Hybrid GA Based Online Support Vector Machine Model for Short-Term Traffic Flow Forecasting

Haowei Su and Shu Yu

College of Computer Science and Engineering, South China University of Technology, Guangzhou 510640, P.R. China
gzsuhw@tom.com, yushu_scut@126.com

Abstract. In this paper, a hybrid genetic algorithm (GA) based online support vector machine (OSVM) prediction model for short-term traffic flow forecasting is proposed, according to the data collected sequentially by the probe vehicle or the loop detectors, which can update the forecasting function in real time via online learning way, and the parameters used in the OSVM were optimized by GA. As a result, it is fitter for the real engineering application. The GA based OSVM model was tested by using the I-880 database, the result shows that this model is superior to the back-propagation neural network (BPNN) model.

1 Introduction

Intelligent Transportation System (ITS) emerges as the times requires. The basic idea of the ITS is: on the precondition that the current road situation is not varied, and recur to the high precision and real time predictive method of traffic flow [1-4], so as to achieve effective control of traffic and transport guide from the largest extent to ease the problem of traffic jams.

According to the theory, the traffic flow forecasting research can be divided into two types. One type is determined based on the mathematical model [5], but the short-term traffic flow prediction is more influenced by the stochastic interferential factors than the long-term one, the uncertainty is greater and the disciplinarian laws are less obvious. Thus using the short-term traffic prediction models based on the classical mathematical methods, the precision of forecast can not satisfactorily meet the demand of real-time traffic control and guidance in ITS.

The second type is knowledge-based intelligent model of forecasting methods [6], in which the typical representative one is the BPNN model. As BPNN learning algorithm used gradient descent algorithm and weights regulation to minimize the objective function, the objective function was set by the square sum of the margin between the input and output values, so the BPNN led to excessively emphasize the learning mistakes and the over fitting problem appeared inevitably.

Based on the fact that the traffic flow prediction is equivalent to the function estimates and approximation [7], it can be handled as the function estimates issue. Support vector machine (SVM) is a new type of learning machine [8], using structured risk minimization (SRM) principles to perform regression or pattern recognition. SVM

M. Xu et al. (Eds.): APPT 2007, LNCS 4847, pp. 743–752, 2007.

can solve some flaws of the neural networks, and has many unique advantages in the fields of small samples and high-dimensional nonlinear manifested [9-10].

Currently, the traffic flow forecast based on support vector machine is a new and hot research field, but still at a preliminary stage. In 2004, a traffic forecasting method was proposed by Wang Jisheng et al [11], which based on support vector machines theory. The model was solved via the LIBSVM algorithm and the results were better than the BPNN. Zhang Chaoyuan et al [12] proposed a least square support vector machine version model to forecast the traffic flow time series, and the algorithm based on the LS-SVM was presented.

The traffic flow prediction has a major characteristic of real-time nature. On the basis of research on support vector machine learning, Yin Ying et al [13] designed a least square support vector machine simulation and real-time traffic flow forecasting system using Matlab language, which gave a new way for visualization expression of the traffic flow guidance data. A real-time traffic flow prediction model based on support vector machine was given by Xu Qihua and Yang [14]. Through sequential minimal optimization (SMO) algorithm, this model can effectively forecast for noise traffic data. In the literature [15], after the analysis of the characteristics of urban traffic flow, the authors introduced the kernel machine methods, and compared performance of the compound and conventional kernels was also shown. A short-term traffic flow forecasting model based on support vector machine was proposed by Yang Zhaosheng et al [16] in 2006, on the basis of concluding variety of traffic flow forecasting models and the mature thought about the nonlinear, complexity and uncertainty of traffic system. Compared with the BPNN model, the results showed that in the fields of accuracy, convergence time, generalization and optimality, SVM-based model is superior to neural network-based models. SVM-based forecasting method was applied to the field of traffic incident detection [17], and had shown good results.

The parameters selected in the SVM are much important to the efficiency of the prediction model, especially in the real application. In order to gain the optimal parameters, a global optimize process should be implement.

