Short Term Traffic Flow Prediction Based on Improved Support Vector Machine

Yuchi Zhang¹ and Zhixiang Hou²*

¹Hunan Industry Polytechnic, Changsha, Hunan 410082, P.R. China
²College of Automotive and Mechanical Engineering, Changsha University of Science and Technology, Changsha 410114, P.R. China

Abstract

The traditional forecasting methods are not suitable for short term traffic flow prediction, due to strong non-linear, time varying characteristics of urban transportation system. In order to improve forecasting accuracy of short term traffic flow, short term traffic flow prediction model based on support vector machine is presented. The most important parameter of support vector machine is parameter selection including the kernel function parameter and the penalty factor, which has significant influence on the properties of model prediction. Particle swarm optimization is used to optimize support vector machine, and particle swarm optimization is improved by means of adjusting inertia weight and choosing acceleration constant dynamically. Then improved particle swarm optimization is used to optimize support vector parameter. The experiment results show that predictive result based on improved particle swarm optimization LSSVM is closer to the real traffic flow data compared with support vector machine based on basic particle swarm optimization. Short term traffic prediction model based on improved particle swarm optimized support vector machine is feasible.

Key Words: Support Vector Machine, Particle Swarm Optimization, Traffic Flow Prediction, Parameter Selection

1. Introduction

Short term traffic flow forecasting is to set up suitable mathematical model to forecast traffic situation of one place reasonably based on historical and current traffic flow data and other related factors in a short period of time. The dynamic analysis and prediction of short-term traffic flow is an important part of intelligent transportation system. In order to enable users to choose the most appropriate path and avoid the congestion section, traffic flow analysis and prediction can be used in traffic guidance, which can reduce driving loss of the user, reduce traffic load, and use road resources reasonably [1]. Therefore, short-term traffic flow prediction is very important basic theory in the intelligent transportation field.

Support vector machine (SVM) is a kind of machine learning algorithm in recent years, and is the research front of complex nonlinear and artificial intelligence science [2]. Due to its outstanding performance of classification and regression, it is gradually applied in many fields, and has got good application in the field of traffic flow prediction. In theory or practice, there are still many questions worth further researching [3]. For example, the construction of the kernel function and its parameters lacks of theoretical guidance. When dealing with large-scale data set, the training of support vector machines (SVM) is slower and how to improve the generalization ability is a research focus [4]. Expansion of incremental learning ability of support vector machine (SVM) is also a problem.
In practical short-term traffic flow prediction, the collected traffic flow data is limited, how to learn the limited data samples quickly and accurately, and achieve good prediction effect is the key to the short-term forecast. Artificial neural network has been widely used in traffic prediction, but the network convergence speed is too slow. It is easy to fall into local extremum and needs larger data sample [5]. The least square support vector machine (LSSVM) is an improved algorithm based on support vector machine (SVM) method. It is able to get the optimal solution through the learning of small samples, and can turn quadratic programming problem into solving linear equations, which reduces the calculation complexity, and improves the learning efficiency [6–8]. But, parameter selection of both SVM and LSSVM may affect the entire network generalization capability, and influence the accuracy of the short-term traffic flow prediction [9,10]. Therefore, we introduce particle swarm algorithm [11,12] to improve the LSSVM, using the global search ability of particle swarm to determine the optimal parameters of LSSVM forecasting model. Previous studies are aimed at using particle swarm optimization to optimize LSSVM, but we proposed a novel particle swarm optimization based on dynamic inertia weight adjustment and dynamical selection of acceleration constant, which has better convergence performance.

In the next section, short term traffic flow prediction model based on least square support vector machine is investigated. In section 3, optimized least square support vector machine based on improved PSO is put forward. In section 4, in order to test the performance of proposed short term traffic flow prediction model, experiments are done based on five consecutive days belonging to a section of Jinan road transportation network. In the end, some conclusions are given.

