TRANSPORT
ISSN 1648-4142 print / ISSN 1648-3480 online
2011 Volume 26(1): 5–10
doi: 10.3846/16484142.2011.555472

THE USE OF LS-SVM FOR SHORT-TERM PASSENGER FLOW PREDICTION

Qian Chen[1], Wenquan Li[2], Jinhuan Zhao[3]

School of Transportation, Southeast University, Si-Pai-Lou 2, Nanjing, China
E-mails: [1]seu_chq@163.com (corresponding author); [2]wenqli@seu.edu.cn; [3] zhaojinhuan@163.com

Received 11 April 2010; accepted 10 January-2011

Abstract. Transit flow is the basement of transit planning and scheduling. The paper presents a new transit flow prediction model based on Least Squares Support Vector Machine (LS-SVM). With reference to the theory of Support Vector Machine and Genetic Algorithm, a new short-term passenger flow prediction model is built employing LS-SVM, and a new evaluation indicator is used for presenting training permanence. An improved genetic algorithm is designed by enhancing crossover and variation in the use of optimizing the penalty parameter γ and kernel parameter σ in LS-SVM. By using this method, passenger flow in a certain bus route is predicted in Changchun. The obtained result shows that there is little difference between actual value and prediction, and the majority of the equal coefficients of a training set are larger than 0.90, which shows the validity of the approach.

Keywords: short-term passenger flow prediction, least squares support vector machine, genetic algorithm.

1. Introduction

With the development of the Intelligent Public Transport System, the dispatcher can obtain timely passenger flow information. Meanwhile, it is difficult to predict short-term passenger flow exactly and reliably applying a usual forecasting method indicating that the information of passenger flow cannot be effectively applied to real-time scheduling.

At present, there is no full insight into short-term passenger flow prediction, however, there are some prediction methods used in the areas such as traffic flow prediction, passenger flow prediction etc. that apply new ideas for short-term passenger flow prediction. Chen and Grant-Muller (2001) used a sequential learning method for forecasting short-term traffic flow. Dia (2001) advanced a traffic flow forecasting method on Neural Networks used in a certain highway. Then, Vlahogianni *et al*. (2005) modified a genetic algorithm by adopting a multilayer structure strategy, considering the randomicity of traffic. De Gooijer and Hyndman (2006) studied a method for time series prediction and concluded that the method was needed to be proved in some aspects. Hamzaçebi (2008) studied the use of the modified genetic algorithm on a time series prediction method. Tsai *et al*. (2009) studied the use of MUTNN Multiple Temporal Units Neural Network and PENN Parallel Ensemble Neural Network for passenger flow prediction on the railway and considered these two methods to be more exact than the conventional MLPNN (Multi-Layer Perception Neural Network) one. Castro-Neto *et al*. (2009) advanced a new method for predicting short-term road traffic flow using a support vector machine.

2. Summary of Least Squares Support Vector Machine

Support Vector Machine (SVM) is a new pattern recognition method developed in recent years and is based on a statistical learning theory. It has a successful application in pattern recognition, function regression and function approximation. Compared with the neural network algorithm, this method settles an excessive learning problem and a local-minimizer achieving problem. Therefore, the complexity of the SVM algorithm depends on the number of sample data, when quadratic programs are more complicated and calculation speed is slower.

Suykens and Vandewalle (1999), and Suykens *et al.* (2000) put forward Least Squares Support Vector Machine (LS-SVM). The main difference between LS-SVM and standard SVM is that the sum of the squares of errors is added to objective functions and unequal restrictions are changed into equal restrictions. In this way, the speed of the solution procedure is noticeably accelerated.

Here is a training sample set (x_i, y_i), $x_i \in R^n$, $y_i \in R, i = 1,2,\cdots,l$. Mapping the sample set from ingoing space to high dimension characteristic space by non-linear mapping ϕ, the nonlinear fitting problem is shifted to the linear fitting problem and the linear equation is as follows:

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

where: w and $\phi(x)$ are n-dimensional vectors and b is threshold.

