

Route Identification and Travel Time Prediction Using Probe-Car Data

Tomio Miwa^{*1} Takaaki Sakai^{*2} Taka Morikawa^{*3}

*Graduate School of Environmental Studies, Nagoya University^{*1}*

(Furo-cho, Chikusa-ku, Nagoya City Japan 464-8603, +81-52-789-3729,

miwa@trans.civil.nagoya-u.ac.jp)

*PACIFIC CONSULTANTS CO., LTD.^{*2}*

(2-7-1 Nishisinjuku, Shinjuku, Tokyo, Japan 163-0730, +81-3-3344-0074,

takaaki.sakai@tk.pacific.co.jp)

*Graduate School of Environmental Studies, Nagoya University^{*3}*

(Furo-cho, Chikusa-ku, Nagoya City Japan 464-8603, +81-52-789-3564,

morikawa@civil.nagoya-u.ac.jp)

Recently the analysis on road traffic conditions with the use of probe-cars has been paid increasing attention. Probe-cars enable us to obtain the spacious data. However, since probe-cars generally provide the coordinates as vehicle's location point, analysts have to identify the vehicle's cruising route. It is difficult to identify the route at the section where elevated urban expressways are allocated parallel with the other roads. This study develops the method of identifying the vehicle's cruising route even where there are elevated urban expressways. It also develops the travel time prediction method using accumulated probe-car data.

Keywords: *probe-car, route identification, travel time prediction*

1. Introduction

Actual traffic conditions are accumulated drivers' route choice behaviors. Therefore, directing drivers to the optimum routes has an effect on realization of environmentally-friendly urban transportation systems. Most of technologies on ITS (Intelligent Transport Systems) aim at realization of these preferable transportation systems, and probe-car (it is also called probe-vehicle or floating car) system is one of ITS technologies.

Stationary equipments, such as AVI (Automatic Vehicle Identification), loop detector and so on, gather various traffic data, i.e. vehicular velocity, traffic volume and traffic accident. However they can cover only specific road sections on road network. These equipments, therefore, can provide the traffic information only at limited road sections. In contrast, probe-car system which utilizes vehicles running through the traffic network for gathering variable traffic data can cover wider area than these stationary equipments, although the amount of the data from probe-car system is less than that of stationary equipments. Additionally the cost of probe-car system is much cheaper than stationary equipments.

Generally, vehicular trajectory which probe-car provides is a series of GPS (Global positioning System)-based location points with coordinates, such as latitude and longitude. Of course car navigation system captures the vehicular information, such as cruising road and vehicular position on the cruising road, every second in

real time using dead reckoning system. Analysts, however, cannot obtain these information through a car navigation system, because details of car navigation system, i.e. used road map data and vehicular cruising route, are not disclosed. Therefore analysts have to identify the route which probe-car cruises by using GPS-based location points. We call this process "route identification". It is also called "map-matching". In the case that probe-car cruises planned route, analyst can instantaneously identify the vehicular position by dropping a perpendicular line from GPS-based location point to planned route. In the case that probe-car cruises the unspecified route, however, several factors make it difficult for analysts to identify the route which probe-car cruised. Especially the road sections where elevated urban expressways are allocated parallel with the other general roads and the long intervals of data transmissions make it difficult. The first purpose of this study is development of method for identifying cruising route even where there are elevated urban expressways and even when the intervals of data transmission are relatively long.

The probe-car data identified cruising routes provide the link costs, i.e. travel time of each link, over wider area. We propose the "Link Cost Table (LCT)" that is made from probe-car data identified cruising route. Then the second purpose is developing the travel time prediction method using LCT.

The rest of this paper is organized as follows. The outline of field experiment in Nagoya and the study area of this study are explained in the next section. In the

third section we discuss some existing route identification methods using probe-car data and develop new method adapted to the road section where there are elevated urban expressways and the long interval of data transmission. In the fourth section we explain the concept of LCT and develop the travel time prediction method using it. And in the final section we conclude with discussions on future directions.

