PAPER • **OPEN ACCESS**

Traffic flow prediction based on optimized hidden Markov model

To cite this article: Shu-xu Zhao *et al* 2019 *J. Phys.: Conf. Ser.* **1168** 052001

IOP Conf. Series: Journal of Physics: Conf. Series **1168** (2019) 052001 doi:10.1088/1742-6596/1168/5/052001

Traffic flow prediction based on optimized hidden Markov model

ZHAO Shu-xu, WU Hong-wei[a] and LIU Chang-rong

School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

[a] Corresponding author: 2551612565@ qq.com

Abstract. In order to alleviate the current urban traffic congestion pressure and provide accurate and reliable traffic condition information, it is difficult to select the optimal state parameters for the original Hidden Markov model (HMM) and the state number redundancy determined during the training process leads to the model over-provisioning. The problem of weak is integration and generalization. An improved Hidden Markov Model for traffic flow prediction is proposed to more effectively fit the actual urban road intersection traffic flow. In the calculation of the negative log-likelihood function, an Akaike information criterion (AIC) or a Bayesian information criterion (BIC) penalty term is added, and the Baum-Welch algorithm is combined to optimize the optimal state number of the model. Experiments are carried out based on the collected real traffic flow and GPS feature data. The results show that the optimized hidden Markov model is superior to the original model in the accuracy and generalization ability of traffic flow prediction.

1.Introduction

With the rapid development of China urbanization and economy, the number of motor vehicles per capita has grown rapidly. Traffic congestion currently has become one problem to be solved urgently in various cities, bringing about issues of time waste, operation cost raise, traffic accidents increase, aggravation of air and noise pollution, etc. [1] In the case of making full use of existing road resources and facilities, the rational distribution of regional road traffic flow and accurate identification of traffic congestion have become important parts of the intelligent transportation system. By collecting real-time feature data of urban regional intersection, mining the data of the potential law in the transportation system, rational analysis of traffic congestion situation has a positive effect on alleviating the traffic load at intersections, which can be used for bus density adjustment, traffic information release by traffic control department, research and development of vehicle navigation system, etc.

In recent years, domestic and foreign experts and scholars have conducted extensive studies on traffic flow prediction and put forward many practical models and algorithms[2]. They are mainly divided into two categories: a traditional traffic flow prediction model based on mathematical statistics and physical methods, such as time series model[2][19], Kalman filter model[2], parametric regression model[2], etc. Another prediction model is based on machine learning research techniques represented by neural networks, such as SVM support vector machine[6-10], Logistic regression[2], LSTM-RNN model[5], etc. A large number of actual traffic surveys

show that road traffic flows are characterized by randomness, complexity and uncertainty, with a high fluctuation. Traditional models and combined methods cannot efficiently reflect this traffic flow nonlinear spatial structure. Although the neural network structure effectively avoids the above disadvantages and improves the traffic flow prediction accuracy to some extent, the neural network structure has too many parameters and a complicated training process, which leads to a defect of too high calculation cost. In order to balance the advantages and disadvantages of various models, their advantages are given fully to play.

Considering the similarity between stock price and traffic flow data feature, Hidden Markov Model (HMM) is a probability framework structure based on time series, achieving good results in stock forecasting[12-15], phonetic recognition, and DNA sequence detection. Among them, in the field of stock forecasting market, Hassan and Nath[13] utilized a fixed state to predict some aviation stocks through similar patterns in historical data, and Aditya Bhuwan[15] also based the possibility of MAP maximization observation state sequence of the fixed-state HMM. Based on the previous work of Hassan and Nath, Nguyet Nguyen used the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to test the performance of the model in terms of the number of states.

For the purpose to further improve other methods and hidden Markov models in their deficiencies of predicting traffic flow, by adding a Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) penalty term to the negative log-likelihood loss function, The Baum-Welch algorithm was fully utilized to optimize the optimal state number and parameters of the model. According to the comparison of performance indicators before and after improvement, the optimized HMM model has better generalization ability of traffic flow prediction, and higher practical applicability.

