2012 15th International IEEE Conference on Intelligent Transportation Systems
Anchorage, Alaska, USA, September 16-19, 2012

A *k*-Nearest Neighbor Locally Weighted Regression Method for Short-Term Traffic Flow Forecasting

Shuangshuang Li, Zhen Shen, *Member, IEEE* and Gang Xiong, *Senior Member, IEEE*

***Abstract*—In this paper, a *k*-nearest neighbor locally weighted regression method (*k*-LWR) is proposed to forecast the short-term traffic flow. Inspired by *k*-nearest neighbor (*k*-NN) method, the traffic flows which have the same clock time with the current traffic flow are viewed as neighbors. The traffic flows which have the same clock time with the predicted traffic flow are viewed as the outputs of the neighbors. The neighbors most similar to the current traffic flow are viewed as nearest neighbors. It is observed that each nearest neighbor has different similarity with the current traffic flow, and the similarity is relevant to the contribution of the nearest neighbor's output to predicted traffic flow. The greater the similarity is, the greater the contribution is. These contributions of the nearest neighbors' outputs are obtained by the locally weighted regression (LWR) method. In this way, *k*-LWR uses less data, but uses it more effectively. We use the root mean square error (RMSE) between the actual traffic flow and the predicted traffic flow as the measurement. The proposed method is tested on the actual data from Xingye intersection and Feihu intersection in Jiangsu Province in China. The experimental results show that *k*-LWR has 20% and 24% improvement over the pattern recognition algorithm (PRA), 26% and 30% improvement over *k*-NN, for the two intersections, respectively.**

I. Introduction

The performance of a traffic flow prediction is closely related to the effectiveness of an urban traffic signal control which is an important topic in Intelligent Transportation Systems (ITS)[1][2][3][4]. With a predictive capability, ITS can provide proactive control and management. Otherwise there will be time lag between the collection of traffic data and the implementation of traffic control strategies[5][6]. The traffic control system can make reasonable control strategies in advance according to the predicted traffic flow. So the short-term traffic flow forecasting is a major requirement for providing a dynamic traffic control.

A variety of methods and techniques have been developed to forecast the short-term traffic flow, for example, the historical average method[7], the time-series method[8], the local regression method[9], the Kalman filter method[10], the nonparametric regression method[5], the pattern recognition method[6][11], the neural network method[12] and so on. Most of these methods are data-driven methods. The primary task of the data-driven methods is to determine how to select out useful information from a great amount of traffic flow data for prediction. The useful information refers to the information contained in the historical traffic flows which have relevant contributions to the predicted traffic flow.

Usually the traffic flow is described as a vector with its components being the numbers of vehicles passing in a fixed and given time interval, says 5 minutes. The *k*-NN[5] takes the traffic flows which have the same clock time of the current traffic flow as neighbors. The useful information is found in these neighbors by selecting the nearest neighbors. The predicted traffic flow is obtained by using these nearest neighbors. Actually, the distances, which are the Euclidean distances between the nearest neighbors and the current traffic flow, contain useful information for prediction. It is observed that the smaller the distance is, the greater the contribution is. Some researchers have done some further research based on *k*-NN. Kim *et al.*[11] proposed a pattern recognition algorithm (RPA) by finding out the useful information of different clock time in different days in the historical traffic database. Li *et al.*[6] proposed a weighted pattern recognition algorithm(WPRA) which was an improvement of *k*-NN and PRA by weighting the useful information according to its clock time.

The traffic flows at the next time period of the neighbors are viewed as the outputs of the neighbors. Moreover, the neighbors' outputs have the same clock time with the predicted traffic flow. The contributions of the neighbors' outputs can be regarded as the mappings from the neighbors to their outputs. The predicted traffic flow is obtained by the mappings. The Locally Weighted Regression (LWR) method[13], which is used to obtain the mappings from the historical input vectors to the historical outputs, shows that the values of the distances between the historical input vectors and the current input vector contain useful information for calculating the current output. For short-term traffic flow forecasting, the historical input vectors which are mentioned above correspond to the neighbors. The historical outputs correspond to the outputs of the neighbors. Atkeson *et al.*[13] proposed the locally weighted regression (LWR) method. Sun *et al.*[14] applied LWR to forecasting

This work was supported in part by the National Natural Science Foundation of China under Grant 60921061, Grant 70890084, Grant 90920305, Grant 90924302, Grant 60904057, Grant 60974095, and Grant 31170670 and in part by the Chinese Academy of Sciences under Grant 2F09N05, Grant 2F09N06, Grant 2F10E08, Grant 2F12N02, Grant 2F11D03, Grant 2F11E08, and Grant 2F10E10.

