Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 4, No. 2, 2022

Direction Prediction of Traffic Flow in Vissim Simulation Based on K Nearest Neighbor Algorithm

Fuchao Geng, Yufan Ren

School of Energy Science and Engineering, Henan Polytechnic University, Jiaozuo 454002, China

Abstract: In order to make short-term prediction of the direction of traffic flow in urban roads, a short-term prediction method of urban road travel time based on K nearest neighbor algorithm and vissim simulation is constructed. First, the intersection of Shiji Road and Yingbin Road was selected as the survey site, and the number of vehicles in each direction of each entrance lane of the intersection was investigated using manual counting, and the signal timing of each phase of the intersection was investigated. Input the survey data into the vissim simulation software to get the travel time of each entrance lane in each direction. Then build a vissim simulation traffic flow direction prediction model based on the KNN algorithm, including the construction of feature vectors, cross-validation methods to determine K values, and local estimation methods. The experimental results show that the average relative error between the predicted traffic flow direction and the actual traffic flow direction tends to 0.27. Due to the small amount of data, the prediction result is more accurate.

Keywords: K nearest neighbor algorithm, Vissim simulation, Direction of traffic flow.

1. Introduction

With the economic and social development, the number of cars in large and medium-sized cities in China began to rise sharply in recent years, and the resulting environmental pollution, traffic accidents, especially traffic congestion, are becoming more and more serious [1]. The development of intelligent transportation system is one of the effective methods to solve the problems of urban road traffic congestion and environmental pollution [2]. As an important research field of intelligent transportation system, urban road traffic flow direction prediction can realize traffic guidance and effectively alleviate the traffic congestion of urban roads.

The data sources of research related to short-term traffic flow prediction mainly include GPS Floating Car Data[3-4], fixed loop detector data[5] and vehicle electronic identification data. However, at present, the short-term prediction of travel time at home and abroad mostly focuses on the data based on GPS floating car and fixed loop detector. GPS floating car only includes taxi and bus data, but does not include a large number of private car travel data, It has the disadvantage of small sample size. The fixed loop detector is usually only deployed at the intersection, which does not have the characteristics of universality in space. For the vehicle electronic identification data, it is mostly used in the related research of short-term traffic flow prediction, which has a certain novelty. At the same time, because it includes the travel data of all vehicles, including private cars, it has the advantage of large sample size and extensive spatial distribution.

The methods of short-term traffic prediction mainly include linear regression model, time series model, Kalman filtering model, neural network model, nonparammetric regression model, etc. compared with the above methods, the main advantage of k-nearest neighbor (KNN) algorithm is that there is no complex parameter estimation and the model training time is short, It is suitable for processing a small amount of data to realize real-time prediction.

At present, KNN algorithm has been applied to short-term traffic flow prediction in literature. Shlim et al[6] used highway toll data to apply KNN algorithm to real-time prediction of travel time. Steve Robinson et al[5] proposed the strategy of selecting local estimation method and K value in KNN algorithm, and carried out experimental analysis on the proposed strategy.

In conclusion, it can be found that most of the existing studies are the prediction of travel time based on GPS floating vehicle data and fixed ring coil detection data, and there is little research on the prediction of vehicle flow direction. Based on this, this paper predicts the direction of intersection traffic flow based on KNN algorithm, based on the actual home socket traffic flow data and VISSIM simulation software.

2. Data Survey

The survey site selected in this subject is the intersection of Yingbin Road and Shiji road in Jiaozuo, Henan Province. The road plan of the survey site is shown in Figure 1.

Figure 1. Intersection of Ying bin Road and Century Road

2.1. Intersection Traffic Volume Survey

Manually count and observe the traffic flow at the intersection. The design of the questionnaire is shown in Figure 2 (taking the east entrance as an example).

Manual Observation Traffic Record Form (East Entrance)

	Turn left	straight	Turn right
7:00~7:15			
7:15~7:30			
7:30~7:45			
7:45~8:00			

Figure 2. Manual counting questionnaire design table

The flow in the morning peak period was investigated, and the number of motor vehicles and non motor vehicles were counted at an interval of 15 minutes. The survey time is one hour, from 7:00 to 8:00. The hourly traffic volume of each lane and direction is obtained by summarizing the survey data, as shown in Table 1.

