CLASSIFICATION OF TRIP TRAJECTORIES MEASURED BY POSITION DETECTION TOOLS

Takamasa Iryo

Research associate, Department of Civil Engineering, Kobe University

1-1, Rokkodai-cho, Nada-ku, Kobe, Japan

Tel/Fax: +81-78-803-6360, E-mail: iryo@kobe-u.ac.jp

Edward Chung

Visiting professor, Center for Collaborative Research, the University of Tokyo

Masao Kuwahara

Professor, Institute of Industrial Science, the University of Tokyo

Hiroyuki Oneyama

Deputy Head, National Institute for Land and Infrastructure Management, Ministry of Land, Infrastructure and Transportation

SUMMARY

This study proposes a methodology which classifies trajectories of travelers' movements and examines its validity by an experiment. These days, technologies like GPS can provide highly detailed data of travelers' trajectories, but they do not directly provide information of which modes or services travelers use. This method classifies travelers' trajectories into modes and transportation services with a prior spatial and temporal knowledge of transportation systems. We have conducted a pilot study to test this method in the center of Tokyo.

INTRODUCTION

This study proposes a methodology which classifies time-space trajectories of travelers' movements measured by GPS or other position detection tools and shows results of an experiment on a real network in the center of Tokyo.

These days, GPS or other tools based on radio wave technologies which measure precise and highly detailed data of travelers' movements are available, and several small and handy equipments are available in the commercial market. This means that we can obtain complete information of traveler's time-space trajectory from his/her origin to destination.

These measurement methods provide us with information about the travelers' time-space trajectories; however, the data they provide are expressed in terms of the coordinates such as latitude and longitude, and time. Such information is sufficient to identify the traveler's position itself. However, it is not sufficient to obtain other information such as the mode of travel or the routes or services he/she used.

This study uses the information of space and time to make an accurate estimation of travelers' modes or services. Previous studies (1) and (2) proposed the methods to know the mode or service with traveler's spatial trajectory with spatial information of the mode or service (position of roads, for example), however, in the cities with dense transportation systems, such methods sometimes do not work well due to the limitation of the accuracy of position detection tools. This study intends to use the time information to avoid the problem.

This study adopts the concept of “classification” to obtain extra information other than coordinates. Here, classification means classifying travelers’ time-space trajectories measured by position detection tools into categories made by prior knowledge of transportation systems they use. The category is, for example, mode of travel, name of the road where travelers traverse or train number which travelers use.

HOW TO CLASSIFY TRAJECTORIES

Description of trajectory obtained by position measurement tools

The concept of “measured path” is defined to describe the trajectory obtained by position detection tools. The "measured path" is defined as a set of measured points, each of which includes space coordinates (longitude and latitude where a traveler was) and the time coordinates (time stamp recorded by the tool). Example of a measured path is shown in figure 1.

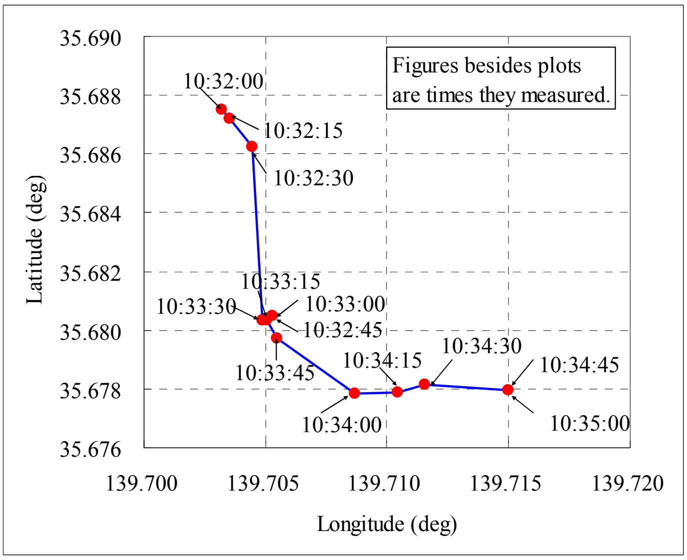

Figure 1: An example of measured path

Description of “a prior knowledge”

The concept of “planned path” is defined to describe the prior knowledge of transportation systems. The “planned path” is defined as an expected time-space trajectory of a person on some transportation system, and ideally this trajectory is expressed as a continuous curve, called a “planned curve”. This curve can be determined with no error if the prior knowledge of the transportation system is perfect. However, in almost all cases, such perfect knowledge cannot be obtained due to many reasons, and as a result, the planned path may have some uncertainty.