Genetic algorithm (GA) [18] is a global stochastic algorithm, which derives from the ideas of selection process in nature and genetic mechanism. GAs were successfully applied to many fields [19], because of their superiority of implementation on massively parallel architecture, compared with traditional optimization methods in searching for the global optimum of complex problem. However, it is not easy to regulate GA's convergence so that GA often suffers from premature convergence [20].

By contrast with GA, simulated annealing (SA) [21] algorithm employs certain probability to escape from local optima and the search process can be controlled by the cooling schedule. Since the complementary strengths of GA and SA, how to integrate GA with SA to achieve more efficient optimization results was widely studied in both theory and application areas [22].

Therefore, in this paper, an online learning approach using support vector machine model was proposed, which can dynamically update the forecasting function via the data collected continuously by the probe vehicle or loop detectors. And the parameters used in the OSVM were determined by the hybrid GA. The goal is to develop a rapid, real-time, and optimal short-term traffic flow prediction model with high generalization ability.

2 Hybrid GA Based Online Support Vector Machine

2.1 Least Square Support Vector Machine

In 1999, Suykens [23] presented the least square_support vector machine algorithm, in which the main idea is to introduce the least square system into the standard SVM.

Assuming study set is

$$\boldsymbol{S}=\{s_i \mid s_i=(\boldsymbol{x}_i, y_i), \boldsymbol{x}_i \in R^n, y_i \in R, i=1,2,\cdots,l\},$$

where the regression function expressed as:

$$y(\boldsymbol{x})=\boldsymbol{w}\cdot\varphi(\boldsymbol{x})+b \tag{1}$$

The LS-SVR proposed by Suykens et al is to solve the following problem:

$$\begin{cases} \boldsymbol{min}\ Q(\boldsymbol{w},\boldsymbol{e})=\frac{1}{2}\|\boldsymbol{w}\|^2+\frac{\gamma}{2}\sum_{i=1}^{l}e_i^2 \\ s.t.\ y_i=\boldsymbol{w}\cdot\varphi(\boldsymbol{x}_i)+b+e_i, i=1,2,\cdots,l \end{cases} \tag{2}$$

It can be known from formula (2), that the balance equation is expressed as:

$$\begin{bmatrix} 0 & \vec{\boldsymbol{1}}^T \\ \vec{\boldsymbol{1}} & \boldsymbol{ZZ}^T+\gamma^{-1}\boldsymbol{I} \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix}=\begin{bmatrix} 0 \\ \boldsymbol{y} \end{bmatrix} \tag{3}$$

where $\boldsymbol{Z}=[\varphi(\boldsymbol{x}_1),\varphi(\boldsymbol{x}_2),\cdots,\varphi(\boldsymbol{x}_l)]^T$, $\boldsymbol{y}=[y_1,y_2,\cdots,y_l]^T$,

$\vec{\boldsymbol{1}}=[1,1,\cdots,1]^T$, $\boldsymbol{e}=[e_1,e_2,\cdots,e_l]^T$, $\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_l]^T$.

From the Mercer condition:

$$\varphi(\boldsymbol{x}_i)\cdot\varphi(\boldsymbol{x}_j)=k(\boldsymbol{x}_i,\boldsymbol{x}_j)\equiv\Omega_{ij},\ i,j=1,2,\cdots,l, \tag{4}$$

here $k(\cdot,\cdot)$ is a kernel function, which is often used as Gauss kernel

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{exp}(-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2/(2\sigma^2))$$

$\boldsymbol{\Omega}+\gamma^{-1}\boldsymbol{I}$ is called as kernel correlation matrix, let $\boldsymbol{A}\equiv\boldsymbol{\Omega}+\gamma^{-1}\boldsymbol{I}$, then formula (4) can be written as:

$$\begin{bmatrix} 0 & \vec{\boldsymbol{1}}^T \\ \vec{\boldsymbol{1}} & \boldsymbol{A} \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix}=\begin{bmatrix} 0 \\ \boldsymbol{y} \end{bmatrix} \tag{5}$$

the regression function in formula (1) is:

$$y(\boldsymbol{x})=\boldsymbol{w}\cdot\varphi(\boldsymbol{x})+b=\sum_{i=1}^{l}\alpha_i\varphi(\boldsymbol{x}_i)\varphi(\boldsymbol{x})+b=\sum_{i=1}^{l}\alpha_i k(\boldsymbol{x}_i,\boldsymbol{x})+b \tag{6}$$

$\boldsymbol{\alpha}, b$ are called the regression parameters.