Table 1. Hourly traffic volume in all directions at the intersection of Century Road and Yingbin Road

30min	peak	motor vehicle (pcu/h)
east import	Turn lift	71
	Straight	312
	Turn right	24
West import	Turn lift	68
	Straight	306
	Turn right	18
south import	Turn lift	65
	Straight	320
	Turn right	21
North import	Turn lift	69
	Straight	298
	Turn right	19
total	1591	

2.2. Investigation on the Current Situation of Intersection Signal Timing Control

Through observation, it can be seen that there are four phases in the intersection. The first phase is straight from north to south, turning left from north to south is the second phase, straight from east to west is the third phase, and turning left from east to west is the fourth phase. The schematic diagram is shown in Figure 3.

By recording the red light time, yellow light time and full red time of each phase, the signal timing of each phase of a signal cycle is obtained, as shown in Figure 4.

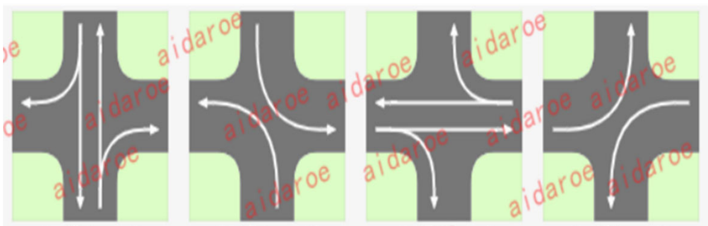

Figure 3. Schematic diagram of signal phase at the intersection of Yingbin Road and Shiji Road

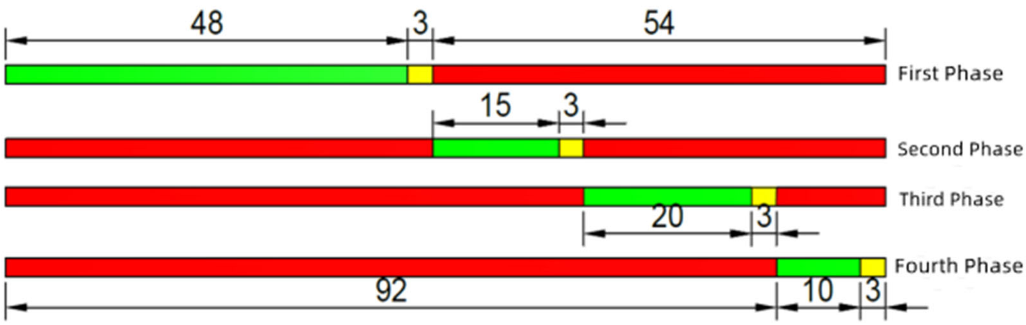

Figure 4. Schematic diagram of signal timing at the intersection of Ying bin Road and Shi ji Road

3. Traffic Flow Simulation

3.1. Simulation Road Network Drawing

After observing the actual road situation at the intersection of Shiji road and Yingbin Road, it can be seen that there is one left turn lane, two through lanes and one right turn lane at the south entrance, one left turn lane, two through lanes and one right turn lane at the north entrance, one left turn lane, two through lanes and one right turn lane at the east entrance, one left turn lane, two through lanes and one right turn lane at the west entrance, The width of each lane is 755 m, as shown in the figure. The width of each lane at the entrance and exit of each lane is 754.5 m.

3.2. Signal Timing and Input of Hourly Traffic Volume

It can be seen from 2.1 that the hourly traffic flow in all directions of each entrance lane at the intersection of Ying bin Road and Shi ji road is shown in Table 1, and the input VISSIM simulation model is shown in Figure 6.

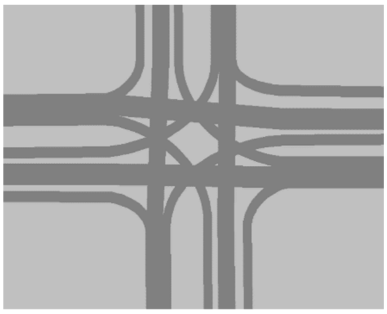

Figure 5. Schematic diagram of simulated intersection of Yingbin Road and Shiji Road

