■   1947

Research on Short-term Traffic Forecast Algorithm Based on Cloud Model

Huai-kun Xiang
Shenzhen Polytechnic, School of Automotive and Transportation Engineering
St. Xili, Nanshan District, Shenzhen 518055, China
e-mail: xianghuaikun@szpt.edu.cn

Abstract

Short-term traffic flow is difficult to predict because of high uncertainty. This paper proposes a short-term traffic forecast algorithm based on cloud similarity. By taking advantage of quantitative and qualitative cloud model mutual conversion function and traffic flow predictability, the historical traffic data can be processed with cloud transformation. Set the current traffic cloud as a standard, traverse the historical traffic cloud to find the best traffic flow period which is with best similarities to the current traffic clouds. Set the future short-term traffic flow of this very period of time as the prediction result of the current period of time. Experiments show that the average prediction error was 3.25 (vehicles) and the prediction error distribution probability was 0.29.

Keywords: traffic flow, cloud model, cloud transformation, traffic forecast, forecast accuracy

1. Introduction

Short-term Traffic Flow Forecasting (STFF) is one of the most popular parts of Intelligent Transportation System (ITS). Researches show that there exists highly uncertainty in short-term traffic sequence. The time-scale is smaller, the time-dependent nature, the nonlinearity and the uncertainty of the traffic flow statistics perform stronger. So far the general forecast time-scales are selected to be 5 min, 10 min or 15min, such as Figure 1 and Figure 2.
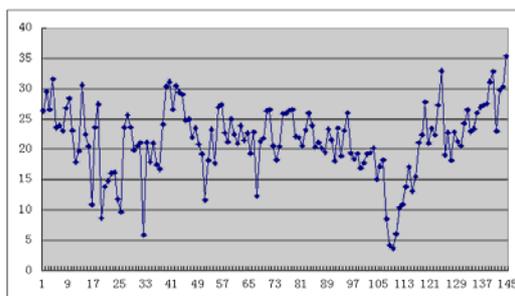

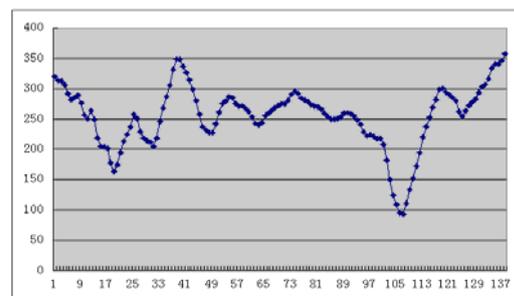Figure 1. Traffic Velocity Distribution of 5min Granularity

Figure 2 .Traffic Volume Distribution of 30min Granularity

For the past few years, both domestic and overseas scholars have come up with numbers of short term traffic flow forecasting models and algorithms. Ross used index filter technology for traffic flow forecast [2]; G.A.Davis et al used K-nearest neighbor algorithm of nonparametric prediction model for short term forecast of expressway traffic flow [3]; Kalman filter has also been used to forecast the traffic flow; M.S.Dougherty et al used BP neural network model for traffic flow prediction [4]; Ahmed et al applied the auto regression moving average model of Box and Jenkins to fast road traffic flow forecast [5]; S.D.Clark et al made comparative study between the time series model and neural network model by means of 5 minutes traffic flow data.

With the deepening of the research, more complicated and accurate forecast methods have been put forward. Dieter Wild [7], a German, introduced a short term traffic flow forecast method which is to transform and classify based on the time series of traffic flow. In America, BrianL.Smith [8] et al use nonparametric regression to test the data of two detectors on M25 loop line of London. They also compared the periodic parameter regression model with non-parametric regression model used to forecast field. SangsooLee [9] et al forecasted the short term traffic flow of the urban expressway by means of subset of ARIMA. Yue Hou et [10] proposed a prediction algorithm for traffic flow prediction of BP neural based on differential evolution. Lingli Li, Hongxia Xia et [11] create the SVM classification model to predict the traffic flow. Hussein Dia [12] studied goal-directed neural network which is suitable for short term traffic flow forecasting. They forecasted the running speed and the travel time of the expressway for a minute for a continuous short term prediction, and achieved good performance.