Shuangshuang Li, Dr. Zhen Shen(Corresponding author) and Prof. Gang Xiong are with State Key Laboratory of Management and Control for Complex Systems, Beijing Engineering Research Center of Intelligent Systems and Technology, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with Dongguan Research Institute of CASIA, Cloud Computing Center, Chinese Academy of Sciences, Songshan Lake, Dongguan 523808, China. (e-mail: zhen.shen@ia.ac.cn.)

short-term traffic flow. Zhong *et al.*[15][16] improved LWR by using genetic algorithms(GAs) to choose the inputs for LWR. Each input is regarded as a gene. According to the GA method, the inputs which have maximum correlation with the output are selected as the inputs for LWR.

As long as the historical traffic flow database is big enough to contain enough traffic scenes, a data-driven method usually can achieve good results. However, if the size of the traffic flow database is very large, it may be a very time-consuming in inverting matrices in using LWR. Based on *k*-NN and LWR, we select the nearest neighbors as the historical input vectors of LWR. According to the distances between the nearest neighbors and the current traffic flow, the predicted traffic flow which is viewed as the current output of LWR is obtained. In this paper, we propose a *k*-nearest neighbor Locally Weighted Regression (*k*-LWR) method by arguing that both the nearest neighbors and their distances contain useful information and are enough for short-term traffic flow forecasting.

The *k*-NN uses the information contained in the nearest neighbors. The LWR uses the information from different distances between the historical input vectors and the current input vector. Inspired by the *k*-NN method, the *k*-LWR improves LWR by choosing the nearest neighbors in order to make good use of the information from nearest neighbors and their distances. In this way, the *k*-LWR makes a prediction by using the information given by the nearest neighbors rather than all neighbors, so that it reduces the computation burden. This is the contribution of the paper.

The structure of this paper is organized as follows. After introducing the methods for short-term traffic flow forecasting, some related works are described and analyzed in Section 2. In Section 3 we describe some notations and the *k*-nearest neighbor locally weighted regression method. In Section 4 we show the experimental results of the proposed method and compare it with some other methods. Section 5 is the conclusion of this paper.

II. Related Works

In this part, we briefly describe four data-driven short-term traffic flow forecasting methods, *k*-NN, LWR, PRA and WPRA. Then a briefly comparative analysis on *k*-NN, PRA, WPRA and *k*-LWR is given.

A. Descriptions for four related data-driven methods

1) k-NN method

The *k*-NN nonparametric regression method[5] is a classic data-driven method for short-term traffic flow forecasting. This method argues that the traffic flow vector at the same clock time of the current traffic flow vector is viewed as a neighbor. According to the distances between the neighbors and the current traffic flow, the *k*-nearest neighbors are picked and the predicted traffic flow is obtained by using the corresponding outputs of these neighbors.

2) Locally weighted regression method (LWR)

Locally Weighted Regression(LWR)[13] was proposed in 1997. This method is used for learning the mappings from historical input vectors to the historical outputs based on the distances between the historical input vectors and the current input vector. According to these mappings, the current output is obtained. The contributions of the historical input vectors to the current output are related with the distances between the historical input vectors and the current input vector. The longer the distance is, the smaller the contribution is. Before computing the current output, a matrix inversion is used for obtaining these mappings.

3) Pattern recognition algorithm (PRA)

The pattern recognition algorithm [11] defines a pattern as a sequence of “increasing”, “equal” or “decreasing” of the traffic flow. These traffic flows which have the same pattern as the current traffic flow pattern are used for forecasting the traffic flow. The PRA assumes that each matched traffic flow has the same contribution to the predicted traffic flow. The experimental results in [11] show that the PRA is about 20% improvement compared with *k*-NN.

