Research on Multi-source Traffic Flow Data Fusion Based on Linear Regression: A Case Study in Mega-city

Xiao-Quan WANG[1,a,*] , Chun-Fu SHAO[1,b] ,Xun JI[1,c] ,Zong-Jie LIU[1,d] ,Yuan YUAN[1,2,e]

[1]MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, China

[2]School of Automotive and Transportation Engineering, Shenzhen Polytechnic (East Campus), Shenzhen, China

[a]15120886@bjtu.edu.cn, [b] cfshao@bjtu.edu.cn, [c]13114241@bjtu.edu.cn,[d]liuzongjie@bjtu.edu.cn, [e]13114240@bjtu.edu.cn

*Xiao-Quan Wang

Keywords: Multi-source data fusion, Regression, Traffic parameters, Case study.

Abstract. Data fusion of traffic parameters such as speed and density is a crucial study in intelligent transportation system, yet complicated to formulate mathematically. The aim of this paper is to study and model the multi-source fusion traffic flow data. Based on the data detected in a mega-city, the regression model is developed to solve the defects in the traditional models in this paper. The result demonstrates that the proposed models can pass the effectiveness test and have a favorable fusion effect, whose fusion result can meet the demand for precision. Apart from the accuracy demand, the proposed models can solve the fusion problem briefly and effectively so that it can be used in the practical engineering application. This study concludes that the proposed model has expressed well the data fusion for the practical multi-source data detected in the mega-city, which proves the model has good effectiveness and applicability.

Introduction

With the rapid development of traffic big data, the data fusion technology can be used to obtain more accurate and valuable traffic information by data fusing. And it can be more scientific and rational to complete the traffic prediction, traffic guidance and traffic organization so that the traffic emergency can be coped with. Furthermore, it can provide basic information and suggestion for the traffic organization and management intelligently. Some scholars have made thorough research to traffic data fusion and made breakthrough in both theoretical research and practical application. The studies combine neural network and fuzzy algorithm to fuse traffic data. The studies can offer a technological support for traffic management and decision. The traffic data fusion experiment has been completed comparatively early and it is proved that it is feasible to apply the data fusion technology in the traffic study [1]. Cao and Li fused the data from experiment simulation and sensors collection and two types of algorithms and models have been applied. The result shows the effectiveness of both algorithms is verified. A fusion system has been established to fuse multi-source and heterogeneous data so that valuable information can be obtained. The fusion system is designed to discover a solution to the problem that the standard is not unified and interaction is poor in the domestic traffic sensors field [2].

In domestic research, the studies of traffic information fusion focus on the estimation of the road operation state and traffic accident detection [3].What is proved that fusion structure is one of the influences on traffic state by the BP neural network fusion experiment [4]. Zhao [5] improved the D-S evidence theory and fused the video information and magnetic sensor information. The data fusion is used to forecast the traffic state and a new type of fusion frame is designed [6,7]. Qiu [8] improved the traditional BP neural network model to establish a new GA-BP model, which is verified by experiments.

In foreign research, Henry put forward some ideas about how to apply data fusion in traffic and transportation, which promoted the development of a new field. Although the paper only focuses on the traffic about the road, but a type of multi-source data fusion technology was described. However there are also some difficulties to be solved including accurate demand, real time and dynamic performance and the guarantee for data quality [9].A new type of algorithm is designed by improving the kalman filter algorithm according to the characteristics of the freeway and urban expressway [10]. A fusion algorithm for multi-source data is designed and verified by experiments [11].Chou [12] designed a data model including different module function, which can improve the precision of the fusion with their interaction.

This paper presents a new model based on dual regression models to fuse the data from winding sensors and geomagnetic sensor in one mega-city. The regression function of license plate recognition given winding sensors and geomagnetic sensors is solved and the regression function is used for accuracy test as the calibration value.