2. Nagoya field experiment and study area

2.1. Outline of Nagoya field experiment

A large scale experiment in the Nagoya metropolitan area for two months (Jan. 28, 2002 – Mar. 1, 2002) was conducted by “Internet ITS Project (Main body: Ministry of Economy, Trade and Industry)” [1]. This field experiment is conducted in cooperation with 32 taxi companies located around Nagoya. On-board GPS are equipped on 1,570 taxis and data transmission is event-based. Table 1 shows the events when data are transmitted and their composition ratio. Events of “Running distance”, “SS” and “ST” consist of about 30-35% respectively and they cover more than 96% of all events.

Table 1. Events of data transmissions

Events	Remarks	Composition ratio
Running distance	when no events during 300m running	35.1%
SS (Short Stop)	when vehicle starting	31.3%
ST (Short Trip)	when vehicle stopping	29.8%
Others	Time period (550sec), Engine turning on/off, Unsafely behavior and so on	3.8%

On-board GPS equipments can be divided into three types. Types 1 and 3 (915 vehicles) are simple mobile servers. On the other hand, Type 2 (655 vehicles) is equipped with car navigation systems that make positioning more precise. The equipment with car navigation system, which has autonomous navigation, is well known for its capability to capture highly accurate positioning information. We verified that type 2 rarely causes positioning errors and errors happen frequently for Types 1 and 3 (Figure 1). Here, positioning error happens when GPS doesn't detect vehicles' position. Positioning error leads to imprecise travel time and fewer data usable. In this study, therefore, we treat the data from type 2 equipments.

It is noted that taxis' running speeds are different between in-service taxis and out-of-service taxis. In-

service taxis travel faster than out-of-service taxis [2]. It is because out-of-service taxis are supposed not to have specific destinations. In this study the data from out-of-service taxis are removed from the analysis and we treated the data of in-service taxis.

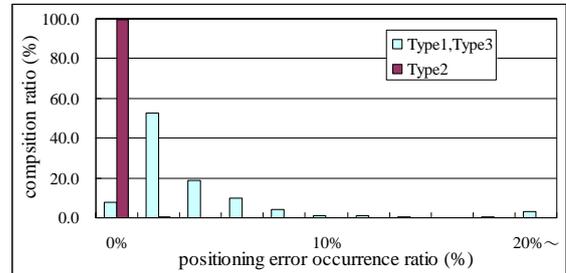


Figure 1. Positioning error occurrence ratio

2.2. Study area

In this study, we treat the O-D pair between the airport (Nagoya airport) and the central station (Nagoya station). The distance is approximately 12-18km. This pair is one of the most heavily taxi traveled O-D pairs in Nagoya metropolitan area and provides lots of data for this analysis. In this O-D pair there are two road sections that most drivers use. One is elevated urban expressway and the other is Route 41 which is national highway (Figure 2). They are allocated parallel and their position coordinates are almost the same.