Based on the principle of structural risk minimization, the regression problem is expressed as a problem of constrained optimization:

$$\begin{aligned} \min_{w,b,e} \quad & J(w,\xi) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{l} \xi^2{}_i; \\ s.t. \quad & y_i = w^T \phi(x) + b + \xi_i. \end{aligned} \tag{2}$$

Parameter ξ_i represents error and $\gamma > 0$ represents penalty term ($\gamma > 0$).

To solve the above optimization problem, the Lagrange function is built as:

$$\begin{aligned} & L(w,b,\xi,\alpha) = J(w,\xi) - \\ & \sum_{i=1}^{l} \alpha_i [w^T \phi(x) + b + \xi_i - y_i]. \end{aligned} \tag{3}$$

Based on conditions $\partial_L / \partial_w = 0$, $\partial_L / \partial_b = 0$, $\partial_L / \partial_\xi = 0$, $\partial_L / \partial_\alpha = 0$, we can deduce the following function:

$$\begin{aligned} & w = \sum_{i=1}^{n} \alpha_i \phi(x_i), \sum_{i=1}^{n} \alpha_i = 0, \alpha_i = \\ & \gamma \xi_i, \phi(x_i) w + b + \xi_i - y_i = 0. \end{aligned} \tag{4}$$

Eliminating parameters ω and ξ, we get a system of linear equations:

$$\begin{bmatrix} 0 & e^T \\ e & GG^T + \gamma^{-1} \end{bmatrix}_{I(l+1)\times(l+1)} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \tag{5}$$

where: e is vector (l×1) and each element is 1. I is identity matrix (l×1), α and y are kernel parameter matrixes ($\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)^T$, $y = (y_1, y_2, \cdots, y_n)^T$, $G = (\phi(x_1)^T, \phi(x_2)^T, \cdots, \phi(x_l)^T)^T$.

According to Mercer's condition, we define nuclear function $K(x_i, x_j) = \phi(x_i) \times \phi(x_j)$:

$$K(x_i, x_j) = \phi(x_i) \times \phi(x_j)\,.$$

Then, the linear decision function is presented below:

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b\,. \tag{6}$$

3. Parameter Optimization of LS-SVM Based on the Genetic Algorithm

3.1. Evaluation Indicators

The value of penalty parameter (γ) and kernel parameter (σ) in LS-SVM is related to its generalization ability. In actual application, the grid method is a common method to select parameters. It is a direct and simple method to research parameters. Nevertheless, when there are too many parameters or search range is inappropriate and step length is uncertainty, an inaccurate result will be obtained and search time becomes very long. Under such consideration, a new evaluation indicator as a new fitness function in the genetic algorithm is proposed in this paper.

Here, the RMSE of the predicted value and actual value is regarded as the evaluation indicator:

$$F = -\sqrt{(1/N) \sum_{i=1}^{N} (y_i - \overline{y}_i)}\,, \tag{7}$$

where: F is the evaluation indicator; N is the number of the training sample; y_i is the actual value; $\overline{y}_i$ is the predicted value.

The effect of the training sample on the parameters is considered, but the effect of the verification sample is not taken into account. Thus, the evaluation indicator (U) of training performance and the evaluation indicator (V) of testing performance are put forward.

Sampling set T is decomposed into P disjoint subsets T_i and each subset contains m samples. That is $T_i = \{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \cdots, (x_{im}, y_{im})\}, i = 1,2,\cdots,p$. Each time, a subset is selected as a testing sample set and other subsets ($p-1$) are regarded as training sample sets. Suppose that a testing subset is T_i in i experiment, the evaluation indicator (U_i) of training performance and the evaluation indicator (V_i) of testing performance are defined as follows:

$$U_i = \sqrt{\frac{1}{n-m} \sum_{\substack{k=1 \\ k\neq i}}^{p} \sum_{j=1}^{m} (y_{kj} - f(x_{kj}))^2}\,, \; i = 1,2,\cdots p\,; \tag{8}$$

$$V_i = \sqrt{\frac{1}{m} \sum_{j=1}^{m} (y_{ij} - f(x_{ij}))^2}\,, i = 1,2,\cdots p\,. \tag{9}$$

