IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

A Fitting Approach to Mend Defective Urban Traffic Flow Information Based on SARBF Neural Networks

Ning Chen, Weibing Weng and Xing Xu

School of Mechanical and Automotive Engineering, Zhejiang University of Science and Technology
Hangzhou, 310023, P.R.China

Abstract

Data-defectives are always occurred during collecting urban traffic flow information due to all kinds of sensors' failures. To mend the defective urban traffic flow information data, a new approach named SARBF neural network fitting is presented. It combines analysis based on spatial autocorrelation and RBF neural network fitting method. The complete data is determined to mend the defective data according to the spatial autocorrelation of traffic grid. Not only the mending precision is improved and also the limitation of regression analysis is avoided by using RBF neural network. Finally, the experiment to mend the defective traffic flow data in Hangzhou City is shown that the method is practicable.

Keywords: *Defective-data Mending, Spatial Auto-correlation, SARBF Neural Network, Fitting Approach, Urban Traffic Flow Information*

1. Introduction

At present, most of the traffic flow data are collected by the loops buried closely to the intersections of urban arteries in China. However, there is data-defective occurred frequently because of the sensor failure while constructing road, transmitting signal or processing data. It is impeded to process and analysis traffic data by the issue. Consequently, it is demanded imperatively to mend the defective data[1]. There exists research on mending methods including the cluster analysis[2], principal component analysis[3], stepwise regression analysis[4], EM(expectation maximization), DA(data augmentation) [5][6], predicting on grey system theory[7] and artificial neural network [8], etc.

There have been both principle guidelines among these methods. Firstly, some methods such as cluster analysis, principal component analysis, stepwise regression analysis, EM, DA merely focus on the relationship of the historical traffic data between the data-defective intersection and the data-complete intersections, but pay little attention to their spatial autocorrelation of these intersections in the whole traffic grid. Secondly, other methods such as traffic prediction based on grey system and the artificial neural network concentrate on the historical data just from the data-defective intersection, without considering the data of other intersections' effect in the traffic grid, which leads to the reflecting disability of the peripheral-zone real-time traffic change to mend the defective data of a sensor-failure intersection.

On the methods of cluster analysis, principal component analysis, stepwise regression analysis, EM, DA, the defective data are mended by the relation between the historical data from the data-defective intersection and the complete traffic flow data in other intersections, without considering the spatial autocorrelation of the intersections in the traffic grid. On the methods of predicting on grey system theory and artificial neural network, the defective data is mended by predicting with the historical data from the data-defective intersection, without considering the data of other intersections' effect in the traffic grid, so the real-time influence of the traffic flow breakdown can't be reflected on the defective data mending.

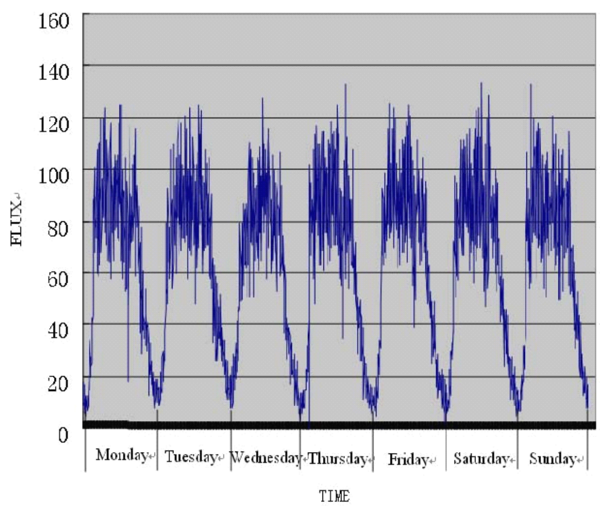
Therefore, here is proposed a new neural network fitting approach named SARBF, abbreviated from Spatial Autocorrelation and Radical Basis Function, to mend the defective data. The approach is composed of spatial autocorrelation based analysis method[9] and RBF neural network fitting method[10][11]. In our research, the precision and calculating speed for mending the defective data is improved by fully using the traffic flow data both from the historical data of a data-defective intersection and its auto-correlated data-complete intersections.