2.HMM Theory

2.1 HMM Construction

HMM is a mathematical statistical model that can be used to mark problems, describing the process of randomly generating observation state sequences from hidden Markov chains, of which essence is a sequential probability generation model. At present, due to the rapid rise of big data and machine learning, HMM serves as a very mature classifier in matching state sequences, consistent with randomness, uncertainty and complexity of traffic flow in the actual transportation system.

In the actual urban road, the traffic flow between adjacent intersections is diffused in time series in the form of traffic waves from the upstream direction to the downstream, hence resulting in regional road intersections present different traffic states, which are transferred to each other with time and generate visible traffic flows in a certain period. In this paper, the data of the four entrances at the east, west, south and north of the intersection were collected, modeled by HMM, and finally the total traffic flow at the entrances of the actual road intersection was fitted.

Considering that the current traffic state is only related to the state at the previous moment, and that the observation at any time only depends on the Markov chain state at that moment, it is conformed to HMM and Markov hypotheses and observational independence hypotheses, providing a reliable theoretical basis for traffic flow prediction.

2.1.1 HMM basic parameters

The hidden Markov traffic model is determined by the initial probability distribution π, the state transition probability matrixes *A* and *B*. HMM is defined as follows:

Let π be the set of all possible traffic states, and V be the set of all possible observations.

$Q=\{q_1,q_2,\cdots q_N\}$ $\quad$ $V=\{v_1,v_2,\cdots v_M\}$

where N is the number of possible traffic states, M is the number of possible observations, I is the traffic sequence at a hidden state with length of T , O is the corresponding feature observation sequence. $I=\{i_1,i_2,\cdots i_T\}$ $\quad$ $O=\{o_1,o_2,\cdots o_T\}$

A is state transition probability matrix: $A=[a_{ij}]_{N\times N}$, where

$$a_{ij}=P(i_{t+1}=q_j\mid i_t=q_i),\ \ i,j=1,2,\ldots N$$

B is the observation probability matrix : $B = [b_j(k)]_{N\times M}$,where

$$b_j(k) = P(o_t = v_k \mid i_t = q_j), k = 1,2...M; j = 1,2...N$$

π is initial state probability vector: $\pi = (\pi_i)$, where $\pi_i = P(i_1 = q_i)$, $i = 1,2,...N$

So the hidden Markov traffic model λ can be represented by these three symbols: $\lambda = (A, B, \pi)$, the three parameters satisfying the following conditions that the probabilities sum of each one is 1, not less than 0.

2.1.2 Three basic matters of HMM

(1)Probability calculation matter. Given the model λ and the feature data of the observation sequence region, the probability of the observation sequence O occurring under model λ is calculated.

(2)Training matter. Knowing observation sequence O and the estimation model parameters $\lambda = (A, B, \pi)$, the parameters are estimated by maximum likelihood estimation (MLE) method to achieve the maximum probability of the feature data observation sequence, and the acquired four-dimensional intersection data are used to train the HMM.

(3)Prediction matter. Namely decoding matter. According to the trained HMM, the feature data of the observation state collected in real time are utilized to predict the state at the next moment, and the state sequence with the largest conditional probability $P(I \mid O)$ is obtained.

2.1.3 Time series analysis of traffic flow based on HMM

HMM exhibits good results in the analysis of non-stationary and obviously fluctuating feature data systems, and the actual road traffic system is also a complex and non-stationary system. Essentially, traffic flow as an observation sequence is continuous numerical data. Considering that is the four-dimensional vector traffic flow at four entrances of an intersection, while is the traffic state at the intersection in a certain period, the state of is one of all hypothetical states. Fig. 1 presents the stochastic process of classic HMM.

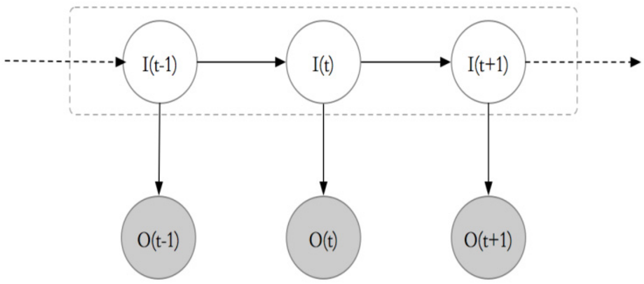

Fig.1. Process of classic HMM

Since the vector takes a real value, the observation can be modeled as a multivariate Gaussian distribution. The observation state is independent, relying on whether the observations are correlative or not, while HMM is a finite-state machine, so I_t can only take discrete values. The solution of the second matter is carried out by the Baum-Welch algorithm.