Now, we define λ as the equilibrium coefficient of U_i and V_i, then we get the evaluation indicator (F_i) of the whole performance:

$$\begin{aligned} & F_i = \lambda U_i + (1-\lambda) V_i = \\ & \lambda \sqrt{\frac{1}{n-m} \sum_{\substack{k=1 \\ k\neq i}}^{p} \sum_{j=1}^{m} (y_{kj} - f(x_{kj}))^2} + \\ & (1-\lambda) \sqrt{\frac{1}{m} \sum_{j=1}^{m} (y_{ij} - f(x_{ij}))^2}\,, \; i = 1,2,\cdots p. \end{aligned} \tag{10}$$

According to the above statement, an optimization problem is described as:

$$\min_{\gamma,\sigma} \quad F = (F_1 + F_2 + \cdots + F_p) / p. \tag{11}$$

3.2. Design of the Genetic Algorithm

The optimal solution to the problem is the optimum value of γ and σ. The genetic algorithm is designed in the paper to solute the optimization of parameters in LS-SVM. The algorithm includes several parts as follows:

1. ***Coding***: In this solution, a real-coded schema is used to avoid affecting the performance of the evolutionary algorithm and precise calculation because of repeated decoding, coding operation and the length of the binary character string in the binary code.
2. ***Original colony initialization:*** Parameters γ and σ must be greater than zero in the theory, and therefore population size is proposed as M=80.
3. ***Adaptive function***: In this research, the adaptive function is defined as formula 10.
4. ***Selection method:*** Roulette is used in the algorithm, in which the selection probability P_i of unit i is described as:

$$P_i = f_i / \sum_{k=1}^{M} f_k, \tag{12}$$

where: f_i is the fitness of unit i; M is population size.

5. ***Crossover and variation:*** When the numbers of iteration are small, a higher crossover and mutation rates are used to enlarge the scope of searching and keep population diversity effectively. In a contrary manner, when the numbers of iteration are large, a lower crossover and mutation rates are suggested to avoid good individuals being destroyed. Some improvement is put forward on crossover and mutation rates:

$$\begin{cases} Dc \le 0.1 \times Nc, \ H_c = 0.9, H_m = 0.05 \\ 0.1 \times Nc < Dc \le 0.9 \times Nc, \ H_c = 0.7, H_m = 0.01 \\ 0.9 \times Nc < Dc \le Nc, \ H_c = 0.5, H_m = 0.001 \end{cases} \tag{13}$$

where: Dc presents iteration times; Nc presents maximum iteration times; Hc shows crossover rates and Hm presents mutation rates.

6. ***Termination criteria:*** In general, searching ends as the value of iteration times is over default limit or the best individual is found according to testing performance. In this research, Nc =500 is used.

The procedure of LS-SVM parameter optimization based on the genetic algorithm is shown in Fig. 1.

