2015 IEEE International Conference on Multimedia Big Data

Analysis and Visualization of Traffic Conditions of Road Network by Route Bus Probe Data

Dechen LIU, Yusuke KITAMURA, Xin ZENG,
Shunsuke ARAKI and Ken'ichi KAKIZAKI
Faculty of Computer Science and Systems Engineering
Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan
{liu-dechen, kitamura, so-kin}@macross.cse.kyutech.ac.jp
{araki, kakizaki}@ci.kyutech.ac.jp

***Abstract*— In recent years, route bus service systems have been collecting and accumulating bus probe data such as GPS positions and traffic speeds from all route buses, day after day. In the present paper, we propose a method to analyze the conditions of a road traffic network from a huge amount of the route bus probe data. First, we show a conversion method from the probe data to a road segment data set. Second, we introduce analysis methods that compute the degree of congestion and reachable areas from the road segment data set. Finally, we discuss road segment data structures for rapid calculations of these analytical methods.**

Keywords-probe data; degree of congestion; reachable area; time-position information of bus; big data

I. INTRODUCTION

The MLIT [1] (Ministry of Land, Infrastructure, Transport and Tourism) of the Japanese government reported that about 3.8 billion hours, whose economic worth equals about 9 trillion yen, are wasted due to traffic congestion in urban areas of Japan every year. Therefore, it is urgent and necessary to solve or relieve the problem of traffic congestion. To formulate a road development plan, firstly, the problems of the present road traffic network in the area must be understood. In order to analyze current traffic conditions, we must prepare some basic information such as records of where, when and what level traffic jams occurred. However, collecting such basic information of an entire region solely by humans is not realistic. For that purpose, we adopt methods of recording vehicle data such as coordinates and vehicle speeds by probe cars, which mount GPS, and we convert this road segment data from the probe data. Probe cars can be classified into three types: general vehicle, taxi and bus. There are some researches using private cars data to analysis traffic conditions. This research [2] employs 17,241 private cars to collect data over a period of one week. Even though it can collect data from all over the urban area, there is a problem of whether the data of one week could reflect the traffic conditions of one year, aside from problems of private information and cost. Therefore, it is difficult to use the probe data of private cars for traffic conditions analysis. As another source of probe data in [3] taxies were employed. There taxies can also be used to collect probe data, but the traveling situation of taxis depends on the passengers. The taxi has a trend toward avoiding routes with traffic jams, and probe data by taxis may still not cover the whole of urban area. It is also unrealistic to obtain data on all main roads over several years since we want to analyze the traffic conditions over less than one year. Therefore, taxis are also excluded.

On the other hand, probe data collected by buses have no problems with private information and cost. They travel almost all main roads over the entirety of urban areas on a fixed trajectory again and again, especially those main roads in densely populated areas. Therefore, data collected by buses is more effective for the analysis of traffic conditions. Another reason that bus data is used in our method is that our method contains almost all situations that appear over several years. Therefore, analysis of this data can reflect the traffic conditions of a whole year accurately rather than speculate the possible situations through the data of a period of time, such as one week.

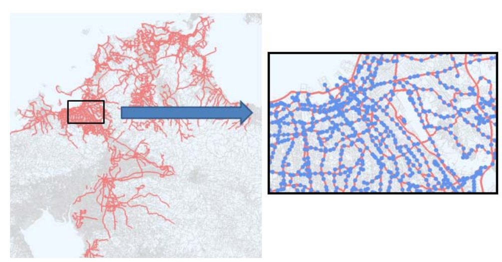

Figure 1. Traveling information of route bus.

In [4], Matsunaka and others showed an analysis method of the traffic conditions by using the bus-probe data. However, their method only converts the original probe data to average speed data of each road segment and use this information to analysis the traffic conditions. For the constantly updated data every day which made the data more complicated, their method cannot retrieving data quickly.

Information about traveling conditions can quickly become a huge amount of data. As shown in Fig. 1, we assume that traveling information of route bus is collected by Nishi-Nippon Railroad Co., Ltd. The company which is one of the largest bus companies in Japan has 2873 buses running on 3229

978-1-4799-8688-0/15 $31.00 © 2015 IEEE
DOI 10.1109/BigMM.2015.75

bus routes, and 7772 bus station on these bus routes, running a distance of 160 million kilometers, daily running distance can circle the earth 11 rounds. The big data is accumulated with such a large amount of data by every day.

We showed an outline of this method of analyzing traffic conditions by using accumulated time-position data recorded by buses in [5].

