Research on Traffic Data Fusion Based on Multi Detector

Suping Liu

Department of Computer Science, Guangdong University of Science and Technology, Dongguan, China
457789090@qq.com

Abstract. In order to alleviate urban traffic congestion, it is necessary to obtain roadway network traffic flow parameters to estimate the traffic conditions. Single-detector data may not be sufficient to obtain a comprehensive, effective, accurate and high quality traffic flow data. The neural network and regression analysis data fusion methods are employed to expand data sources as well as improve data quality. The multi-source detector data can provide fundamental support for traffic management. An empirical analysis is conducted using Beijing urban expressway traffic flow parameters acquisition technology. The results show that the proposed data fusion method is feasible and can provide reliable data sources.

Keywords: Data fusion · Neural network method · Regression analysis · Multi-source data

1 Introduction

With the increase in car ownership, the urban traffic congestion situation also increasingly serious [1]. By obtaining traffic status, it can reflect the intersection of cities. Traffic flow speed is the best way to reflect traffic flow characteristics, and a traffic flow parameter that is easier to obtain.

Obtaining traffic flow speed can well reflect the traffic status, thus providing reliable parameter data for traffic managers. After the data fusion, the multi-source traffic information can be more authentic and reliable traffic information [2], which provides strong support for traffic operation and management. In order to better serve the traffic management, control and guidance, it is necessary for the multi detector to detect the traffic flow speed.

In order to better serve the induction in traffic management, control and multi detector on traffic flow detection has become inevitable. Using information fusion method source detector collected traffic flow data with the data fusion are more abundant and high quality traffic information, improve traffic mobility, safety and orderly. By using the information fusion technology, the traffic flow velocity data collected by multi-source detectors can be obtained by means of data fusion, and more abundant and high-quality traffic information can be obtained, so as to improve the mobility, safety and orderliness of traffic.

© Springer Nature Singapore Pte Ltd. 2018
https://doi.org/10.1007/978-981-13-1648-7_39
The processing object of data fusion is the speed of data from different sources, and the core of data fusion is to coordinate data optimization and comprehensive treatment. It is necessary to carry on traffic flow speed acquisition integration, the advantages of data fusion, the characteristics of the intelligent transportation system is the key technology and traffic flow and joint decision. Data fusion technology can get more accurate traffic information through comprehensive processing of traffic data from different sources.

As the center of the city within the scope of the road network in the main framework, Beijing city fast road network system without traffic lights, although the length of the total length of road network in the city accounted for just 8% days, but the traffic volume in the future to bear at least counted for the city’s total traffic volume 50% [3]. In the city of Beijing Expressway Development in city traffic plays an important function, the quality of its operation, operation will directly affect the overall road network in Beijing City, the supply capacity of expressway is a freeway traffic flow characteristics reflected, whether in traffic planning or control the daily traffic management, how to correctly analyze determine the traffic flow characteristics of the expressway, will determine the running state of city traffic network.

2 The Model of Data Fusion

2.1 B-P Model

The advantages of B-P algorithm: The research theory is mature, has the rigorous derivation process, the fault-tolerant ability is strong, the versatility is good, so far is the main algorithm of the forward network learning. The disadvantages of B-P algorithm: (1) the learning efficiency is low, the training time is long, the convergence speed is slow, and the increase of the sample dimension will make the network performance worse. (2) Greedy algorithm is especially easy to form local minimum, so that it can’t get the global optimum. (3) The selection of hidden nodes in network is lack of theoretical support. (4) In the course of training, learning the new sample will forget the trend of the previous samples [4, 5].

In the application of B-P model to solve practical problems, better data normalization method can be used to avoid convergence or slow convergence, and to a large extent improve the performance of the network.