The concept of “planned plane” is introduced to express the uncertainty of planned paths due to imperfect transportation system information. The planned plane shows the area where the planned curves of some transportation service can pass. Figure 2 shows an example. A train service between two stations (station A and B) is assumed here. Train departure timetables at these stations are obtained beforehand (10:42 at station A and 10:46 at station B) and the coordinates of the railway track are measured precisely. In such a case the planned path of this train can be determined as the path which passes through a time-space point “a”, whose coordinates are the position of the station A and the departure time of the station A, and a point “b”, whose coordinates are the position of the station B at the departure time of the station B. These points “a” and “b” can be determined precisely from the information of the track and train timetable, however the shape of the path between two points cannot be determined because of the lack of the information about the details of train operation, such as the acceleration rate or brake timing and so on. So, the all paths which connect the point “a” and “b” can be the planned paths of this train service. The “planned plane” of this train is determined as a square of point “a”, “a*”, “b”, and “b*”, which can cover all possible paths of the service. When the train which passes through several stations is considered, the planned plane for this train can be determined like as figure 3. Note that this plane should be perpendicular to the space plane because the spatial information of the track can be obtained precisely in almost all cases.

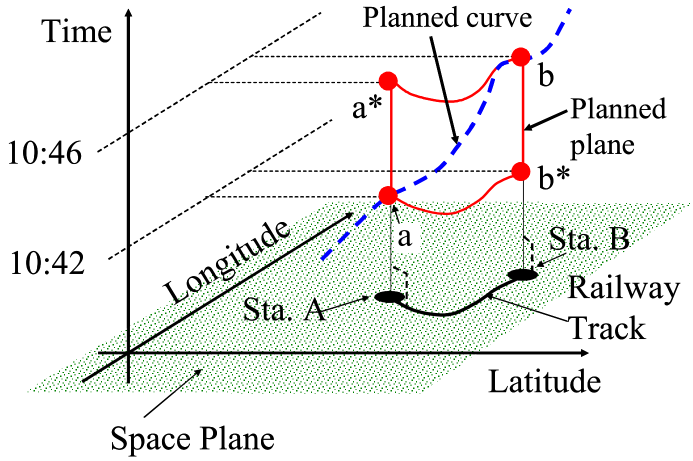

Figure 2: Planned plane

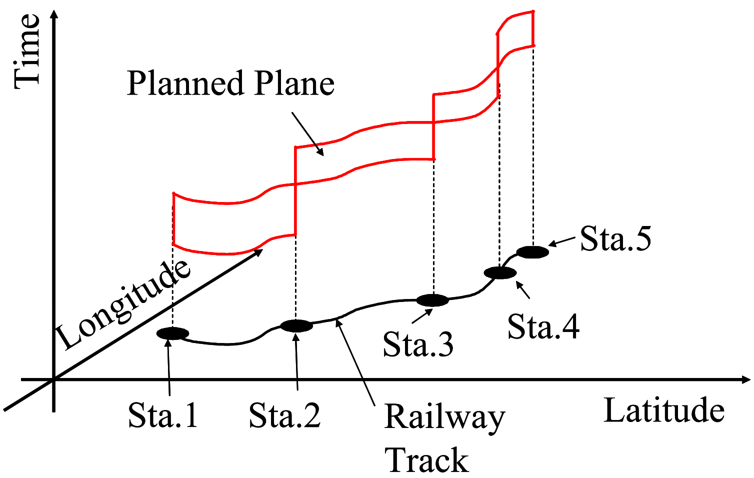

Figure 3: Planned plane along many stations

Definition of distance between planned and measured paths

Here the concept of "distance" is defined. This concept is used to evaluate the similarity between a measured path and a planned path.