4) Weighted pattern recognition algorithm (WPRA)

The weighted pattern recognition algorithm[6] is proposed by the same authors of Li *et al.* in 2011. This method takes the same definition of “pattern” as in [11]. The WPRA divides a whole day into several intervals and weights matched patterns according to the intervals at which the matched patterns are. The experimental results in [6] show WPRA significantly outperforms PRA, about 20% improvement.

B. Comparative analysis

The *k*-NN finds the useful information only from the traffic flows at the same clock time of different days as the current traffic flow. In PRA method, the *k*-NN is improved by using the information from the traffic flows of different clock time of different days. In WPRA method, the PRA is improved by considering the information of time intervals. The traffic flows which have the same pattern as the current traffic flow are weighted according to the clock time in database. In this paper, according to our experiments the improvement of WPRA to *k*-NN is about 20% and to PRA is about 12%. The WPRA and PRA find the matched traffic flows by comparing the historical traffic flow patterns with the current traffic flow pattern. Only those traffic flows which have the same pattern as the current traffic flow are viewed as the matched traffic flows and used for prediction. However, the distance between the matched traffic flow and the current traffic flow may be very large. Some matched traffic flows may have little contribution to the predicted traffic flow. The *k*-LWR makes better use of the information of the nearest neighbors and the distances between them and the current traffic flow. In this paper, based on our experiments, the improvement of *k*-LWR to WPRA is about 10%.

III. A *K*-NEAREST NEIGHBOR LOCALLY WEIGHTED REGRESSION METHOD

A. Notations

At first, we define some notations. The traffic flow at the n-th sampling period in the l-th day is $x_l(n)$. The current traffic flow vector can be written as

$$\boldsymbol{x_c} = \left[x_c(n-v), x_c(n-v+1), \cdots, x_c(n-1), x_c(n)\right]^T \tag{1}$$

where $x_c(n)$ is the traffic flow at the n-th sampling period in current day, $x_c(n-1)$ is the traffic flow at one lag sampling period, and so on. The traffic flow to be predicted at the (n+1)-th sampling period in the current day is $x_c(n+1)$. Similarly, the historical traffic flow vector corresponding to $\boldsymbol{x_c}$ in the l-th day is defined as $\boldsymbol{x_l}$. $\boldsymbol{x_c}$ is a column vector. We denote its length as g, where $g = v+1$.

B. The k-LWR method

We use k-NN to find the k-nearest neighbors which are viewed as the inputs of LWR. We use LWR in order to obtain the information from the different distances between historical input vectors and the current input vector, and find the mappings from historical input vectors to historical outputs. In the k-LWR method, we use the information not only from nearest neighbors but also from the different distances between these nearest neighbors and the current traffic flow vector for prediction.

Suppose that there are m pairs of input vectors and output vectors in database. We set that the input vectors are $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_m$ and the corresponding output vectors are $\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_m$. The current input vector is denoted as $\boldsymbol{q}$ and the corresponding output vector is denoted as $\boldsymbol{p}$. Both $\boldsymbol{q}$ and $\boldsymbol{p}$ are column vectors. The length of $\boldsymbol{q}_i$ is g as we defined above. We denote the length of $\boldsymbol{p}_i$ is s. The Euclidean distances between $\boldsymbol{q}$ and each historical input vector in database are $d_1, d_2, \cdots, d_m$. The value of d_i contains information that $\boldsymbol{p}_i$ contributes to $\boldsymbol{p}$. The contribution has the inverse ratio with the distance d_i. w_i is defined as the expression (i=1,2,…,m)

$$w_i = f(d_i) \tag{2}$$

where $f(\cdot)$ is a decreasing function. This function shows the relationship between the contribution and the distance.

The mappings from the input vectors to their corresponding output vectors can be expressed as

$$\boldsymbol{P} = g(\boldsymbol{Q}) \tag{3}$$

where $\boldsymbol{Q}$ is a matrix with the i-th row as $\boldsymbol{q}_i^T$ and $\boldsymbol{P}$ is a matrix with i-th row as $\boldsymbol{p}_i^T$.