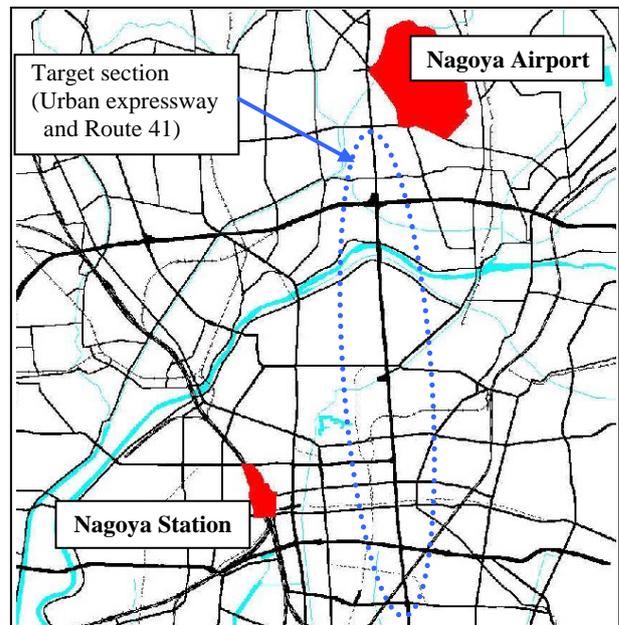


Figure 2. Study area

3. The methodology of route identification

3.1. Existing route identification methods

Many route identification methods (or map-matching algorithms) have been developed recently. Most methods of them, however, are developed for real time map-matching which is applied to car navigation with dead reckoning system [3][4]. They utilize vehicular mean speed of the front wheels, running distance, vehicular heading angle and so on. In contrast probe-car data provide a few information, such as data transmitting time, GPS-based location points (latitude, longitude) and running speed, because of cost for data transmission.

In this section, we review some of them which are applied to probe-car data and discuss the problem for applying them to this study case.

3.1.1. PROLIMAS. PROLIMAS (Probe car Link Matching System) is the map-matching algorithm that was developed by the Japan's Ministry of Land, Infrastructure and Transportation [5]. This method was developed for route identification of the probe-car data that are stored on recording media (memory cards) every second. The algorithm of this method is as follows.

Step 1: Separate the continuous probe-car data into origin and destinations. For separation, direction data are used. Origins and destinations are the points where the direction changes suddenly.

Step 2: Using each second data from origin to destination, extract only data where there was a change in direction data and join these data with a straight line.

Step 3: Extract sub network links from DRM (Distal Road Map) network. Sub network links are located within a radius of 30m from these joined route data.

Step 4: Search for the minimum distance route in sub network from origin to destination. And identify this route with the cruising route.

This method utilizes the minimum distance route algorithm for identifying the cruising route of probe-car. Therefore this method often misidentifies the general road section with the cruising route instead of elevated expressway section, when they are allocated parallel like this study area. Because elevated expressway section which has on-ramp and off-ramp is almost always longer than general road section. Additionally this method was developed for very frequent probe-car data (every second). Therefore we cannot apply this route identification method without modification, because the data transmission intervals of our probe-car are very large (often 300m distances).

3.1.2. The method of Asakura et al. Asakura et al. developed the route identification method, and applied to the data of location positioning system of PHS

(Personal Handy-phone System) [6] [7]. They analyze the travel behavior of an individual such as the staying time and point and moving route. This system gathers the positioning data (coordinates of PHS user) every two minutes interval. The algorithm of this method is as follows.

Step 1: Extract the trip data that are from the stay point (origin) to next stay point (destination).

Step 2: Extract sub network links from full network. Sub network links are located within a threshold distance from the location point.

Step 3: Search the minimum distance route from origin to destination.

Step 4: Search the multiple routes using screening method and identify the route that has the minimum error indicator (\bar{D}) with the route used by traveler.

The threshold distance, which is noted in step 2, is determined by considering both measurement error of the positioning system and the number of links in sub network. Screening method, which is noted in step 4, is the algorithm that searches the multiple routes [8]. When the minimum distance route is set as the first route, screening method searches the $k+1$ th route diverging from k th route. Details of this method were presented in [6].

In this method, the route with minimum value of \bar{D} calculated by equation (1) is identified with the traveler's used route.

$$\bar{D} = \sum_{i=1}^I d_i / I \quad (1)$$

where d_i means the distance from location point i to the route, and I means the number of location data included in trip data.

This method can treat with the long interval of data transmission (every two minutes), and make an adjustment of the traveler's route.

3.2. Development of route identification method

In this section, we develop the route identification method which can be applied to both our probe-car data and study area. First, we summarize the controversial points of our data and study area as follows.

1. The maximum interval of data transmission is long distance (300m).

2. There are two road sections that most drivers use. One is elevated urban expressway and the other is general road (Route 41). They are allocated parallel and their position coordinates are almost the same.

3. It is not necessarily the case that probe-cars cruise the minimum distance route. They often cruise the longer distance route, such as expressway section with on-ramp (off-ramp) loop.

To deal with these problems, we develop new route identification method by improving the method of Asakura et al. The methodology of our route identification is as follows.

Step 1: Extract trip data from probe-car data. It is a series of data from passenger's getting on to getting off. That is, origin is passenger's getting on point and destination is passenger's getting off point.