2.2 The Method of Speed Fusion Based on Linear Regression

Regression analysis is a statistical analysis method that determines the degree of influence or interdependence between the changes of one or more variables to another specific variable. In this paper, the relationship between the velocity data and the microwave velocity data (independent variable) and travel time velocity data (dependent variable) from the same cross section is studied, and the relation is expressed by regression method.
2.3 The Model of Two Dimensional Regression and Regression Equation

Set 2 independent variables are $x_1$, $x_2$, and the dependent variable is $y$. The equation describing the relationship between the dependent variable, the independent variable and the error term $\varepsilon$ is called bivariate regression model. Its general form as follow:

$$
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon
$$

Where, $\beta_0$, $\beta_1$, $\beta_2$ are parameters of data fusion Model, $\varepsilon$ is the error term. Formula (1) indicates that $y$ is the sum of the linear functions of $x_1$, $x_2$ and error term $\varepsilon$ reflects the influence of random factors other than independent variables on $y$, which is not explained by the linear relationship between independent variables and dependent variables. The parameter $\beta_0$, $\beta_1$, $\beta_2$ in the regression equation is unknown. $\beta_0$, $\beta_1$, $\beta_2$ are estimated by sample statistic $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and finally the bivariate regression equation is obtained. Its general form as follow:

$$
\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2
$$

Where, $\hat{y}$ is the estimate of the dependent variable $y$; $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ is the estimated value of parameter $\beta_0$, $\beta_1$, $\beta_2$. $\hat{\beta}_1$ means the average mean amount of change of dependent variable $y$ that when $x_2$ remains unchanged, $x_1$ changes 1 units. $\hat{\beta}_2$ means the average mean amount of change of dependent variable $y$ that when $x_1$ remains unchanged, $x_2$ changes 1 units.

2.4 Least Square Estimation of Parameters

The $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ in the regression equation is obtained by the least square method, The purpose is to minimize the sum of squares of residuals [7].

$$
E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y - \hat{\beta}_0 - \hat{\beta}_1 x_1i - \hat{\beta}_2 x_2i)^2
= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2
$$

The standard equations for solving $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ can be obtained.

$$
\begin{align*}
\frac{\partial E}{\partial \beta_0} & \bigg|_{\beta_0 = \hat{\beta}_0 = 0} = 0 \\
\frac{\partial E}{\partial \beta_1} & \bigg|_{\beta_1 = \hat{\beta}_1 = 0} = 0 \\
\end{align*}
$$

where, $i = 1, 2$
3 Data Acquisition and Preprocessing of Beijing Network

This paper mainly studies and analyzes the data from 1 directions of 1 license plate detectors in express way, namely Guanghui Nanli district to SanKuaiban village. Beijing expressway traffic flow detector mainly refers to the microwave detector and the license plate recognition detector. According to statistics, there are 829 microwave detectors and 231 license plate detectors in the expressway.

3.1 Selection of Fusion and Forecast Periods

In this paper, the traffic flow velocity data of the whole day are fused and forecasted through 3 periods of morning peak, flat peak and evening peak.

To study the fusion of the 3 detector speeds, data for the 3 detectors must be present for one day’s data. In accordance with this principle, the data obtained from the Beijing Traffic Management Bureau screening, to find eligible only on June 20, 2011, 21, and 22 days of data more appropriate. Because there is a large number of data missing in the license plate recognition detector, the data of early peak, flat peak and evening peak can be used for fusion and prediction. The data for these 3 periods are predicted by the morning peak model, using data from June 22, 2011 5:00–10:00, and model validation using 10:00–11:00 data. The model of predictions of flat peak using June 20, 2011 12:00–17:00 data, model validation using 17:00–18:00 data. Model predictions for evening peaks were trained using data from June 22, 2011 15:00–20:00, and model validation using 20:00–21:00 data.

3.2 Acquisition of Raw Data

In this paper, the data of license plate recognition detector which locate from Guanghui Nanli district to Sanbankuai village are fused.

The data used include floating vehicle detector data, microwave detector data, and license plate detector data, all from the Beijing Traffic Authority Database. In the choice of data, based on to the integrity of the data, the license plate recognition detector data is large enough basis for data screening. In data time selection, data fusion of morning peak and flat peak periods is performed separately. Research on the road is a highway between Guanghui Nanli district to Sanbankuai village. In this section, the expressway is crossed with the East Third Ring Expressway.

The road is Hebei highway road Tonghui between X cell and y a, on the road, Hebei road and East sanhuan Tonghui Expressway Interchange. Study on the road between Guanghui Nanli district to Sanbankuai village section of the road. The main reason for choosing this section is that, on the one hand, the coverage length of the fixed detector can be guaranteed. On the other hand, the number of floating vehicles on the road can also be satisfied. After comparing the original data of the expressway, the data of the license plate recognition detector is relatively sufficient to ensure that the following fusion can be carried out smoothly. The attributes of the data are shown in Table 1.
Method for obtaining data and collating data

(1) Requirements are floating vehicles larger amount of data, license detector data missing less sections, in accordance with this requirement to find the road section.