To define the distance between a measured path and a planned plane, the distance between two time-space points should be defined first. The time-space distance between point A, whose coordinates is (x_1, y_1, t_1), and point B, whose coordinates is (x_2, y_2, t_2), is defined as

$$d_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (\alpha(t_2 - t_1))^2} \tag{1}$$

where α is a constant. This definition depends on the value of the constant α, however it will be eliminated later and need not to be determined.

The distance between a measured point and a planned plane is defined as the shortest distance between the measured point and the planned plane. The actual planned path is expressed as a planned curve on the planned plane and the distance between the planned path and the measured point should be defined as the shortest distance between two in ideal situations. However, the actual shape of the planned curve is not defined now and the concept of the planned plane which has already explained means that any shape of the planned curve on the planned plane can be considered. So we can assume that the planned curve is passing through the nearest point to the measured point and this is why the distance between the planned path and the measured point can be considered as the shortest distance between the measured point and the planned plane. As all planned plane are perpendicular to the space plane, the shortest distance is the same as the distance between the measured point and the point on the planned plane whose time coordinates is equal to the time at the measured point. An example is shown in figure 4. Because of this, the distance between the measured point and the planned plane can be calculated without considering the $(t_2 - t_1)$ term in equation (1) and this eliminates undetermined constant α from calculation of the distance. The distance between the measured point and planned path is ignored when the measured point is outside the planned plane (with reference to time).

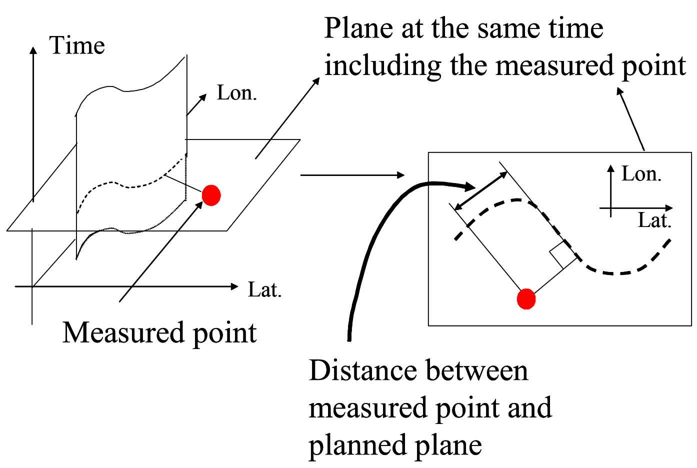

Figure 4: Definition of distance between measured point and planned plane

The distance between a measured path and a planned plane is defined as the square-root mean of all distances between measured points which overlap with the planned plane temporary (see figure 5) and the planned path. This is defined as an equation of

$$d = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d_i^{\,2}} \qquad (2)$$

where d_i is the distance between the measured point i and the planned plane and N is the number of the measured points.

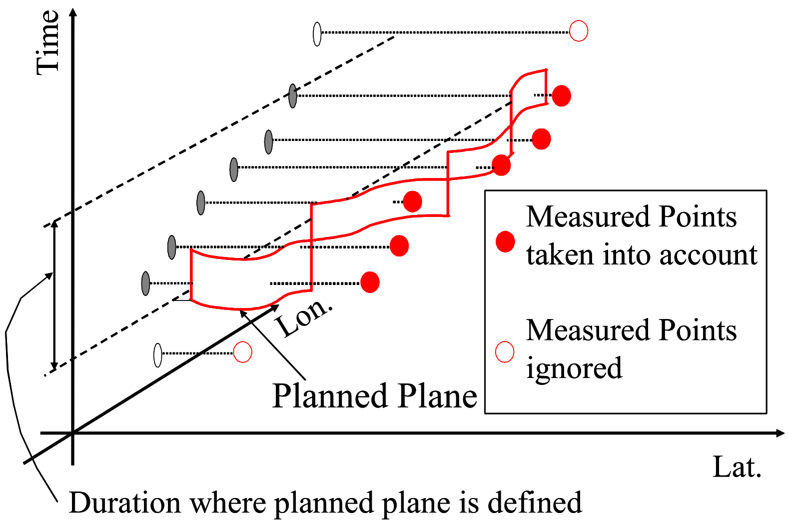

Figure 5: How to calculate distance between measured paths and planned plane

The planned path whose distance to the measured path is shorter than a threshold will be chosen as the planned path of the used service. The threshold distance should be determined to choose the used planned path. It will depend on conditions of the experiment and should be determined with a preliminary experiment.