Model Formulation

The regression analysis is to explore the follow relation of the variables and uses the mathematical expression to describe the relation. Furthermore, it can confirm the influence degree of one or some variables on the specific variable. The paper studies the quantitative relation among winding sensor data, geomagnetic sensor data (independent variable) and the license plate recognition data (dependent variable) on the same cross section and uses the regression model to expresses the relation. The regression analysis can solve the problem as follows:

(1)The relation between the independent variable and dependent variable can be expressed and the functional relationship can be obtained.

(2)It can confirm the influencing factors of the variables whose contribution values are different.

(3)The functional relationship can be used to forecast and analyze the reliability according to the statistics principle.

Suppose y is the dependent variable and x_1, x_2 are the independent variables. The equation that describes how the dependent variable depends on the independent variables and the error term ε is called binary regression model. The general form is as follows [13]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{1}$$

Where β_0, β_1 and β_2 are the parameters of the model; ε is the error term.

The function shows that the dependent variable is equal to the sum of the linear function of the independent variables and the error term. The error term reflects the influences of random factors on the dependent variable except the linear function of the independent variables, which cannot be expressed by the linear function.

In the model, three basic assumptions are set about the error term:

(1)According to the statistics principle, the error term ε is a mutually independent variable and distributed normally.

(2)For all the independent variables, the variance of the error term ε is the same.

(3)The expected value of the error term ε is 0.The expected value of dependent variable can be obtained as the follow: $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

The parameters β_0, β_1, β_2 in the regression model can be calibrated according to the sample data. The calibrated parameters β_0, β_1 and β_2 can be used to estimate the sealed values. The general form is as follows [13]:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \tag{2}$$

Where β_0, β_1 and β_2 are the estimated value of β_0, β_1 and β_2; $\hat{y}$ is the estimated value of y.

The parameters β_0, β_1 and β_2 are calculated according to the least square method, whose goal is to minimize the residual sum of squares. The form of the residual sum of squares is as the follow [14]:

$$E=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y-\hat{\beta}_0-\hat{\beta}_1x_1-\hat{\beta}_2x_2)^2=\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_{1i}-\hat{\beta}_2x_{2i})^2 \tag{3}$$

The standard equation to solve the parameters β_0, β_1 and β_2 can be obtained, whose form is as the follow:

$$\begin{cases} \left.\dfrac{\partial E}{\partial \beta_0}\right|_{\beta_0=\hat{\beta}_0}=0 \\ \left.\dfrac{\partial E}{\partial \beta_1}\right|_{\beta_i=\hat{\beta}_i}=0 \end{cases} \tag{4}$$

Where probability assignment of i can be 1 and 2.

Case Study

The model is verification by a case study based on the multi-source data of a mega-city in China. A certain arterial road is divided into some independent road sections in the mega-city. The road section selected is 1.45 kilometer long and includes one license plate recognition, one geomagnetic sensor and two winding sensors.

The distances of all detections on the road section selected are shown in Table 1.

Table 1. Location Information of the Detections [m]

Detection	Distance	Cumulative Distance
winding sensor No.1	——	——
license plate recognition	275	275
geomagnetic sensor	358	633
winding sensor No.2	578	1211

Considering the accuracy of the speed information, the speed data detected by license plate recognition is selected as the calibration data. The data from one geomagnetic sensor and two winding sensors is fused in this paper.

A regression model variable is developed, in which the speed data of the license plate recognition is the dependent variable and the speed data of the geomagnetic sensor and winding sensors is independent variables. MATLAB is used to calibrate the parameters in the regression model.

The regression equation obtained is as followed: $y=0.37255+0.65875x_1+0.37255x_2$.

According to the basic data, the effectiveness judgment formula is used to obtain the LSE values of different detections in Figure 1.