Step 2: Extract sub network links from DRM links. Sub network links consist of links locating within a radius of longer distance between adjacent two GPS-based location points from each GPS-based location point.

Step 3: Search the minimum cost route from origin to destination in sub network. The cost is the original one and presented later.

Step 4: Search the multiple routes with original cost using screening method.

Step 5: Identify the cruising route of the probe-car among the multiple routes using equation (1).

Step 6: Identify vehicular position by dropping a perpendicular line from GPS-based location point to identified route.

When sub network is extracted at step 2, we use longer distance between adjacent two location points from each location point as a radius. We intend the size of sub network to be light and to improve computing speed. DRM used in this study is based on 1:25000 topographic maps and includes all arterial roads and the other roads with width of 5.5m or more.

We apply three original costs that are used in step 3 and step 4. They are presented as follows.

<Link cost 1>

$$c_k = l_k \quad (2)$$

where, c_k is the cost of link k and l_k is the length of link k .

<Link cost 2>

$$c_k = l_k \times d_k \quad (3)$$

where, d_k is the distance from the link k to the nearest GPS-based location point.

<Link cost 3>

$$c_k = l_k \times d_k \times w_k \quad (4)$$

where, w_k is the weight determined based on both vehicular velocity and road category of the link. This weight is applied for improving the accuracy of route identification. And the value of this weight is presented in table 2, which are set by trial-and-error according to slowing of the velocity on expressway at evening peak period.

**Table 2. Value of the weight, w_k
During evening peak (16:00 – 19:00)**

Link i is expressway and velocity of probe-car is more than 50km/h	0.1
Link i is expressway and velocity of probe-car is less than 30km/h	10.0
Otherwise	1.0

During other time periods (0:00 - 16:00, 19:00 – 24:00)

Link i is expressway and velocity of probe-car is more than 80km/h	0.1
Link i is expressway and velocity of probe-car is less than 60km/h	10.0
Otherwise	1.0

In our route identification method, the screening method is applied to adjust the on-ramp and off-ramp that probe-car passed. In particular, the screening method with link cost 3 is applied to make an adjustment of the cruising route based on vehicular velocity and expressway section.

3.3. Validation of route identification accuracy

In this section we validate the route identification methods developed in the previous section. For the validation of each method, we use the indicator, presented by equation (5). We call it “CDR” (Correct Distance Rate).

$$CDR (\%) = \frac{\{\text{Sum of distance of correct identified links}\}}{\{\text{Distance of true cruising route}\} * 100} \quad (5)$$

True cruising routes, sampled 200 routes, were made by hand. Table 3 presented the results based on each link cost.

Table 3. Route identification accuracy

	Link cost 1	Link cost 2	Link cost 3
CDR (%)	47.2	84.0	94.2

The method using link cost 1 is almost the same as the method of Asakura et al. The result of method 2 using d_k , which is the distance from the link k to the nearest GPS-based location point, is more accurate than that of method 1. The application of d_k dramatically enhances the accuracy. This result is attributed to the fact that even detour routes in distance, which are cruised by prove-cars, can be searched by the effect of link cost 2. The result of method 3 using both d_k and w_k is the most accurate among three methods. The weight

w_k serves as making an adjustment to passing urban expressway section according to vehicular velocities.

The main causes of the failure of route identification are as follows.

1. Heavy traffic jams at expressway
2. Passing through the narrow pathways which are not included in DRM.
3. No data transmission near the turning point

In these cases, it is very difficult to verify the true cruising route, even if we carefully examine raw data with eyes. For more precise identification, improvement of hardware is necessary, such as introduction of trigger of direction indicator lamp and altitude information.