It can be seen from formula (6) that the key to get the regression parameters is the computation of the matrix, $\boldsymbol{A}^{-1}$.

2.2 Online SVM Learning Algorithm

Conducting short-term traffic flow forecasting, the system continuously collects the data on the flow and return to the prediction model in sequence, after that the model should be adaptively adjusted according to the new collected sample.

When new samples (x_{l+1}, y_{l+1}) were added to the learning set, the kernel correlation matrix was changed into:

$$A_{l+1} = \begin{bmatrix} A_l & B^T \\ B & c \end{bmatrix} \tag{7}$$

where A_l, A_{l+1} were kernel correlation matrix of learning set S and $S \bigcup \{ s_{l+1} \}$

$$B = \lfloor y_{N+1} \quad \Omega_{N+1,1} \quad \Omega_{N+1,2} \quad \cdots \quad \Omega_{N+1,N} \rfloor$$

If A_l^{-1} can be used to obtain A_{l+1}^{-1} without the totally recalculating, then the online SVM learning task is done. In fact, this is the improvement to the online LS-SVR algorithm which was the LS-SVR based research result of liu et al [20]. In paper [21] there can be shown:

$$A_{l+1}^{-1} = \begin{bmatrix} A_l & B^T \\ B & c \end{bmatrix}^{-1} = \begin{bmatrix} \left[A_l - \frac{1}{c} B^T B \right]^{-1} & A_l^{-1} B^T \left[B A_l^{-1} B^T - c \right]^{-1} \\ \left[B A_l^{-1} B^T - c \right]^{-1} B A_l^{-1} & \left[c - B A_l^{-1} B^T \right]^{-1} \end{bmatrix} \tag{8}$$

$$\left[A_l - \frac{1}{c} B^T B \right]^{-1} = A_l^{-1} - A_l^{-1} B^T \left[-c + B A_l^{-1} B^T \right]^{-1} B A_l^{-1} \tag{9}$$

From (8-9), on the basis of the original results, the prediction function should update as following, according to the new samples added:

$$f(x) = \sum_{i=1}^{l+1} \alpha_i k(x_i, x) + b \tag{10}$$

2.3 Hybrid GA

The hybrid GA was used to optimal the parameters γ and σ in the OSVM, the object value is the training accuracy.

GA initializes a population. The population evolves from parent generation to child generation by three basic operators: the selection, the crossover and the mutation. Then the parent and child populations are united as an extended population. SA is used as a Bolzmann reduction operator to reduce the extended population to original size, so a reduced population comes into being. To control the population diversity, the diversity function is used to quantify the degree of parent and reduced population diversity, and in terms of which the one with superior diversity must have a higher probability to be chosen as the new population. The structure of the presented algorithm can be seen in the follows:

Begin
While not stop **do**
Initialize Population (P_0)
$k = 0$
While not stop **do**
Do n/2 times
{
Select two parents from P_k
Generate two children by using crossover
Mutate the two children
Introduce the children in the child population CH
}
Make the extended population $P_k \cup CH$
Introduce the n/10 best individuals of the extended population in the elitist pool
Make the reduced population P'' by reduce the extended population to the original size
Choose P_k or P'' to be P_{k+1} in terms of their degree of population diversity
Modify temperature (c_k)
$k = k + 1$
End

End

End

Both the variable and object function are continuous. In order to improve the accuracy and the convergence speed of GA, we use real coding. So one representation of the solution of an individual can be $X = [x_1, x_2] = [\gamma, \sigma]$.