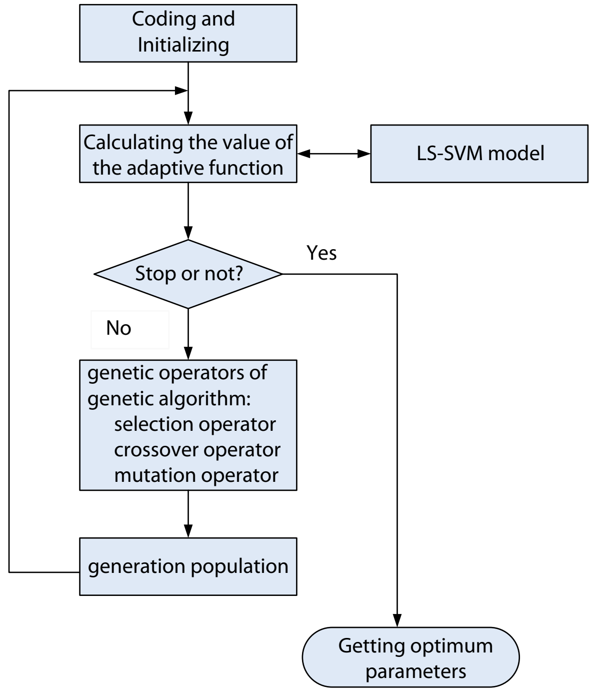

Fig. 1. The process of LS-SVM parameter optimization based on the genetic algorithm

4. LS-SVM Model for Passenger Flow Prediction

4.1. Analysis of Influencing Factors on Passenger Flow

Short-term passenger flow varies randomly because of the weather effects, due to competitive routes etc. It also has a certain law varying at a cycle of seven days.

For example, passenger flow on Monday is relevant to that on Mondays of m weeks before. There is a connection between passenger flow for the day and that in the past few days. Taking it into account the received information, passenger flow can be predicated according to the sequence of passenger flow in n days before. Now we can define that the time unit for short-term prediction is T_0 (minute). If the operation time of a certain bus is T (minute), then it can be divided into n_0 observation time units (n_0=T/T_0). Suppose that passenger flow in a certain time unit is closely related to the value in s time units before, then passenger flow F_d^t is closely related to those, such as $F_{d-7*m}^t, \cdots, F_{d-14}^t, F_{d-7}^t, F_{d-n}^t, \ldots, F_{d-2}^t, F_{d-1}^t$, $F_d^{t-s}, \ldots, F_d^{t-2}, F_d^{t-1}$, among which d represents date and t represents time unit. When T_0=60(min), the corresponding prediction value is passenger flow in an hour.

4.2. Building a Prediction Model

According to the above analysis, passenger flow in a certain time of some day is relevant to that in s time units before the same day, to that in the same time in n days before and also to that in the same time in the same work days in m weeks before.

If F_d^t represents passenger flow in time t in some day d, passenger flow in the whole time units of the day is expressed as: $F_d = [F_d^1, F_d^2, \cdots, F_d^{n_0}]$.

y_i is defined as passenger flow in time $s+i$ in some day and impact factor of y_i is expressed as:

$$x_i = \Big[F_{d-7m}^{s+i}, \cdots, F_{d-14}^{s+i}, F_{d-7}^{s+i}, F_{d-n}^{s+i}, \cdots,$$
$$F_{d-1}^{s+i}, F_d^i, F_d^{i+1}, \cdots, F_d^{s+i-1} \Big], \ i = 1, 2, \cdots, l.$$

Then, we can get the training set $\{(x_1, y_1), \cdots, (x_i, y_i), \cdots, (x_l, y_l)\}$, in which l is the number of samples. It can be shown as follows:

$$x=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_l \end{bmatrix}=$$

$$\begin{bmatrix} F_{d-7m}^{s+1} \cdots F_{d-14}^{s+1} & F_{d-7}^{s+1} & F_{d-n}^{s+1} \cdots F_{d-1}^{s+1} & F_d^1 & F_d^2 & \cdots & F_d^s \\ F_{d-7m}^{s+2} \cdots F_{d-14}^{s+2} & F_{d-7}^{s+2} & F_{d-n}^{s+2} \cdots F_{d-1}^{s+2} & F_d^2 & F_d^3 & \cdots & F_d^{s+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{d-7m}^{s+l} \cdots F_{d-14}^{s+l} & F_{d-7}^{s+l} & F_{d-n}^{s+l} \cdots F_{d-1}^{s+l} & F_d^l & F_d^{l+1} & \cdots & F_d^{s+l-1} \end{bmatrix}.$$

The passenger flow is expressed as:

$$y=\begin{bmatrix} y_1 & y_2 & \cdots & y_l \end{bmatrix}^T.$$

Put the training set into LS-SVM and compute Lagrange multiplier α and threshold b. Then, we get linear decision function $f(b)$ (formula (6)).

