Jestr Journal of Engineering Science and Technology Review 6 (2) (2013) 95-99

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Research Article

Investigation on Comprehensive Multi-point GPS-based Traffic Information Treatment

Guanli Huang[1,2], Meng Zhou[1,*] and Jiangyi Lv[2]

[1]*School of Mathematics and System Science, Beihang University, Beijing100191, China*
[2]*Beijing Polytechnic,Beijing, China*

Received 15 May 2013; Accepted 25 July 2013

Abstract

GPS data sets of several vehicles on the road are obtained via wireless network, from which specific sections of data are picked up by filtering algorithm. Then, a variety of fusion algorithms are applied to the several sets of specific sections of data to remove the error data, to guarantee the veracity of the fused result. On the basis of valid data, traffic conditions of this section can be acquired through the smart identification algorithm, which will provide the real-time traffic information. The method presented in this paper is simple and reliable. With good performance on the experimental data, it opens a new door for information collection of urban road.

Keywords: traffic condition determining, Data fusion, identification algorithm

1. Introduction

With city expansion and traffic load increasing, traffic condition of urban road is becoming much more complicated, which makes inconvenient drive within city. Therefore, a method, capable of traffic condition real-time acquisition and feedback, is necessary to improve the efficiency of urban traffic. Traditional methods, based on information collection via surveillance cameras and radar system mostly, generate traffic condition available to users by artificial interpretation, which are costly and difficult to cover the entire traffic network completely. With development of the Internet of things and its correlative technologies [1], it is more convenient to acquire the node data of wireless sensor, the same as GPS data sets of several cars on the road. Meanwhile, widely used digital map also make it possible to generate traffic conditions of vehicle-related road automatically from GPS data. On the basis of mentioned above, it is possible to realize users' complete understanding of real-time traffic condition under relative low costs of operating and maintenance.

2. Data collection and storage

2.1 Data collection

Since the United States stopped executing SA (Selective Availability), civilian used GPS accuracy has been stable at about 10 meters. This accuracy is the foundation of investigation on multi-point GPS-based traffic information data processing, as accurate traffic information obtained only from accurate GPS data. With GPS platforms installed on the vehicles to form multiple GPS sensors, information of time, vehicle location, speed and direction and so forth is acquired by receiving GPS signals at any moment. In general, these GPS sensors isolate from each other can only passively receive GPS information, without communicating and transferring information with outside. However, the internet of things and wireless communication technology enable GPS sensor to send data to a server on the Internet, making use of GPRS, CDMA and other mobile communication technology. By establishing a database on the server, data can be stored and used for further processing. Although 3G mobile communications era has been coming, GPRS of 2G, outperformed 3G on width of coverage area, stability of communication quality, cheapness of communication cost and reliability of data transmission, is still the preferred choice [2].

When there is a communication connection established between GPS sensor and database server, data transmit via a custom protocol, which includes header, address, command, data, checksum and trailer, etc., to ensure the reliability and accuracy of data transmission effectively. When the number of sensor nodes is relative small, a single database server can achieve data collection and storage. While the number of sensor network nodes is large, up to levels of ten thousands, the amounts of data will be huge. With data continuously increasing, it is inevitable to result in inefficient storage and operation of the database, which leads to real-time damage. To solve this problem, a method based on distributed data processing (DDP) for data storage and query is adopted in this paper, shown as Figure 1.

Distributed databases are established on multiple servers. GPS cars are divided into several groups, each group corresponding with a different database server. Then, GPS data from all cars are stored in the distributed databases,

* E-mail address: zhoumeng1613@hotmail.com

which can provide data query and other operations through a Web service. When data is demanded for processing by user service program, central database sends a request of query to network via a Web service reference. The request of query is distributed processed [3], and only the query results return to the central node to complete the subsequent calculations. During the non-request data period, the central database can periodically get latest data from the distributed databases and store them in the central database.