Before introducing the flows of the LWR, let's give a brief description to the intuition behind LWR. The mappings $\boldsymbol{R}$ from the input vectors to their corresponding output vectors is found by using a weighted regression to minimize the criterion

$$C = \sum_{i=1}^{m} w_i^2 \left\| (\boldsymbol{q}_i^T \cdot \boldsymbol{R} - \boldsymbol{p}_i^T) \right\|^2 \tag{4}$$

So w_i gives larger weight to the term $\left\| (\boldsymbol{q}_i^T \cdot \boldsymbol{R} - \boldsymbol{p}_i^T) \right\|^2$ in the minimization of (4) when w_i is big.

According to [13], $\boldsymbol{R}$ can be express as

$$\boldsymbol{R} = \left[(\boldsymbol{WQ})^T(\boldsymbol{WQ})\right]^{-1}(\boldsymbol{WQ})^T(\boldsymbol{WP}) \tag{5}$$

where $\boldsymbol{W}$ is a diagonal matrix with diagonal elements $w_{ii} = w_i$. The dimensionality of $\boldsymbol{R}$ is $g \times s$.

Based on $\boldsymbol{R}$, the $\boldsymbol{p}$ can be calculated by the following expression.

$$\boldsymbol{p}^T = \boldsymbol{q}^T \boldsymbol{R} \tag{6}$$

For short-term traffic flow forecasting, if we regard $\boldsymbol{x}_c$ as the $\boldsymbol{q}$, $\boldsymbol{x}_l$ as $\boldsymbol{q}_i$, $x_l(n+1)$ as $\boldsymbol{p}_i$, $x_c(n+1)$ as $\boldsymbol{p}$, the LWR can be used for traffic flow prediction. The process of picking traffic flow data for LWR is shown in Table I. The s is the total time periods of prediction at the current day.

TABLE I
THE PROCESS OF PICKING TRAFFIC FLOW DATA FOR LWR

Prediction sequence	**The *i*-th day in database**		**The current day**	
	$\boldsymbol{q}_i$	$\boldsymbol{p}_i$	$\boldsymbol{q}$	$\boldsymbol{p}$
1	$[x_i(1),x_i(2),\ldots,x_i(g)]^T$	$x_i(g+1)$	$[x_c(1),x_c(2),\ldots,x_c(g)]^T$	$x_c(g+1)$
2	$[x_i(2),x_i(3),\ldots,x_i(g+1)]^T$	$x_i(g+2)$	$[x_c(2),x_c(3),\ldots,x_c(g+1)]^T$	$x_c(g+2)$
…	…	…	…	…
s	$[x_i(s),x_i(s+1),\ldots,x_i(s+g-1)]^T$	$x_i(g+s)$	$[x_c(s),x_c(s+1),\ldots,x_c(s+g-1)]^T$	$x_c(g+s)$

TABLE II
THE STEPS OF FINDING *k*-NEAREST NEIGHBORS

Given the total number of neighbors N and the distance vector $\boldsymbol{e}$ which is used for saving the distances of the k-nearest neighbors, the length of $\boldsymbol{e}$ is k:

1. Compute Euclidean distances between the front k neighbors in database and the current traffic flow vector $\boldsymbol{x}_c$, $d_1, d_2, \ldots, d_k$. The front k neighbors refers to the neighbors with order numbers 1-st, 2-nd, …,k-th in the database.
2. Set the i-th element e_i=d_i and find the maximum value $d_{\max}$ in $\boldsymbol{e}$.(i=1,2,…,k)
    For j=k+1:N
        Calculate the Euclidean distance d_j between j-th neighbor and $\boldsymbol{x}_c$;
        If d_j<$d_{\max}$, remove $d_{\max}$ from $\boldsymbol{e}$, then add d_j to $\boldsymbol{e}$ and find the new $d_{\max}$ in $\boldsymbol{e}$, and record the index of the traffic flow vector with this d;
    End
3. Obtain the k-nearest neighbors in $\boldsymbol{e}$.

