Available online at www.sciencedirect.com
ScienceDirect
Transportation Research Part A 41 (2007) 684–690
TRANSPORTATION RESEARCH PART A
www.elsevier.com/locate/tra

Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument

Yasuo Asakura [a,*], Takamasa Iryo [b]

[a] *Graduate School of Science and Technology, Kobe University, 1-1 Rokkodai-cho, Nada, Kobe 657-8501, Japan*
[b] *Department of Architecture and Civil Engineering, Faculty of Engineering, Kobe University, 1-1 Rokkodai-cho, Nada, Kobe 657-8501, Japan*

Received 24 November 2005; received in revised form 17 May 2006; accepted 18 July 2006

Abstract

Recently, tracking-type travel data collection methods using mobile communication instruments have developed rapidly. The detailed and accurate travel data are useful for understanding travel behaviour and evaluating the actual effects of transport policy such as traffic demand management. This paper studies a simple index of a tourist behaviour using tracking data collected with a mobile instrument. Based on the proposed index and actual data collected in Kobe, cluster analysis is applied to find the topological characteristics of tourist behaviour.

Keywords: Travel behaviour; Mobile instrument; Tracking data

1. Background and research objectives

The behavioural data of individual travellers in urban areas are essential for travel demand analysis and prediction required in the planning and management of urban transport systems. However, behavioural surveys based on questionnaires are not always sufficient for measuring microscopic travel behaviour in space-time dimensions. Recently, mobile communication technologies have advanced rapidly. They have a great potential to be used as survey instruments for observing individual travel behaviour. As discussed in the second chapter, various data collection instruments have become available for travel data collection.

The travel behaviour data collected using mobile instruments is sequential 'dot' in space-time dimensions. Each dot has longitude, latitude and time of day labels. The approach for travel behaviour analysis involves transferring 'dot' to 'trip', and then applying traditional analytical tools such as trip based models. However, the original characteristics of the dot data may be lost in this data transfer process. Therefore, it is necessary to analyse the dot data in their original form.

* Corresponding author. Tel./fax: +81 78 803 6208.
E-mail addresses: asakura@kobe-u.ac.jp (Y. Asakura), iryo@kobe-u.ac.jp (T. Iryo).

0965-8564/$ - see front matter
doi:10.1016/j.tra.2006.07.003

In addition to understanding travel behaviour, the detailed and accurate characteristics of dot data are useful for evaluating the effects and performance of transport policies such as traffic demand management. It is also very important to develop analytical tools for dot data directly so that the original location characteristics of each dot might be retained to the maximum extent. This implies that the topological characteristics implicitly involved in the dot data should be fully analysed.

Tourist behaviour involves a continuous trip and is more complex than ordinary commuting or shopping behaviour. The trajectory of a tourist route contains various items of information that could be used for developing travel guidance policies and marketing services. This paper studies a simple index of tourist behaviour using tracking data collected with a mobile instrument. Based on the proposed index and actual data collected in Kobe, cluster analysis is applied to find the topological characteristics of tourist behaviour.

2. Travel data collection with mobile instruments

Various mobile communication tools can be used for travel data collection: these include GPS (Global Positioning System), cellular phones (including personal handy phone systems or PHS), RF (radio frequency) ID tags, etc. GPS-based travel data collection systems have been available since the middle of the 1990s. Zito et al. (1995) applied GPS to 'travel time survey by floating vehicles'. D'Este et al. (1999) developed a GPS-based system to measure the performance of traffic systems. Murakami and Wagner (1999) studied the results of a GPS-based experiment for 100 households in the USA. Draijer et al. (2000) showed the results of a pilot study of a GPS-based behavioural survey for all travel modes in the Netherlands. Doherty et al. (2001) described a comprehensive approach that combined GPS and GIS (Geographic Information Systems) technologies with a recently developed computerized activity scheduling survey. Wolf et al. (2001) demonstrated that it would be feasible to derive the purpose of the trip from the GPS data by using a spatially accurate and comprehensive GIS. Forrest and Pearson (2005) found that the number of trips in the GPS data was much greater than the number reported in the computer-assisted telephone interview (CATI) data. Wolf (2004) presented an overview of the recent applications of various location collection technologies in travel behaviour surveys and technology trends for the future.