In overall term, it is divided into two categories of the present short-term traffic forecast research fruits. One is the traffic forecast based on certain mathematic model, such as historical average method, exponential smoothing and time series method. The other is based on uncertain mathematic model, such as neural network and nonparametric regression. As the distribution of the short-term traffic flow parameters are highly uncertain, the forecast based on mathematic model is not a good choice. The forecast based on uncertain mathematic model needs complicated training and it is suitable for the on-line forecast application. As the disadvantages mentioned above, this paper deeps into the short-term traffic forecast algorithm based on Cloud Model.

2. Cloud Model and the Numerical Characteristics

Cloud model is a mathematic model used for the uncertainty transformation between qualitative concepts and quantitative values [15]. Its basic definitions are listed below.

Definition 1: a fuzzy set A which belongs to domain X, to any element $x \in X$, there is a random $\mu_A(u)$. $\mu_A(u)$ is stable and it is called the degree of membership of x to a special fuzzy set A. If the elements in domain are simply sequential or it can be transformed to be well organized if it is not, then the membership $\mu_A(u)$ is called Cloud Model. It is Cloud for short [16].

Figure 3 and Figure 4 are the Cloud distribution diagrams for one dimension and two dimensions respectively.
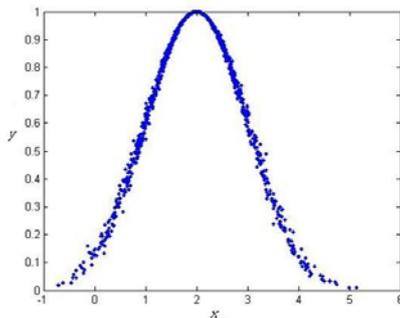

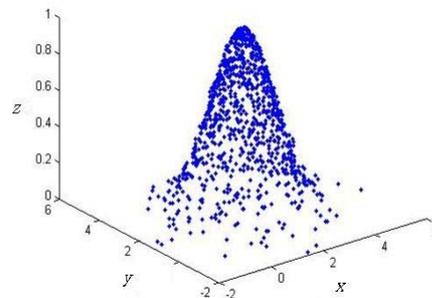Figure 3. One Dimension Normal Cloud Model

Figure 4. Two Dimensions Normal Cloud Model

There are three numerical characteristics of Cloud Model: Expected value, Entropy and Hyper entropy [16-22]: (1) Expected value perfectly presents the value accepted. Generally it corresponds to the x value for cloud center which reflects the information center value of corresponding qualitative concepts. (2) E_n is a measure for the fuzzy degree of qualitative concepts. Its value straightly determines the number of elements which can be received through qualitative concepts in domain. It embodies the both-and margin of the qualitative concepts. (3)

Hyper entropy is the entropy of differentiable entropy which reflects the dispersion degree of cloud. The value of the indirectly determines the thickness of cloud.

The transformation between data and Cloud Model is through the positive cloud and the reverse cloud generator. Positive cloud generator generates cloud droplets by use of construction algorithm and completes the transformation from the qualitative concepts to quantitative expressions according to the three numerical characteristics [23-24]. The reverse cloud restores the three numerical characteristics of the Cloud Model form the given quantitative cloud droplets. It completes the transformation from quantitative values to qualitative linguistic terms.

3. Feature Extraction of Short-term Traffic Flow Based on Cloud Model

Traffic flow characteristics mainly refer to the change of numerical distribution in unit time of the traffic flow parameters, such as size, frequency, periodic and so on. Studies have shown that traffic flow characteristics changes with the following rules:

(1) Continuity

There is continuity of anything in terms of development in time. If we know the past and the present state of the traffic flow, it can be calculated the future.

(2) Analogy principle

There are similarities of the different things in the express form. Using the feature, some kind of the change rule of previous traffic flow we had held can be analogized to similar thing, and to make prediction.