2. Short Term Traffic Flow Prediction Model Based on Least Square Support Vector Machine

Least squares support vector machine (LSSVM) was proposed by Suykens in order to improve the training efficiency of support vector machine, which applied the least square method in the optimization goal and constructed the quadratic loss function. Inequality constraints in the standard support vector machine are replaced by equality constraints. On this condition, the quadratic programming problem is transformed into the problem of solving linear equation, and it has high solving speed. Suppose the given learning sample set is \( \{(x_i, y_i) | i = 1, 2, \ldots, l\} \), \( l \) represents the number of sample. The nonlinear function expression is

\[
 f(x) = w \cdot \varphi(x) + b 
\]  

(1)

\( \varphi(x) \) represents nonlinear mapping function, \( w \cdot \varphi(x) \) represents inner product of the two and \( b \) represents offset. The regression problem is described as

\[
 s.t. \ y_i = w^T \varphi(x_i) + b + e_i 
\]  

(2)

The lagrange equation is set up to solve the optimization problem.

\[
 L(w, b, \alpha_i, e_i) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{l} e_i^2 - \sum_{i=1}^{l} \alpha_i [w \cdot \varphi(x_i) + b + e_i - y_i] 
\]  

(3)

\( e_i \) represents error variable, \( C \) represents penalty factor and \( \alpha_i \) represents lagrange multiplier. The optimization problem is transformed into solving linear equations.

\[
 \begin{bmatrix} 0 & I^T \\ I & K + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} 
\]  

(4)

\[
 \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_l]^T, \ y = [y_1, y_2, \ldots, y_l]^T, \ I = [1, 1, \ldots, 1]^T, \ K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j). \]  

The regression function is

\[
 f(x) = \sum \alpha_i K(x_i, x) + b 
\]  

(5)
Similar to SVM, LSSVM is also confronted with the choice of kernel function [13,14]. Different kernel function is adopted to establish the LSSVM regression model, so that the structure of model, learning and generalization ability is different. For short-term traffic flow prediction, the choice of kernel function may directly affect the prediction precision of LSSVM model. As a kind of global kernel function, polynomial kernel function has stronger generalization ability, but their learning ability is poor in the process of model training [15]. Sigmoid kernel function is not positive definite, the parameters need to meet certain conditions in practice. Radial basis kernel function not only has a strong nonlinear processing ability and wide adaptability, but also its parameters are the least. Therefore, we select the radial basis kernel function to set up LSSVM regression model for short-term traffic flow prediction.

When the selection of kernel function is completed, parameters influencing LSSVM regression model include penalty factor and kernel parameter [16]. Penalty factor is used to measure error tolerance of the model, which can adjust the space incredible scope and the proportion of empirical risk, and make the actual risk minimum, so as to improve the generalization performance of LSSVM. Nuclear parameter determines the distribution scope and complexity of input samples in the feature space, and the size of the kernel parameter also can affect the related degree between support vector at the same time. Above all, penalty factor and kernel parameter influences the performance of LSSVM. In the short-term traffic flow prediction, in order to improve prediction accuracy of LSSVM, the reasonable selection of penalty factor and kernel parameter is very important. Penalty factor and kernel parameter selection is essentially an optimization process which searches the optimal solution in the complex space through the competition and collaboration between the individual [17]. Each potential solution of optimization problem is taken as a particle in search space, and each particle has a fitness value, which is decided by optimized objective function [18]. Each particle has a velocity vector that determines the distance and direction of the particles in search space. In each iteration, the particle update itself by tracking individual and global optimal value. In the D dimension of objective search space, K number of particle forms one swarm. \( x_i = (x_{i1}, x_{i2}, \ldots, x_{iD}) \) represents the position of the \( i \)-th particle, \( i = 1, 2, \ldots, K \). Its speed \( v_i = (v_{i1}, v_{i2}, \ldots, v_{iD}) \) is a vector of \( D \) dimension. The optimal position of the \( i \)-th particle at present is labeled as \( pbest_i = (pbest_{i1}, pbest_{i2}, \ldots, pbest_{iD}) \). The optimal position of the whole swarm at present is \( Gbest = (Gbest_1, Gbest_2, \ldots, Gbest_D) \). The position and speed update formula is

\[
\begin{align*}
\dot{v}_{i+1} &= w \cdot v_i + c_1 \cdot rand(0,1)(pbest_i - x_i) \\
&\quad + c_2 \cdot rand(0,1)(Gbest_i - x_i) \\
x_{i+1} &= x_i + v_{i+1}
\end{align*}
\] (6)

\( v_i \) is limited to \([-v_{\text{max}}, v_{\text{max}}]\), \( w \) represents inertia weight, \( c_1 \) and \( c_2 \) are constants. In order to get the better result, dynamic inertia weight [19] is used.

\[
w = \max \{ w - curcount \cdot \frac{\max w - \min w}{looptime} \}
\] (7)

