Research on Prediction of Traffic Congestion State

Jun Zhao

Jiangsu Food & Pharmaceutical Science College, Huai'an, Jiangsu, China

ABSTRACT: This method of prediction using the data mining to analyze huge amounts of data as a preferred tool has been widely used in various fields. In the midst of it, the routine traffic data exists in a large number of isolated data in real time without establishing relationships with other data, and detects the amount of data which is greater than that at present. The usage of these data which is relatively shallow requires an in-depth analysis of its data model. Therefore, this paper uses a fuzzy clustering analysis method of feature points to study the traffic flow, uses a Markov decision chain model to study traffic jams, uses quantitative sample points based on the information entropy to calculate traffic flow trends and uses a heuristic prediction model to predict the road congestion. Through the simulation experiment which verifies the correctness of the model, this research is to advance the development of the road and to provide a basis for a dredging plan.

Keywords: intelligent transportation systems; data mining; fuzzy clustering; traffic

1 INTRODUCTION

Nowadays, the database system has been widely used; meanwhile, the storage capacity of much database management software has been increased a lot. In this age of big data, the mass data using should not only stay in the query function, but also be used to more effectively analyze, process and use data by advanced tools. How to use these data effectively will be the key to the data mining.

Nowadays, to provide decision support for requirement of intelligent transportation business, we process and analyze data through the data mining in the intelligent transportation field, and we also continue to process the heterogeneous data and multivariate data by using a large number of intelligent algorithm and platform, such as the prediction of the traffic intersection congestion or change of some regional traffic flows.

2 TRAFFIC FLOW PARAMETERS

In the field of intelligent transportation, the main research object is information which includes the analysis of density, speed and individual number of the traffic flow and people flow on the road, and some aspects such as city traffic scheduling tasks, traffic control, signal lamp arrangement, real-time data monitoring and special service assistant and so on. Due to the limited space, only a few typical examples are selected. The traffic flow [1] which is a description of a person on the road or the vehicle flow form is a classification of the floating object since the analysis of the pedestrians in intelligent traffic isn't very important. Thus, we usually carry out a research on the motor vehicle.
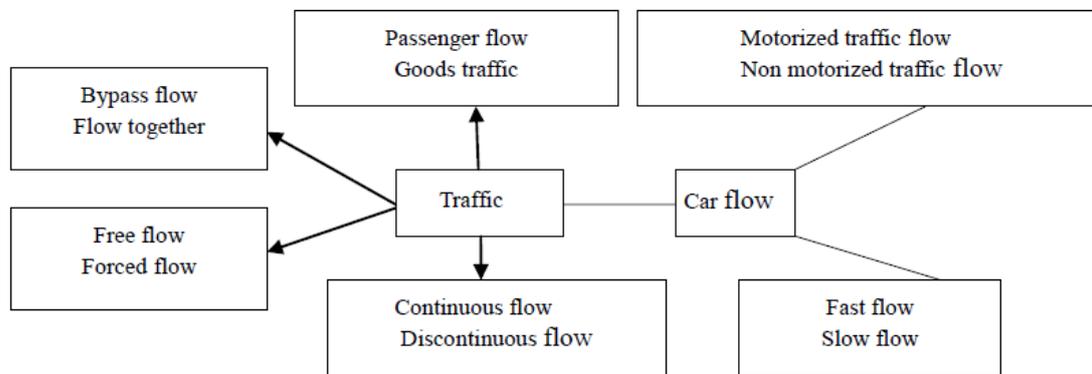

Figure 1. Description of the traffic flow classification diagram.

Since the traffic flow is a motorial state, it needs some parameters, such as the dynamic vehicle or people's quantity and the speed of individual flow (which can acquire individual speed, and also use average speed to calculate area density), regional individual density and so on. We can further analyze the relationship between the classification on traffic through this three physical description parameters.

(1)Traffic

Traffic flow indicates the number of vehicles on a road section, and we can acquire the number of vehicles through the camera which is installed above the road or the ground sensing coil which is deeply buried under the ground. In a statistical process, we need to analyze an important parameter, namely the headway of H, which means that it requires time for vehicle to pass the same place. When we collect data for traffic, we often use the parameter h=1/H as the traffic volume statistics[2].