(3) Correlation theory

There are some direct and indirect connections between anything. After known the connections between the subject and other thing, it can be calculated the future.

The smaller the traffic observation time window, the less obvious the distribution regularity of traffic flow characteristics is, the randomness gets stronger. So the traditional mathematical statistics method is hard to express the feature effectively. According to the theory of cloud model, a fuzzy subset A exists in the domain, can be expressed as a collection of a number of atomic concepts, fuzzy subset A is expressed as following:

$$A\{A_1(Ex_1, En_1, He_1),..., A_i(Ex_i, En_i, He_i)\}$$

(1)

In the form of (1), A_1 to A_i are the atomic concepts. Through a series of atomic concepts, a set of irregular data can be classified softly, the real distribution of the data in fuzzy subset can be reflected by a group of atomic concepts, the practice is known as cloud transformation.

For an irregular data distribution, according to some mathematical transformation principle, the core content of cloud transformation is to become the superposition of a number of different sizes of cloud. As shown in the following form.

$$g(x) \approx \sum_{i=1}^{n} c_i \times f_i(x)$$

(2)

In the form of (2), $g(x)$ is the distribution function of the data, $f_i(x)$ is the expected probability density function of the i concept, c_i is the corresponding coefficient, n is the number of atomic concept after soft classification [16-17]. Because $f_i(x)$ is determined by the numerical characteristic of cloud, the soft classification process of fuzzy subset is the process to solve the numerical characteristic of every cloud which has been divided.

After cloud transform, the distribution characteristic of transport flow can be transported to the corresponding cloud numerical characteristic, and prepares for the cloud similarity predicts. There are a lot of basis to extract the traffic flow distribution characteristic of cloud transform, the cloud transformation algorithm base on traffic flow parameters change peak value is used in this paper. The specific steps are as follows:

Step 1: For each possible x for property X (traffic volume or speed), generate the real number y in data base. There is going to be the frequency distribution function which is the frequency distribution curves for traffic volume or speed.

Step 2: Smooth the frequency distribution function for original traffic flow parameters. Then do the interpolation to get $g(x)$. Set the step size as 0.01.

Step 3: Find the current peak location of $g(x)$. It is the maximum of the current traffic flow parameter frequency. Define the corresponding x_i (the specific value of traffic flow parameters as Ex_i, the expected value of cloud. Define $g(x_i)$ as c_i.

Step 4: Fit $g(x)$, Ex_i to get the En_i of Cloud Model.

Step 5: Calculate the fitting residue, $g'(x) = g(x) - c_i \times \exp[-(x - Ex_i)^2 / 2(En_i)^2]$

Step 6: Return to step3 and replace $g(x)$ with $g'(x)$, repeat this step until the error is acceptable to get the superposition of several cloud clusters.

Step 7: Calculate the entropy of the super cloud by use of $g'(x)$ and original $g(x)$ through the last step and Ex_i、En_i of every cloud.

Step 8: Each set of the numerical characteristics (Ex_i, En_i, He_i) for every cloud present presents a qualitative concepts corresponding to the traffic flow parameter.

By the above cloud transform algorithm, the digital characteristics of short time traffic flow can be divided quickly and transformed as cloud digital characteristics to express the concept of traffic flow characteristics.

4. The Algorithm Design of the Short-term Traffic Forecast Based on Cloud Model

The general idea about short term traffic flow prediction algorithm design based on cloud model is: After getting the cloud transform distribution sequence of traffic flow, the current time (like 5min) of traffic flow characteristics is set as a standard, and compares with the historical traffic cloud distribution characteristics. The method is to set a cloud similarity criterion, while founding the historical traffic cloud transformation characteristics of sequence is satisfied with the similarity criterion, extract the historical traffic cloud numerical characteristics period corresponds to the next period of time (like 5min) as the forecast values of the future short term of the current period of time(like 5min). The positive cloud generator completes the transformation from qualitative concepts to quantitative data to obtain forecast traffic flow data. The algorithm chart is seen in Figure 5.