Selection: the selection operator used is "roulette wheel selection". The probability of reproduction for the individual i is given by: $P(i) = F(i) / \sum_{i \in I} F(i)$.

Crossover: the crossover operator used is convex crossover.

$$\begin{matrix} X_1' = \lambda_1 X_1 + \lambda_2 X_2 \\ X_2' = \lambda_2 X_1 + \lambda_1 X_2 \end{matrix}, \lambda_1 + \lambda_2 = 1, \quad \lambda_1 > 0, \lambda_1 > 0$$

Mutation: The mutation operator used is as followed: $X' = X + r \cdot d$, d is the approximately grads, the i th variable of d can be computed by

$$d_i = \frac{f(x_1, \cdots x_i + \Delta x_i, \cdots, x_n)}{\Delta x_i}$$

r is a non- negative random real number.

The reduction operator consists of sampling a Boltzmann probability distribution in the extended population (the union of the parent and child populations). The value of this probability distribution depends on the fitness function of the individuals in the population. If the size of the population is n , in the case of maximization, it works as follows:

While not n individuals have been selected **do**
{
Choose randomly an individual i from the extended population
If $F(i) > \overline{F}$ then
Select individual i for the reduced population
Else
Select individual i with probability equal to $\exp(\frac{F(i)-\overline{F}}{c_k})$
}
End

Where $F(i)$ is the fitness value of the ith individual and $\overline{F}$ is the mean value of fitness function F in the parent population.

3 Hybrid GA Based OSVM Model for Short-Term Traffic Flow Forecasting

The probe vehicle or loop detectors can be used to collect the traffic information [24]. Subject to random factors (e.g. transmission errors, etc.), it can not be avoided to lose data accuracy such as data errors and data loss, so the data preprocessing need to be implemented to correction errors, of which the threshold test and traffic flow theory-based check are two commonly used methods. At last, the data should be normalized treatment to improve the efficiency of computation.

The short-term traffic flow forecasting process can be described as followed:

Step1: Chose N traffic flow samples as initial training set

$$\boldsymbol{S} = \{s_i \mid s_i = (\boldsymbol{x}_i, y_i), i = 1,2,\cdots, N\},$$

According to formula (1)-(6) of LS-SVM algorithm, the initial prediction regression function is obtained. Then the prediction value for traffic flow y_j is:

$$y_j = \sum_{i=1}^{N} \alpha_i k(\boldsymbol{x}_i, \boldsymbol{x}_j) + b \tag{11}$$

Step2: Alone with the data collected continuously, the prediction regression function should be updated in time. Assuming a new samples are chosen to add in the training set, so

$$\boldsymbol{S} = \{s_i \mid s_i = (\boldsymbol{x}_i, y_i), i = 1,2,\cdots, N, N+1\}$$

According to formula (7)-(10) of OSVM algorithm, the regression function can be reconstructed, the prediction value for traffic flow y_j is:

$$y_j = \sum_{i=1}^{N+1} \alpha_i k(\boldsymbol{x}_i, \boldsymbol{x}_j) + b \tag{12}$$

The Guass kernel

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{exp}(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 /(2\sigma^2))$$

was used as the kernel function.

Step3: Use hybrid GA to optimal the parameters of the online support vector machine.

4 Model Validation and Comparison

In order to analyze the forecast results better, five error indicators are introduced:

Relative Error:

$$rerr = \frac{v_{pred}(t) - v_{real}(t)}{v_{real}(t)},$$

Mean Relative Error:

$$mrerr = \frac{1}{N} \sum_t \frac{v_{pred}(t) - v_{real}(t)}{v_{real}(t)},$$

Mean Absolute Relative Error:

$$marerr = \frac{1}{N} \sum_t \frac{|v_{pred}(t) - v_{real}(t)|}{v_{real}(t)},$$

Root Mean Square Relative Error:

$$rmrerr = \sqrt{\frac{1}{N} \sum_t \left(\frac{v_{pred}(t) - v_{real}(t)}{v_{real}(t)} \right)^2},$$

Equalization Coefficient (EC):

$$EC = 1 - \frac{\sqrt{\sum_t (v_{pred}(t) - v_{real}(t))^2}}{\sqrt{\sum_t (v_{pred}(t))^2} + \sqrt{\sum_t (v_{real}(t))^2}}$$

EC means the difference between the predicted and actual fit in a well fitting, when it is above 0.90.