2.2 Prediction of HMM traffic flow

Prediction of the traffic flow at the intersection in the next period is mainly to calculate the maximum likelihood values of the traffic flow of the first N observation states, and then to compare the maximum likelihood values of the observation state data in all the same data dimensions along a certain time window. Finally, the maximum maximal likelihood value of the observation sequence segment in the next time period is calculated as the predicted traffic flow probability. The formula is as follows:

$$j = argmin_i(| P(O_t, O_{t-1}, O_{t-2}, ..., O_{t-n} | \lambda) - P(O_{t-i}, O_{t-i-1}, O_{t-i-2}, ..., O_{t-i-k} | \lambda) |) \quad i = 1, 2, ..., T/N \tag{2-1}$$

In order to ensure the continuity of the traffic flow value at a certain moment, we must calculate the change of the traffic flow from the previous time period to the next time period, and then add this traffic volume change into the traffic volume in the current time period to predict the traffic volume in the next time period. The formula is as follows:

$$O_{t+1} = O_t + (O_{t-j+1} - O_{t-j}) \tag{2-2}$$

Subsequently, after obtaining the results of the real traffic observation state, we add them to the training data set, and at the meanwhile, adjust the HMM parameters through continuous training iteration until achieve stability. That means, after determining the size of the observation sequence, a subsequence of a set of stable model parameters is trained in the training set, regarded as a subsequence model parameter of the prediction model.

2.2.1 Training of optimized HMM traffic models

In order to select the optimal traffic states number, in general we will select different states numbers in the state space Q to train a series of parameter models, and then calculate the negative logarithmic maximum likelihood of the training set data, selecting the corresponding parameter whose derivative extreme point is zero as its model. However, the selection of increasing state number makes the model structure more complicated, which leads to the model overfitting. To avoiding this problem, we added a penalty term to the maximum likelihood function of the models. In practice, two indicator criteria are selected to correct the model state parameters, namely Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Both of them are based on the optimal likelihood function value and related parameters to select the model, and in the calculation the introduction of hidden states number of BIC training samples is superior to that of AIC. Therefore, this paper chooses BIC as the performance metric of the models to select the optimal state number of the function. The formulas are as follows:

$$BIC = -2log(P(O_{train|\lambda})) - p.log(T) \tag{2-3}$$

$$AIC = -2log(P(O_{train|\lambda})) - 2p \tag{2-4}$$

$$L(O, I | \lambda) = -logP(O, I | \lambda_k) + BIC \quad k = 0,1,2...N \tag{2-5}$$ where $p = N^2 + 2N - 1$.

(1) Baum-Welch algorithm training process based on optimized HMM

For the collected traffic flow and GPS observation data in Linzi district of Zibo City, Shandong Province, without corresponding state sequences, the goal is to learn the HMM parameters $\lambda = (A, B, \pi)$ during the training process and to continuously iterate over the corresponding parameters for the estimation of the maximal value. We now treat the observation sequence data as observation data O, and the state sequence data as unobservable hidden data I, and then the probability traffic model of the hidden variable is shown as below:

$$P(O | \lambda) = \sum_I P(O | I, \lambda) P(I | \lambda) \tag{2-6}$$

Thereinto, its parameter learning can be realized by the EM algorithm. Firstly, the log likelihood function of the HMM traffic model in the regional feature data is determined, and the complete historical feature data $(O, I) = (O_1, O_2 \cdots O_T, i_1, i_2 \cdots i_T)$ are collected by the regions. Then the step E of the EM algorithm is used to solve the function Q, with the corresponding formula as below:

$$Q(\lambda, \bar{\lambda}) = \sum_I l\,og P(O, I | \lambda) P(O, I | \bar{\lambda}) \tag{2-7}$$