Cloud similarity is defined as L. The forecast core is to identify by way of cloud similarity. Its definition is below.

Definition 1: Supposed $C_1(E_{x_1}, E_{n_1}, H_{e_1})$ and $C_2(E_{x_2}, E_{n_2}, H_{e_2})$ were cloud models of the same domain U. Define $L(C_1, C_2)$ as the similarity of C1 and C2 shown as follows:
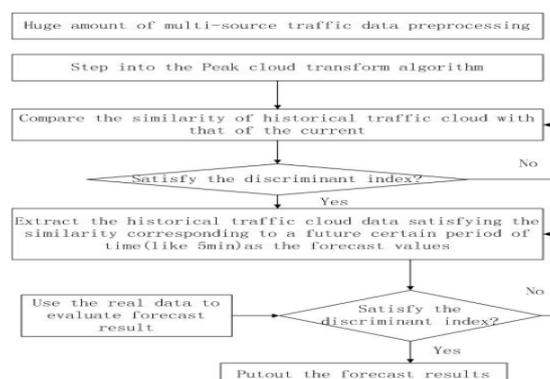

Figure 5. The Forecast Algorithm Diagram for Short-term Traffic Flow Based on Cloud Model

$$L(C_1, C_2) = \begin{cases} 0 & \left| E_{x_1} - E_{x_2} \right| \geq 3(E_{n_1} + E_{n_2}) \\ \dfrac{\min(E_{n_1}, E_{n_2})}{\max(E_{n_1}, E_{n_2})} & \left| E_{x_1} - E_{x_2} \right| = 0 \\ \exp \dfrac{-(E_{x_1} - E_{x_2})^2}{2(\min(E_{n_1}, E_{n_2}))^2} & 0 < \left| E_{x_1} - E_{x_2} \right| < 3(E_{n_1} + E_{n_2}) \end{cases}$$

(3)

When the inequality satisfies $L < \varepsilon$ (ε is a given discriminant threshold), it shows that among the historical sequential data are traffic flow distribution characteristic. It has the highest similarity with the current period of time. If enlarging the traffic flow distribution tendency of this historical panel point to 5min, according to the similarity of traffic flow distribution tendency, it can be considered that the traffic flow in the future 5min is the future 5min traffic flow distribution of the current period of time. The data can be used as the forecast result.

To identify the performance of the traffic flow forecast algorithm, here are two generally used evaluation indexes:

Average forecast error (E_{MAPE}):

$$E_{MAPE} = \left(\sum_{j=1}^{N} \left[\frac{q_{tij} - Q_{tij}}{Q_{tij}} \right] \times 100\% \right) / N$$

(4)

Error distribution probability (E_{PPE}):

$$E_{PPE} = f(\Delta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\Delta^2}{2\sigma^2}}$$

(5)

The former is to forecast the total average value of the percentage errors relative to real traffic volume. It stands for the forecast accuracy of the model. While the latter is to ensure the probability is in the error arrange (like $\pm$ 10%), the reliability of forecast.

5. Verification of Algorithm

Table 1. Cloud Cluster Numerical Characteristics Of Average Travel Speed

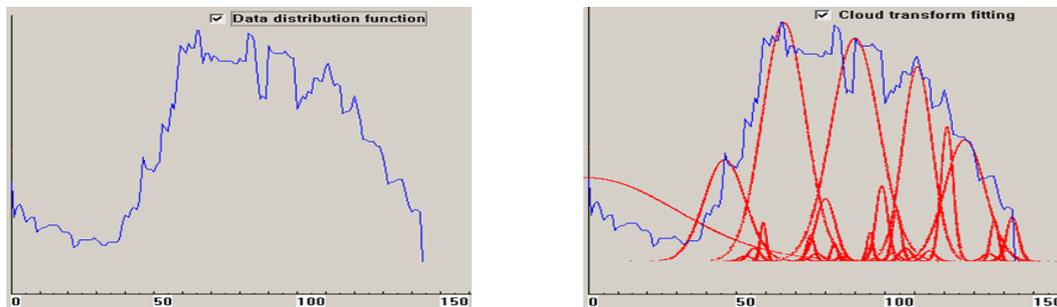|					
Ex	66	90	111	46	…
En	7.349999905	8.109999657	4.96999979	6.900000095	…
He	0.100000001	0.100000001	0.100000001	0.100000001	…