In order to compare the prediction results of SVM based model with that of popular neural network based method, we construct the BPNN model for the traffic flow forecast. The BPNN-based traffic flow prediction model is composed of data processor, input layer, output layer and hidden layer. It can be seen in the Fig.1.:

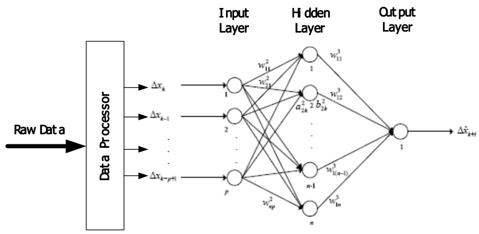

Fig. 1. BPNN-based traffic flow prediction model

The formula of the prediction model is as following:

$$Q_t = \left[\frac{\boldsymbol{exp}(-\sum_{j=1}^{i} v_{jt})}{1+\boldsymbol{exp}(-\sum_{i=1}^{n} w_{ij} x_i \theta_j)} + \gamma_i \right]^{-1} \tag{13}$$

here, Q_t is the forecast value, w_{ij} is the weight of connection between input layer and hidden layer, θ_j is the threshold of hidden layer unit, v_{jt} is the weight of connection between hidden layer and output layer, γ_i is the threshold of output layer unit.

To validate the correctness and accuracy of the forecasts model, the I-880 database [25] was used to test. Since there were large amount of raw data in this database, we simply randomly selected some samples in a short period in each day, and the data were regulated for computational convenience. The BPNN-based short-term traffic flow prediction model was programmed by using Matlab7.0.1 neural network toolbox, the hybrid GA based OSVM forecast model was programmed by using Microsoft Visual C++ 6.0 complier. The operating environment is: CPU Pentium 1.5 MHz, Memory 1G, Microsoft Windows XP operation system.

Table 1. Results of two models

Model	Evaluate Indicators			
	mrerr%	marerr%	rmrerr%	EC
OSVM	0.51	5.25	6.84	0.97
BPNN	0.77	1.46	13.28	0.933

According to the results of Table 1, the hybrid GA based OSVM model gave higher accuracy than the BPNN-based model, and reached a high value of EC fitting. Besides the OSVM updated forecast function via online learning algorithm, which was more suited for the practical application. In generalized performance, the OSVM model is also superior to the BPNN-based model. The basis of OSVM was support vector machines, so the forecast error is relatively stable even the fitting error was large. But neural network appear over fitting easily, when fitting errors are reduced the forecast error will get larger soon. Moreover, the training of OSVM in fact is a convex quadratic programming problem, which can obtain the global optimal solution with hybrid GA optimized parameters. Contrarily, gradient descent algorithm was used for neural network, which often result in local optimal solution.

5 Conclusions

The performance of many components in intelligent transportation systems depends heavily on the quality of short-term traffic forecasts. Considering the real-time, nonlinear, complexity and uncertainty in traffic problems, a new hybrid GA based online support vector machine was proposed for designing short-term traffic flow prediction model. This method was based on least square support vector machine, and used online learning strategy to dynamic update forecast function, which was more suited for the practical application. Moreover, the designed online learning algorithm used hybrid GA to optimal the parameters. The I-880 database was used to test the hybrid GA based OSVM model and the BPNN-based model, and the results shown that the hybrid GA based OSVM was superior to the BPNN work in accuracy, generalized ability and optimization. The further work is to find an effective strategy to dump the useless history data in the training set, in order to avoid reducing the convergence speed when the learning samples increase in number.