Long planned path and short planned path

Here we explain the difference between a "long planned path" and a "short planned path". The long planned path is the path which is defined for the long duration, whereas the short planned path is the path which is defined for the short duration.

Generally speaking, distance to the long planned path has higher accuracy than to the short planned path. One reason is that the number of measured points which matched with the long planned path is much greater than the short planned path. Furthermore, the distance between the measured path and the planned path is defined as the square-root mean of distances to the measured points and statistical error may be larger when the number of measured point is small. The other reason is the number of service operation information available such as operation timetable and service platform, for long planned path is likely to be more than that of the short planned path. The above factors make the accuracy of the long planned path much higher than the short planned path.

However, if the length of the long planned path is excessive, the number of traveler's travel pattern that could be extracted may be limited as only measured path that matched the planned path precisely over the whole section of the planned path will be selected. Hence the length of planned paths should be determined according to the purpose of the analysis.

Making a set of planned paths for classification

The methodology for making a set of planned paths for classification is explained here. The all planned paths of the transportation mode or service which travelers have some possibility to use should be listed up when we want to know what transportation mode or service they use. The set of planned path is the list of all planned paths which belong to some category of transportation mode or service. For example, the set of planned paths which includes all train services on some railway route should be considered when we want to classify travelers' trajectories into the category of railway and other categories. We should make this set(s) to classify traveler's trajectory into categories, and can determine which category the trajectory belongs to with considering the distance between the measured path of the traveler and one of the planned paths in the set.

The easiest way to make the set(s) is gathering short planned paths which belong to the category to be considered. The concept of this way is shown in figure 6. Such short planned paths may have short distances to the measured paths not only when the traveler use the whole section (from station A to station F, for example) of the service but also when he/she use the short section of the service (from station C to station D, for example). This means that such kind of sets can describe almost all cases of travelers' trajectories which are made by the service; however, larger errors are expected because of the disadvantage of the short planned path which has been explained.

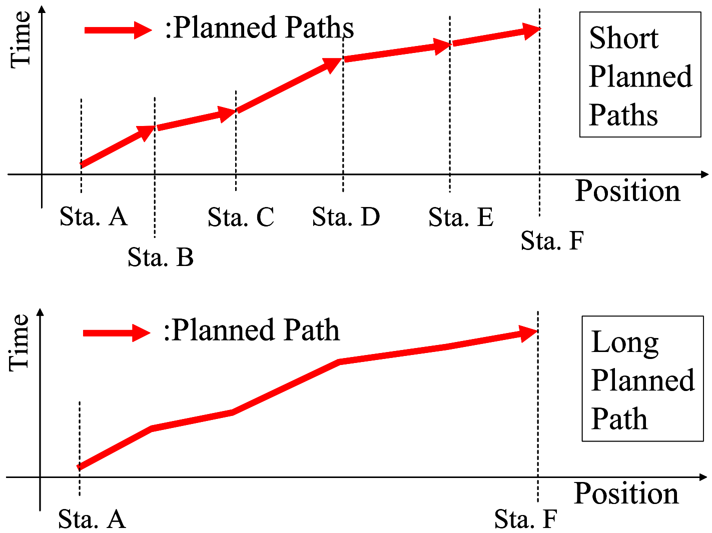

Figure 6: Short planned paths and long planned path

To avoid the error, making the set(s) with the long planned paths is a good way. The concept of this method is shown in figure 6. Many kinds of paths are needed to describe all planned paths connecting any two stations, so, limiting stations where travelers get on or off is a good idea to make realistic strategy for the classification. The long planned path shown in figure 6, for example, limits the stations to station A and station F.

Classification can be made with considering the distance between the measured path and planned paths in the set of the class. If the measured path has short distance to the planned path in the set, we can determine that this measured path belongs to the class. Simultaneously, the detailed information of the service, like train number and so on, can be obtained. One measured path can belong to many classes if the traveler uses many types of transportation services.