Using constructive impact factors as input variables and putting them into decision function $f(x)$ we can get the prediction value of passenger flow.

4.3. Example

To test the efficiency of the method, it was realized with the help of the GA Toolbox of MATLAB. Research referred to the data applied by Yang (2008) that represented passenger flow on Route 6 in Changchun, at the peak time on May 8–10, 2007. The basic data is shown in Table 1 and Table 2.

The procedure of short-term passenger flow prediction is as follows:

1. First, passenger flow between 8:20 and 8:30 on May 10 was regarded as a testing set, whereas the other was regarded as a training set. Because of limited data, passenger flow was predicted according to passenger flow in the same time set two days before and that in two time sets in the same day. Namely, F_{d-2}^t, F_{d-1}^t, F_d^{t-3}, F_d^{t-2}, F_d^{t-1} were used to predict F_d^t.
2. The equilibrium coefficient $\lambda=0$ was initialized as $\lambda=0$.
3. Parameters γ and σ were optimized applying the LS-SVM method based on the genetic algorithm shown as Fig. 1.

Table 1. The actual value of passenger flow on Route 6 in Changchun at the peak time on May 8, 2007 (certain direction)

Time	The number of bus stops						
	1	2	3	4	5	6	7
6:30–6:40	39	17	15	29	15	13	0
6:40–6:50	26	28	25	41	21	8	1
6:50–7:00	23	12	24	15	25	6	4
7:00–7:10	23	30	5	20	35	15	1
7:10–7:20	6	17	25	42	34	8	0
7:20–7:30	44	20	10	8	11	13	1
7:30–7:40	16	18	16	21	0	1	2
7:40–7:50	5	9	25	17	13	1	1
7:50–8:00	14	0	0	6	11	9	0
8:00–8:10	9	6	6	4	12	12	11
8:10–8:20	5	0	12	14	16	17	26
8:20–8:30	6	0	5	3	3	5	12

Table 2. The actual value of passenger flow on Route 6 in Changchun at the peak time on May 9, 2007 (a certain direction)

Time	The number of bus stops						
	1	2	3	4	5	6	7
6:30–6:40	36	16	16	30	16	12	1
6:40–6:50	25	30	25	38	22	9	0
6:50–7:00	23	12	25	16	22	7	5
7:00–7:10	22	28	6	22	30	14	2
7:10–7:20	6	19	22	40	35	9	0
7:20–7:30	40	25	11	7	10	12	2
7:30–7:40	14	22	15	20	1	2	2
7:40–7:50	6	8	24	18	14	1	1
7:50–8:00	13	1	0	7	12	8	0
8:00–8:10	10	7	7	5	11	11	13
8:10–8:20	6	0	11	11	15	16	24
8:20–8:30	6	1	5	4	3	4	13

4. Having applied the trained LS-SVM, passenger flow between 7:00-8:30 on May 10 was predicted and normalized.
5. Searching ended as $\lambda = 1$, otherwise $\lambda = \lambda + 0.1$ return to step (3).
6. By comparing different indicator F at different equilibrium coefficient $\lambda = 0$, the corresponding equilibrium coefficient $\lambda = 0$ to minimal F was found out as an optimal equilibrium coefficient. Accordingly, the optimal values of parameter γ , σ and predicting passenger flow were obtained.

Considering the above iteration process, the best testing performance was discovered when $\lambda = 0.1$ and $\gamma = 127.39, \sigma^2 = 238.69$ were the optimal value. By inserting γ and σ in LS-SVM, predicting passenger flow was got, shown as Table 3, Fig. 2 and Fig. 3.