(2) In order to make the speed of the road closer to the real value and reflect the real-time traffic condition, the data in the basic time interval of different types of detectors are processed differently. The weighted average of the speed of the different moving vehicles in the same direction is weighted equal to the reciprocal of the flow, which is equivalent to the license detector section. The speed of different lanes detected by the coil detector is weighted by the reciprocal of the flow, and the average speed of the detector is obtained by weighting the average speed of all lanes detected by the detector. Then, the average speed of different

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-06-02</td>
<td>12:00–18:00</td>
<td>Flat peak</td>
</tr>
<tr>
<td>(Monday)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-06-22</td>
<td>5:00–11:00</td>
<td>Morning peak</td>
</tr>
<tr>
<td>(Wednesday)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-06-22</td>
<td>15:21:00</td>
<td>Evening peak</td>
</tr>
</tbody>
</table>

Fig. 1. Expressway section

Fig. 2. Detector distribution of section
detectors in the same direction is weighted by the reciprocal of the corresponding flow, and weighted average to the license detector section. The average value is calculated after the bad value is removed from the travel time speed within the basic time interval of 2 min. Figure 1 is a screenshot of a quick section on Baidu maps. Figure 2 shows the distribution of detectors in the study section.

3.3 Data Preprocessing

The data fusion and analysis carried out in this paper are based on 2 min as the basic data time interval.

3.3.1 Microwave Data Preprocessing

The raw data obtained by microwave detector include flow rate, lane number, time occupancy, speed, etc. The road section studied in this paper is a section with license plate recognition as the basic unit. It includes a microwave detector with lanes numbered 23005 and 23006. The preprocessing of microwave data needs to be done in 2 steps.

The first step: obtaining weighted average speed data for each lane speed of a sub detector with an interval of 2 min.

Because of the reciprocal 1/k of the vehicle traffic density, Q can be used as the weight to weigh each lane speed to obtain the weighted velocity data of the microwave detector number 23005 (or 23006). The process is as follows:

\[
\bar{v}_{kl} = \frac{\sum_{i=1}^{n_l} \frac{v_{kli}}{q_{kli}}}{\sum_{i=1}^{n_l} \frac{1}{q_{kli}}} \quad i = 1, 2, \ldots, n_l (i = 11, \ldots, n_l); \quad l = 23005, 23006
\]  

(4)

where, The \(\bar{v}_{kl}\) represents the weighted average of the speed of each lane of the Lth microwave detector within the kth interval. \(v_{kli}\) indicates the speed of the ith Lane detected by the Lth microwave detector within the kth time interval. \(q_{kli}\) indicates the ith Lane flow detected by the Lth microwave detector within the kth time interval.

The second step: Obtain weighted average speed of equivalent license plate recognition section at 2 min intervals. The average speed of the 2 detectors is weighted by the reciprocal of the flow at the microwave detector. The processing method is as follows:

\[
\bar{v}_k = \frac{\sum_{l=23005}^{23006} \frac{v_{kl}}{q_{kli}}}{\sum_{l=23005}^{23006} \frac{1}{q_{kli}}} \quad i = 1, 2, \ldots, n_l (i = 11, \ldots, n_l); \quad l = 23005, 23006
\]  

(5)
\[
\bar{v}_k = \frac{\sum_{i=1}^{3} \frac{v_{ki}}{q_{ki}}}{\sum_{i=1}^{3} \frac{1}{q_{ki}}}, \quad i = 1, 2, \ldots, n_l (i = 11, \ldots, n_l); \quad l = 23005, 23006
\] (6)

where, \(\bar{v}_k\) represents the weighted average of vehicle speed detected by 2 microwave detectors within Kth time interval. \(\bar{v}_{kl}\) represents the weighted average of the speed of each lane of the Lth microwave detector within the Kth time interval. \(q_{kli}\) indicates the ith Lane flow detected by the Lth microwave detector within the kth time interval.