EVALUATION OF ACCURACY WITH EXPERIMENTAL DATA

We carried out an experiment to evaluate the method for classification. Here we explain the details of the experiment and evaluate the accuracy and the practicability of the method.

Outline of the experiment

The experiment is carried out in the central area of Tokyo. Tokyo has many railway services including subways. Density of train services is high and two or more types of services (local and rapid services, for example) are operated on some lines. It also has dense road network and some roads runs parallel to railway tracks and sometimes the distance between two is very near. These factors make it difficult to distinguish transportation services with the spatial trajectory data.

We selected "Chuo Line" and "Sotobori-dori" as experimental transportation services. Chuo line is a railway service which runs through center of Tokyo and has four tracks and includes two types of services, local services and rapid services. Duration between services is up to 5 minutes during the day time. "Sotobori-dori" is an arterial road which is close to the Chuo line and is used to check whether the proposed method can distinguish travelers on trains and on cars. A map is shown in figure 7.

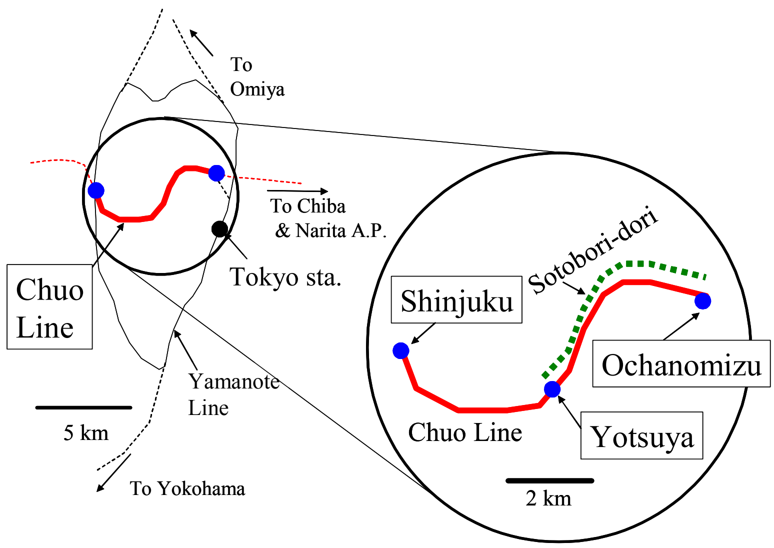

Figure 7: Map of Chuo Line

A system called "PEAMON" was adopted as a position detection tool. "PEAMON" is a position detection system which is based on a network of radio bases, which is originally provided for the handy phone system called PHS. The radio bases of PHS are distributed densely because the strength of radio wave used by this system is weaker than normal cellular phone, so it is more suitable to use the position detection.

Eight sample measured paths (seven are made by train services and one is made by car) are obtained. The travelers who made these paths were requested to use rail services or road services according to the pre-determined plans. The duration of each path is about six hours including rest time. The train travelers traveled on Chuo line between Ochanomizu station and Shinjuku station, whereas the car traveler drove along Sotobori-dori between Ochanomizu and Yotsuya.

Measured paths obtained by the experiment

Spatial distribution of all measured points made by train travelers is shown in figure 8. Figure 9 shows a detailed map near Chuo line and Sotobori-dori with measured points by train travelers and the car traveler. This figure indicates that we cannot determine whether a traveler use a train or a car with spatial coordinates data.