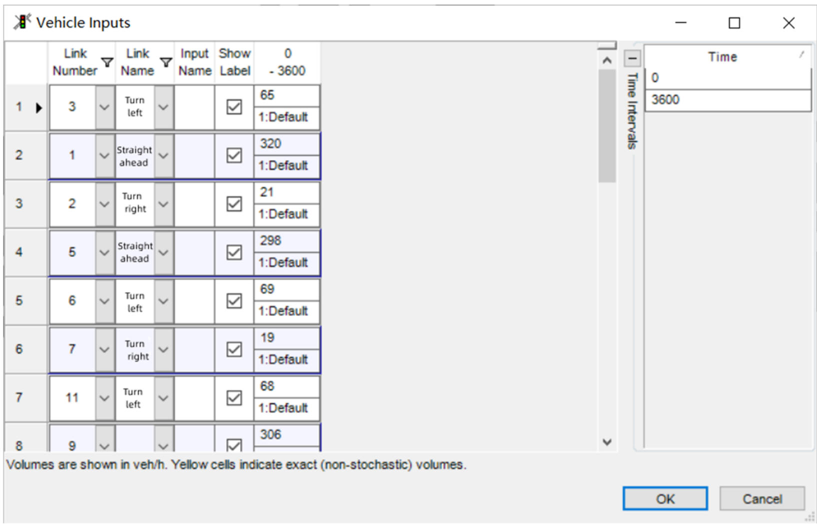

Figure 6. Hourly traffic input simulation model

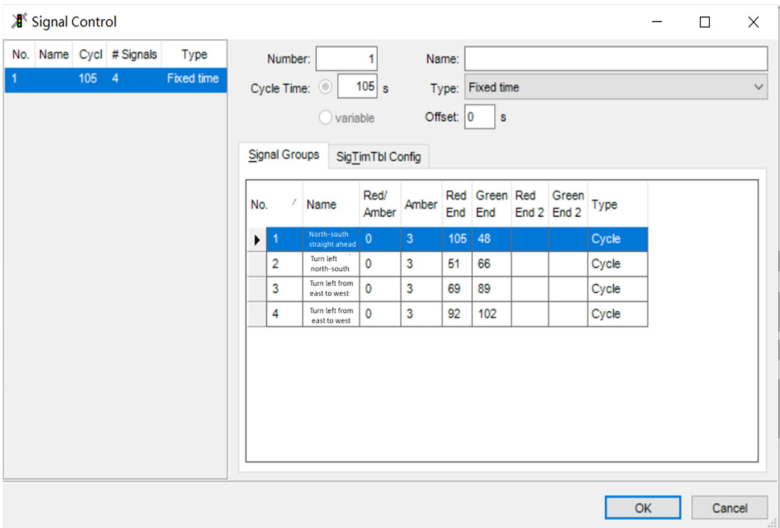

Figure 7. Signal timing input simulation model

It can be seen from 2.2 that the signal timing of each phase at the intersection of Ying bin Road and Shi ji road is shown in Figure 4, and the input VISSIM simulation model is shown in Figure 7.

3.3. Output of Link Travel Time and Number of Vehicles

After the simulation road network is drawn in chapters 3.1 and 3.2, and the hourly traffic volume and signal timing in all directions of each entrance road are input into the simulation model, the simulation road network will be built successfully. Click the start button to run the intersection simulation, as shown in Figure 8.