2. Algorithm Design of SARBF Neural Networks Fitting Method

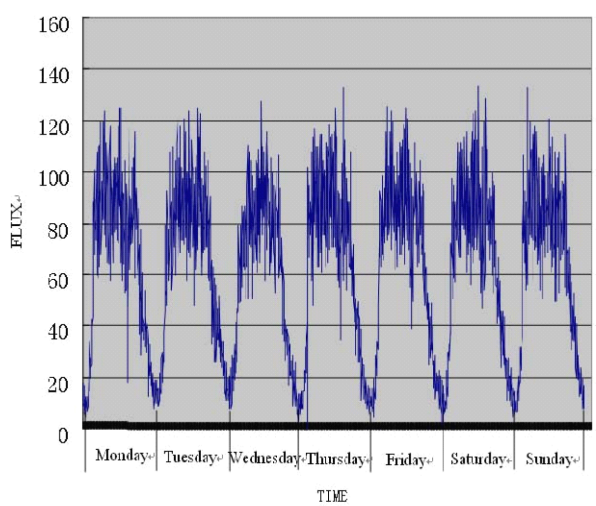
Our research focuses on mending all the defective data of each transportation intersection in urban area. As shown in Fig.1, by the analysis of the historical records, there are both salient features concerning traffic flow data.

1) ***The time periodic characteristic.*** It is evident that the traffic flow data of each intersection changes periodically.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org



(a) Intersection A



(b) Intersection B

Fig. 1 Two adjacent intersections traffic flow data.

For example, Fig.1 shows the changes of the traffic flow data collected at two intersections during one whole week in Hangzhou city.

2) ***The spatial correlation.*** There is correlation of the traffic flow data among the proximity intersections, because the change regularity of the adjacent intersections is usually very similar.

According to the SARBF neural network fitting method, the defective data can be mended resorting to processing the complete data from other intersections. Prior to mending the defective data by means of the neural network, the data-complete corresponding intersections should be selected firstly to improve the mending speed. As shown in Fig.1, the data-complete intersections can be selected by the spatial autocorrelation of the intersections from the traffic grid.

Further more, the defective data can be mended in terms of RBF neural network fitting method based on the historical records of the selected data-complete intersections and the occurred data-defective intersections. The algorithm flow is shown in Fig.2.

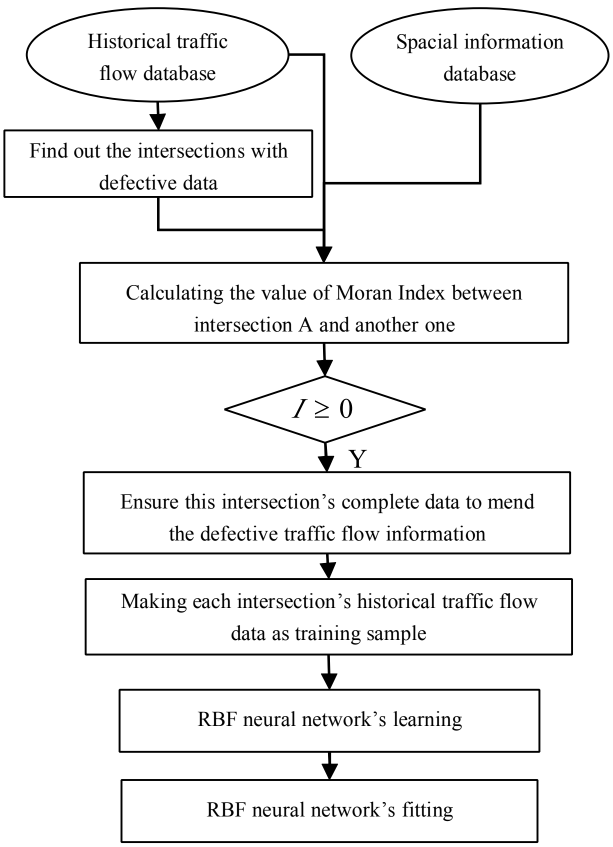


Fig. 2 Diagram of SARBF neural networks fitting method

2.1 Spatial Auto-correlation Analysis

The first step of the SARBF neural networks fitting method for mending defective traffic flow data is to seek for the relating data-complete intersections, which is the basis to establish a neural network model. The urban traffic grid is an organic whole with the intersections being connected by sections. The traffic flow of most adjacent intersections is closely related because of their similarity of the travelers' regularity and travel-mode. The data-complete intersections which are strongly correlated to the data-defective intersection can be chosen by Moran's I spatial autocorrelation analysis method at the benefit of reducing data redundancy.

The spatial autocorrelation in urban traffic flow reflects their correlative degree among the different intersections in the traffic grid. The spatial autocorrelation theory tell us that the closer, the more similar between both objects. The judging procedure by the Moran's I spatial autocorrelation is as follows:

1) Generate a data table with the position of every node and their connections in the traffic grid by GIS software. The intersections are numbered from 1 to *n*.