3.3.2 Data Preprocessing of Floating Vehicles

Average traffic, average speed, and average occupancy rate are the original data of the floating vehicles obtained in this paper. The data feature is data that has been standardized to be 2 min intervals. The road section “Guanghuinanli district-SanbanKuai village” studied in this paper is a section with license plate recognition as the basic unit. It is divided into 3 floating vehicle sections according to the direction of traffic flow, in which the number of 3 floating vehicles in the direction of “Guanghuinanli district-SanbanKuai village” is 86569662 and 9663 respectively. Because of the \(1/k = 0\) of the traffic density, \(1/Q\) can be used as the weight to weigh the speed of the 3 floating vehicles in each direction, and the weighted average speed data of the floating vehicle of the section “Guanghuinanli district-SanbanKuai village” is obtained on the average.

The preprocessing procedure is as follows:

\[
\bar{v}_k = \frac{\sum_{i=1}^{3} \frac{v_{ki}}{q_{ki}}}{\sum_{i=1}^{3} \frac{1}{q_{ki}}}, \quad i = 1, 2, 3
\] (7)

Where, \(\bar{v}_k\) represents the weighted average value of the 3 floating vehicles in the same direction in the Kth interval. \(v_{ki}\) represents the detection speed of the ith floating vehicle section in the kth time interval. \(q_{ki}\) represents the detection flow of the ith floating vehicle section within the kth time interval.

3.3.3 The Preprocessing of the License Plate Identification Detector Data

The original data obtained by the license plate recognition detector mainly include vehicle entry time and exit time, and the average speed of the road can be obtained with the length of the section. Its data characteristics are data that has not been standardized processed for 2 min intervals.

The section “Guanghuinanli district-SanbanKuai village” studied in this paper is the section with the license plate recognition as the basic unit. The data preprocessing method eliminates the large deviation value of the data in 1 unit time intervals (2 min) with the confidence of \(a = 0.05\), and then obtains the travel time data of the 2 min intervals on the remaining data. Since the data obtained within 2 min are generally less than 20, are small samples that can be considered as t distributions subject to the degree
of freedom of $n - 1$ ($N$ is the sample size of travel time data for the first $k$ intervals), i.e. $V-T$. The confidence interval of travel time speed can be obtained under the condition that the significance level is $a = 0.05$.

The average speed in a $Kth$ interval:

$$\bar{v}_k = \frac{\sum_{i=1}^{n_k} v_{ki}}{n_k}$$  \hspace{1cm} (8)

Confidence interval of velocity in the first $k$ interval.

$$\left[ \bar{v}_k \pm t_{z/2}(n_k - 1) \frac{s_k}{\sqrt{n_k}} \right]$$  \hspace{1cm} (9)

where, $\bar{v}_k$ represents the average speed in the $Kth$ time interval; $v_{ki}$ indicates the average speed of the $ith$ cars detected by the license plate identification method within the $kth$ time interval. $n_k$ represents the sample size of the license plate identification method detected within the $kth$ time interval. $s_k$ indicates the speed standard deviation detected by the license plate recognition method within the $Kth$ time interval. In the confidence interval, external velocity data should be eliminated. Velocity data within the confidence interval should be retained, and velocity data outside the confidence interval should be eliminated.

4 Example Analysis

4.1 Fusion of B-P Neural Network Method

The model training is divided into 2 parts according to the different research periods: rush hour and peak hour. The rush hour is divided into early rush hour and evening rush hour. A total of 3 model training and 3 validity judgments were tested. Considering the average speed of the road section, it is possible to stagger the morning peak and the evening peak of the traffic to a certain extent. The solution of the model is forward 1 h and pushes backward 2 h on the basis of flow early peak and evening peak. That is, the data time for model training is 5 h, and the time of data verification is 1 h.

Taking into account the non-coincidence of the peak period of speed and flow, the data come from June 22, 2011 5:00–10:00 were used for trained; and the data come from 10:00–11:00 were used for validation in morning peak stage. The data come from June 20, 2011 12:00–17:00 were used for trained; and the data come from 17:00–18:00 were used for validation in flat peak stage. The data come from June 22, 2011 15:00–20:00 were used for trained; and the data come from 20:00–21:00 were used for validation in evening peak stage.