The LWR aims to make use of the all inputs and outputs in database. It may be a very time-consuming processing for matrix-inverse when the number of input-output pairs m is large. Considering this, in k-LWR only the k-nearest neighbors of $\boldsymbol{q}$ are used for prediction. Table II shows the process of searching the k-nearest neighbors.

There are also other simple and useful methods to find these neighbors. For example, we can calculate all distances and sort these distances then choose the k traffic flow vectors with the shortest distances. Another method is that we can set a suitable threshold. Only the inputs whose distances are less than the threshold are regarded as the nearest neighbors. In order to calculate $\boldsymbol{p}$, the k-LWR should choose the suitable two parameters. One is the dimension g of $\boldsymbol{q}$ and the other is k.

Some historical traffic flow data is used for training and choosing optimal g and k. The steps are shown in Table III.

TABLE III
THE STEPS OF CHOOSING THE PROPER g AND k

1. Start with a minimal neighborhood size, k=2.
2. Start with a minimal dimension size, g=1.
3. Find the k-nearest neighbors according to Table 2,$\boldsymbol{q}_1$, $\boldsymbol{q}_2$,…,$\boldsymbol{q}_k$.
4. Calculate the predicted traffic flow $x_c(n+1)$ using LWR:

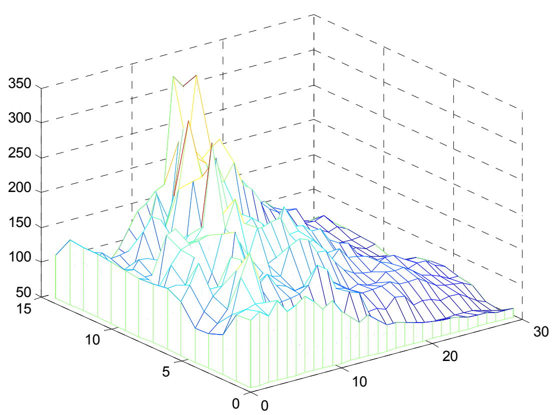
$$x_c(n+1) = \boldsymbol{x}_c^T \boldsymbol{R} \quad (7)$$

where $\boldsymbol{R}$ can be obtained by (5) and its dimensionality is $g \times 1$.

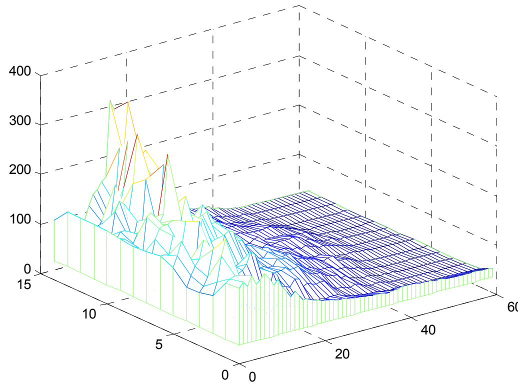
5. Repeat step 3 to step 4 until the prediction for a day is finished.
6. Compute the root mean squared error (RMSE) between the actual values and predicted values.
7. Repeat step 3 to step 5 for dimension size g+1,g+2,…,g_{max}.
8. Repeat step 2 to step 6 for neighborhood size k+1,k+2,…,k_{max}.
9. Choose the proper g and k for k-LWR which has minimal RMSE.

After obtaining the proper g and k, according to k-LWR, $x_c(n+1)$ is obtained by Step 3 and Step 4 in Table III.