2) Find out the data-defective intersection No.i from the traffic flow database, then building a spatial adjacency matrix W_{ij}.

If the intersection No.j is close to intersection No.i, then $W_{ij}=1$, else $W_{ij}=0$.

Meanwhile $i \neq j$, $W_{ii}=0$.

3) Calculating the decision index I for the spatial autocorrelation:

$$I=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}\times(X_i-\overline{X})(X_j-\overline{X})}{\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}\times\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2} \tag{1}$$

In which, W_{ij} is expressed as spatial adjacency matrix, X_i is the traffic flow data from the intersection No. *i*, and X_j is the traffic flow data from the intersection No. *j*.

4) Determine the corresponding data-complete intersections on the value I. The value range of I come from the Moran's I formula is between -1 to 1. It is called positive correlation if $I>0$. In contrast, it is negative correlation if $I<0$. The greater value I is, the higher spatial correlation it is between both intersections. Simultaneously, the lower value I is, the weaker spatial correlation it is between both intersections. The spatial distribution would be at random, when I tends to be 0.

2.2 RBF Neural Network Fitting

Because of the random variation of the traffic flow, the relation between both intersections is always nonlinear. However, it can be pursued easily by RBF neural network method.

The radical basis function neural network, i.e. RBFnn, is a feed-forward neural network model with the capacity of local approximation, which can be used in mending the defective traffic data. The weight from input layer to hidden layer is 1 all the time, only the weight from hidden layer to output layer is adjustable. The radial basis function has a local response to the input signal, and the large output would be produced when the input signal closes to the central range of the radial basis function.

Generally, the Gaussian kernel function is selected as the transform mode in the hidden layer due to its preferential properties as simplicity, radial symmetry, smoothness and accurate analyses, the form show as:

$$G(\|x-x_c\|)=\exp\left\{-\frac{\|x-x_c\|^2}{2\sigma^2}\right\} \tag{2}$$

In which, x_c is the center of kernel function; σ is the width parameter of kernel function which controls the radial action sphere and could be determined by k-means clustering algorithm; " $\|.\|$ "is norm which usually takes Euclidean norm.

The RBF network of multi-parameter input and single item output is designed to mend the defective data. Meanwhile, the historical traffic data of the data-defective intersection and the corresponding data-complete intersections turns to be training samples. The weight vector W of the hidden layer to output layer can be obtained by interpolation algorithm when x_c is determined, because the input vector must be mapped to the hidden layer.

3. CASE STUDY

Currently, the urban traffic flow data in urban Hangzhou is usually collected by SCATS system with loops. The deployment situation of SCATS system in Xihu ward of Hangzhou city is shown in Fig.3. The GIS application platform is developed by Map-info system. Besides, the traffic flow data is collected in 5 minutes interval. Here the defective traffic flow data is occurred because of the loops failure, which could be mended by using the SABRF neural network fitting method.

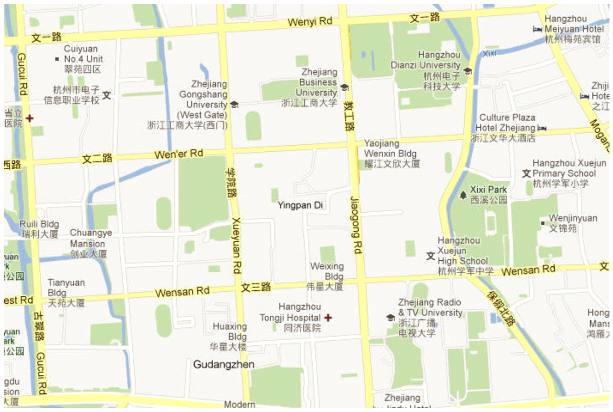


Fig. 3 The deployment of SCATS in Xihu ward of Hangzhou.

Take the Wensan-Xueyuan east intersection encoded 871 as an example, in which the digit 87 stands for 87th intersection and the digit 1 means the east direction. It is obvious the defective data was occurred as shown in the table. There is some traffic flow data in Xihu ward as shown in table 1. The first row data in table I is the intersection number, in which the last number "1", "3", "5", and "7"stands for the directions of East, South, West and North separately. In addition, the character "/" represents for defective data.