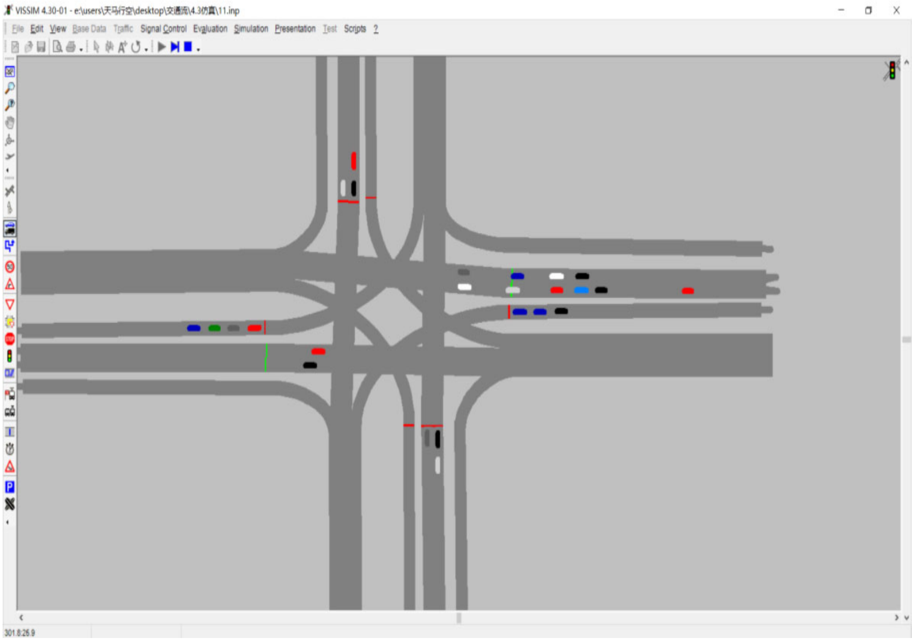

Figure 8. Simulation running effect

After the simulation operation, one hour traffic volume data of the intersection will be obtained, including road network data such as travel time, number of vehicles passing through the road section, delay time and queue length. Since this paper predicts the traffic flow direction of the intersection based on the travel time of the road section and the number of vehicles passing through the road section combined with KNN algorithm, this paper outputs the road network data in a time period of 5min, The travel time of the output section and the number of vehicles passing through the section are shown in Figure 9 (only some data are intercepted).

Name	Section numbe	Journey time	vehicle throughput
300	1	59.9	17
600	1	55.3	29
900	1	55.9	26
1200	1	69.6	23
1500	1	49	27
1800	1	60.1	28
2100	1	56.2	30
2400	1	51.1	30
2700	1	37.8	21
3000	1	59.2	33
3300	1	55	25
3600	1	51	30
300	2	51.2	6
600	2	60.7	2
900	2	45.8	6
1200	2	53.7	3
1500	2	49.8	7
1800	2	72.9	3
2100	2	47.5	8
2400	2	48.9	4
2700	2	63.2	9
3000	2	73.1	8
3300	2	43	4
3600	2	51.3	7
300	3	9.3	2
600	3	9.4	2
900	3	9.2	2

Figure 9. Simulation output data

Taking 300s as a time period, the data obtained are as shown in Figure 9. Each line includes the section number, the travel time of this time period and the number of vehicles passing through this section in this time period. Section numbers 1, 2 , ..., 12 successively represent west east straight line, west east right turn, west east left turn, North South straight line, North South right turn, North South left turn, East West straight line, East West right turn, east west left turn, North South straight line, North South left turn, North South right turn, Take the west east direct behavior as an example, which means from west import to East export.

4. Build Model

4.1. K Nearest Neighbor Algorithm

The basic idea of k nearest neighbor (KNN) algorithm is to match the current input variable with the historical data value with similar input variables, in which the input variables are usually called eigenvectors. Each feature vector describes a point in the multivariable space called feature space. If the feature vector contains n attribute values, the feature space is n-dimensional. The output value of KNN algorithm is defined as a function related to the known output history of input eigenvectors with similar eigenvectors.

Y_t is the predicted value of the set period of t, which is represented by a predicted value of the set period of t. In KNN algorithm, for a given input eigenvector X_t, K historical eigenvectors X_{h1}, X_{h2}, $\cdots$, X_{hk}, closest to the input eigenvector Y_t are searched in the historical data set, and then the K eigenvectors are solved by weighted estimation.

4.2. Construct Feature Vector

The composition of the feature vector needs to have a certain correlation with the travel time of the road section, and can reflect the differences of the same road section at different times. In order to meet the needs of real-time prediction, the feature vector also needs to be easily calculated from the historical data set.

Due to the characteristics of traffic flow, the travel time has a certain autocorrelation in the time series. Generally, the travel time of the first few time periods of the current time period is selected as a part of the feature vector. Due to data limitations, only the travel time of the current time period is selected as a part of the feature vector.

The number of vehicles passing through this section in a period of time also reflects the traffic situation of the current road, which is greatly related to the travel time of the current road. For two acquisition points a and B at both ends of a road, the number of vehicles passing through acquisition points a and B in time period p is taken as a part of the feature vector in time period P, and the expression of the feature vector in time period p: [T, n]

4.3. K Value Determination

K value is the only parameter in KNN algorithm, and its selection is directly related to the accuracy of prediction results. This paper uses cross validation to determine the optimal K value. The specific steps are as follows:

(1): the minimum and maximum values of K_{min} and K_{max} are selected respectively.