There is an evident difference between the predicted and observed (8:20–8:30) passenger flow at bus stop 5. It is relevant for the following reasons:

1. There was less data on training and the performance of passenger varying at a cycle of seven days was not considered in the prediction.
2. The randomness and complexity of passenger flow is an important reason for deviation. An equal coefficient was put forward:

$$EC_t = 1 - \frac{\sqrt{\sum_{k=1}^{ZS} (T_t(k) - P_t(k))^2}}{\sqrt{\sum_{k=1}^{ZS} T_t(k)^2} + \sqrt{\sum_{k=1}^{ZS} P_t(k)^2}}, \qquad (14)$$

where: $T_t(k)$ presents actual passenger flow on board at bus stop k , $P_t(k)$ presents predicted passenger flow on board at bus stop k and ZS presents the whole number of bus stops along the route. The coefficient presents the fitness of the model, which is between 0 and 1.The larger is the coefficient, the better is the result.

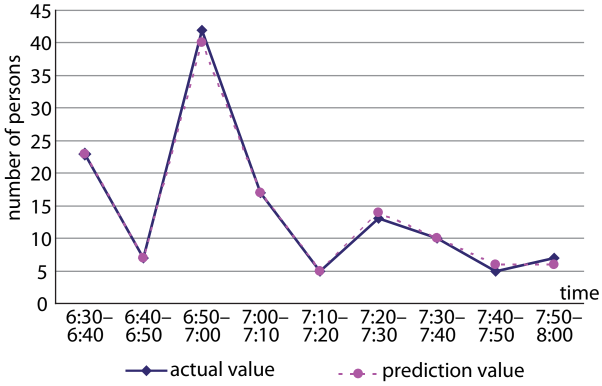

Fig. 2. The actual value and the prediction value of passengers boarding on the bus from 7:00 to 8:30 at bus stop 1 on Route 6 in Changchun

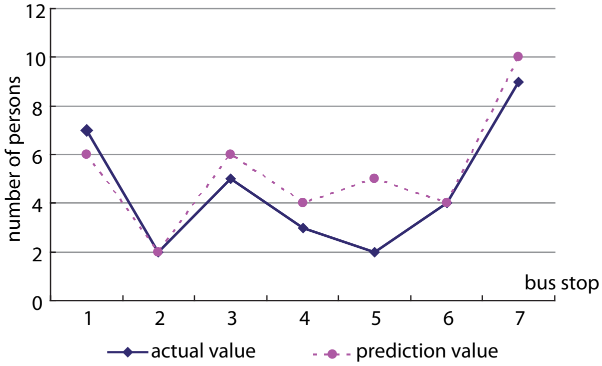

Fig. 3. The actual value and the prediction value of passengers boarding on the bus from 8:20 to 8:30 at each bus stop on Route 6 in Changchun

There is little difference between the actual value and prediction. The equal coefficient of the training set is larger than 0.94 and the equal coefficient of prediction (8:20–8:30) is 0.876 (Table 4), the accuracy of which meets the requirements of an actual application.

Table 3. The actual value and the prediction value of passenger flow on Route 6 in Changchun at the peak time on May 10, 2007 (a certain direction)

Time/	The number of bus stops													
	1		2		3		4		5		6		7	
	A.V.	P.V.	A.V.	P.V.	A.V.	P.V.	A.V.	P.V.	A.V.	P.V.	A.V.	P.V.	A.V.	P.V.
6:30–6:40	37	–	15	–	18	–	25	–	13	–	13	–	1	–
6:40–6:50	25	–	27	–	22	–	40	–	20	–	9	–	1	–
6:50–7:00	23	–	13	–	24	–	16	–	22	–	7	–	5	–
7:00–7:10	23	23	29	27	6	5	20	21	36	32	13	12	1	2
7:10–7:20	7	7	19	18	20	21	42	40	32	33	7	8	0	0
7:20–7:30	42	40	21	21	11	12	8	8	12	12	12	11	2	3
7:30–7:40	17	17	20	20	14	14	21	20	1	2	2	2	2	3
7:40–7:50	5	5	7	8	27	25	17	18	13	12	1	2	1	1
7:50–8:00	13	14	0	1	0	1	6	6	12	12	9	9	0	0
8:00–8:10	10	10	7	6	5	6	4	6	14	14	16	15	15	13
8:10–8:20	5	6	1	1	15	15	14	13	14	16	17	17	29	28
8:20–8:30	7	6	2	2	5	6	3	4	2	5	4	4	9	10