In this paper, we propose a method to analyze the road traffic network by using bus probe data such as the degree of congestion and reachable areas. Our method uses a huge amount of probe data collected by buses. As a result, we converted the collected time-position data into basic data that is more easily usable. Then, if road segment data is defective, for instance there is no road segment data for one road segment or more than one road segment data for the same place, we convert the road segment data into statistical data. Based on this, we did further analysis of statistical data that was obtained to extract the intensive data which can be used directly.

II. Brief Overview of Bus Probe Data

In the existing systems, devices installed on buses record statuses such as the opening of doors, stopping with time-positions by GPS and vehicle's speeds, and sends the data to some servers during business hours [6]. The time-position data of bus is recorded by two factors. One is generated on a certain time interval. Another is generated at some events such as station stop, etc. The data generated by two different reasons, but they were mixed recorded in the same format. The main data of record as follows:

- The number of bus route system
- Bus route number
- Date and time of the GPS
- Location information of the GPS
- The running status of bus
- The running speed of bus

III. Analysis Method for Road Traffic Network by Time-Position Data of Bus

In this section, we show a method for converting time-position data of buses into road segment data that can analyze road traffic networks easily. After this process, the degree of traffic congestion that can represent the congestion situation of an identified road section, and reachable areas that can represent the distance from a specified position to the reachable position are also shown.

A. Calculation of Road Segment Data

As we know, it is necessary to obtain a data set that can represent the traffic situation of individual road sections to analyze the degree of traffic congestion overall. However, it is impossible to extract the essential data from the time-position data of the individual bus directly. For this reason, we need to convert the time-position data into individual road segment data which can represent the situation of each road segment, and then extract the data associated with each road segment.

In our research, road segments mean the road sections that use intersections to divide the road.

The constitution of road segment data is as follows:

- Identifier of road segment
- Identifier of bus route and number
- Year, month and day and day of the week
- Distinguish information of starting node
- The pass time of starting node
- The pass time of terminal node

A method for calculating the passage of time of the starting node and the terminal node is shown as follows:

1) Specify the number the route bus and read the specified data from the time-position data of bus.
2) Match the map with the location of the data.
3) Extract a pair of time-position data points which were close to the focused intersection.
4) Calculate the passing time of intersection by using pairs of time-position data which cross the intersection.
5) Save the data as road segment data.
6) Repeat from Step 1) to Step 5) .

Due to the time-position data obtained by GPS receiver, the allowable error of the latitude and longitude would be about 10m. Position information about the road would deviate from the actual road.

In Step 2) map-matching is used to match the deviated position data to the correct position in real situations. One of the existing methods [7] matches the maps by comparing the distance of the road and the position of the vehicle and divides the map into meshes to reduce the calculation of the distance. However, our method only calculates the distance between time-position data and road segments that need be compared to each distance.

The proportions of the distance between neighboring time-position data are used to calculate the pass time at each intersection in Step 4). For example, the time-position data of A, C, E shown in Fig. 2 is used to calculate the pass time of intersections B and D. The process is shown as follows:

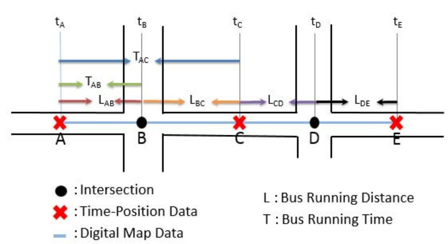

Figure 2. Matching time-position data to the road segment data.

4.1) Extract the neighboring time-position data of A and C that cross intersection B.

4.2) Calculate the through time T_{AC} between A and C by Eq. (1).

$$T_{AC} = t_C - t_A \quad (1)$$

4.3) Calculate the through time T_{AB} between A and B by Eq. (2).

$$T_{AB} = \frac{L_{AB}}{L_{AB} + L_{BC}} \times T_{AC} \quad (2)$$

4.4) Calculate the passing time of B by Eq. (3).

$$t_B = T_{AB} + t_A \qquad (3)$$

4.5) Calculate the passing time of D to be the same manner of the above step.

The times of B and D can also be calculated by the proportion of the distance between time-position data A and E.

B. Convert Road Segment Data to Statistical Data

Road segment data comprises a large volume of data. However, buses run in different time intervals. The passing time recorded in road segments would be biased discretely. For this reason, even the time-position data obtained by buses is huge. Time-position data on focused road segments may not exist. In contrast, if more than one time-position data exists on a targeted road segment over a period of time, there is also another problem of deciding which data would be used as a representative value.

