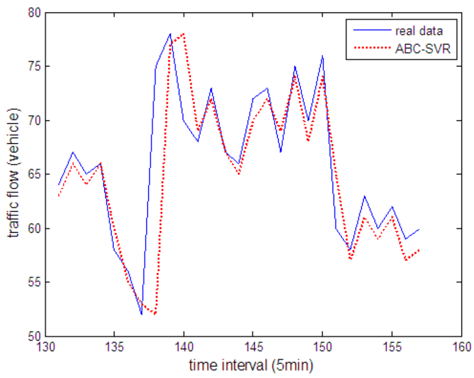
Parameter	PSO-SVR	GA-SVR	ABC-SVR
C	14.31	20.19	48.22
ε	0.10	0.16	0.23
σ	0.64	0.68	0.56



(a) PSO-SVR



(b) GA-SVR

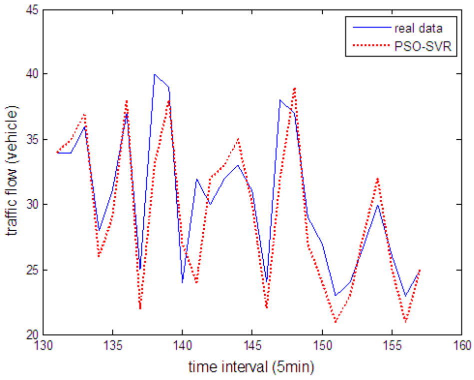


(c) ABC-SVR

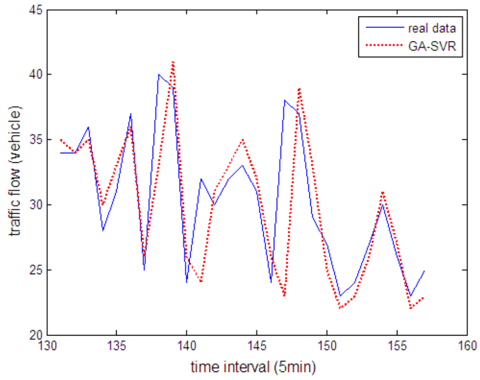
FIGURE 2. Prediction result of each model on 7th September

4.2.2. *Experimental analysis.* By analyzing Figures 2-4, it can be observed the curve fitting between PSO-SVR and the real value, that between GA-SVR and the real value, and that between ABC-SVR and the real value are better. And it proves that the prediction accuracy of all three algorithms is higher than that of SVR model prediction results. From the evaluation indexes of the prediction results of each optimization algorithm, it is demonstrated that the ABC-SVR algorithm has much higher prediction accuracy, and the mean relative error (MRE) of each model prediction result is shown in Table 3.

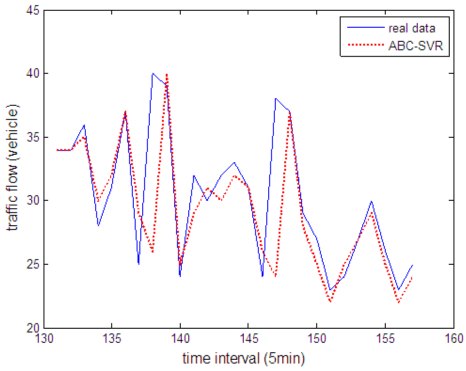
Through the analysis of experimental results, it can be confirmed that the ABC-SVR algorithm of the paper has more precise prediction accuracy as well as higher prediction property.