\( \max w \) represents the maximum inertia weight value and \( \min w \) represents the smallest inertia weight value. Generally, the maximum inertia weight value is 0.9 and the smallest inertia weight value is 0.4. \( looptime \) represents the maximum iteration times of the particle swarm and \( curcount \) represents the current iteration times. When \( c_1 \) is 0, particle only depends on the mutual...
information transmission to find the optimal solution, and has slow convergence speed. When $c_2$ is 0, it is difficult to get the optimal solution. Previous researches [13,20] just considered dynamical inertia weight or acceleration constant. Here, we take two factors into account. Dynamical acceleration constant is also introduced, which makes the particle have better space searching ability in the early stages of the search. At the end of the search, the convergence rate of the optimal solution is improved. The acceleration constant [21] is

$$c_1 = r_1 + \frac{r_2 \cdot \text{curcount}}{\text{loopcount}}$$

$$c_2 = r_3 - \frac{r_4 \cdot \text{curcount}}{\text{loopcount}}$$

In the experiment, $r_1$ is 1, $r_2$ is 0.5, $r_3$ is 6 and $r_4$ is 2. In order to prevent the particle flying off the search space in the iterative process, the speed of particle is limited.

$$v_i = (1 - \frac{\text{curcount}}{\text{loopcount}})^k \cdot v_{\text{max}}$$

$$v_i > (1 - \frac{\text{curcount}}{\text{loopcount}})^k \cdot v_{\text{max}}$$

$$v_i = (1 - \frac{\text{curcount}}{\text{loopcount}})^k \cdot v_{\text{min}}$$

$$v_i < (1 - \frac{\text{curcount}}{\text{loopcount}})^k \cdot v_{\text{min}}$$

$k = 0.05$. When iteration starts, because $\text{curcount}$ is smaller, particle flying speed is limited in a large speed range, and with the iteration going on, $\text{curcount}$ increases gradually, the particle is limited in the smaller speed range. The particle can accurately search near the optimal solution. Support vector machine is optimized based on improved particle swarm optimization. The optimized are penalty parameter and kernel parameter, which are taken as the position of the particles. Each group of penalty and kernel parameter is the position of particles in the particle swarm. The best parameter combination to be found is global optimal value of particle swarm optimization. The particle swarm initialization is to initialize the particle’s position and velocity. It also calculates the fitness value of all particles initialization position, records the best fitness value at present and the best position corresponding to the best fitness value. The fitness function should be calculated.

In order to avoid particle searching in the solution space blindly, it is necessary to limit the particle velocity and position. When particle velocity exceeds the maximum speed limit and the minimum speed limit, the current speed should be assigned again. If the particle current speed is greater than the maximum speed limit, the current speed value is replaced by the maximum speed limit value. If the current speed is less than the minimum speed limit, the current speed is replaced by the minimum speed limit. In an iterative update, calculate fitness value of each particle, fitness value of each particle in the current generation is compared with fitness value of the last generation, update the individual optimal value of each particle, and the optimal value from the optimal value of each particle individual is found to be compared with the optimal fitness value of the last generation. Then the global optimal value is updated. So, the corresponding penalty parameter and radial basis kernel function value is found. When updating individual optimal and global optimal value, the phenomenon that the difference between the last and current generation of fitness value is small can turn up. The too big penalty parameter will lead to over learning. When absolute value which is equivalent to the difference between the $i$-th generation of global optimal fitness value and the global optimal fitness value is less than 0.001, parameter combination with smaller penalty parameter is selected. The existence of singular sample data will cause subsequent prediction model cannot effectively converge. Therefore, before training of measured traffic flow data set, the data is normalized according to formula (11).

$$y = \frac{x - \text{min value}}{\text{max value} - \text{min value}}$$

$x$ represents original value, $y$ represents normalized value, min value represents the minimum value in the data set and max value represents the maximum value in the data set.

The process of proposed short term traffic flow prediction is as follows.
Step 1. Traffic flow training set and testing set data is preprocessed.

Step 2. The penalty factor and RBF kernel parameter are taken as the position of the particles, and each group of penalty factor and RBF kernel parameter is the location of particle in the particle swarm. The best parameter combination is the global optimal value in the particle swarm. Initialize particle swarm scale, position, speed and the maximum iteration times.

Step 3. The testing set is taken as input of least square support vector machine. Calculate fitness value of each particle.

Step 4. Calculate the best fitness value of the i-th generation of particle which is denoted as global_fitness(i). If global_fitness(i) is less than global fitness value, record current global fitness, penalty and kernel parameter. Update particle speed and position according to equation (6) which is limited by equation (10). The updated particle speed and position is taken as the new input parameters of least square support vector machine. Otherwise, turn to step 5.