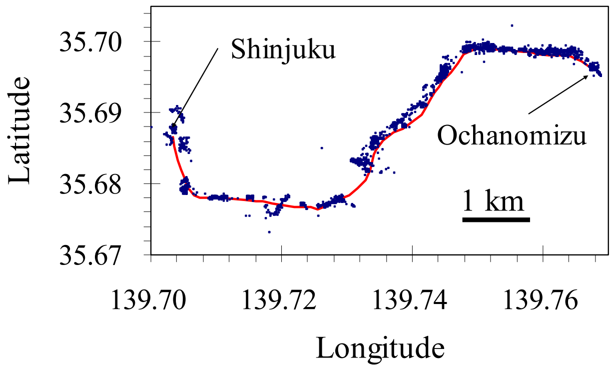

Figure 8: All measured points made by train travelers

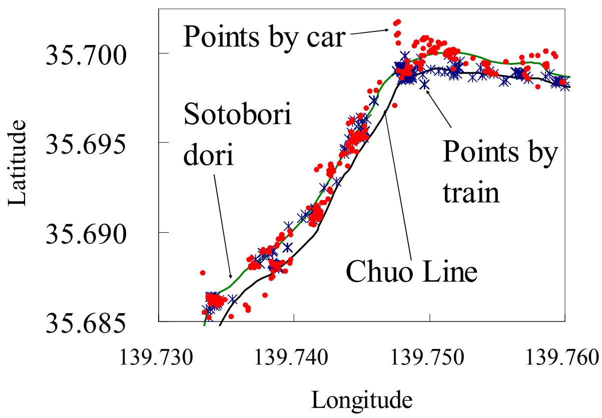

Figure 9: A part of measured points made by car traveler and train travelers

Set of planned paths

The sets of planned paths were made from a timetable of stations on Chuo line. Here we made two types of sets, one includes all train services between Ochanomizu station and Shinjuku station (set 1) and the other includes all train services between Ochanomizu station and Yotsuya station (set 2). Train services operating during the experimental hours were considered in these sets. Planned paths of local trains were made with departure times at all stations. However, paths of rapid services were made with actual departure times at Ochanomizu, Yotsuya, and Shinjuku, where the rapid services stops, and the passing time at intermediate stations are estimated by interpolation with respect to the distance between stations. All planned paths are expressed as planned planes. An example of a planned path is shown in figure 10.

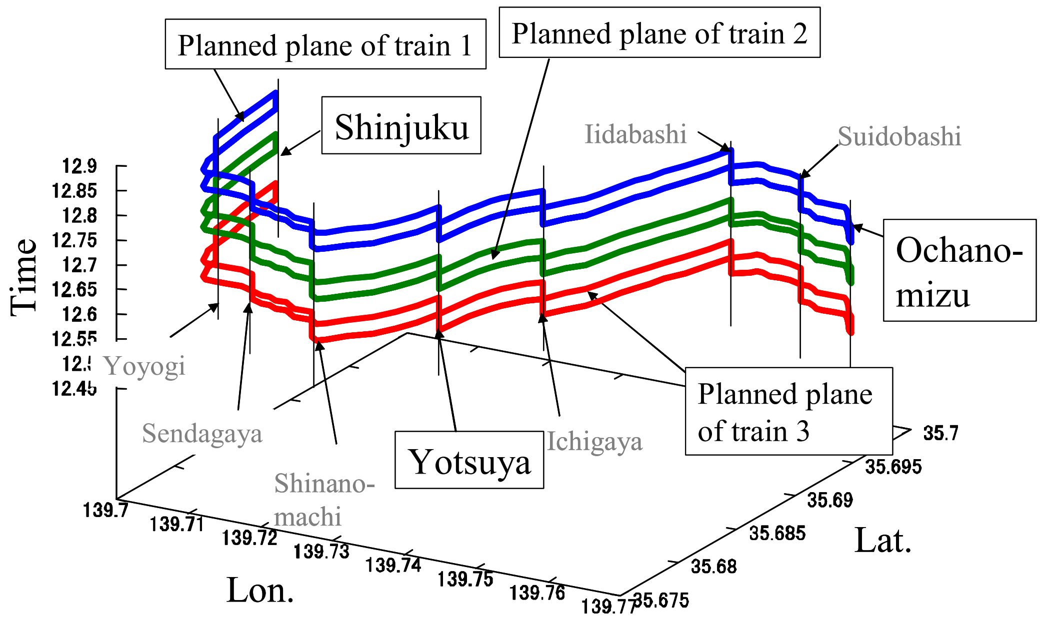

Figure 10: Example of planned path between Ochanomizu and Shinjuku

Classification and its accuracy

Here we classify the measured paths made by train travelers into train services boarded and therefore, train numbers or departure time at some specific station can be determined. The classification was made by considering distances between the measured paths and the planned paths. The planned paths whose distance to the measured path of a traveler is shorter than a threshold distance could be considered as the service he/she has used.