(2)Average speed

The average speed means that in a period of time, all speed of vehicles that go through a section of road will be used to calculate the arithmetic average value, and the calculation formula is shown in formula 1 [2]:

$$v = \frac{1}{n}\sum_{i=1}^{n} v_i \qquad (1)$$

(3)Density

Traffic flow density directly reflects the vehicle congestion in a local area which is usually using the share of the vehicle in a certain area when release time. For a certain level of the fleet, the more time of traffic intersection it possesses, the more congestion it is, and vice versa.

Collecting density information on the highway, we generally use the induction coil as the detection tools of an area (which is usually a section of the road), and arrange ground sensing coil on both sides of the road. First, calculate the quantity and speed of vehicles in the area. Then, take the average value (or difference) as the basis of next step, which is according to the vehicle's number and the speed off the road. Since each variable of traffic flow is affected by many uncertain factors, therefore, we need to calculate a lot of dynamic changes of parameters using average values. For example, the vehicle is random, so when we count up the figure of traffic flow, we can use the average flow over a period of time to present the general situation during this period of time. The methodology which is from random transferring to regular transferring is widely used in the traffic data processing. All this three parameters are obviously random for the traffic volume, speed and density. We can calculate the average value as the index of the traffic flow model, and their relationship can be expressed by the following formula:

$$\rho = q / v_s \qquad (2)$$

3 TRAFFIC FLOW ANALYSIS

According to the massive traffic data, looking for the law of traffic flow is the main purpose of the traffic flow analysis [3].The analysis can be done through two methods. One is the description, which means that it is used to abstract the rules existed in the data or packet the characteristic data; another one is the prediction, which means that it is used to figure the results out through the exact values of data, and obtain the characteristics of the trend of data change. Both two kinds of methods of expression mentioned earlier can be used to explore the data of traffic flow, and this paper adopts the prediction method and uses the method of fuzzy clustering analysis, which is expressed as follows:

The basic ideas of fuzzy cluster analysis:

The data set under the classification is shown as follows:

$$X = \{x_1, x_2, ..., x_k, ..., x_n\} \qquad (3)$$

Set the data mentioned earlier into C classes. If a random sample x_k to the I class's membership is u_{ik}, where $u_{ik} \in [0,1]$ and $\sum_{i=1}^{c} u_{ik} = 1$.

To restrict it in the process of clustering, we need to set the criteria for clustering fuzzy clustering:

$$\min J(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m (d_{ik})^2 \qquad (4)$$

Wherein, m stands for the ownership of the change trend of data clustering; d_{ik} represents the K samples x_k to the I cluster sample of center v_i distance. That is, through the calculation method of center distance of samples, it can be calculated using the following method:

$$(d_{ik})^2 = \| x_k - v_i \|^2 = (x_k - v_i)^T A (x_k - v_i) \qquad (5)$$

In the formula mentioned earlier, there is a sample data existed in symmetric matrix A, which uses the Lagrange multiplication to calculate those two conditions which meet the distance of sample center:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}(\frac{d_{ik}}{d_{jk}})^{\frac{2}{m-1}}} , \ u_i = \frac{1}{\sum_{k=1}^{n}(u_{ik})^m}\sum_{k=1}^{n}(u_{ik})^m x_k \qquad (6)$$

Due to the fuzzy concept's complex multi parameter dynamic changes model, the comparison is applied to the description such as traffic flow, and the intersection traffic jams in different periods will be evaluated by means of fuzzy clustering method. Assume the sample set X consists of C samples, the number of samples per class pi. (i=1, 2, 3…, c), then the total number of samples is P. And the sample set for [5] information entropy is shown as follows:

$$I(X) = -\sum_{i=1}^{c} \frac{p_i}{p} \log_2 \frac{p_i}{p} \tag{7}$$

If the sample sets attribute A as a decision tree roots, then A is with classification V which is $v_1, v_2, ..., v_v$. It divides the samples into v subsets $\{E_1, E_2, ..., E_v\}$. If Ei has a number of J class samples, then the number of samples in the subset E_i is $|E_i|$, and the subset's information entropy is $I(E_i)$, which is expressed as an information entropy with

$$I(E_i) = \sum_{j=1}^{c} -\frac{p_{ij}}{|E_i|} \log_2 \frac{p_{ij}}{|E_i|} \tag{8}$$

