

Short-Term Traffic Flow Forecasting Considering Upstream Traffic Information

Fei Kou^{1, a}, Weixiang Xu^{1, b}, Huiting Yang^{2, c}

¹School of Traffic and Transportation Beijing Jiaotong University, Beijing, China, 100044, China

²School of Transportation, Southeast University, Nanjing, Jiangsu, China, 211189, China

^awgdbjm@163.com, ^bwxxu@bjtu.edu.cn, ^cyanghuiting84@163.com

Keywords: short-term traffic flow forecasting, K-NN model, upstream information

Abstract. Timely and accurately short-term traffic flow forecasting is very important in the application of the Intelligent Transportation System (ITS). In this paper, an improved k-nearest neighbor (K-NN) model considering upstream traffic information was proposed to forecast short-term traffic flow. In this study, the traffic state of a road segment was described as a state matrix with upstream information instead of only a time series vector. And the weighted Euclidean distance gave different weight to target and upstream road segment was used to measure the similarity between the target state matrix and historical state matrix. The K-NN model was trained by the training data to determine the optimal K value. This study used the reverse distance weighted average method based on similarity of the neighbors to generate the forecasting traffic flow in the future time steps. The same traffic data was used to compare the improved model with three models.

1. Introduction.

With the exponential growth of the number of vehicles, there are a series of problems in some big cities all over the world, such as traffic congestion, environmental pollution and frequent accidents. During the peak period some cities are congested extremely serious. Therefore, the problem of urban traffic, especially traffic congestion, has become one of the most important issues which related to the development of urban economy. Timely and accurately short-term traffic flow forecasting is of great significance to solve traffic congestion. K-NN model is a simple, nonparametric model. In recent years, there are a lot of researches in traffic area with K-NN model.

Davis et al [1] examined the validity of the K-NN model by comparing with a simple univariate linear time series prediction method. Haworth J et al [2] studied the nonparametric forecasting model under missing data, which found that K-NN model is effective in forecasting under missing data. Hou X et al [3] proposed a Two-tier K-NN model to improve the speed and precision of the model. Lin L et al [4] proposed a KNN-LLWNN forecasting model based on K-NN. The training set of LLWNN is formed by K-NN model instead of the whole historical data. Wu S et al [5] proposed an improved K-NN model considering the temporal and spatial information. Cai P et al [6] proposed a spatiotemporal correlative K-NN model, which uses the equivalent distance instead of the physical distance, the traffic state is represented by the spatiotemporal state matrix rather than the traditional time series, uses Gaussian weighted Euclidean distance to determine "nearest neighbors", and applies the Gaussian function to set the weight for each nearest neighbor. Habtemichael F G et al [7] improved the naïve K-NN by time-weighted Euclidean distance, loess method, damp the effect of extreme candidate values, rank-based weighted, and so on.

It can be seen that short-term traffic flow forecasting is a hot research topic in recent years, and K-NN model is also a hot research topic in traffic forecasting area.

2. Improved K-NN Model.

Based on the naïve K-NN model, this study has made some improvements to improve the forecasting accuracy. The specific steps of the improved K-NN model are as follows: 1. Use a matrix space considering upstream traffic information to describe the current traffic state and historical

traffic states, use these historical traffic states to build a sample database; 2. Calculate the weighted Euclidean distance of the current traffic state and all historical traffic states, and obtain K nearest neighbors of the current state; 3. Use the reverse distance weighted average of K-NNs to generate the traffic flow of the forecasted road segment in future time steps. Fig. 1 shows the methodological procedure adapted in this paper.