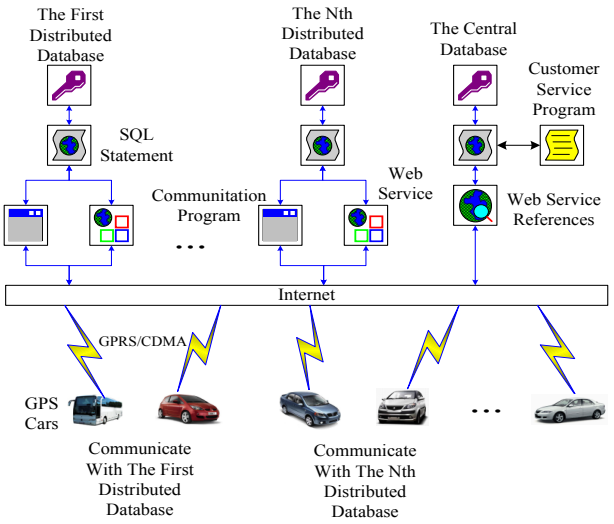

Fig. 1 Distributed data processing diagram

Adapting a strategy of combined changing acquisition time and distance intervals, can reduce the power of data collection and storage to the lowest point. The implementation of this strategy is expected to achieve minimum consumption under sufficient collection information. For stand-alone section (no export between start and end), road length denoted as L_r, it is expected that the sensor travelling in this section collects at least N pieces of information, and updates traffic condition every certain period of time, denoted as t_e. Taking into account the system response speed, the distribution of urban roads, car speed and other factors, distance interval ranges from the minimum 100 meters up to 1000 meters, with initial value set to 200 meters(d_0=200m). And the initial value of time is set to 3min, with minimum interval to 1min and maximum to 30min.

For a single sensor, the relationship between time interval and distance interval is determined on the following formula:

$$V_{\arg} = \begin{cases} V_k & V_{k-1} \le 0 \\ \sum_{i=0}^{m} V_{k-i} / k & m < N \end{cases} \quad (1)$$

$$d = \begin{cases} d_0 & V_{\arg} > 0, V_k > 0, V_{k-1} = 0 \\ d_{\min} & L_r / V_{\arg} < N \\ L_r / N & \end{cases} \quad d \ni [100,1000] \quad (2)$$

$$t = \begin{cases} t_{\max} & V_{\arg} \le 0, V_k = 0 \\ t_e & V_{\arg} > 0, V_k > 0, V_{k-1} = 0 \\ L_r / V_{\arg} & V_{\arg} > 0 \end{cases} \quad t \ni [1,30] \quad (3)$$

In the formula:

V_k -- Collection speed of the current time

V_{k-1} -- Collection speed of the last time

V_{k-i} -- Collection speed of the No. k-i time

$V_{\arg}$ -- Average speed of the current time

d -- Distance interval

t -- Time interval

To reduce the computational resource to achieve the above strategy, conditioned burst mode is adopted. For a single sensor tracking, strategy changes only when both the road (from a section to another) and vehicle status (stop, start, etc.) change. For multiple sensors, strategy within a section changes only when traffic condition of the section is tracked. Besides the two cases mentioned above, default values are used.

It is proved in practice that the strategy can decrease the number of data collection on condition that the data collection is sufficient [4], thereby reducing the transmission and storage consumption.

2.2 Data screening

When the number of data collected from GPS sensor is large, it is necessary to process data screening to facilitate the subsequent calculation of traffic condition. By data screening, only valid data are used for calculation, which can compute more efficiently.

Data screening is actually a process that a central database sends request to distributed databases via web service reference and gets result. Considering that traffic condition is always about specific section, the requested data should be related to that section as well. In this system, section of road is aggregate of line segments depicted by multiple points of a geographic information system database, as line segment OABC shown in Figure 2, where the location of point is described by its latitude and longitude.