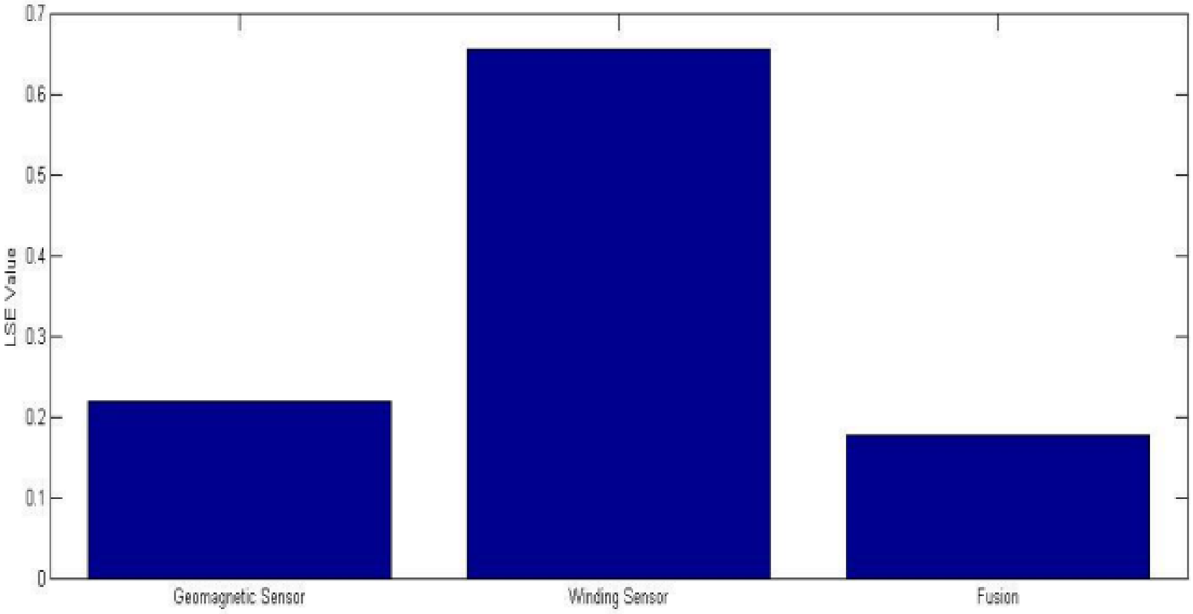

Fig. 1. LSE Values of Different Detections

According to Figure 1, the LSE value of the fusion data is lower than both values of the single-source data, which proves that the fusion model is effective.

Table 2. LSE Values of Different Detections

Type	Geomagnetic	Winding sensor	Fusion
LSE Value	0.219	0.656	0.207

According to the LSE values of different detections in Table 2, the results can be obtained as followed:

(1)The square sum of error of the fusion data is smaller than single detection, which shows that the model meets the effectiveness judgment conditions and it is effective.

(2)By comparing the LSE values of the single detection, the conclusion can be obtained that the speed data of the geomagnetic sensor is more accurate than the winding sensor.

According to the result of the regression model, the fusion data and the exact value can be compared in Figure 2.