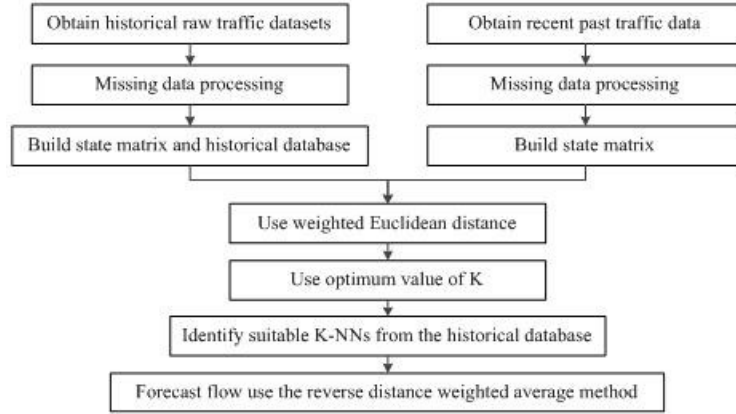


Fig.1 Flow chart showing the improved K-NN model

2.1 The State Matrix.

In this paper, the state is described as a matrix with traffic flow and upstream traffic information of the forecasted road segment instead of a simply time series vector. The state matrix is built with the forecasted road segment state vector and upstream road segments state vectors. The first-row vector of the state matrix is the forecasted road segment state vector.

$$T_1 = \{t_{1,n}, t_{1,n+1}, t_{1,n+2}, \dots, t_{1,n+T}\} \quad (1)$$

Where $t_{1,x}$ is the raw traffic flow of the forecasted road segment in the x time step.

The other row vectors of the state matrix are the upstream road segments state vectors.

$$T_2 = \{t_{2,n}, t_{2,n+1}, t_{2,n+2}, \dots, t_{2,n+T}\} \quad (2)$$

$$T_m = \{t_{m,n}, t_{m,n+1}, t_{m,n+2}, \dots, t_{m,n+T}\} \quad (3)$$

Where $t_{i,x}$ is the calculated traffic flow of upstream road segment in the x time step, show as Equation (4).

$$t_{i,x} = t_{i,x}^{\text{raw}} \times \frac{N_{i1} \times \lambda_{i,1,x}}{\sum_{j=2}^m N_{ij} \times \lambda_{i,j,x}} \quad (j \neq i) \quad (4)$$

The state matrix shows as Equation (5).

$$T = \begin{pmatrix} t_{1,n}, t_{1,n+1}, t_{1,n+2}, \dots, t_{1,n+T} \\ \vdots \\ t_{m,n}, t_{m,n+1}, t_{m,n+2}, \dots, t_{m,n+T} \end{pmatrix} \quad (5)$$

2.2 Weighted Euclidean Distance.

The similarity of historical state and current state in this study was determined by the weighted Euclidean distance. Calculate the Euclidean distance of each row state vector in the two state matrices. Finally weight these rows' Euclidean distance to calculate the final weighted Euclidean distance of two state matrices. The weight of first row is $\frac{1}{2}$, the weight of other rows is $\frac{1}{2(m-1)}$. The Euclidean distance is calculated with Equation (6). The final weighted Euclidean distance in this paper is calculated with Equation (7).

$$D_i(x_{i,t}, y_{i,t}) = \sqrt{\sum_{t=n}^{t=n+T} (x_{i,t} - y_{i,t})^2} \quad (6)$$

$$d = \frac{1}{2} D_1 + \frac{1}{2(m-1)} (D_2 + D_3 + \dots + D_m) \quad (7)$$

2.3 Weight Assignment to the K-NNs.

In this paper, the reverse distance weighted average method is used to forecast the future state, it gives the more similar historical state the higher weight.

$$T_{1,p} = \sum_{q=1}^k \frac{d_q^{-1}}{\sum_{q=1}^k d_q^{-1}} t_{q,p} \quad (8)$$

2.4 Measuring Performance of the Model.

In this paper, three measures are used as performance indicators for accuracy of the improved K-NN model, they are: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

$$MAE = \frac{1}{n} \sum_{i=1}^n |T_i - O_i| \quad (9)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{T_i - O_i}{O_i} \right| \times 100\% \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (T_i - O_i)^2}{n}} \quad (11)$$

Where T_i is the i forecasting value, O_i is the i observed value, n is the number of samples.

3. Case Study.

In this paper, the Gaoliangqiao Xie Street is chosen as the case study forecasted road segment. Several original data instances are shown in Table 1.

Table 1 Original data instances

Sensor ID	Data Time	Lane No.	Vol. long	Speed	Occupancy	Volume
4001	2016-10-08 16:00:02	2	3	44	60	3
4001	2016-10-08 16:00:02	3	4	46	80	4
4001	2016-10-08 16:00:02	4	9	38	160	9
...