Although GPS-based travel data collection systems are suitable for monitoring vehicular traffic, it is not always applicable for observing human behaviour since the GPS requires an open sky environment to detect signals from multiple satellites. This problem may be resolved if auxiliary signal transmitters are distributed in the areas where satellite signals are not available. Apart from GPS-based studies, Asakura et al. (1999), and Asakura and Hato (2001) investigated the application of a cellular phone named personal handy phone system (PHS) to monitor individual travel behaviour in Japan. Further review of the data collection methods using mobile instruments can be found in Asakura and Hato (2004). The Global System for Mobile Communication (GSM) based tracking methods have been developed in Germany. Wermuth et al. (2003) used GSM for automatic location positioning and developed a computer-assisted data collection method named 'tele travel system (TTS)'. Kracht (2004) explained how the parallel use of GPS and GSM tracking technology could produce reliable tracking results. Portable electronic devices equipped with GPS/GSM were used to conduct real-time on-site interviews with the respondents.

Recent technologies in GPS assisted cellular phones enabled development of various data collection methods. Ohmori et al. (2006) presented the results of an activity diary survey using a GPS mobile phone. Hato (2006) developed a wearable data collection instrument named mobile activity loggers (MoALs) supported by GPS cellular phones. Itsubo and Hato (2006) examined the effectiveness of a travel data collection system using GPS equipped cellular phones combined with a web diary. As a result of these developments, mobile communication instruments will be very essential for travel data collection and analysis in the near future.

3. Analysis of tourist behaviour using a route topology index

The dot data of travel behaviour obtained with mobile communication instruments has more detailed space-time information than trip data. It is possible to apply traditional trip-based travel behavioural models after transferring dot data to trip data. Asakura and Hato (2004) proposed the 'stay-and-move' identification and 'route identification' algorithms as data transfer tools for dot data. However, the detailed and original information of dot data may be lost through the data transfer process. In particular, the location specific information of dot data such as spot speed and its fluctuations will not be retained in the trip data. A sequence of dot data represents the trajectory of a traveller in space-time dimensions. Since the transferred trip data can be interpreted as a rough approximation of the continuous movements of a traveller; detailed topological characteristics of the trajectory could not be analysed by using trip data. It is necessary to develop methodologies for analysing the travel behaviour by using dot data directly.

A 'travel tour' is a sequential and complex movement of a tourist in space. It is desirable to describe the tour route as a whole without breaking the tour into independent trips. The movement of a tourist in space has its own topological characteristics. Thus, it is necessary to consider the topological aspects of a tour route when we analyse the travel behaviour of the tourist. In the following sections, a simple index is proposed for describing and analysing a tour route. The topological aspects of a tour are included in the index.

3.1. Route topology index

A tour route is assumed to start from a specific point and finish at the same starting point after travelling in the tour area. One of the simplest shapes of a tour route is a circle. When a circular route is observed at a specific point in the area, one of the three relations can be observed between the circular route and the observation point. A 'reference point' term is used instead of the specific point in the following sections of this paper. The first case is that the direction of the circular route is in a clockwise direction around the reference point. The second case is that the direction of the route is in an anticlockwise direction around the reference point. The third case is that the reference point is not located within the internal area of the circular route.

In order to identify these three possible cases, we propose a route topology index (RTI). It is defined as +1, −1 or 0 for each of the above three cases. The RTI value depends on the reference point that can be located at any place in the area. In order to describe a more complex tour route, the RTI can be evaluated at multiple reference points. The RTI vector of a tour route is then introduced. The *i*th element of the RTI vector denotes the RTI for the *i*th reference point. For example, the RTI of the tour route shown in Fig. 1 changes to −1, 0, +1, −1 and −1 for reference points from 1 to 5, respectively. Then, the RTI vector is $\mathbf{R} = \{-1, 0, +1, -1, -1\}$. When the route does not make a circuit, the route topology index could be defined as the cumulative angle of a tour route rotating around a reference point.

3.2. Clustering tourist behaviour

The route topology index can be used to evaluate the similarity and the dissimilarity of tour routes among different tourists. The difference between two tourists A and B is defined as

$$D^2 = \|\mathbf{R}_A - \mathbf{R}_B\| = \sum_i (r_{Ai} - r_{Bi})^2 \tag{1}$$

$\mathbf{R}_A = \{r_{Ai}\}$ and $\mathbf{R}_B = \{r_{Bi}\}$ denote the RTI vectors for tourists A and B, respectively. This index is called 'distance'; however, this has no unit like metres. The element of the vector is the RTI of the *i*th reference point.