The attributes of A root classification are shown as follows:

$$I(A) = \sum_{i=1}^{v} \frac{|E_i|}{|E|} I(E_i) \tag{9}$$

The information gain is shown as follows:

$$G(A) = I(S) - I(A) \tag{10}$$

Select properties A^-. Make $G(A)$ become the maximum, and take this attribute as the root node of classification, repeat the process mentioned earlier, and form a decision tree [4]. The specific algorithm is described as follows:

(1)Create a tree structure's root node N;

(2) In the calculation of the leaf node, which is in accordance with the method of information entropy comparison mentioned earlier, clusteredly calculate the sample according to the fuzzy clustering algorithm, and take the same type of data as a leaf node which is linked to the root node;

(3)If the composition of the sample data node on node list is empty, then the current node is independent of the presence of isolated point;

(4)Carry out a data test on the tree list, detect the sample node's all successor pointers of leaf nodes, and take the current node as the final node. If the succeeding pointer is not null, then return to the step (2) and continue to calculate the next node until traverse all the samples' dates and establish the decision tree.

Use the method of preorder traversal to read the node in the decision tree, calculate from the decision tree's root node to the leaf node in the way of cluster nodes using IF-THEN structure.

Take the fuzzy clustered results by the calculation method mentioned earlier as the motion mode of vehicle on the road. Analyze behavior in different sections and time domain, predict next behavior of vehicles based on the decision tree traversal results, and provide the basis for forecasting the traffic blocking condition.

4 ESTABLISHMENT OF THE CONGESTED TRAFFIC STATE FORECASTING MODEL

4.1 Establishment of the congested traffic state forecasting model

Through the definition of node information in traffic flow, we can know if we give all the vehicles around intersections the information increment which can reflects the congestion state, then we can assess the local traffic state. This paper uses the collection function and the Markoff decision model as a description of this state.

Define the traffic state of local environment as, each vehicle s $d_i \in D$, wherein D represents a collection of all the vehicles in the local environment. In order to calculate the characteristics of local traffic flow, it is defined as the collection function that followed to indicate the local traffic flow's relieved or tense conditions.

At t time, the traffic network G's utility is. We use it to indicate the number of vehicles waiting in the intersection at every moment. The traffic light signal's change is defined as. According to the "red light stop, green line" rule, the utility of the traffic network is shown as follows:

$$EU(G,t) = \sum_{v_i \in V} \sum_{v_j \in n(v_i)} \sum_{v_k \in \{n(v_i - v_j)\}} T\zeta_i^j(k, t+1) \tag{11}$$

Among them, $n(v_i)$ represents the number of vehicles waiting at the intersection.

According to the characteristics of the Markoff decision chain, $EU(G,t)$ can be expanded into the form $(S_{d_i}, A, T, Tf(v_i))$. This paper uses the following to define the Markoff decision chain elements:

(1)The state $S_{d_i}(t)$ represents vehicle $d_i \in D$ in the moment t where the state of the junction, including density, velocity state and so on.

(2)According to the traffic lights, action A represents vehicle's behavior in the crossing v_i, including the stop and the form.

(3)The vehicle state conversion method is defined as the moment of the traffic situation, which is expressed by the calculation of the t+1 state.

4.2 Establishment of the prediction model

Though vehicle clustering, we can make the description of the biggest intersection node $G(A)$ as a quantitative congestion. In the Markoff decision chain, we take the judgment of $G(A)$ as the target of traffic condition prediction. This paper uses a heuristic prediction model to analyze it.

First, define the time of a traffic light, the number of vehicles through a junction is θ. Then, the whole vehicle state can be obtained with the transfer function which is shown as follows:

$$Td_i^j(k,t) = \begin{cases} Tf_i^j(k,t), Tf_i^j(k,t) < \theta \\ \theta, others \end{cases} \tag{12}$$

Second, the establishment of t t+1 from time to time the traffic flow changes state, the main analysis in unit time of vehicle through the intersection number and again gathered the number of intersections. The calculation method according to the definitions of vehicle distance in Formula 6 can be calculated per unit time. As the number of vehicles is gathered, the information entropy is increased as (Formula 10.) calculation $\Delta G(A)$.