(Remarks: A.V. – actual value; P.V. – prediction value)

Table 4. Equal coefficients (EC) in different time sets

Time set	EC
7:00–7:10	0.956
7:10–7:20	0.977
7:20–7:30	0.974
7:30–7:40	0.976
7:40–7:50	0.960
7:50–8:00	0.959
8:00–8:10	0.943
8:10–8:20	0.969
8:20–8:30	0.876

5. Conclusions

The paper presents a new transit flow prediction model based on Least Squares Support Vector Machine (LS-SVM) according to non-linear, stochastic and complex flow characteristics. The improved Genetic Algorithm (GA) is used for optimizing the penalty parameter and nuclear parameter. An actual example showed the validity of the approach.

Acknowledgements

This work has been supported by the National High Technology Research and Development Program of China under Grant No. 2007AA11Z210, the Nature Science Foundation of China under Grant No. 50978057 and No. 61074141, the National Eleventh Five-Year Science and Technology Supporting Plan under Grant No. 2009BAG13A06, and the Science and technology fund of Southeast University under Grant No. KJ2010432.

References

Castro-Neto, M.; Jeong, Y.-S.; Jeong, M.-K.; Han, L. D. 2009. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions, *Expert Systems with Applications* 36(3): 6164–6173. doi:10.1016/j.eswa.2008.07.069

Chen, H.; Grant-Muller, S. 2001. Use of sequential learning for short-term traffic flow forecasting, *Transportation Research Part C: Emerging Technologies* 9(5): 319–336. doi:10.1016/S0968-090X(00)00039-5

Dia, H. 2001. An object-oriented neural network approach to short-term traffic forecasting, *European Journal of Operation Research* 131(2): 253–261. doi:10.1016/S0377-2217(00)00125-9

De Gooijer, J. G.; Hyndman, R. J. 2006. 25 years of time series forecasting, *International Journal of Forecasting* 22(3): 443–473. doi:10.1016/j.ijforecast.2006.01.001

Hamzaçebi, C. 2008. Improving artificial neural networks' performance in seasonal time series forecasting, *Information Sciences* 178(23): 4550–4559. doi:10.1016/j.ins.2008.07.024

Suykens, J. A. K.; Lukas, L.; Vandewalle, J. 2000. Sparse approximation using least squares support vector machine, in *Proceedings of the IEEE 2000 International Symposium on Circuits and Systems*, 28–31 May 2000, Geneva, Switzerland. Vol. 2, 757–760. doi:10.1109/ISCAS.2000.856439

Suykens, J. A. K.; Vandewalle, J. 1999. Least squares support vector machine classifiers, *Neural Processing Letters* 9(3): 293–300. doi:10.1023/A:1018628609742

Tsai, T.-H.; Lee, C.-K.; Wie, C.-H. 2009. Neural network based temporal feature models for short-term railway passenger demand forecasting, *Expert Systems with Applications* 36(2): 3728–3736. doi:10.1016/j.eswa.2008.02.071

Vlahogianni, E. I.; Karlaftis, M. G.; Golias, J. C. 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach, *Transportation Research Part C: Emerging Technologies* 13(3): 211–234. doi:10.1016/j.trc.2005.04.007

Yang, X. 2008. *Research on public Transport Scheduling Optimization Based on the Information of Quasi-real-time Flow*. A Thesis Submitted to Jilin University For the Academic Degree of Master of Engineering, 43–46 (in Chinese).