To determine the threshold, we need to consider some indexes to evaluate the quality of the classification. The threshold should be set so as to maximize the quality. Two indexes were considered to describe accuracy of the classification. One is "cover ratio", which is defined as

$$\text{Cover ratio} = \frac{\text{Number of planned paths used by the traveler and chosen by the method}}{\text{Number of all planned paths used by the traveler}}.$$

This ratio can be 1 if the threshold is quite large, however, too large threshold may choose wrong planned paths which were not used by the traveler. The other index, which is called "precise ratio", can be used to evaluate how many wrong planned paths were chosen. Precise ratio is defined as

$$\text{Precise ratio}=\frac{\text{Number of planned paths used by the traveler and chosen by the method}}{\text{Number of all planned paths chosen by the method}}.$$

"All planned paths chosen by the method" may include wrong planned paths, which were not used by the traveler, so the ratio may be under 1 when the method chooses some wrong planned paths. Precise ratio increases if the threshold is smaller because only planned paths whose distances to the measured path meet the more stringent criteria are selected.

If the situation of choosing is quite ideal, we can set some threshold which achieves a perfect cover ratio (=1) and a perfect precise ratio (=1). However, in most cases, such perfect ratios cannot be obtained due to noise of measurement tools and imperfect knowledge of transportation system. The cover ratio would be large with a large threshold but the precise ratio would be small, and vice versa. So we should decide some realistic threshold which obtains sufficient cover ratio and precise ratio.

We examined the relation between two ratios and the threshold in our experience. Figure 11 shows the relationship in the case of set 1 (Ochanomizu – Shinjuku) and figure 12 shows the relationship in the case of set 2 (Ochanomizu – Yotsuya) for all train travelers. In the case of set 1, threshold of 600 m achieves 80% cover ratio and 80% precise ratio. In the case of set 2, threshold of 400 m achieves 70% cover ratio and 70% precise ratio. These results mean that to some extent these thresholds can be adopted for classifying a traveler's trajectory by train service.

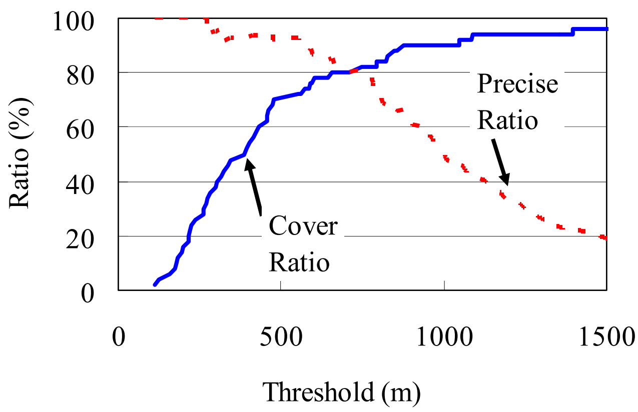

Figure 11: Relation between threshold and cover and precise ratios
(Planned paths along Ochanomizu – Shinjuku)

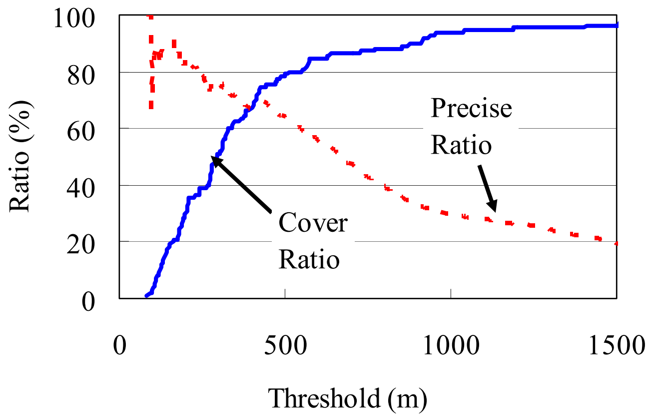

Figure 12: Relation between threshold and cover and precise ratios
(Planned paths along Ochanomizu – Yotsuya)

This study also examined whether this method can classify the measured paths into types of modes, train and car. The measured path made by car is also examined by the method to check this. In an ideal situation, distances between the measured path by car and all planned paths of train services would be longer than the threshold because it must not be classified into train services. Figure 13 shows the number of planned paths in the set 2 (train service between Ochanomizu – Yotsuya) whose distances to the measured path made by car is lower than the threshold. For comparison, the figure also shows the number of planned paths which have shorter distance to the measured path of a train traveler. The result indicates that a few planned paths of train services are selected with 400 m threshold when the path of a car is considered but the number is much smaller than the number of planned path selected for the train traveler. We can confirm that the measured path made by train has many selected planned paths (16 train services), whereas the path made by car has much smaller selected planned paths (3 train services).