Table 1: Part of traffic flow data in Xihu district

time	…	451	871	891	923	455	…
07:05	…	56	/	95	44	71	…
07:10	…	47	/	92	85	99	…
07:15	…	24	/	109	77	68	…
07:20	…	84	/	78	86	74	…
07:25	…	26	/	85	84	55	…
…	…	…	…	…	…	…	…
time	…	451	871	891	923	455	…

The intersections information table can be acquired by Map-info system to build the spatial weight matrix shown as formula (3).

$$\begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 1 & 1 & 0 & \cdots \\ \cdots & 0 & 1 & \ddots & 1 & 0 & \cdots \\ \cdots & 0 & 1 & 1 & 1 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{3}$$

The next step is to hunt for the corresponding data-complete intersections to mend the defective data of the Wensan-Xueyuan east intersection by calculating the Moran's *I*-value from the prospective relevant intersections such as Wener-Xueyuan intersection, Wensan-Jiaogong intersection, and so on. The training samples for the established Neural Network model originate from the historical flux data of the various intersections including not only the data-defective intersection but also data-complete intersections. It is concluded that the regularity of traffic flow usually varies with different periods by analyzing the historical data. Thus in order to guarantee the mending outcome precision, the training samples should belong to the different periods. Meanwhile, it is important to distinguish the different kind of time segment for the RBF neural networks mending model. The time segments may be short ones or long ones The long ones can be further divided into several types such as Spring Festival period i.e. the holiday of Chinese lunar new year, other legislated festival period such as The Labor Day, The National Day, or New Year's Day, weekends, normal time from Tuesday to Thursday, special time such as Monday and Friday. Similarly, the short ones can be divided into morning rush hour, evening rush hour and others.

To mend the defective data using RBF neural network fitting method, it is primary to determine which time segment the defective data locates. Only after the determination could the defective data be mended resorting to the corresponding neural network model. For example, if the defective data of the intersection encoded 871 occurred in December 1st , Saturday in 2012, It demands to establish three different neural network models using the historical traffic data in different time segments that is morning rush hour, evening rush hour and others on Saturday. The mending traffic flow data are shown in table 2.

The MATLAB software, a popular calculating toolkit system can help us to realize the computing process to mend the defective data in terms of RBF neural network fitting method[12-14]. During the process, the traffic flow data of the data-complete intersections being correlated to the data-defective intersection is taken as input vector, and the traffic flow data of the data-defective intersection is taken as object vector. It is proposed that the GOAL of the mean square error is 0, and SPREAD index is 1 in this algorithm. The mean square error of network output will be decreasing continuously by automatically increasing the radial nerve, and the training of the network will be kept until the error attains to the GOAL.

It is the time to use it after the neural network being trained. The defective data can be mended by the below simulation function of the Matlab software:

$$y = \mathrm{si}\,m(net, P) \tag{4}$$

To demonstrate the advantage of our proposed method, it is compared with another data mending method, regression analysis abbreviated RA. The comparing result is shown in table 3 .

The table tells the relative error with the RBF neural network fitting method is about 5%, while the result with regression analyzing method is above 10%. It is apparent that the mending precision with the RBF neural network fitting method is much higher.

Table 2: Mending result（2012.12.1）

Period	Real data/cars	Mending result /cars	Relative error%
12:00~12:05	48	51	6.25
12:05~12:10	70	60	14.3
12:15~12:20	60	49	18.3
12:20~12:25	53	58	9.43
12:25~12:30	65	76	16.9
12:30~12:35	38	40	5.26
12:35~12:40	58	66	13.8
12:45~12:50	82	90	9.75
12:50~12:55	60	68	3.33

Table 3: Comparing result（2012.12.1）

Period	Real data/cars	Mending result /cars		Relative error%	
		RA	RBF	RA	RBF
07:00~07:05(AM)	100	82	95	18.0	5.0
08:00~08:05(AM)	168	192	159	14.0	5.4
08:50~08:55 (AM)	161	189	154	17.4	4.3
09:10~09:15 (AM)	193	219	183	13.5	5.2
07:00~07:05(PM)	146	120	139	17.8	4.8
08:00~08:05(PM)	108	92	103	14.8	4.6
08:50~08:55 (PM)	101	85	97	15.8	4.0
09:10~09:15 (PM)	99	84	95	15.2	4.0

4. Summary

In our research and experiment, the SARBF neural network fitting method is proved to enable speeding process to mend defective traffic flow data by selecting relevant historical data as training samples for neural network by means of spatial autocorrelation analysis. Furthermore, the mending precision can be also improved by RBF neural network method comparing with traditional regression analyzing method. Besides, the mending experiment has given us an acceptable mending precision practically in XiHu ward of Hangzhou city.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (Grant No.61273240), and the Key Project of Chinese Ministry of Education (Grant No.211070), and Zhejiang Provincial Natural Science Foundation of China (Grant No. LY12F03024), and the Project of International S & T Cooperation Program of China (Grant No. 2013DFA31920), as well as the National

High-tech R&D Program of China (863 Program, under Grant No. 2012AA101704).