Finally, the step M of the EM algorithm is to maximize the function Q to find the model parameters A, B, π. Lagrangian multiplication is expanded to derivation, obtaining the HMM parameters as follows:

$$\pi_i = \frac{P(O, i_1 = i \mid \bar{\lambda})}{P(O \mid \bar{\lambda})} \tag{2-8}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j \mid \overline{\lambda})}{\sum_{t=1}^{T-1} P(O, i_t = i \mid \overline{\lambda})} \tag{2-9}$$

$$b_j(k) = \frac{\sum_{t=1}^{T} P(O, i_t = j \mid \overline{\lambda}) I(o_t = v_k)}{\sum_{t=1}^{T} P(O, i_t = j \mid \overline{\lambda})} \tag{2-10}$$

where the sum of the corresponding parameter coefficients is 1, satisfying the probability condition. The above formulas are implemented according to the Baum-Welch algorithm, which is a specific application of the EM algorithm in HMM Learning, proposed by Baum and Welch. The specific pseudo codes of the algorithm are presented as below:

Baum-Welch algorithm

Data: $o_{1:T}$, initial parameters $\lambda_0 = (A_0, B_0, \pi_0)$

***Result*:** Final model parameters $\lambda = (A, B, \pi)$

Repeat

// E-step

$\alpha, \beta \leftarrow$ Forward and backward feedback $(o_{1:T}, \lambda)$

$\tau_t(i) \leftarrow p(i_t = q_j \mid o_{1:T}; \lambda) = \frac{1}{i} \alpha_t(i) \beta_t(i)$

$\tau_t(i, j) \leftarrow p(i_{t-1} = i, i_t = j \mid o_{1:T}; \lambda)$

$= \frac{1}{i} \alpha_{t-1}(i) A_{ij} p(o_t \mid i_t = j; \lambda_j) \beta_t(j)$

// M-step

$\pi_i \leftarrow \tau_1(i)$

$A_{ij} \leftarrow \frac{\sum_{t \geq 2} \tau_t(i, j)}{\sum_{j'} \sum_{t \geq 2} \tau(i, j')}$

$B_i \leftarrow \frac{\sum_{t \geq 1} \tau_t(i) o_i}{\sum_{t \geq 1} \tau_t(i)}$

***Until convergence*;**

3.Instance verification of traffic feature data

3.1 Traffic characteristics data source and evaluation criteria

In order to better verify the performance superiority of the optimized HMM in traffic flow prediction, relying on the study of traffic service computing system for urban growth and the application project, we conducted the simulation comparison between the traffic flow and GPS data collected in Linzi district of Zibo City. According to the data collected in Linzi District of Zibo City, taking advantage of 12600 traffic flows data of Qidu Road and Xinguang Road intersections and the GPS data points of the main roads, we selected 70% of the data to train the optimized model, and verified the generalization ability of the model with the other 25% of the data. The first four lines of the intersection and main road data are shown in Tables 1 and 2 below:

References are cited in the text just by square brackets [1].(If square brackets are not available, slashes may be used instead, e.g./2/.)Two or more references at a time may be put in one set of brackets [3,4].The references are to be numbered in the order in which they are cited in the text and are to be listed at the end of the contribution under a heading *References*, see our example below.

Table 1. Traffic flow of Qidu Road and Xinguang Road entrances (Veh/5 Minutes)

Time	Lane1	Lane2	Lane3	Lane4	Total
0:00	31	41	31	19	122
0:05	36	48	35	14	133
0:10	32	31	27	15	105
0:15	19	26	23	8	76

Table 2. GPS data of main roads in Linzi District (m)

No.	Longitude	Latitude	Name of main road
1	118.344043774	36.8540643831	Qidu Road
2	118.344018285	36.8545040898	Qidu Road
3	118.239692809	36.7391341777	Xinguang Road
4	118.344021067	36.8545162136	Qidu Road