(2):divide the road section R data set into M parts at random to obtain the data set D_1, D_2, $\cdots$, D_M, and then successively take the data set D_i (i = 1, 2,..., m) as the test data set and the other M-1 data sets as the historical data set.

(3) Taking $K = K_0$, K_0 is between K_{min} and K_{max}. Calculate the average absolute error percentage of the test data set D_i.

$$E_{(K_0,M_i)} = \frac{100\%}{N_p}\sum_{r=1}^{N_p}\left|\frac{A_r - A_p}{A_r}\right| \tag{1}$$

Where: N_p is the sample size in the test data set D_i; A_r is the true value of record r in the test data set; when K takes K_0, A_p is K_0. The predicted value obtained by applying the prediction method in the test data set D_i.

(4) when K takes K_0, Calculate the average absolute error percentage corresponding to M data sets according to the above formula, and calculate the mean value of these m values:

$$\bar{E}_{(K_0)} = \frac{1}{M}\sum_{i=1}^{M} E_{(K_0,D_i)} \tag{2}$$

(5) When $\bar{E}_{(K_0)}$ takes the minimum value, the corresponding K_0 value is the optimal K.

4.4. Distance Measurement

Because the feature vector includes two parts, travel time and the number of vehicles, the distance between feature vectors is calculated by Euclidean distance formula.

$$d_{ab}^2 = (x_a - x_b)(x_a - x_b) \tag{3}$$

4.5. Traffic Flow Direction Prediction

Firstly, the K feature vectors closest to the feature vectors of the predicted value are found in the historical data, and then the value corresponding to the feature vector is weighted and estimated to obtain the predicted value. Since the travel time of the historical data has a greater impact on the predicted value when the eigenvector of the historical data is closer to the eigenvector of the value to be predicted, the section number is obtained by weighting.

$$T_{P(p)} = \sum_{u=1}^{K} \omega\left(X_{P(p)}, X_{H(u)}\right) T_{K(\omega)} \tag{4}$$

Where: $T_{K(p)}$ is the number of road sections to be predicted. $T_{K(u)}$ is the link travel time value corresponding to the feature vector X_u (u = 1, 2,..., k) among the K adjacent historical feature vectors. $\omega\left(X_{P(p)}, X_{H(u)}\right)$ is the function used for weighted estimation,

$$\omega\left(X_{P(p)}, X_{H(\omega)}\right) = \frac{\exp[-d(X_{P(p)}, X_{H(\omega)})]}{\sum_{u=1}^{K}\exp[-d(X_{P(p)}, X_{H(\omega)})]} \tag{5}$$

Where: $X_{P(p)}$ is the eigenvector corresponding to section r in time period P; $X_{H(\omega)}$ represents one of the k nearest historical eigenvectors; $d\left(X_{P(p)}, X_{H(\omega)}\right)$ represents the distance between $X_{P(p)}$ and vector $X_{H(u)}$ obtained by the eigenvector similarity measurement method.

5. Prediction of Traffic Flow Direction Based on K-nearest Neighbor Algorithm

5.1. K Value Determination

In KNN algorithm, because the selection of K value has a great impact on the experimental results, the optimal K selection is different in different sections and at different times. Therefore, this section first carries out relevant experiments on the calibration of K value. The K value is selected from 0 to 21, and the average relative error of the experimental results is calculated. The results are shown in Figure 10.

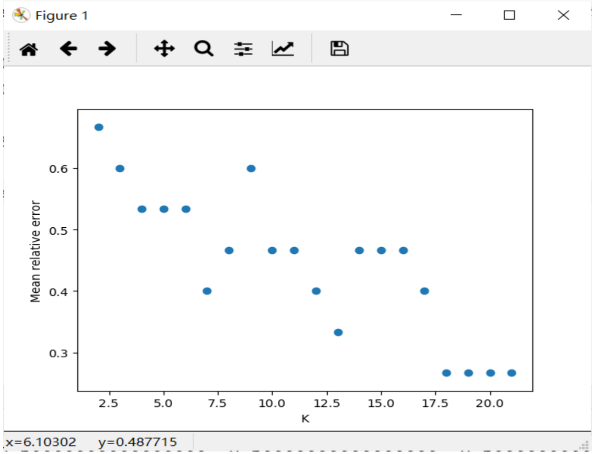

Figure 10. The average relative error varies with the value of K

Taking the average relative error as the Y-axis and the k-value as the x-axis, it is displayed on the image, as shown in the figure. It can be seen that when k = 18, the value of the average relative error converges and tends to be stable, and the average relative error tends to 0.27. Therefore, K is taken as 18.