Thanks for the suggestions and helps from Miss Bingjing Xiang to process all kinds of traffic data involved in the figures and tables in this paper.

Thanks for the technical supporting from Zhejiang Provincial Key Laboratory for Food Logistics Technology and Equipment (under cultivation).

References

[1] Michael Kyte, "Capacity analysis of unsignalized intersections" ,Transportation Research: Part A, Vol.43, No.6,2009, pp. 157-160.

[2] He. Zhang, Zhaosheng. Yang, and Yiwu. Li, "Study on forecast of traffic volume at non-detector intersections", Journal of Highway and Transportation Research and Development, Vol. 29, No. 1, 2012, pp. 91-95.

[3] Zhaosheng. Yang, He. Zhang, Juan. Li, "Step-by-step regression used in traffic flow prediction of non-detector road intersections", Journal of Jilin University (Engineering and Technology Edition), Vol. 42, No. 4, 2012, pp. 16-19.

[4] He. Zhang, "Application of cluster analysis and stepwise regression in predicting the traffic volume of lanes," Journal of Southeast University (English Edition), Vol. 21, No. 3, 2005, pp. 359-362.

[5] Weiguo. Han, Jinfeng. Wang, Yige. Gao, and Jianjun. Hu, "Forecasting and analysis of regional traffic flow in space and time," Journal of Highway and Transportation Research and Development, Vol. 24, No. 6, 2007, pp. 92-96.

[6] Weiguo. Han, Jinfeng. Wang, and Jianjun. Hu, "Imputation methods for missing values in traffic flow data," Journal of Communications and Computer, Vol. 23, No. 1, 2005, pp. 39-42.

[7] Yan. Sun, Senfa. Chen, and Zhenguo. Zhou, "Application of grey models to traffic flow prediction at non-detector intersections," Journal of Southeast University (Natural Science Edition), Vol. 32, No. 2, 2005, pp. 151-153.

[8] Xinquan. Chen, Zhixiang. Hou, Yihu. Wu, and Zhenwen. Liu, "A gray neural network model for traffic flow prediction at nondeterctor intersections," Journal of System Simulation, Vol. 16, No.12, 2004, pp. 2655-2656.

[9] Zhonggang. Liu, Manchun. Li, Jianfeng. Liu, and Yan. Sun, "Analysis and implement of spatial weight matrix based on discrete points," Journal of Geography and Geo-Information Science, Vol. 22, No. 3, 2006, pp. 53-56.

[10] Jun. Zhang, Entao. Yao, Guofen. Ni, and Juan. Ji, "Application of EMD-RBF and image processing to weighing-in-motion of vehicles," Journal of Mechanical Science and Technology for Aerospace Engineering, Vol. 26, No. 6, 2007, pp. 701-704.

[11] Kai. Zhou, "A new isolated point detecting algorithm based on statistical clustering RBF neural neural network," Journal of Computer Science, Vol. 33, No. 10, 2006, pp. 196-197.

[12] Jinpei. Wu and Deshan. Sun, Analysis on Modern Data, Beijing: China Machine Press, 2006.

[13] Liangjun. Zhang, Jing. Cao, and Shizhong. Jiang, Practical Course on Neural Network, Beijing: China Machine Press, 2008.

[14] Xiangnan. Liu, Fang. Huang, Pin. Wang and Zhijun. Tong,Principle and Method on GIS Spatial Analysis, Beijing: Science Press, 2005.

Dr. Ning Chen received the Ph.D. degree in mechanical engineering from Zhejiang University in 2004. Currently, he is an Associate Professor at Zhejiang University of Science and Technology, and the fellow of Chinese Association on Mechanical Engineering. His research interests include Intelligent Transport System(ITS) and Logistics Engineering. He has published 33 papers around mechanical engineering, transportation engineering, and logistics engineering. His work has been supported by the National Natural Science Foundation of China as well as other academic originations.

Dr. Weibing Weng received the Ph.D. degree in Logistics Engineering from Dortmund University of Technology in 2011. Currently, he is an assistant professor at Zhejiang University of Science and Technology. His interests are in logistics technologies and equipments.

Dr. Xing Xu received the Ph.D. degree in mechanical engineering from Zhejiang Sci-Tech University in 2013. Currently, he is an assistant professor at Zhejiang University of Science and Technology. His interests are in urban transportation system control design and information processing.