We get the volume data of these days between 16:00—19:00, add up the volume per 15 minutes.

3.2. The Optimum K Value.

In this paper, we test the performance of K-NN model with different K in a wide range. The value of K changes from 1 to 20, MAPE of forecasting results in different K value are compared in Figure 2.

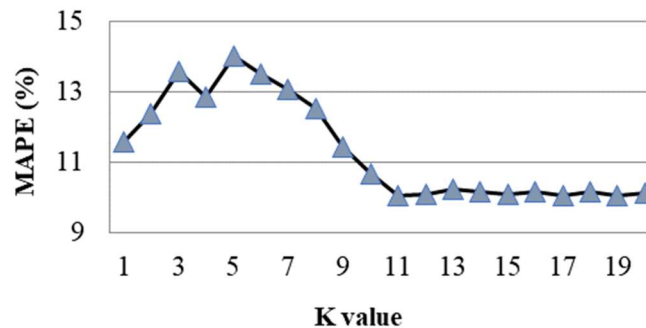


Fig.2 MAPE of different K value

We can find that, the value of K changes between 1 to 5, the accuracy shows up and down. The accuracy is increasing with the increase of K value after 5, and we can get the best accuracy when K is 11. Then with the increase of K value, the accuracy has no significant changes. So we use the 11 as the optimum K value.

3.3 The Comparison of Accuracy.

In order to evaluate the accuracy of the improved K-NN model, the forecasting results are compared with HA, naïve K-NN model, and K-NN model consider the upstream traffic flow. Figure 3 shows the MAE of these four models and Figure 4 shows the MAPE, Figure 5 shows the RMSE.