The morning peak period fusion and prediction results are shown in Figs. 3 and 4 respectively.
At the same time, the data of early morning peak, flat peak and evening peak are used for training and prediction, and the predicted fusion value is compared with the accurate value, as shown in Table 2. Table 2 is a comparison of fusion values with floating vehicle data and microwave detection data. Table 2 shows the following conclusions:

(1) The fusion value LSE is smaller than that of the floating vehicle LSE and the microwave detector LSE, so the model is valid.
(2) The difference between the floating vehicle LSE and the fusion value LSE is less than the difference between the microwave detector LSE and the fusion value LSE, which shows that the floating vehicle detector is more accurate than the microwave detector in speed detection. The reason may be that for the whole section, the microwave detector only detects the average speed of the 2 points, and is not representative of the average speed of the whole road; The floating vehicle detected by the floating car detector is basically distributed evenly on the road section. The average speed of multiple points along the direction of the road is detected, and the average speed of the road is more apparent.
(3) Compared with LSE in 3 periods of morning peak, flat peak and evening peak, it is concluded that the accuracy of the detection data is higher in the flat peak period than in the morning peak period.

Table 2. Comparison of fusion values and two type of detector of expressway

<table>
<thead>
<tr>
<th>Type of peak</th>
<th>LSE of floating car</th>
<th>Microwave detector LSE</th>
<th>LSE of Fusion values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning peak</td>
<td>0.219</td>
<td>0.656</td>
<td>0.147</td>
</tr>
<tr>
<td>Flat peak</td>
<td>0.133</td>
<td>0.470</td>
<td>0.040</td>
</tr>
<tr>
<td>Evening peak</td>
<td>0.176</td>
<td>0.560</td>
<td>0.062</td>
</tr>
</tbody>
</table>

4.2 Regression Analysis for Data Fusion

The regression equation was established with the speed of license plate recognition detector as dependent variable, the speed of floating vehicle and the speed of microwave detector as independent variables.

(1) The data come from morning peak June 22, 2011 5:00–10:00 were regressed to obtain the regression equations.

\[ y = 15.82389 + 0.552847x_1 - 0.00597x_2 \]

Because of the significance \( F = 1.39 \times 10^{-16} < \alpha = 0.05 \), the linear relationship is obvious.

(2) The data come from flat peak June 20, 2011 12:00–17:00 were regressed to obtain the regression equations.

\[ y = 60.13448 - 0.13127x_1 - 0.01601x_2 \]

Because of the significance \( F = 0.022544 < \alpha = 0.05 \), the linear relationship is obvious.

(3) The data come from evening peak June 222, 2011 15:00–20:00 were regressed to obtain the regression equations.

\[ y = 44.12753 - 0.088128x_1 - 0.009443x_2 \]

Because of the significance \( F = 0.017131 < \alpha = 0.05 \), the linear relationship is obvious.

The data of morning peak, flat peak and evening peak are trained and predicted, and the predicted fusion values are compared with the accurate ones. The results are shown in Table 3.
The fusion value LSE is smaller than the floating vehicle LSE and the microwave detector LSE, so the model is valid. The difference between the floating vehicle LSE and the fusion value LSE is less than the difference between the microwave detector LSE and the fusion value LSE, which shows that the floating vehicle detector is more accurate than the microwave detector in speed detection. Compared with the 3 peaks of morning peak, flat peak and evening peak LSE, it is concluded that the accuracy of the detection data is higher in the flat peak period than in the rush hour period. From the analysis of the fusion effect, the fusion effect of the flat peak period is better than that of the rush hour period.

From the above analysis, we can get the same result by the neural network method and the linear regression method. The 2 models have been tested by the effectiveness, and achieved good results.

5 Conclusions

In this paper, the neural network fusion model and the two element regression model are deeply studied. Based on the 2 models, the 3 stages of the early peak, flat peak and evening peak of the traffic flow in the fast section of Beijing are tested. To judge the validity of the model, the fusion value of LSE was less than the floating car LSE and microwave detector LSE, the value of flat peak period is more close to the real value of speed of the fusion seriously. At the same time, the difference between the floating vehicle LSE and the fusion value LSE is less than the difference between the microwave detector LSE and the fusion value LSE, indicating the floating vehicle detector is more accurate than the microwave detector in speed detection. Two models are designed to predict the road speed value, and the results are better. The LSE between the fusion value and the accurate value is smaller than the LSE value between the unit detector and the accurate value, whether in early peaks or flat peaks.

References