Based on the collected traffic feature data above, in order to fully evaluate their effect on traffic flow prediction, commonly-used evaluation indicators include the following two aspects[21], mean square error (MSE) and mean absolute percentage error (MAPE). The structural formulas of these indicators are shown as below:

$$MSE = \frac{1}{N}(\sum_{i=1}^{N}(y_i - y_i^{\wedge})^2)\times 100\% \qquad (2\text{-}11)$$

$$MAPE = \frac{1}{N}(\sum_{i=1}^{N}\frac{|y_i - y_i^{\wedge}|}{y_i})\times 100\% \qquad (2\text{-}12)$$

$$Accuracy = 1 - \frac{\sum_{Train,Test}(y_i - y_i^{\wedge})^2}{(y_i - y_i^{\wedge})^2} \qquad (2\text{-}13)$$

where y_i represents the original traffic flow data, $y_i^{\wedge}$ represents the predicted traffic flow data. The smaller the value of these indicators, the higher the accuracy, and the stronger the generalization ability of the model.

3.2 Prediction results and analysis of traffic flow data

This study was implemented on the window7 system, with running memory of RAM 8.00GB, in anaconda python 3.6 and Google Colab platform. Our goal is to use the optimized HMM to accurately predict traffic flow on the training set, simultaneously to improve the generalization ability of the model on the test set.

First, the mean square error (MSE) and the mean absolute percentage error (MAPE) were iterated through calculation, and then the HMM with the lowest value of Bayesian criteria was selected to predict the traffic flow, and finally the predicted results of the traffic flow on the corresponding training set and test set were visualized, shown as Fig. 2 and Fig. 3. In order to more visually reflect the effects before and after optimization, we plotted the mean absolute percentage errors (MAPE) in Fig. 4, and the accuracy on the test set and the training set in 5. The figures reflect a good improvement of the optimized HMM effect on the test set.