The distance defined by Eq. (1) is used for the clustering method when there are a certain number of tourists. This study adopts a hierarchical clustering method proposed by Johnson (1967). For the initial setting, the distance between each pair of tourists is calculated. The pair of tourists with the minimum distance consists of

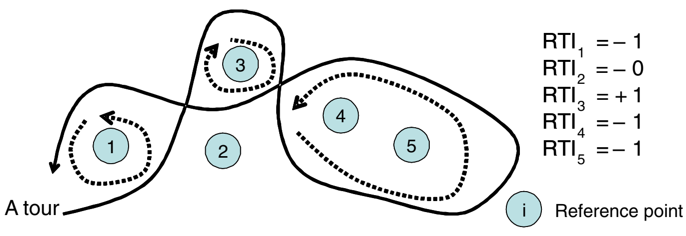

Fig. 1. Route topology index.

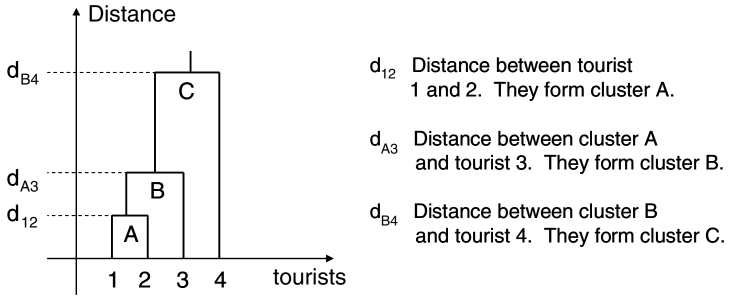

Fig. 2. Schematic explanation of a dendrogram.

the first (lowest) cluster. The average of the two RTI vectors of the tourists in the first cluster is calculated. This represents the updated RTI vector of the first cluster. The next step is to calculate the distance between each pair consisting of the first cluster and/or the rest of the tourists. The pair with the minimum distance is chosen as the second cluster. This procedure is repeated until all the tourists are included in one cluster (the highest cluster). A dendrogram is drawn to recognize clusters and visualize its similarity. The dendogram describes the similarity among travellers and clusters by a tree whose lowest branches represent each traveller and the highest root shows the highest cluster. The distances where clusters are formulated are shown as the height of the nodes where two lower branches merge. An example of a dendrogram is shown in Fig. 2. When the RTI vectors of some tourists are similar, these tourists will be grouped in the lower cluster. This implies that the clustering would be equivalent to finding the market segment of tourists who have shown similar touring behaviour. If there is a cluster that involves many tourists in a relatively lower stage, this cluster indicates a large group of tourists that show a similar sightseeing pattern. Thus, dominant touring movements in the area could be identified in this manner.

Clustering of sub-areas is also possible. A sub-area is assumed to correspond to a reference point. The difference between two sub-areas is evaluated similar to Eq. (1) and spatial clustering becomes possible.

4. Kitano area tourist survey and analysis

4.1. Location positioning test and tourist survey

The proposed method was examined in an actual urban area. Kobe Kitano, a popular sightseeing area, was chosen as the test site. The test site area was approximately 300 m × 300 m. Although this area may not be substantially large, it includes many sightseeing spots, and the area is suitable for a half-day travel experiment. Detailed results of the location positioning test were presented in Asakura and Iryo (2005). A wearable data logger named personal activity monitor (PEAMON) was developed by the authors and was used as the data collection instrument. PEAMON recorded the signal strength of the neighbouring base stations and their identification numbers in a sequence of time. The data collection interval was set at 15 s. The location of the position was calculated by using the signal strength pattern. This procedure was described in Asakura and Iryo (2005). In order to evaluate the location positioning error, a test walker with PEAMON was asked to move in the area along the predetermined route and time. The dot data obtained by the test walker and this predetermined trajectory (true data) were compared. The averaged error distance of the location positioning test was 28.8 m. Table 1 shows the distribution of the error distance. The error distance that was less than 40 m was approximately 80%. The performance of the location positioning method was almost satisfactory when we applied this method to trace the movement of tourists in the area.