5.2. Traffic Flow Direction Prediction

Due to the small amount of data, 90% of the data is used for training, and 10% of the data is used to test the correctness of the predicted traffic flow direction. After allocation, there are 14 groups of test data. After training, input the feature vectors of the test data into the prediction model, and output the predicted traffic flow number of the feature vectors of the test data, as shown in Figure 11.

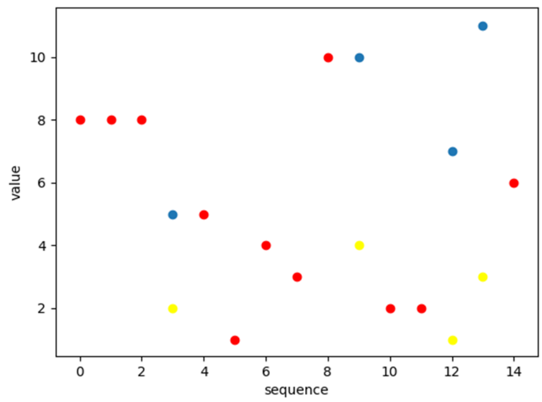

Figure 11. Traffic flow direction can be predicted

The predicted traffic flow number and the actual traffic flow number are displayed on the image, as shown in Figure 11. The x-axis represents the group and the y-axis represents the traffic flow direction number, in which blue represents the actual traffic flow direction and yellow represents the predicted traffic flow direction. When the predicted traffic flow direction is the same as the actual traffic flow direction, it is displayed in red. It can be seen from the figure that most of the 14 groups of data are predicted accurately.

6. Conclusion and Prospect

Firstly, this paper investigates the actual intersection, obtains the hourly traffic volume in all directions of each entrance lane of the intersection and the signal timing of each phase of the intersection, then inputs the survey data into the VISSIM simulation model, simulates the motor vehicles at the intersection for one hour based on VISSIM simulation, obtains the simulation data, and puts forward the traffic flow direction prediction method based on KNN algorithm, The results show that the predicted traffic flow direction has achieved good results. The prediction results of this paper can provide reference information for a large number of urban driving road selection, timely understand the direction of traffic flow, reduce the probability of entering congested traffic flow, and provide relevant decision-making information for urban traffic managers.

Due to the small amount of data, although the prediction results of this paper have been considerable, they are not particularly reliable. In the next research, we can use a large number of traffic flow data, such as traffic flow data for 3 ~ 5 days, to further improve the authenticity of the prediction results. This paper forecasts the traffic flow direction based on VISSIM simulation. Due to the difference between the actual traffic condition and the simulated traffic condition, the actual traffic flow data can be used to predict the traffic flow direction in the next research, so as to further improve the practicability of the prediction results.

References

[1] Yang, Hai, Wolfson, et al. Urban Computing: Concepts, Methodologies, and Applications[J]. ACM transactions on intelligent systems and technology, 2014, 5(3):38.1-38.55.

[2] Pan G , Qi G , Zhang W , et al. Trace analysis and mining for smart cities: issues, methods, and applications[J]. IEEE Communications Magazine, 2013, 51(6):120-126.

[3] X Zheng, W Chen, P Wang,etc. Big Data for Social Transportation[J]. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 2016, 17 (3): 620-630.

[4] Jenelius, Erik, Rahmani, et al. Non-parametric estimation of route travel time distributions from low-frequency floating car data [J]. Transportation Research Part C Emerging Technologies, 2015.

[5] Robinson S , Polak J . Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method[J]. Transportation Research Record Journal of the Transportation Research Board, 2005, 1935:47-56.

[6] LIM S H, LEE H M, PA R K S L, et al. A Study of TravelTime Prediction using K-Nearest Neighborhood Method [J]. Korean Journal of Applied Statistics, 2013, 26(5) :835 -845.