4. Travel time prediction

4.1. Link Cost Table

The probe-car data identified cruising routes enable analysts to utilize the link costs of each link (i.e. travel times of each link) along the identified route. The link cost is calculated by estimating the inflow and outflow time of target link. Inflow and outflow time of target link are calculated by equation (6), (7), and link cost is calculated by equation (8).

$$\text{Inflow time} = (t_a * l_2 + t_b * l_1) / (l_1 + l_2) \quad (6)$$

$$\text{Outflow time} = (t_c * l_4 + t_d * l_3) / (l_3 + l_4) \quad (7)$$

$$\text{Link cost} = \text{Outflow time} - \text{Inflow time} \quad (8)$$

where, t_a is the last data transmission time before flowing into the target link. t_b and t_c are the first and the last data transmission times on the target link respectively. And t_d is the first data transmission time after flowing out the target link.

After we carried out the route identification of all probe-car data in Nagoya metropolitan area, we estimated "Link Cost Table (LCT)". LCT is composed of the link costs (mean travel times of each link) of all links in the network at every five minutes time period. And it has the information of transition of traffic conditions every five minutes time period. LCT is one idea for

accumulating the probe-car data. We estimated four types of the tables, weekday or holiday and fine weather or wet. Table 4 shows the sample format of LCT.

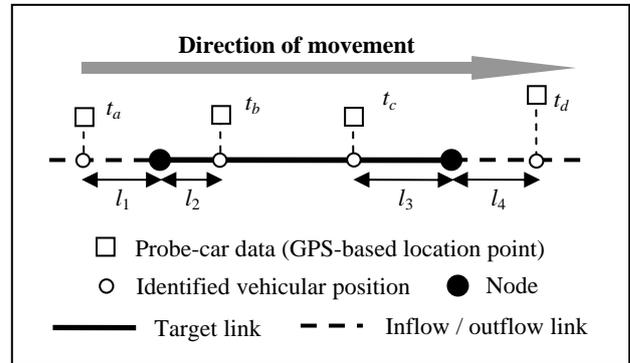


Figure 3. Calculation of link cost

4.2. Method of travel time prediction

In this section, we develop the route search method considered the transition of traffic condition using LCT. The algorithm of developed travel time prediction is as follows.

- Step 1: Set the input data, such as nodes of origin and destination, departure time, weather and a day of the week. Set t = departure time + $mod(\text{departure time}, 5\text{min.})$. Here " $mod(A, B)$ " means remainder of A/B .
- Step 2: Apply the link cost of the time period corresponding " $t-5 \sim t$ " to all links having nonarrival outflow node. Using Dijkstra algorithm, search the minimum cost route from origin node to destination node.
- Step 3: Under searching the minimum cost route, if maximum cost of arrival node exceeds t , suspend searching calculation.
- Step 4: If destination node is not arrived, set $t = t+5\text{min.}$ and go to Step 2.

Table 4. Sample format of Link Cost Table (LCT)

Node1	Node2	Road category code	Length of link (m)	Time period 1 (0:00 – 0:05)	Time period 2 (0:05 – 0:10)	...	Time period 288 (23:55 – 24:00)
29001	29025	7	112	17.5 sec	15.8 sec	...	19.2 sec
29003	29012	4	200	19.1 sec	17.6 sec		20.3 sec
29006	30108	3	136	12.8 sec	12.0 sec		14.54sec
...
54004	54013	3	68	6.4 sec	5.1 sec		7.0 sec

changes slightly. In the next section we analyze the characteristics of predicted routes at the time period when traffic condition transits greatly.

4.3. Analysis on characteristics of predicted routes

In this section, we analyze the characteristics of predicted route using two prediction methods presented in above section. The time period of evening peak (16:00 – 19:00) and the trip from Nagoya Airport to Nagoya Station are treated. In evening peak period traffic conditions change greatly. Especially during the time period after 17:30, the traffic conditions are heavy congestion and travel time saving by expressway is little in this O-D pair.

4.3.1. Predicted route. Figure 6 shows the distances of the minimum cost routes predicted by each method. In this study area, all available routes are divided into the two groups with respect to the used road section. One is the expressway routes and the other is the general road routes. Moreover expressway routes are longer than general road routes because the conformation of expressway make users detour. In this figure routes with distance of about 18km are expressway routes. This figure shows that the routes predicted by each method are different in many cases. Especially the time instants using expressway routes are different in almost all cases.