When we evaluate the route topology index of a tourist by using his/her dot data, it is not always necessary to replace an observed dot on a corresponding road section. This is because the route topology index for a reference point represents the relation between the location of the reference point and the direction of a tour route. Thus, it is sufficient to know whether the direction of the tour route around the reference point is clockwise or anticlockwise. This implies that the map matching of dot data to a road map is not always necessary.

Table 1
Distribution of error distance in the field test

Error distance (m)	Number of observation	Frequency (%)	Cumulative frequency (%)
0–10	94	18.2	18.2
10–20	114	22.1	40.3
20–30	119	23.0	63.3
30–40	74	14.3	77.6
40–50	51	9.9	87.5
50–60	23	4.4	91.9
60–70	27	5.2	97.1
70–80	4	0.8	97.9
>80	11	2.1	100
Total	517	100	

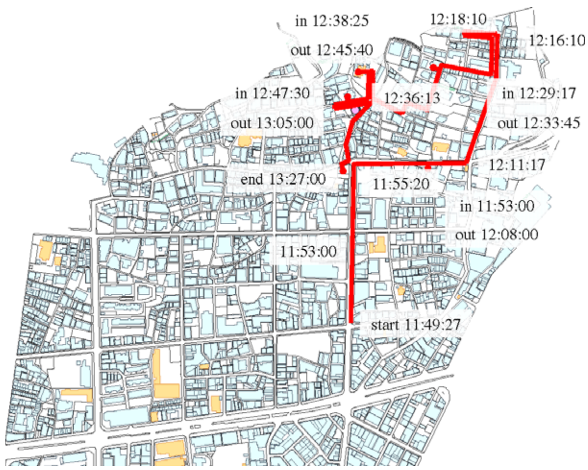

Fig. 3. Example of a tour route.

However, in the following case study, the dot data were replaced by the nearest road links or sightseeing facilities in order to evaluate other tour characteristics, such as travel distance and staying time at facilities. The map matching methods proposed by previous studies (Asakura and Hato, 2004; Marchal et al., 2005) could be applied.

A one day travel survey in the Kitano area was carried out in November 2003. The tourists selected were 56 pairs of female students. They were asked to carry the mobile communication instrument during the tour. Fig. 3 shows the tracking route of an example pair. In addition, the arrival and departure times at each location in the area and passing time at the intersection are also shown in the figure. 'In' and 'out' indicate their arrival and the departure times, respectively. This pair walked a distance of 1700 m and visited four locations. The staying time at those locations was 38 min, and the time spent for moving from one location to another was 53 min. For all the pairs, the average distance travelled, the average staying time and the average number of visited spots facilities in the area were 2105 m, 207 min and six locations, respectively.

4.2. Clustering behaviour of tourists with RTI

When we apply the clustering method using a route topology index (RTI), it is necessary to first allocate reference points in the area. Two different allocation patterns have been examined – one is a grid and the other is a polygon (street block). In this paper, the results of polygonal allocation are shown because it would be more appropriate to reflect the characteristics of each block. There were 21 street blocks in the area, and the average size of the block was approximately 70 m × 70 m. The route topology indexes are calculated at all reference points.

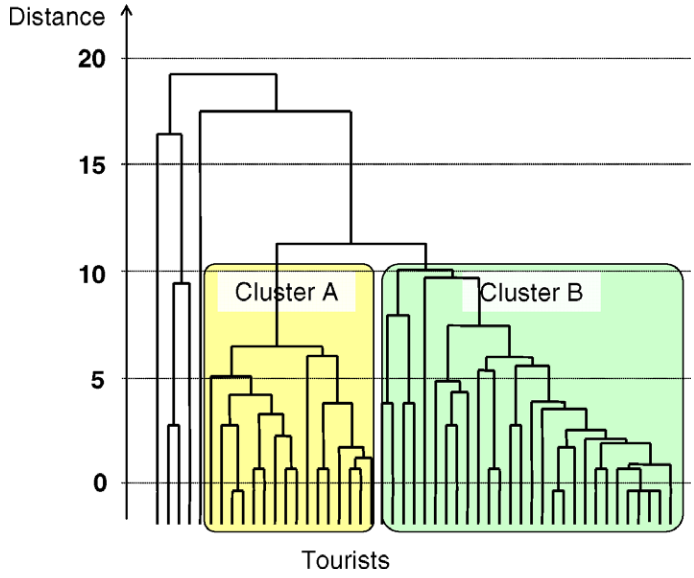

Fig. 4. Cluster analysis dendrogram based on the route topology index.