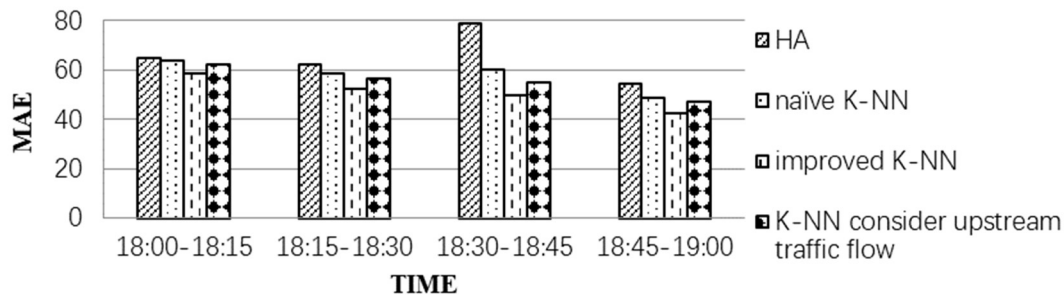


Fig.3 Comparison of the MAE of four models

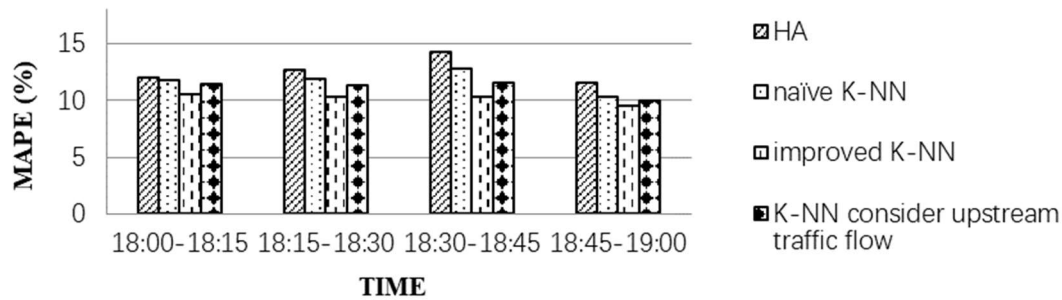


Fig.4 Comparison of the MAPE of four models

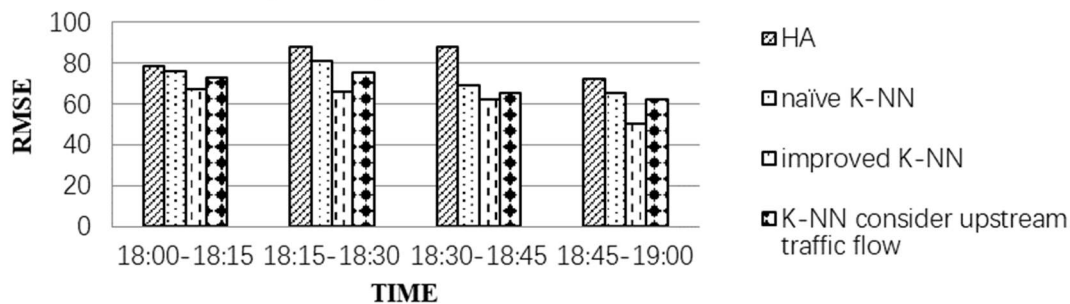


Fig.5 Comparison of the RMSE of four models

In these three charts, we can find that the improved K-NN model performs better than other three models. HA model gets the worst accuracy because the model used the simple historical flow averaging results as the forecasting volumes. It doesn't have a response to different traffic conditions, but we all know that, the traffic conditions have become more and more complex and varied. The naïve K-NN model performs better than HA model, because it selects the K neighbors which are most similar with the target's traffic condition and excludes the impact of the days whose traffic situations completely different with target. But the naïve K-NN model didn't consider the impact of the similarity of the neighbors and gives all neighbors the same weight. The K-NN model consider upstream traffic flow performs better than naïve K-NN model. In this model, more information is considered, and the traffic flow in this time step of upstream road segments are part of traffic flow in next time step of forecasted road segment, so it found the more similar neighbors than naïve K-NN model. The improved K-NN performs best, because the state matrix makes it describe the state more realistic, and the neighbors are more similar with target. It also finds the traffic flow of upstream road segments which are more related with forecasted rod segment by considering division of lanes function and signal timing. At the same time, the improved K-NN model gives higher weight to the neighbors more similar with target.

4. Discussion and Conclusion.

In this paper, the upstream road segment traffic information is used to build the state matrix to describe the current state and the historical state, and the weighted Euclidean distance is used to determine the distance (similarity). Then select the K neighbors by the weighted Euclidean distance,

use the inverse distance weighted average method to generate the future one-way traffic flow. However, the study in this paper also has some shortcomings. First, some data come from VISSIM simulation, we didn't consider the impact of these data. Then, because of the limitation of data, we didn't get more evidence to address the strength-the proposed method can handle the sudden change of lanes function and the timing of signal although it is credible in theory. Next, the impact of weather condition and other factors are not considered in this paper. The model can be improved by solving these problems in further study.

Acknowledgments

This work was supported by National Natural Science Foundation of China (41701120, 61672002). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

- [1]. Davis G A, Nihan N L. (1991). Nonparametric Regression and Short - Term Freeway Traffic Forecasting. *Journal of Transportation Engineering*, 117(2), pp. 178-188.
- [2]. Haworth J, Cheng T. (2012). Non-parametric regression for space-time forecasting under missing data. *Computers Environment & Urban Systems*, 36(6) , pp. 538-550.
- [3]. Hou X, Wang Y, Hu S. (2013). Short-term Traffic Flow Forecasting based on Two-tier K-nearest Neighbor Algorithm. *Procedia-Social and Behavioral Sciences*, 96, pp. 2529-2536.
- [4]. Lin L, Li Y, Sadek A. (2013). A k Nearest Neighbor based Local Linear Wavelet Neural Network Model for On-line Short-term Traffic Volume Prediction. *Procedia-Social and Behavioral Sciences*, 96, pp. 2066-2077.
- [5]. Wu S, Yang Z, Zhu X. (2014). Improved k-nn for Short-Term Traffic Forecasting Using Temporal and Spatial Information. *Journal of Transportation Engineering*, 140(7), pp. 0401-4026.
- [6]. Cai P, Wang Y, Lu G. (2016). A spatiotemporal correlative k -nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C Emerging Technologies*, 62, pp. 21-34.
- [7]. Habtemichael F G, Cetin M. (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation Research Part C Emerging Technologies*, 66, pp. 61-78.