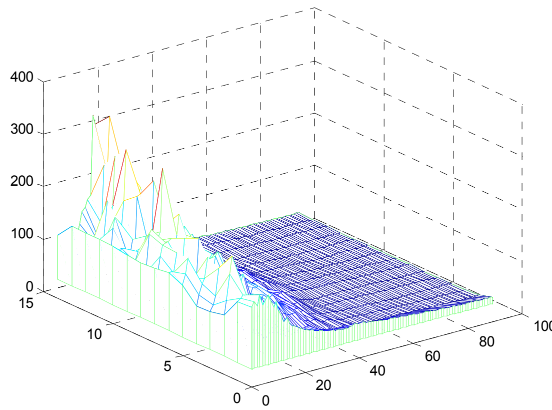
Fig. 1 Spatial location of study site



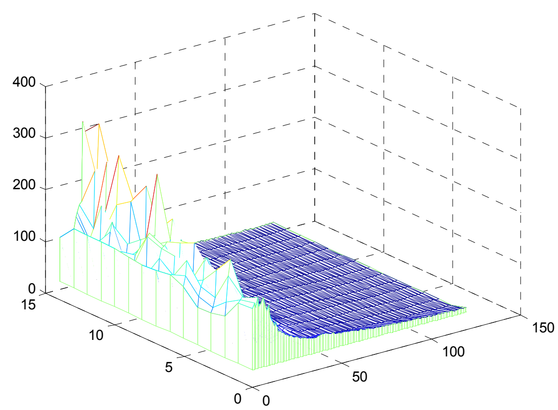
(a)



(b)



(c)



(d)

Fig. 2 The values of RMSE with different k and g. These panels represent (a) k_{max}=30, (b) k_{max}=60, (c) k_{max}=90 and (d) k_{max}=120

IV. Experimental Results

In order to evaluate the performance of k-LWR, we need traffic data. In our experiment, we test the performance of the k-LWR by real traffic flow data. And we compare the performance of k-LWR, WPRA, PRA and k-NN by studying the RMSE between actual traffic flow and the predicted traffic flow.

A. The data

The traffic flow data was collected from Suzhou City in Jiangsu Province in China by Beijing Engineering Research

Center of Intelligent Systems and Technology, Institute of Automation, Chinese Academy of Sciences. We choose the traffic flow data which is collected in 2010 at Feihu intersection and Xingye intersection to test the proposed method. The spatial location of Feihu intersection and Xingye intersection is shown in Fig. 1. The sampling period is 5 minutes and there are a few periods of missing observations, where data is not available for up to 5 minutes. The historical traffic flow data at these links is divided into two sets: the training data set and the test data set. The training data is used for determining the parameters and the test data is used for testing the performance of the proposed method.

B. *Experiments for k-LWR at Xingye intersection*

1) *Parameter determination* The traffic data of six months at Xingye intersection is used for obtaining the proper g, k and the suitable function for (2). In this experiment, we set $g_{\max}$=15. In order to determine the proper g and k, (2) is the Equation 8(a) with h=1.

According to Table III, the results with different $k_{\max}$ are shown in Fig. 2. In Fig. 2, the values of RMSE fluctuate strongly when the value of k is less than 15. With k over 15 and going up, the amplitude fluctuation decreases moderately. The values of RMSE almost remain stable when the k is more than 40. The minimal RMSE is 37.63 when the corresponding g equals 12 and the k_{pr} equals 52. k_{pr} is the proper value chosen for the number of the nearest neighbors. Fig. 2 indicates that the k-nearest neighbors contain almost all useful traffic information for current prediction. It can avoid the large time consumption of predicting traffic flow by using all historical traffic data. In this experiment, the traffic flow data in six months is used for training. The neighbors' number is 180, but the number of nearest neighbors used is 52 for k-LWR. The k-LWR, which uses fewer neighbors, empirically achieves the same performance as LWR.

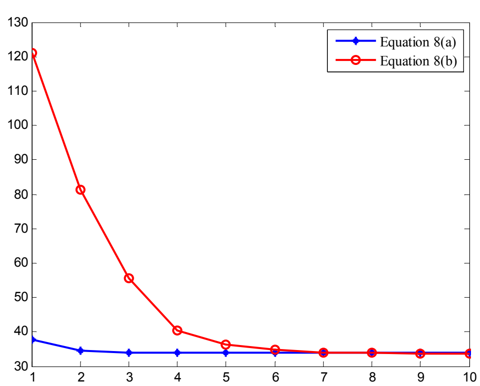


Fig. 3 The RMSE with different h for Equation (8)

After determining the g and k, we should choose the function for w_i in (2). The distance d_i is calculated by using Euclidean distance. In this experiment, two functions are picked to (2) and expressed as follows

$$\begin{cases} w_i = d_i^{-\frac{1}{h}} & 8(a) \\ w_i = \exp(-d_i^{\frac{1}{h}}) & 8(b) \end{cases}$$

where h is the adjustable parameter.