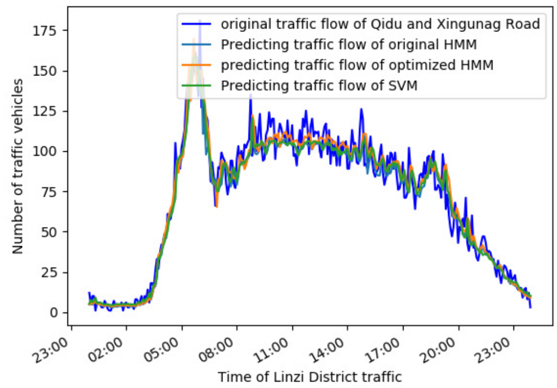

Fig. 2. Traffic flow fitted by training set

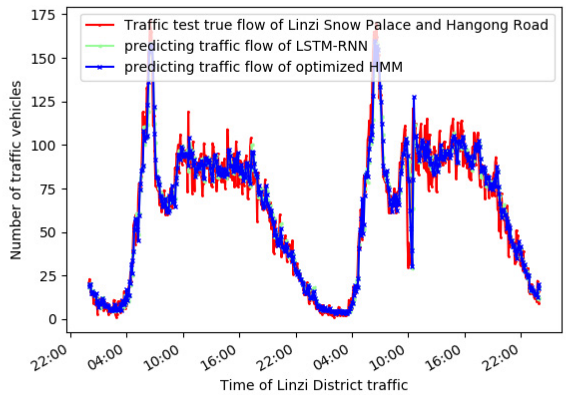

Fig. 3. Traffic flow predicted by training set

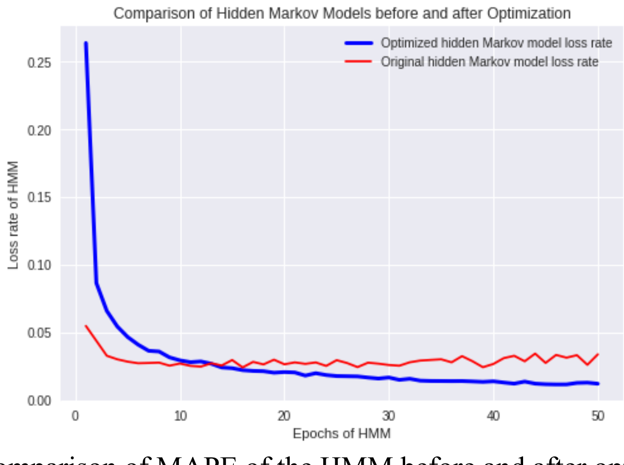

Fig. 4. Comparison of MAPE of the HMM before and after optimization

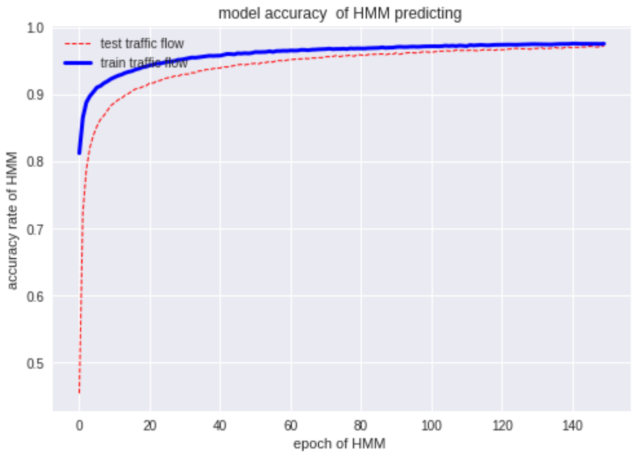

Fig. 5. Accuracy of optimized HMM on the test and the training sets

Through the analysis of Fig. 4 and Fig. 5, it is found that the absolute percentage error loss rate of the model in the traffic flow prediction based on the trained and optimized HMM is significantly superior to that of the original model, and with the increase of the number of iterations, the stability performance is better. At the same time, the fitted accuracies on the test set in the traffic flow prediction are also closely consistent. For the purpose to better illustrate the performance of the predicted traffic flow in the actual main road traffic state, we clustered the test set GPS data collected in one-day period on the main road network[16] , shown as Fig.6. As can be seen, Qidu Road and Xinguang Road are in a smooth state from 10:00 pm to 4:00 am. Consistent with the actually-predicted traffic flow in Figure 4, it basically conforms to the forecasting situation, indicating that the optimized HMM model has realistic significance in the accurate prediction of urban road traffic flow for the traffic management department to coordinate urban traffic management planning.

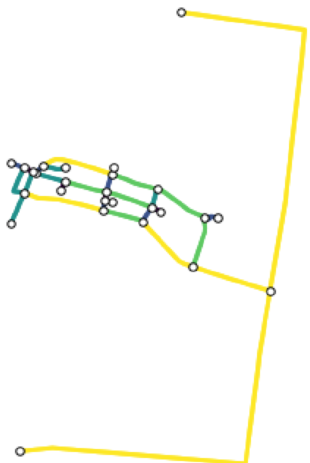

Fig. 6. Clustered traffic state map of the main road network in Linzi district of Zibo City

Finally, we compare the above two evaluation indicators of the support vector machine, LSTM-RNN, and optimized HMM in traffic flow prediction. As shown in Table 3, the comparison of the three models indicates that though they all exhibit good performance in traffic flow prediction, SVM is sensitive to its missing values in predicting traffic flow, while LSTM-RNN is prone to gradient disappearance during training. But the prediction errors of the optimized HMM are relatively smaller, and once the above defects have been made up, its stability is getting better as the number of training iterations increases.

Table 3. Performance indicators analysis of three models in traffic flow prediction

Error indicator	SVM	LSTM-RNN	optimized HMM
MSE	0.0537	0.0593	0.0426
MAPE(%)	0.0087	0.0085	0.00521

4.Summary

This paper based on the road traffic flows collected in Linzi district of Zibo City, Shandong Province and the GPS data of urban main roads, Bayesian Information Criterion (BIC) was served as a penalty item to select the traditional HMM model and the optimal state number, thereby reducing the complexity of the model in predicting time series structure data, and the structural optimization of the original HMM model was achieved.

The experimental results showed that the optimized HMM had a significant improvement in traffic flow prediction performance, and the generalization ability on test set data was obvious. Compared with the predicted traffic flow data, the GPS data of the actual roads were basically consistent with the traffic state of the clustered main roads in a certain period of time,demonstrating the feasibility and practicability of the optimized HMM model in traffic flow prediction.

Acknowledgement

This research was financially supported by the Science and Technology Ministry Foundation and Department of Transportation of Gansu Province (Grant NO. 1504GKCA018;Mission Book Number: 2017-012).