Figure 6. Clouds Created by Quantitative Data

To verify the design ideas, part of the data is selected in a section of Chuangye road in Nanshan district of Shenzhen city of Guangdong province, China. Rode traffic flow parameters selected average travel speed. The data collection interval is 5 minutes. The time range is form August 16 in 2012 to May 25 in 2013. The original data is fifty thousand approximately. The data after processing of the recognition of fault data and recovery is 17280 approximately.

In the peak cloud transform algorithm, the error range is set to 0 to get the number of clouds of three sections in both directions and the corresponding number of clouds digital features $(E_{x_1}, E_{n_1}, H_{e_1})$. Table 1 shows the average travel speed clouds digital features of part sections, Figure 6 is the sections of the rate of the average travel speed distribution and its corresponding cloud.

Based on the Cloud similarity identification, extract the history cloud is the most similar with the current period's traffic clouds, which determine its traffic flow data of next five minutes. And it is used as the traffic flow parameters predicted value for the next five minutes. Figure 7 shows the vuerlay of the measured and predicted values, through calculation, The five minutes driving speed prediction error mean based on cloud mode is 3.25 (vehicles), the error distribution probability is 0.29.
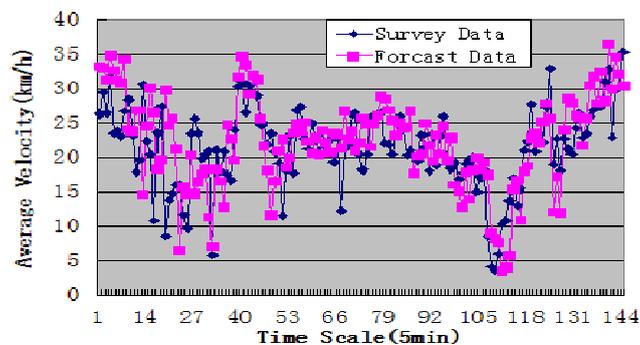

Figure 7. The Average Travel Speed Forecast Based on Cloud Model

6. Conclusion

This paper puts forward and verified the short time traffic flow forecast algorithm based on the cloud model, which is supported by the Shenzhen basic research project (the project's number is JCYJ20120617144302660, JC201006030851A). Its core is based on the continuous principle of the traffic flow distribution, using the unique of cloud model advantage in processing huge amounts of data, to process the traffic flow feature with cloud transform. In real term, although the algorithm is effective, the following two points need to be investigated and resolved further. In the peak cloud transform, the entropy is based on fitting the positive and negative residuals to determine the end of fitting process, the next step can research from the perspective of nonlinear fitting, which will help to improve the fitting accuracy. The predicted value is based on historical traffic cloud point of time at the highest similarity to a short period of expansion, the validity and reliability of this extension is necessary to demonstrate further.

References
[1] Shuntian zhu, Hong Wang, Hongyan Xiang. Traffic flow parameters and traffic incident dynamic prediction method. Southeast University Press. 2008: 21-22.
[2] P Ross. Exponential Filtering of Traffic Data. In Transportation Research Record 869, TRB, National Research Council, Washington, D.C. 1982: 43-49.
[3] GA Davis, NL Nihan. Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *Journal of Transportation Engineering*. 1991; 177(2): 178-188.
[4] MS Dougherty, HR Kirby, RD Boyle. The Use of Neural Networks to Recognize and Predict Traffic Congestion. *Traffic Engineering and Control*. 1993; 34(6): 311-314.
[5] MS Ahmed, AR Cook. Analysis of Freeway Traffic Time-Series Data by using Box-Jenkins

Techniques. In Transportation Research Record 722,TRB,National Research Council, Washington. D.C. 1979:1-9.