References

1. Zhu, Z., Yang, Z.S.: Real time traffic flow prediction model based on ANN. China Journal of Highway and Transport 11(4), 89–92 (1998)
2. Yang, Z.S., Gu, Y.: Real time and dynamic traffic flow forecast research. Journal of Highway and Transportation Research and Development 15(3), 4–7 (1998)
3. Yang, Z.S.: Theories and models of urban traffic guidance system. China Communications Press, Beijing (2000)
4. Li, C.J., Yang, R.G., Zhang, J.S.: Traffic flow forecasts based on wavelet analysis. Journal of Computer Applications 23(12), 7–8 (2003)
5. He, G.J., Li, Y., Ma, S.F.: Short-term traffic flow prediction based on mathematic models. Systems engineering-theory & practice 12, 51–56 (2000)
6. He, G.Q.: Introduction to ITS systems engineering. China railway publishing house, Beijing (2004)
7. Yang, Z.S., Jiang, G.Y.: Theories and models of urban traffic guidance system based on high-order neural networks. Journal of Highway and Transportation Research and Development 6, 16–19 (1998)
8. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995)
9. Cheng, H., Cheng, L.H., et al.: The Appliance of BP-Network and SVM in Approach of Non-linear Function. Aeronautical Computer Technique 34(3), 27–30 (2004)
10. Xu, Q.H., Shi, J.: Aero-Engine Fault Diagnosis Based on Support Vector Machine. Journal of Aerospace Power 2, 298–302 (2005)
11. Wang, J.S., Gao, B.C., Shi, L.P.: Application of support vector machines in traffic volume forecast. Information Technology 4, 8–10 (2004)
12. Zhang, C.Y., Hu, G.H., Xu, T.Z.: Traffic flow time series prediction based on LS-SVM. Journal of Yunnan University 26, 19–22 (2004)

13. Yin, Y., Zhang, C.Y., Hu, G.H., Xu, T.Z.: Design of Real-Time Traffic Flow Simulating and Forecasting System Based on SVM. Journal of Computer Engineering and Applications 10, 197–199 (2005)
14. Xu, Q.H., Yang, R.: Traffic Flow Prediction Using Support Vector Machine Based Method. Journal of Highway and Transportation Research and Development 22(12), 131–134 (2005)
15. Jiang, G., Xiao, J.: Real-time Forecast of Urban Traffic Flow Based on Kernel Machine Method. Computer Engineering 32(17), 48–51 (2006)
16. Yang, Z.S., Wang, Y., Guan, Q.: Short-term traffic flow prediction method based on SVM. Journal of Jilin University 36(6), 881–884 (2006)
17. Qin, P.P.: Comparison of SVM and Neural Network Model for Incident Detection. Journal of Computer Engineering and Applications 34, 214–232 (2006)
18. Holland, J.H.: Adaptation in natural and artificial system. MIT Press, Cambridge, MA (1975)
19. Joine, J., Houck, C.: On the use of non-stationary penalty functions to solve nonlinear constrained optimization problems with GAs. In: Proceedings of the First IEEE Conference on Evolutionary Compution, pp. 548–579. IEEE Press, Los Alamitos (1994)
20. Kirpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science 220(4598), 671–680 (1983)
21. Kovac, A., Glavic, P.: Retrofit of complex and energy intensive processes-I. Computers & Chemical Engineering 19, 1255–1270 (1995)
22. Kralj, A.K., Glavic, P.: Retrofit of complex and energy intensive processes. Computers & Chemical Engineering 21(Suppl.), 517–522 (1997)
23. Suykens, J.A.K., Vandewalle, J.: Least Squares Support Vector Machine Classifiers. Neural Process Letter 9, 293–300 (1999)
24. Zhang, C.B., Yang, X.G., Yan, X.P.: Traffic Data Collection System Based on Floating Cars. Computer and Communications 5(24), 31–34 (2006)
25. Petty, K.: The analysis software for the FSP project, http://ipa.eecs.berkeley.edu/ pettyk/FSP