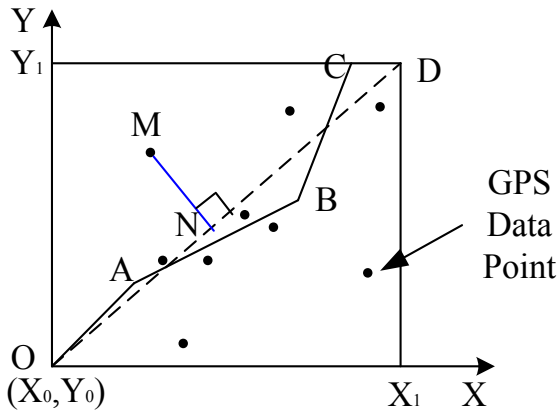

Fig. 2 Schematic diagram of the data screening

In order to improve computational efficiency of data screening, section depicted by multiple line segments or

curves will be replaced by a fitting line segment according to Least-Square Method (LSM). Line segment OD in Figure 2 is a fitting road section by applying LSM on four points of line segment OABC, section L denoted as:

$$Y = aX + b \quad X \in [X_0, X_1] \tag{4}$$

Where a and b is fitted by the coordinates of O, A, B, and C. We use d_{max} denote the maximum distance from the fitted points O, A, B and C to the fitting line, α_{max} the maximum angle between fitting line OD and line segments OA, AB and BC, α_{OD} the angle between fitting line OD and X axis. Then, a, b, d_{max}, α_{max}, α_{OD}, $[X_0, Y_0]$ and $[X_1, Y_1]$ will be stored in database as parameters describe road section.

Considering the width of the road (denoted by r_w) and the GPS measurement errors (denoted by σ_{GPS}) , for any point M(X, Y), it will be screened out once satisfying the equation (5).

$$d_M = \frac{|aX - Y + b|}{\sqrt{a^2 + 1}} \le d_{max} + r_w + \sigma_{GPS}$$
$$X \in [X_0, X_1], \quad Y \in [Y_0, X_1] \tag{5}$$

In addition, traffic condition is generally determined on two directions respectively, so that the direction of GPS cars should be taken into consideration. The space angle of certain GPS car is given by sensor, and the difference angle of a road is theoretical to be 180 degrees. Comparison between space angle and angle of road can determine data from different directions. Let ∇_{GPS} denote integrate angle error, which include GPS measurement angle error and errors induced by road and vehicle variation. Assuming that angle of a certain section is α , its presented driving direction can be determined as follow:

$$\begin{cases} |\alpha - \alpha_{OD}| < \alpha_{max} + \nabla_{GPS} & Direction1 \\ |\alpha - \alpha_{OD}| < 180 + \alpha_{max} + \nabla_{GPS} & Direction2 \end{cases} \tag{6}$$

As there is usually a time constraint for determining traffic condition, data screening should also comply with a time constraint. In practice, several factors should be considered for road division, not only calculation, but also actual situation, especially for section consisting of multiple large angle corners.

When data is requested, central database transmits the parameters and time constraint of corresponding road section via a web reference to distributed databases, which screen out data via equation (5) and (6) and return results to central database for storage.

2.3 Abnormal data processing

After the screening, the data meet the requirements in time, space and direction aspect. But abnormal and unreasonable data from levity of vehicle status will lead to the error of subsequent state judgment [5]. So, to make sure the veracity, it needs to effectively eliminate the abnormal part through the data consistency check. The check grounds on the relativity and complementarities of multiple GPS data in spatial and time domain.

When determining traffic condition of a certain section, there are multiple sets of data from some GPS cars, which characterize the same section. As high correlation of GPS cars' speeds and little variance during short interval, C-means clustering algorithm is performed to eliminate the abnormal. Given that speed is positive, three clusters, down abnormal clustering, effective clustering and up abnormal clustering, are divided. Assuming the length of data is N and speed of each data is Vi，i=1,⋯,n, the initial clustering center of the three type of clustering is given by:

$$\begin{cases} E_d = \min(V_i) \\ E_e = \sum V_i / N \\ E_u = \max(V_i) \end{cases} \tag{7}$$

In the formula:

E_d --Down Abnormal Clustering

E_e --Effective Clustering

E_u --Up Abnormal Clustering

When the initial clustering centers are given, the rest of data will be assigned to similar clustering respectively according to their similarity with the clustering centers (the similarity is presented by the speed difference), and the new center of each clustering is updated by the mean of all speeds in the clustering. Repeat the process until convergence of standard measure function to the minimum. The standard measure function is the sum of standard deviation of three clustering, demonstrated as:

$$E = \sum_{i=1}^{3} \sum_{p \in C_i} |p - m_i|^2 \tag{8}$$

Where E is Sum of all mean square error, p is a point in the clustering, C_i is data aggregate of each clustering and m_i is the mean of clustering C_i. Here p and m_i are variable speed.