The (2) shows the relationship between the distance and its contribution. The relationship is that the smaller the distance is, the greater the contribution is. So (2) is a decreasing function and the degree of its attenuation is dependent on h. The RMSE is different with different h, as shown in Fig. 3. There is almost no difference between Equation 8(a) and 8(b) when the value of h is larger than 6. That is to say, as long as the h is more than 6, either of them can be selected for (2). In this case, we select h=6.

2)*Results for k-LWR* After determining the parameters and function, the proposed method is tested by using the test data set. We compare the performance of the proposed method with that of some other methods, such as WPRA, PRA and k-NN. In this experiment, the parameters for k-LWR, WPRA, PRA and k-NN are shown in Table IV. For WPRA and PRA, the k_p is the pattern size. For k-NN, the k_n is the nearest neighbor number. k_{pr} is the proper number for the nearest neighbors in k-LWR, as we defined above.

TABLE IV
THE PARAMETERS FOR *k*-LWR, WPRA, PRA AND *k*-NN

Parameters	*k*-LWR	WPRA	PRA	*k*-NN
k_{pr}	52	N/A	N/A	N/A
k_p	N/A	6	7	N/A
k_n	N/A	N/A	N/A	65
g	12	5	4	7

The performance of these methods is evaluated by studying the RMSE between the actual and the predicted traffic flow. These methods are used for predicting 7 days traffic flow. The RMSE of the each method at each day is shown in Table V. We can see that the proposed method is about 10% improvement compared with WPRA, about 20% improvement compared with PRA and about 26% improvement compared with k-NN.

The sampling period is 5 minutes. There are 288 points for traffic flows in a day. According to the above analysis, s=276, k_{pr}=52, $g_{\max}$=15, m=180 and $k_{\max}$=120. The k-LWR method only consumes 49% of the time of LWR in this case.

TABLE V
THE RMSE FOR EACH METHOD IN EACH DAY AT XINGYE INTERSECTION

Day	*k*-LWR	WPRA	PRA	*k*-NN
1	35.18	39.52	44.01	47.59
2	35.08	39.63	44.52	48.01
3	35.57	38.89	43.65	47.25
4	34.99	39.34	43.80	48.37
5	35.10	39.24	44.26	48.64
6	35.07	38.79	43.61	46.83
7	35.50	38.64	43.08	46.95

C. *Experiments for k-LWR at Feihu intersection*

The traffic flow data of two months at Feihu intersection is used for obtaining the proper g, k and the suitable function for (2). The steps are the same as that of experiments at Xingye intersection. The optimal g equals 8 and k_{pr} equals 32. The suitable function for (2) is Equation 8(a) or 8(b) with h=5. In this experiment, the parameters for k-LWR, WPRA, PRA and k-NN is shown in Table VI.

TABLE VI
THE PARAMETERS FOR k-LWR, WPRA, PRA AND k-NN

Parameters	k-LWR	WPRA	PRA	k-NN
k_{pr}	32	N/A	N/A	N/A
k_p	N/A	5	6	N/A
k_n	N/A	N/A	N/A	35
g	8	5	4	5

TABLE VII
THE RMSE FOR EACH METHOD IN EACH DAY AT FEIHU INTERSECTION

Day	k-LWR	WPRA	PRA	k-NN
1	21.90	23.04	27.35	30.71
2	20.46	24.64	27.81	31.63
3	21.21	23.89	28.87	30.02
4	20.97	24.23	28.83	30.28
5	21.78	24.58	27.82	30.41
6	21.52	24.84	28.78	30.40
7	20.91	24.47	27.11	31.21

We choose the seven days traffic flow data in Feihu intersection for testing the proposed method and compare its performance with the performance of WPRA, PRA and k-NN. The RMSE of the each method at each day is shown in Table VII. We can see that the proposed method is about 12% improvement compared with WPRA, about 24% improvement compared with PRA and about 30% improvement compared with k-NN. The k-LWR takes about 69% of the time of LWR in this testing case.