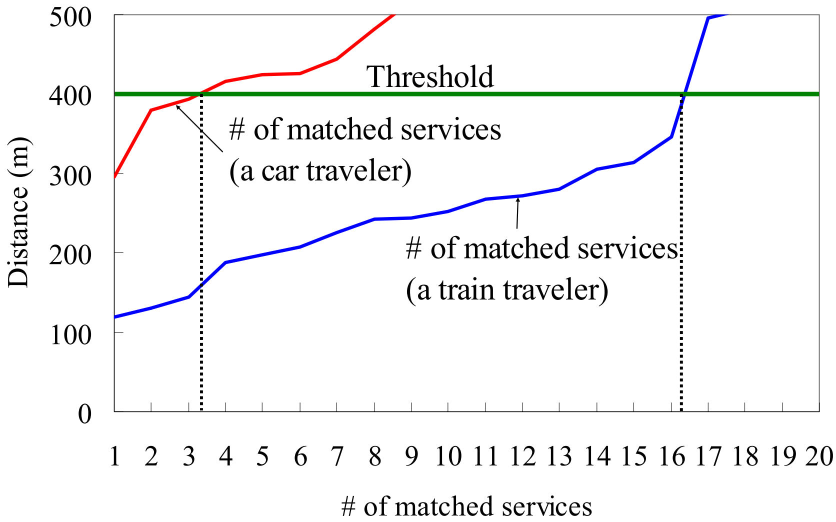

Figure 13: Number of matched train services between Ochanomizu and Yotsuya

DISCUSSION

This study proposes a method for the classification of traveler's trajectory and tests whether the method working. We conclude that

* This method can classify traveler's trajectory into each train service indicated by train number if the service has accurate timetable. It also can distinguish whether the traveler used a train or a car in the situation where spatial trajectory of a train and of a car are completely overlapped, and

* This method has high accuracy when longer planned paths are considered, whereas it has lower accuracy with shorter planned paths.

It means that this method may be used to analyze traveler's behavior in dense cities even when the position detection tools do not have sufficient spatial resolution. Technologies of position detection are progressing day by day and the resolution will be high enough to analyze behavior in congested cities. However, this method will be valuable even in the future due to some limitation of the measurement imposed by social reasons, such as privacy issue, and so on.

One of the problems of this method is how to determine the threshold which decides whether the measured path is near to the planned path or not. It cannot be decided just with the prior knowledge because it depends on the characters of position detection systems and transportation systems. Some preliminary experiments should be carried out before conducting the main experiment for each survey. Making some database which holds the results of former position detection experiment may reduce the time for such preliminary experiments.

Another problem is how to classify trajectories made by car or some other transport systems that do not have a schedule, are not punctual or not reliable. In the case of Tokyo, where the experiment was held, does not have such a problem because most people uses train services to travel and punctuality and reliability of operation are high. However, some cities may not have reliable railway systems because most people travel by car, or some cities may not have accurate public transportation systems. The method should be extended to take into account of these cities with an inefficient public transportation services, but accuracy would become lower.

REFERENCES

(1) Y. Asakura, E. Hato, Y. Nishibe, T. Daito, J. Tanabe, and H. Koshima (1999) Monitoring Travel Behaviour using PHS Based Location Positioning Service System. *Proc. of the 6th ITS World Congress in Toronto*, CD-ROM

(2) Y. Asakura and E. Hato (2000) Analysis of Travel Behaviour using Positioning Function of Mobile Communication Devices. Travel Behaviour Research; the Leading Edge, Hensher, D. (ed.) Pergamon, pp. 885-899.