References

[1] Liu Xiao. A literature review of urban traffic congestion research [J]. Economic Research Guide, 2010 (4):102-103.
[2] LIU Jing, GUAN Wei. Overview of Traffic Flow Prediction Methods[J]. Journal of Highway and Transportation Res;earch and Development, 2004, 21(3): 82-85.
[3] Songlin Qiao. Short-Term Traffic Flow Forecast Based on Parallel Long Short-Term Memory Neural Network[A]. The Institute of Electrical and Electronics Engineers、IEEE Beijing Section.Proceedings of 2017 IEEE 8th International Conference on Software Engineering and Service Science[C].The Institute of Electrical and Electronics Engineers IEEE Beijing Section:,2017:5.
[4] Xu D, Shi Y. A combined model of random forest and multilayer perceptron to forecast expressway traffic flow[C]// IEEE International Conference on Electronics Information and Emergency Communication. IEEE, 2017:448-451.
[5] Wang Xiangxue, Xu Lunhui. Short-term traffic flow prediction based on deep learning [J]. Transportation Systems Engineering and Information, 2018, 18(1): 81-88.
[6] Deng Y, Xiang J, Ou Z. SVR with hybrid chaotic genetic algorithm for short-term traffic flow forecasting[C]// Eighth International Conference on Natural Computation. IEEE, 2012:708-712.
[7] Wang H, Wang Y. Vessel traffic flow forecasting with the combined model based on support vector machine[C]// International Conference on Transportation Information and Safety. IEEE, 2015:695-698.
[8] Lou W, Zhou Y, Sheng P, et al. An improved least square support vector regression algorithm for traffic flow forecasting[C]// IEEE, International Conference on Intelligent Transportation Systems. IEEE, 2014:2379-2384.

[9] Yang Y N, Lu H P. Short-Term Traffic Flow Combined Forecasting Model Based on SVM[C]// International Conference on Computational and Information Sciences. IEEE Computer Society, 2010:262-265.

[10] Fu Gui, Guo Qiang, Yan Feng, et al. Short-term traffic flow prediction model based on support vector machine regression[J]. Journal of South China University of Technology(Natural Science Edition), 2013, 41(9): 71-76.

[11] YU Bin, GENG Shan-hua, WANG Ming-hua, et al. Short-term traffic flow prediction model for K-nearest neighbors[J]. Journal of Traffic and Transportation Engineering, 2012, 12(2): 109-115.

[12] Poonam Somani. Stock Market Prediction Using Hidden Markov Model[A]. IEEE Beijing Section.Proceedings of 2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference[C].IEEE Beijing Section:,2014:4.

[13] Hassan M R, Nath B. StockMarket Forecasting Using Hidden Markov Model: A New Approach[C]// International Conference on Intelligent Systems Design and Applications, 2005. Isda '05. Proceedings. IEEE, 2005:192-196.

[14] Nguyen N, Apergis N. Hidden Markov Model for Stock Trading[J]. 2018, 6(2):36.

[15] Gupta A, Dhingra B. Stock market prediction using Hidden Markov Models[C]// Engineering and Systems. IEEE, 2012:1-4.

[16] Constantinescu M M, Constantinescu M P, Vălean H. GPS collars data visualization for wildlife management[C]// System Theory, Control and Computing. IEEE, 2013:146-151.

[17] Koki C, Meligkotsidou L, Vrontos I. Bayesian analysis of predictive Non-Homogeneous hidden Markov models using Polya-Gamma data augmentation[J]. 2018.

[18] Ouyang Yu. Research on vehicle travel time prediction method based on hidden Markov model [D]. Wuhan University, 2015.

[19] Wang J H, Leu J Y. Stock market trend prediction using ARIMA-based neural networks[C]// IEEE International Conference on Neural Networks. IEEE, 1996:2160-2165 vol.4.

[20] Zhang Jinqiu,Zhao Yuxu,Qu Ruitao.Visualization of Traffic Time and Space Data Based on Point and Heat[J]. Journal of Lanzhou Jiaotong University, 2017, 36(3): 63-69.