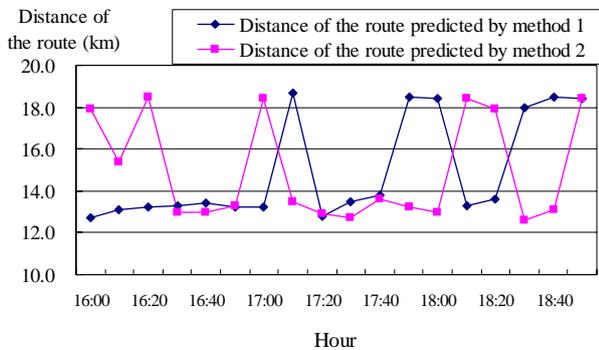


Figure 6. Difference of predicted routes

4.3.2. Predicted travel time. Figure 7 shows that the travel times predicted by each method. The predicted travel times are different in almost all cases. While method 2 predicts travel time considering the condition only at the departure time, method 1 predicts considering the transition of traffic condition after departure. This result shows that predicted travel times vary according to the different usage of LCT.

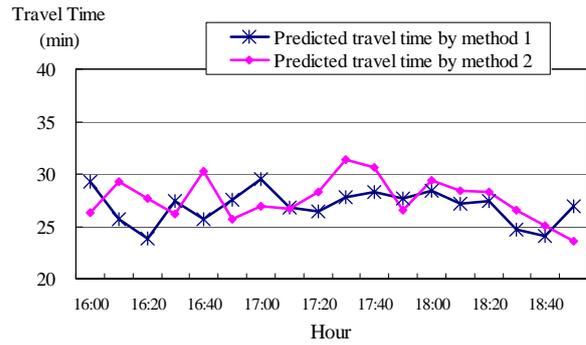


Figure 7. Predicted travel time

4.3.3. Prediction error of method 2. Figure 8 shows the prediction error of method 2. Dotted line shows the actual travel time needed for the route predicted by method 2. At almost all times method 2 provides mis-predicted travel time because of the transition of traffic conditions after departure. Especially, during the congested time, after 17:30, actual travel times are larger than predicted travel times.

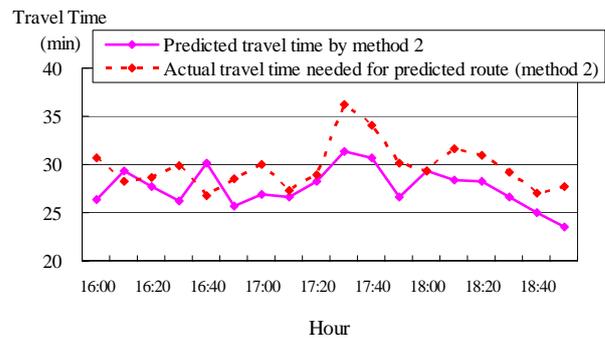


Figure 8. Prediction errors of method 2

4.3.4. Actual travel time needed for predicted route. Figure 9 shows that the difference of actual travel times needed for predicted routes between two methods. The actual travel times of method 1 are consistently less than that of method 2. It is because method 1 considers the transition of traffic conditions after departure at the origin. As just described, the method using LCT can provide the ex-post minimum cost route.

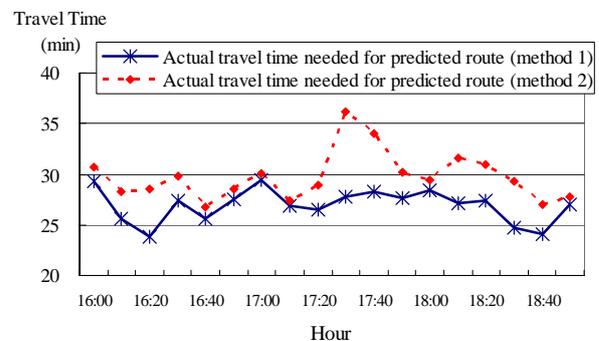


Figure 9. Actual travel times needed for each method