Step 5. If it achieves the maximum iteration times, the optimal parameters and equation (1) to (5) is used to predict short term traffic flow. If it does not achieve the maximum iteration times, inertia weight and acceleration constant are adjusted by equation (7) to (9) and the algorithm turns to step 2 to go on.

4. Verification

In order to test the performance of proposed short term traffic flow prediction model based on improved support vector machine, experiments are done. We choose traffic flow data for five consecutive days belonging to a section of Jinan road transportation network in January 2015. Data acquisition is from 7:00 AM to 18:30 PM, acquisition interval is 5 minutes. Among them, there are 601 training samples and 60 testing samples. All samples are normalized, the value of which belongs to [0,1], $r_1$ is 1, $r_2$ is 0.5, $r_3$ is 6 and $r_4$ is 2, $k = 0.05$. The swarm scale is 10, the maximum iteration times is 50, penalty parameter belongs to [0,100], and kernel function parameter search scope belongs to [0.1,50]. The maximum speed of particle is $v_{\text{max}} = 4$ and the minimum speed of particle is $v_{\text{min}} = -4$.

Schaffer function is used to test the performance of improved particle swarm optimization.

$$F(x, y) = 0.5 - \frac{\sin(\sqrt{x^2 + y^2})^2 - 0.5}{1 + 0.001(x^2 + y^2)^2}$$  \hspace{1cm} (12)

The convergence performance is shown in Figure 1. It can be seen that the proposed PSO converges to the optimal value in the 35-th generation. PSO based on dynamical inertia weight and acceleration constant can converge to the optimal value in the 90-th generation. Standard PSO cannot converge to the optimal value. So, we can conclude that the proposed PSO has the best convergence performance. Then, the optimal fitness value curve of LSSVM based on PSO and improved PSO are worked out. With the increasing of the number of iterations, the swarm evolves following the change of overall fitness value. When improved particle swarm optimization is used to optimize the parameters of LSSVM, the 65th generation of optimal fitness value is 0.2209, namely...
the global optimal position of the whole swarm. The optimal fitness value of basic PSO in the 45th generation is 1.5607. We can conclude that the optimal fitness value of improved particle swarm optimization is superior to basic PSO. Basic PSO is easy to fall into the local extreme value point, and does not converge to the global optimal position. Therefore, PSO adopting adaptive inertia weight has strong global search ability, and can effectively seek global optimal solution of punishment factor and kernel parameter of LSSVM. Root mean square error and equalization coefficient are used to measure performance of proposed scheme, which are defined as follows.

\[
RMSE = \sqrt{\frac{1}{L} \sum_{i=1}^{L} (y_i - \overline{y}_i)^2} \tag{13}
\]

\[
EC = 1 - \frac{\sqrt{\sum_{i=1}^{L} (y_i - \overline{y}_i)^2}}{\sqrt{\sum_{i=1}^{L} y_i^2 + \sum_{i=1}^{L} \overline{y}_i^2}} \tag{14}
\]

where

\[ L \] represents the length of predictive sample, \( y_i \) represents actual value and \( \overline{y}_i \) represents predictive value. Predictive result of two algorithms is shown in Figure 2 and predictive result comparison of two algorithms is shown in Table 1. Root mean square error and equalization coefficient of improved PSO-LSSVM is 4.51 and 0.9803 respectively. Root mean square error and equalization coefficient of PSO-LSSVM is 5.56 and 0.9641 respectively. It can be concluded that predictive result based on improved particle swarm optimization LSSVM is closer to the real traffic flow data.

5. Conclusions

Least squares support vector machine has strong learning and generalization capability, and it can dispose small sample and nonlinear problems. To improve the forecasting precision of short-term traffic flow, forecasting method based on improved LSSVM is put forward. A novel particle swarm optimization is presented to optimize LSSVM, in which the optimal penalty factor and kernel parameter of LSSVM are selected by improved particle swarm optimization. Experiment results indicate that the forecasting model based on improved LSSVM not only has better forecasting capability, but also is effective and feasible in the forecasting of short-term traffic flow.

Acknowledgments

The paper was financially supported by the China National Natural Science Foundation (No. 51678077).

References


[18] Zheng, B. and Cui, B. T., “A Routing Protocol for WSN Based on Chaotic PSO and Ant Colony Algo-


Manuscript Received: Dec. 27, 2016
Accepted: Apr. 13, 2017