(a) PSO-SVR

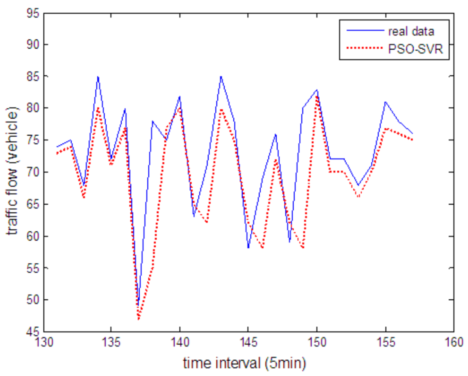


(b) GA-SVR

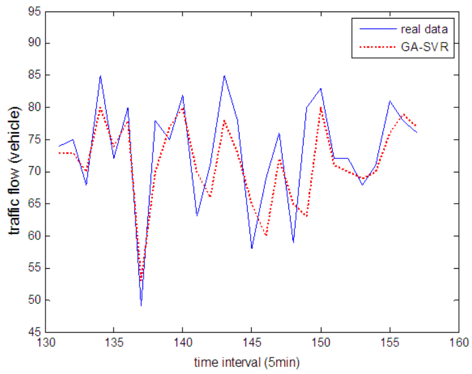


(c) ABC-SVR

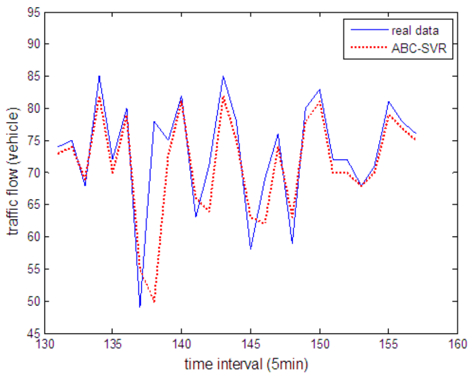
Figure 3. Prediction result of each model on 8th September



(a) PSO-SVR



(b) GA-SVR



(c) ABC-SVR

Figure 4. Prediction result of each model on 9th September

Table 3. MRE of each model prediction result

Date	MRE (PSO-SVR)	MRE (GA-SVR)	MRE (ABC-SVR)
7th September	3.82%	4.12%	3.58%
8th September	7.18%	7.55%	6.41%
9th September	5.98%	5.77%	4.85%

5. **Conclusions.** The SVR algorithm has the capacity of solving the problems of non-linearity, small sample and high dimension, which conforms to the characteristics of the short-term traffic flow prediction research. Therefore, SVR algorithm is adopted to construct the short-term traffic flow prediction model. Aiming at the SVR prediction model parameter selection problem, ABC algorithm is used to acquire the optimal parameter, and the short-term traffic flow prediction algorithm by support vector regression based on artificial bee colony optimization is presented. The simulation experimental results prove the feasibility and accuracy of the proposed algorithm. The intelligent combination prediction algorithm is a hot issue at home and abroad, and the SVR algorithm proposed in the paper can also be applied to other fields. The optimal algorithm will show more important and practical value in the future. Meanwhile, the adaptability in different fields of prediction algorithm, the public data set construction of traffic flow, the computational complexity of prediction process as well as the own optimization of optimization algorithm need further investigations and explorations in the future research.

Acknowledgment. This work is partially supported by the Education Department of Heilongjiang Province of China, No. 135209234 and No. 135209243, and the Bureau of Science and Technology of Qiqihar of Heilongjiang Province of China, No. GYGG-201612. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] Z. S. Yang, Y. Wang and Q. Guan, Short-term traffic flow prediction method based on SVM, *Journal of Jilin University (Engineering and Technology Edition)*, vol.10, no.6, pp.881-884, 2006.

[2] J. Liu and W. Guan, A summary of traffic flow forecasting methods, *Journal of Highway and Transportation Research and Development*, vol.21, no.3, pp.82-85, 2004.

[3] B. M. Williams and L. A. Hoel, Modelling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results, *Journal of Transportation Engineering*, vol.129, no.6, pp.664-672, 2003.

[4] I. Okutani and Y. J. Tephanedes, Dynamic prediction of traffic volume through Kalman filtering theory, *Transportation Research Part B: Methodological*, vol.18, no.1, pp.1-11, 1984.

[5] L. I. Qi and G. Y. Jiang, Bi-level model of multi-step forecasting for short-term data of loop in Sydney coordinated adaptive traffic system, *Journal of Jilin University (Engineering and Technology Edition)*, vol.43, no.4, pp.854-860, 2013.

[6] M. L. Huang, Intersection traffic flow forecasting based on v-GSVR with a new hybrid evolutionary algorithm, *Neurocomputing*, vol.147, pp.343-349, 2015.

[7] B. Abdulhai, H. Porwal and W. Recker, Short-term traffic flow prediction using neuro-genetic algorithms, *Intelligent Transportation Systems Journal*, vol.7, no.1, pp.3-41, 2002.

[8] R. X. Fang, W. H. Yao and X. D. Zhang, Short-term traffic flow forecasting based on SVR with improved artificial fish swarm algorithm, *Applied Mechanics & Materials*, vol.734, no.8, pp.508-514, 2015.

[9] M. Qin, G. F. Yang, et al., Short-term traffic flow forecasting based on exponential smoothing and Kalman filter, *Journal of Beihua University (Natural Science)*, vol.16, no.6, pp.814-817, 2015.