Through this clustering analysis, three types of clustering are obtained. When distance between centers of up abnormal clustering or down abnormal clustering and effective clustering exceeds a certain value (experiential value), data in that clustering is eliminated as abnormal, otherwise retained. For some special cases, such as the amount of data less than three, the nearest time node is chosen as effective data via recent priority principle [6].

2.4 Data fusion and condition determining

As data from each GPS sensor has different characteristics, fusion technology may take full advantage of these resources. Through reasonable control and use of these sensors and measurement information, by combining redundant or complementary information of multiple sensors in space and time domain in terms of a certain criteria, interpretation or description concerning consistency of the measured object can be obtained, thereby enhancing determining effectiveness of traffic conditions. In this paper, data fusion is achieved by a weighted algorithm base on improving average speed.

In order to reduce the determining errors resulted from individual events, a small time period (e.g. 5 minutes) is set to collect all GPS data in one section in this paper. Data

fusion and integrate processing are performed on these data, to accomplish traffic conditions determining.

In this paper, once section and direction of road is selected, collected GPS information can be denoted by vector P (R, V, t), where R is the position of test time, V is the speed of the test time, t is the test time. Direction of all the vectors in the section is indicated by Figure 3, where each line segment represents a vector P, locations of black points represent R and t in the vector P, and the length of line segment represents the speed V.

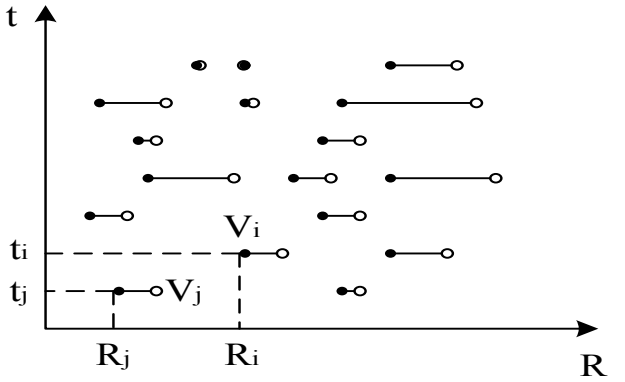

Fig.3 distance between two points at different time and different locations

As data for condition determining may not be in the same time and same location, analytical method of random process is applied to data, which gives us equivalent data at the time and location needing traffic condition determining. That is to say, examine the probability of possibility that these data move to the test time or test location under current speed. Then the probability will be used for weighted average as the weight [7]. Hence, it is requisite to acquire the probability of possibility that a single vehicle drive to test point under current speed.

For condition of a single vehicle on urban road with initial speed denoted as V, in the time range $[t^0, t^0 + \Delta]$, consider a situation that the vehicle can not travel in speed V for road congestion and other reasons, but travel back to normal speed after a good turn. It makes the incident that the vehicle can not drive with speed V due to traffic condition a random point process. If state 0 indicates that vehicle travels with speed V and state 1 indicates vehicle can not travel with speed V, a counting process ($\{N^{t_2}: t \in [t^0, \infty]\}$) can be used to describe the incident that vehicle can not travel with speed V, which is characterized as follow:

(1) $P\{N_{t0} = 0\}$=1, it is under state 0 at the initial time with probability equal to 1.

(2) For any positive integer n and $t_i \in [t_0, \infty]$, i=1,2,…,n, $t_1 < t_2 < \ldots < t_n$, if increments of random variables $\{N_t\}$, denoted as $N(t_2) - N(t_1)$, $N(t_3) - N(t_2)$,…, $N(t_n) - N(t_{n-1})$, are mutual independent random variables, and the probability distribution of each increment has nothing to do with t, $\{N_t\}$ is a smooth process of independent increments.

(3) For any $t > t_0$, we have $0 < P\{N_t > 0\} < 1$. It means that in the time interval $[t_0, t)$ the probability of "random points" occurrence is positive, but not necessarily "random points" occur.