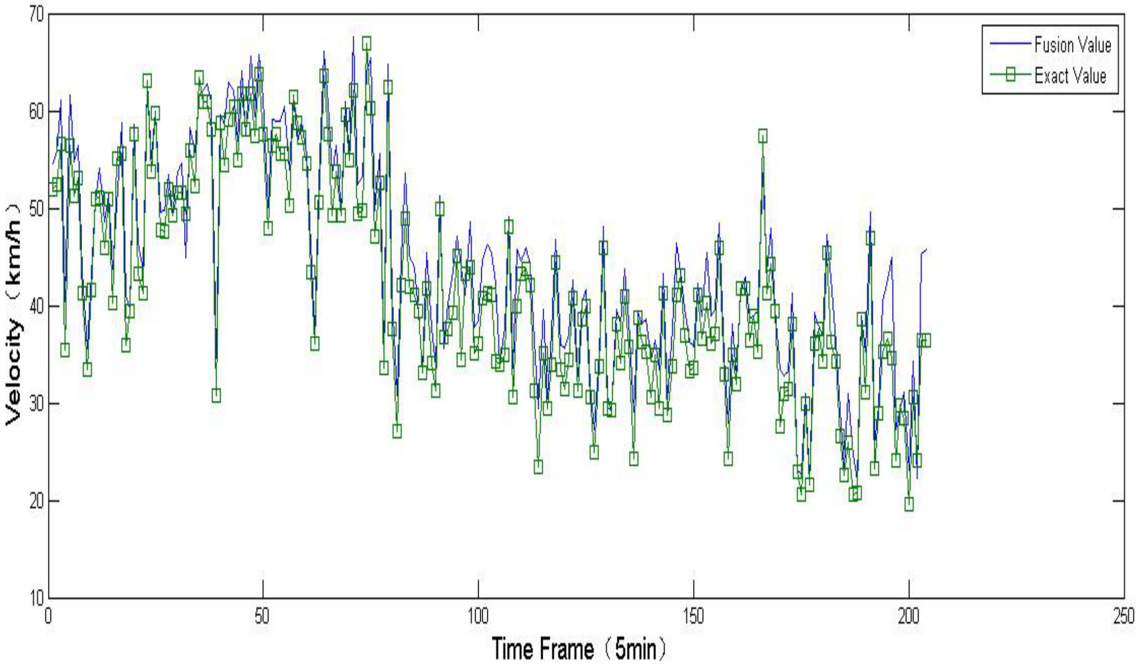

Fig. 2. Comparison of Calibration Value and Fusion Value

According to the result of the regression model, the error of fusion data and the exact value can be obtained in Figure3.