V. Conclusion

By taking virtues from both the k-NN and LWR methods and we design the k-LWR method. The k-LWR uses both the nearest neighbors' information and the information in the distances between the neighbors and current traffic flow vector. The proposed method is tested on the traffic flow data at Xingye intersection and Feihu intersection, respectively. According to our experiment, the performance of k-LWR is about 26% and 30% improvement for the two testing cases over k-NN, 20% and 24% improvement over PRA, when measuring by the root mean square errors between the predicted traffic flow and the actual traffic flow. And the time consumptions are only 49% and 69% of the ones of LWR on the two cases.

Acknowledgment

We would like to thank Prof. Fei-Yue Wang for his guidance and helpful discussions.

References

[1] F.-Y. Wang, "Parallel Control and Management for Intelligent Transportation System: Concepts, Architectures, and Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol.11, no.3, pp.630-638, September, 2010.

[2] F. H. Zhu, Z. J. Li, "A Case Study for Traffic Signal Control System Evaluation Based on Artificial Transportation Systems," *IEEE International Conference on Service Operations, Logistics, and Informatics*, July 10-12, Beijing, 2011, pp.347-352.

[3] K. Wang, Z. Shen, "Artificial Societies and GPU-Based Cloud Computing for Intelligent Transportation Management," *IEEE Intelligent Systems*, 2011, vol.26, no.4, pp.22-28.

[4] Z. Shen, K. Wang, F. H. Zhu, "Agent-based Traffic Simulation and Traffic Signal Timing Optimization with GPU," *14th International IEEE Conference on Intelligent Transportation Systems (ITSC 11)* Washington D. C., 2011, pp.145-150.

[5] B. L. Smith, B. M. Williams and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol.10, no.4, pp. 303-321, August 2002.

[6] S. S. Li, Z. Shen, F.-Y. Wang, "A Weighted Pattern Recognition Algorithm for Short-Term Traffic Flow Forecasting," *9th IEEE International Conference on Networking, Sensing and Control*, Beijing, 2012, pp.1-6.

[7] V. J. Stephanedes, P. G. Michalopoulos and R.A. Plum, "Improved Estimation of Traffic Flow for Real-time Control," *Transportation Research Record No.*795, Washington DC, 1981,pp.28-39.

[8] C. K. Moorthy and B. G. Ratcliffe, "Short term traffic forecasting using time series methods," *Transportation Planing and Technology, vol.* 12, no. 1, pp. 45–56,1988.

[9] G. A. Davis, "Adaptive forecasting of freeway traffic congestion," *Transportation Research Record, n*o. 1287, pp. 29–33, 1990.

[10] I. Okutani and Y.J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filter theory," *Transportation Research, Part B: Methodological,* vol. 18 B, no. 1, pp. 1–11, Feb.1984.

[11] T. Kim, H. Kim and D. J. Lovell, "Traffic Flow Forecasting: Overcoming Memoryless Property in Nearest Neighbor Non-Parametric Regression," *Proceeding of the 8th International IEEE Conference on Intelligent Transportation System*, Vienna, Austria, September 13-16, 2005, pp. 965-969.

[12] J. Hall and P. Mars, "The limitations of artificial neural networks for traffic prediction," *Proceeding of the 3rd IEEE Symposium on Computers and Communications,* Athens, Greece, 1998, pp. 8–12.

[13] C. G. Atkeson, A. W. Moore, S. Schaal, "Locally Weighted Learning," *Artificial Intelligence Review*, vol.11, no.1-5, pp.11-73,1997.

[14] H. Y. Sun, H. X. Liu, H. Xiao, B. Ran, "Short-Term Traffic flow Forecasting using the Local Linear Regression Model," *Transportation Research Record: Journal of the Transportation Research Board*, vol.1836, pp.143-150,2003.

[15] M. Zhong, S. Sharma, P. Lingras, "Refining Genetically Designed Models for Improved Traffic Prediction on Rural Roads," *Transportation Planning and Technology*, vol.28, no.3, pp.213-236, 2005.

[16] M. Zhong, S. Sharma, P. Lingras, "Short-Term Traffic Prediction on Different Types of Roads with Genetically Designed Regression and Time Delay Neural Network Models," *Journal of Computing in Civil Engineering*, vol.19, no.1, pp.94-103, 2005