(4) For any $t \geq t_0$, we have

$$\lim_{h \to 0} \frac{P\{N_{t+h} - N_t \geq 2\}}{P\{N_{t+h} - N_t = 1\}} = 0 \tag{9}$$

Namely in sufficiently small time interval, at most a "random points" occur.

Thus, according to probability theory, incident $\{N_t\}$ that a vehicle can not travel with speed V is a Poisson process, meaning

$$P\{N_{t+s} - N_s = k\} = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, k = 0,1,2... \tag{10}$$

Where λ denotes the average number of incident occurrence during a unit of time.

Here λ can be determined by historical data. For example, examine a time interval $[t_1, t_2]$, the historical average times and time of road congestion denoted as M and t respectively, then λ can be approximately calculated as follows:

$$\lambda = \frac{M t}{t_2 - t_1} \tag{11}$$

Take GPS data collection time as the initial time, so the time t when the vehicle moving to the test point is determined on the speed, location, time of collection point and location, time of test point and other parameters. There are four cases of relative condition between information of GPS collection point and test point. Assuming that the location and time of test point k is (R_k, t_k), and the vector of GPS collection point is $x_i(R_i, V_i, t_i)$, t can be calculated by equation below:

$$t = [(|R_i - R_k| / V_i) + |t_i - t_k|]/2 \tag{12}$$

The physical meaning is after t time units, the vehicle at the collection point arrives at test point k, or meets the test time. The probability P_i of the possibility that the vehicle drives from GPS collection point to the test point normally should be the probability of no congestion occurring at all during that time.

$$P_i = 1 - P\{N_{t+s} - N_s \geq 1\} = P\{N_{t+s} - N_s = 0\} = e^{-\lambda t} \tag{13}$$

If the total number of collection points is n, average speed V_k of this section in this moment is performing weighted average on the speeds of n-point, which is:

$$V_k = \frac{\sum_{i=1}^{n} P_i V_i}{\sum_{i=1}^{n} P_i} \tag{14}$$

In terms of traffic rule about speed under various traffic conditions in various sections of road, traffic condition can be determined by the speed V_k.

3. Conclusions

As the amount of information of current collected GPS data is relatively small, the algorithms mentioned in this paper are all for a single section. However, with the increasing of GPS cars, more sections covered by real-time collected data, data screening will no longer be aimed at a single road, but directly an area. Therefore, how to obtain regional multiple points' traffic condition, based on information of a large number of GPS network collection points, is still needed further investigation. Such as fuzzy clustering method, select the appropriate initial clustering centers based on information of each section within the region, divide the collection point information into several types of clustering [8], by integrate information from all types of clustering, and consequently achieve a traffic condition determining of several road sections.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (NSFC 11271040) and the 2010 Scientific Fund of Beijing Education Commission (ITEM NO.KM201000002002).

References

1. Guanli Huang, Hongxue Yang, Xiaoming Liu, "Design on Comprehensive Treatment of Multi-point GPS Traffic Information", Computer Measurement & Control，2011,19（12）,pp.3041-3043.
2. Zenghui Liu,"Research development of position sensor technique ", Transducer and Microsystem Technologies,2005,24（10）,pp.5-6.
3. Guanli Huang, Hui Wang, Huaping Xu,"Research on drifting of GPS positioning based on temporal series", Computer Engineering and Applications，2008，44 (31),pp.94-97.
4. Yunli Zhu, "Multi-site Data Collection Monitoring System Based on 802.15.4 Protocol", Instrument Technique and Sensor， 2009,2,pp.65-69.
5. Xiaowu Cheng, Qinping Zhao, "A Method of Generating Multi-resolution Smoke Based on the Obscured Visibility in Virtual Environment", Chinese Journal of Computers，2002（3）,pp.285-291.
6. Yunli Zhu, Yanfeng L, "Research and Implementation on Intelligent Monitoring System for the Ship", Computer Measurement & Control，2009，17(5),pp.893-896.
7. Huaping Xu, "The Equivalent Analysis of Direct Geocoding Model and Spaceborne INSAR Altitude Model in Height Uncertainty", Journal of Electronics & Information Technology，2010，1,pp.49-53.
8. Xiaoyin Bai, "WebResearch on Web Service Testing", Computer Science，2006，33（2）, pp.251-256.