In order to solve this problem, statistical data such as average speed, maximum speed and minimum speed are calculated. When there is more than one road segment data for a period of time, we can use values such as the average speed to represent the value. If there is no road segment data for a period of time, the statistical data can be estimated from the before and after periods of time road segment data. In the analysis, one hour was the unit length of time, as shown in Fig. 3, the data of each day is segmented. With road segment data being contained by time segments as objects, statistical data such as the average, standard deviation, quantity of data, the maximum, minimum, etc. is calculated and saved.

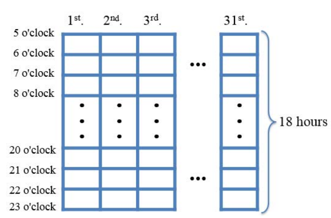

Figure 3. Road segment data converted to statistical data by in hours.

Processing sequence in details is shown as follows:

The following processing sequence is repeated to

1) Extract all road segment data set of the targeted road segment.

2) Repeat the following process for all time periods in a day.

a) Extract the targeted day's data from the road segment data set.

b) Classify the data by hour.

c) Calculate the average, standard deviation, minimum, maximum, quantity of data of the classified data set, and save them all.

3) Repeat Step 1) to Step 2) for all road segments.

C. Degree of Congestion

The degree of congestion means the evaluation of traveling time per unit distance or the traveling distance per unit time.

The degree of congestion of any location at any time can be calculated using statistical data. As a specific treatment method, it firstly extracts the statistical data corresponding to the specified conditions, and then displays the average, minimum and maximum. Display methods are shown as follows:

1) The degree of congestion of the road segments are displayed in different colors on the map when the date and time are specified by the user.

2) The graph of the degree of congestion is displayed for roads designated by the user.

Display method 1) can overview the degree of congestion on the whole road transportation network, and can also confirm the congestion situation of a road segment. Display method 2) can be used to analyze detailed situations, such as what time of day or week the road is crowded.

D. Reachable Areas

Based on the time/distance to pass through each road segment, calculate reachable areas using the shortest path algorithm. The detailed process is the following steps:

1) Obtain the required time for passing a specified location around the road segment.

2) Calculate reachable areas by the shortest path algorithm which takes the shortest possible time.

When exploring the shortest path, the utilized time between nodes is obtained, through the pass time of the concerned road segments that are connected by concerned nodes. Statistical data gathered by every time interval is related to each road segment. The pass time is obtained by the statistical data of the time interval corresponding to the starting time of the concerned nodes. There are two display methods to indicate as follows:

1) Display reachable areas using only existing statistical data on the road.

2) Display reachable areas using a certain speed of bus when the statistical data of the road is not exist.

Representation 1) is only displayed by the existing data of the bus, so it can more clearly express the bus route that comes from the specific region. Therefore, it also can be used as effective parameters for increasing bus route or the times of buses in specific regions. Unlike representation 1), representation 2) is displayed by all roads. In other words, it displays all the regions where ordinary vehicles can be reached at the specified speed in specified time.

E. Speeding Up the Calculation of the Degree of Congestion and Reachable Area

If data collected over a long time is analyzed, even if it is divided into statistical data by time segments such as hours, the cost of calculating the average and standard deviation will be high. Therefore, investigations that have a higher probability of being investigated should be processed in advance. This process is called the intensification of data. Frequent investigations are dependent on time and week should be processed in advance such as "Conditions of the road traffic network on weekdays from 7:00 to 23:00 which is time for school or work" and "The conditions of road traffic networks at noon on Mondays." Therefore, the following intensive data is processed:

1) Statistical data is extracted every month by day of the week (Fig. 4).

2) The statistical data is extracted every year by day of the week.

Statistical data is extracted to reduce the quantity of data, and the data that is processed can be read and processed rapidly.

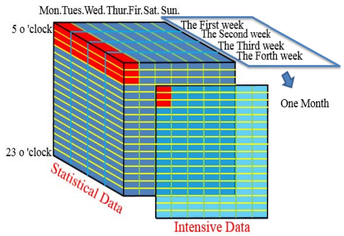

Figure 4. Monthly intensive data.

IV. EXPERIMENTAL PRODUCTION AND EVALUATION OF OUR SYSTEM

In this section we show effectiveness of our proposed methods. First, we produced a trial system experimentally. Second, we prepared test data for verifications. Last, we obtained visualized results, namely the degree of congestion and reachable area.