The number of vehicles in the intersections can be calculated by the following formula:

$$P_i^j(k) \times \sum_{v_i \in n(v_j) and v_i \neq v_j} Td_j^l(i,t) \tag{13}$$

Vehicles through the intersection are based on the method of probability $P_i^j(k)$.

Take the calculation results mentioned earlier into consideration, the traffic congestion at the t+1 moment can be expressed as follows:

$$Tf_i^j(k,t+1) = Tf_i^j(k,t) + Tc_i^j(k,t) - Td_i^j(k,t) \tag{14}$$

4.3 Simulation experiment

This paper uses NetLogo 5.0.3 as the experimental platform, proposes the traffic congestion state forecasting model, and can be used to verify the junction traffic congestion state after a specific reference moment.

In this experiment, the model clock is a second, all the 200 vehicles around the junctions of traffic lights randomly turn left, turn right or move straight at the intersection. And according to the cross roads' red or green lights state, the waiting time is 2 seconds. In order to control this experiment, this paper selects second x second blocks to analyze the traffic state. It also uses the display window to record the second x second blocks at the current moment of the vehicle congestion and moment conditions at the same time. The results of its operation are shown in Figure 2:
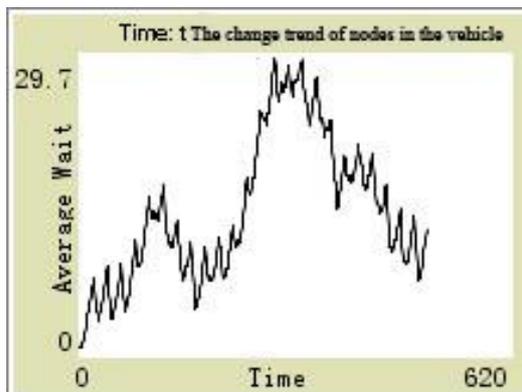
(a) t moment
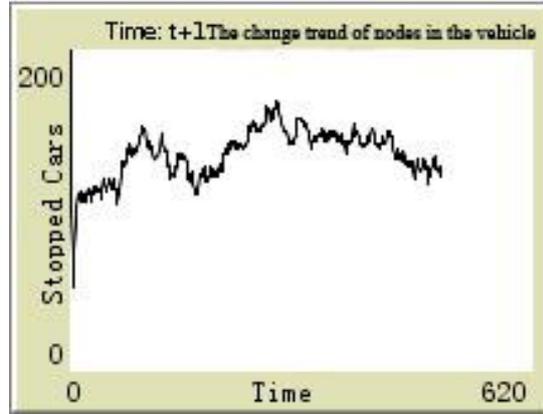
(b) t+1 moment

Figure 2. Current time and time of vehicle congestion contrast

Compared with the current moment and the t+1 moment vehicle congestion, we can see that the node traffic condition and the trend prediction of the t moment and t+1 moment are approximately the same.

5 CONCLUSION

In this paper, we analyze the traffic flow by using fuzzy clustering method to establish the information entropy and the model of traffic congestion on this basis of which uses the Markoff decision chain as a tool of analysis to judge the traffic congestion at current moment and next moment, validate the model by experiment, and achieve a better result. This method can be used to help the intelligent traffic to a certain extent, make the scheme of road guidance in advance. Therefore, it has a significantly practical meaning.

REFERENCES

[1] Wu Zhongze. 2013. The current situation of the development and future trend of intelligent transportation industry in China, *low carbon world*. (11).

[2] Sun Duo. 2011. Study on phase transition phenomena in cellular automata traffic flow, *University of Science and Technology of China*. (10).

[3] Liu Ying et al. 2012. Study on the improvement of the fuzzy C- mean clustering algorithm. *Journal of Jilin Institute of Chemical Technology*. (11).

[4] 2011. The principle of data mining technology in integrated transportation platform and the application in intelligent traffic [DB/OL].

[5] Song Bi, Liqun Han, Yixin Zhong. & Xiaojie Wang. 2011. All-day traffic states recognition system without vehicle segmentation. *The Journal of China Universities of Posts and Telecommunications*, (04).