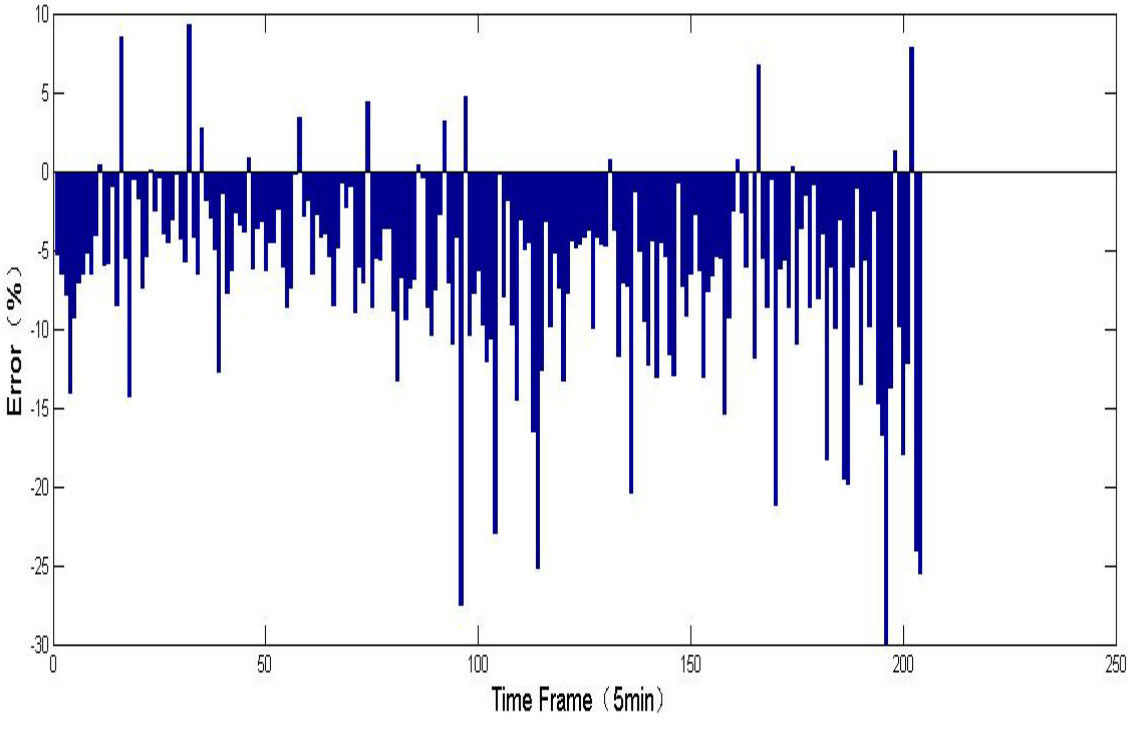

Fig. 3. Error of Calibration Value and Fusion Value

The result shows that the regression model can pass the effectiveness test and have a favorable fusion effect. The fusion result can meet the demand for precision.

Conclusions

In this paper, we study and model the multi-source traffic flow data fusion. Based on the data detected in a mega-city,the regression model is developed and calibrated.The fusion result can meet the demand for precision and the model has good effectiveness and applicability.The proposed models can solve the traffic flow data fusion problem briefly and effectively so that it can be used in the practical ITS,compared with the traditional models. The case study result shows that the model can be applied in the relistic engineering.

Acknowledgement

This paper was supported by the Hebei Natural Science Foundation of China (No. E2016513016)and the Fundamental Research Funds for the Central Universities (Grant No. 2016YJS083).

References

[1] Yang, Z.S., Wang, S., Ma, D.S., Review Of Basic Traffic Data Fusion Methods , J. Journal of Highway and Transportation Research and Development, 2006, 23(3).

[2] Cao, J., Li, Q.Q., The Data Fusion of Probe Vehicle and Winding Sensor In Urban Road Network , J.Traffic and Computer, 2008, 26(4).

[3] Lu, H.P., Intelligent Transportation System Conspectus.Beijing: China Railway Publishing House,2004.

[4] Zhang, X., Fusion of Multi-Source Heterogeneous Traffic Flow Data for the Assessment of Traffic Operational Conditions, D. Beijing: Beijing Jiaotong University, 2008.

[5] Zhao, W.T., Multi-Resources Traffic Information based Data Fusion Research and

Application , D.Shanghai: Shanghai Jiaotong University, 2007.

[6] Lee W H, Tseng S S, Tsai S H. A knowledge based real-time travel time prediction system for urban network , J. Expert Systems with Applications, 2009, 36(3): 4239-4247.

[7] Lee W H, Tseng S S, Shieh W Y. Collaborative real-time traffic information generation and sharing framework for the intelligent transportation system, J. Information Sciences, 2010, 180(1): 62-70.

[8] Qiu, F.C., Study on Data Fusion Technology of Multi-source Traffic Data of Urban Expressway and Arterial Road , D. Beijing: Beijing Jiaotong University, 2012.

[9] El Faouzi N E, Leung H, Kurian A. Data fusion in intelligent transportation systems: Progress and challenges–A survey, J. Information Fusion, 2011, 12(1): 4-10.

[10] Herpel T, Lauer C, German R, et al. Multi-sensor data fusion in automotive applications , C.Sensing Technology, 2008. ICST 2008. 3rd International Conference on. IEEE, 2008: 206-211.

[11] Cheu R L, Lee D H, Xie C. An arterial speed estimation model fusing data from stationary and mobile sensors ,C.Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE. IEEE, 2001: 573-578.

[12] Chu L, Recker W. Micro-simulation modeling approach to applications of on-line simulation and data fusion [J]. California Partners for Advanced Transit and Highways (PATH), 2004.

[13] Zhou, X.B., Semiparametric Estimation And Simulation For The Censored Regression Model , J. Systems Engineering-Theory & Practice, 2012,32(1):60-66.

[14] Zeng, Zh.W., Zh, X.F., Use of Linear Regression and Linear Programming in Optimizing Process Parameters of Diammon Phosphate Production, J. Journal of Fertilizer Industry, 2016,43(1):10-12.