A. The Indication of Degree of Congestion

As shown in Fig. 5, the degree of congestion is represented by different colors. If buses passed this road segment at low speeds, this road segment is congested, so it is marked by red, on the contrary, if buses passed quickly, it is marked by green. We indicated the degree of congestion by average speed, because the average speed as basic information analysis of road condition does not vary with different scenarios. We can choose to use the average speed as a measure of standards. We further visualized the data by the degree of congestion. In this method, users can more intuitively see road conditions. Users also can obtain the change with time of the degree of congestion as shown in Fig. 6. We transformed monotonous figures into graphic images using degree of congestion so that the utilization of information was improved.

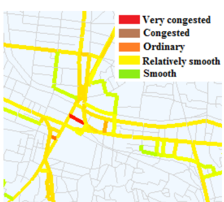

Figure 5. Degree of congestion distinguished with different colors.

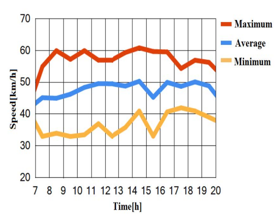

Figure 6. Variation diagram of degree of congestion.

B. The Indication of Reachable Area

In Fig. 7, reachable area is represented with route bus data only. For the road without the route bus data, reachable area is represented by the average speed that was calculated by the route bus data in Fig. 8. In those figures, the reachable areas distinguished the degree of difficulty to reach each different area with different colors. This method to show the reachable areas supposes that the color, from blue to red, changes with the degree of difficulty from easy to hard. In another words, the starting place is marked by blue point. The road segments corresponding color changed with increases in the desired time. And the edge of the reachable area is represented by the red circle. By this method we are able to show the more detailed analysis of the resulting data. We can easily understand the farthest area we can go and the degree of difficulty of the different area from these figures.

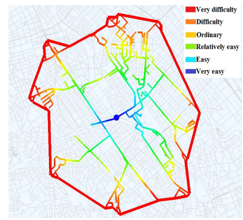

Figure 7. Reachable area: Only route bus data.

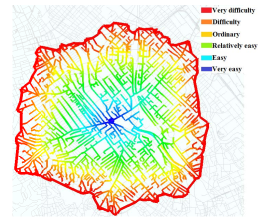

Figure 8. Reachable area: The roads without route bus data represented by average speed.

V. CONCLUSIONS

As an analytical method of traffic conditions of road networks, we proposed a conversion method from the probe data to road segment data set, based on road segments. The major contribution of this method is the rapid analysis of road traffic conditions, with the computed degree of congestion, for cases where the trajectory is determined and reachable areas from the road segment data set are known and, after calculating all relevant data, to provide a visualization of these analyzed results.

REFERENCES

[1] MLIT, "National Land Numerical Information – Bus Routes," (available online at http://nlftp.mlit.go.jp/ksj/gml/datalist/ KsjTmplt-N07.html; visited on Dec. 20, 2014), (in Japanese).

[2] N. Andrienko, G. Andrienko and S. Rinzivillo, "Experiences from Supporting Predictive Analytics of Vehicle Traffic," IEEE VIS 2014 Workshop on Visualization for Predictive Analytics, 2014.

[3] N. Takahashi, K. Munehiro and M. Asan, "Variation of Floating-Car-Data Characteristics: Different Vehicle Types and In-Vehicle Systems," Committee of Infrastructure Planning and Management, JSCE, 2005, (in Japanese).

[4] R. Matsunaka, M. Taniguchi and H. Hanato, "Study on the Estimation Method of Travel Speed for General Vehicle Using Bus Probe Data," Journal of Japan Society of Civil Engineers, Vol.23, No.4, pp. 871-877, 2006, (in Japanese).

[5] X. Wang, X. Zeng, Y. Kitamura, S. Araki and K. Kakizaki, A "Study on Analysis of Public Transportation Network by Using a Huge Amount of Time-position Information of Route Buses," Proc. of IPSJ Kyushu Hinokuni Symposium 2014, 5A-2, 2014, (in Japanese).

[6] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo and R. Trasarti, "Unveiling the Complexity of Human Mobility by Querying and Mining Massive Trajectory Data," VLDB Journal, Vol. 20, No. 5, pp. 695-719, 2011.

[7] T. Imai, K. Fujiyama, K. Kida and N. Nakamura, "Link Matching Method for High Speed Processing of Large Scale Probe-car Information," Proc. of the 70th National Convention of IPSJ, Vol.70, No.3, pp.37-38, 2008, (in Japanese).