[6] SD Clark, MS Doughetry, HR Kirby. *The Use of Neural Networks and Time Series Models for Short-term Traffic Forecasting: A Comparative Study.* Proceedings of Transportation Planning Methods. 1993: 151-162.

[7] D Wild. Short-term Forecasting based on a Transformation and Classification of Traffic Volume Time Series. *International Journal of Forecasting.* 199; 13: 63-72.

[8] Brian L smith, Billy M Williams, R Keith Oswald. *Comparison Of Parametric And Non-Parametric Models For Traffic Flow Forecasting.* Transportation Research Part C. 2002; 10: 303-321.

[9] Sangsoo lee, Daniel B Fambro. *Application Of The Subset ARIMA Model For Short-Term Freeway Traffic Volume Forecasting.* Transportation Research Board, Washington D.C. 1999.

[10] Yue Hou, Yuemei Mai. Chaotic Prediction for Traffic Flow of Improved BP Neural Network. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2013; 11(3): 1682-1690.

[11] Xiaodan Huang, Wei Wang. Traffic Prediction Based on Correlation of Road Sections. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2013; 11(10): 5523-5529.

[12] Hussein Dia. An Object-Oriented Neural Network Approach To Short-Term Traffic Forecasting. *European Journal of Operational Research.* 2001; 131: 253-261.

[13] Davis Gary A, Nihan Nancy L. Nonparametric Regression And Short-Term Freeway Traffic Forecasting. *Journal of Transportation Engineering.* 1991; 177(2): 178-188.

[14] Bin Xia, Qing Dong, Zuo Zhang. The Traffic Flow Prediction Model Study Under the Characteristics of Periodical Similarity. *Highway Traffic Science and Technology.* 2003; 20(2): 73-76.

[15] Ziqiang Luo, Guang-wei Zhang, Deyi Li. Probability Statistics Analysis of One-Dimensional Normal Cloud. *Information and Control.* 2007; 36(4): 471-475.

[16] Shaowei Yu, Kai Cao. Dynamic Soft Classifying Algorithm of Traffic Data Stream Based on Cloud Model. *Computer Engineering and Applications.* 2007; 43(28): 218.

[17] Huaitian Zhang. Data Mining Based on Cloud Model and its Application in Traffic Flow System. Tianjin: Tianjin University. 2006: 28-42.

[18] Lijun Zeng, Jingzhong Liu, Ang Li. Decision Tree Construction Based on Cloud Transformation and Rough Set Theory. *Computing Technology and Automation.* 2010; 29(3): 140-144.

[19] Li DY, Liu CY. Study on the Universality of the Normal Cloud Model. *Engineering Science.* 2004; 6(8): 29-33.

[20] Shaowei Yu, Kai Cao, Mo Zhao. Traffic Information Forecast Algorithm Based on the One-dimension Cloud Model. *Journal of ShanDong University (Engineering Science).* 2007; 37(2): 122-124.

[21] Xiangyu Du, Quanjun Yin, Kedi Huang, Diannong Liang. Transformation Between Qualitative Variables and Quantity Based on Cloud Models and its Application. *Systems Engineering and Electronic.* 2008; 30(4): 773-774.

[22] Qin K, Xu K, Liu FL, et al. Image Segmentation Based on Histogram Analysis utilizing the Cloud Model. *Computers & Mathematics with Application.* 2011; 62(7): 2824-2833.

[23] Li LL, Liu LX, Yang CW, et al. *The Comprehensive Evaluation of Smart Distribution Grid Based on Cloud Model.* Energy Procedia. 2012; 17: 96-102.

[24] Niu DX, Fan LL. *Technology Feasibility Evaluation of Grid-connected Distributed Generation Based on Cloud Model.* System Engineering Procedia. 2012; 4: 305-313.