Fig. 4 depicts the dendrogram from the cluster analysis. A dendrogram is defined as follows: the vertical axis indicates the distance defined in (1). When the threshold distance is set at 10, two major clusters A and B are found. Cluster A involves 16 tourists and cluster B involves 28 tourists. Only 5 tourists are not involved in these two major clusters. Fig. 5 shows the typical tour routes that belong to cluster A and B, respectively. The direction of almost all the tour routes in cluster A is in an anticlockwise direction and these tour routes have a relatively larger loop. The direction of the tour route in cluster B is generally in a clockwise direction and smaller loops are included in the tour route. Some of the tour routes in cluster B just show a simple out-and-back travel without a loop.

Cluster analysis using route topology index is used to understand the topological similarity of tour routes between tourists. It is possible to identify some groups of tourists with similar routing strategies. The tourists within a cluster may have the similar preference of sightseeing behaviour. Thus, the proposed method would be applicable to understand the implicit routing characteristics of tourists. The example also indicates that the representative tour routes and their topological characteristics can be identified through the cluster analysis. This would be useful to discuss the travel information systems for tourists, such as the location of the route guidance in the area.

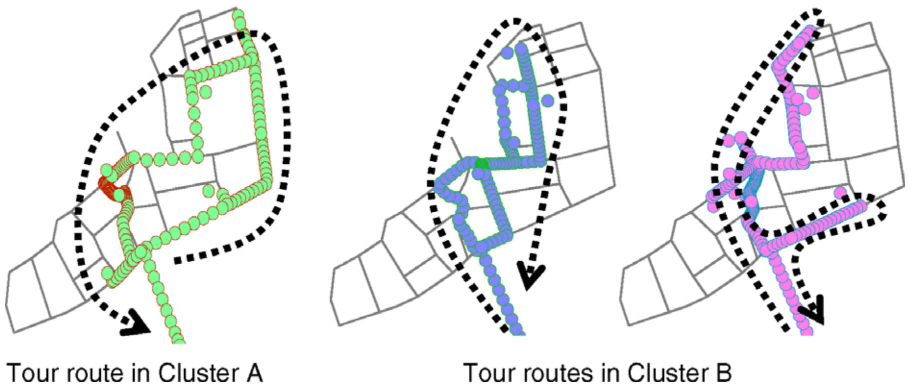

Fig. 5. Typical tour routes in clusters A and B.

5. Discussion

This paper discussed how dot data of travel behaviour can be analysed. The location characteristics of travel routes were partly reflected in the proposed route topology index. Although the reference points should be carefully allocated, the RTI vector appears to be useful for describing a complex tour route. For future improvement, time-based information could also be reflected in the RTI. For example, the product of spot speed and the angle of a loop around a reference point would serve as an improved index.

For marketing purposes, it would be interesting to find the hidden behaviour of tourists, for example, shop A is always visited after a sightseeing visit to location B. Other statistical methods using the RTI would be more useful to explain such tendencies of the tourists. The index and analytical method studied in this paper could be extended to travel behaviour analysis for evaluating the effects of transport policies.

Acknowledgements

The authors show their sincere appreciation to the graduate students of Kobe University, who helped with the field survey and data analysis. This research is partly supported by the Grant-in-Aid for Scientific Research of JSPS (Japan Society for the Promotion of Science) #15360279 and #17206053.

References

Asakura, Y., Hato, E., Nishibe, Y., Daito, T., Tanabe, J., Koshima, H., 1999. Monitoring travel behaviour using PHS based location positioning service system. In: Proceedings of the 6th ITS Conference in Toronto (in CD-ROM).

Asakura, Y., Hato, E., 2001. Analysis of travel behaviour using positioning function of mobile communication devices. In: Hensher, D. (Ed.), Travel Behaviour Research: The Leading Edge. Pergamon, pp. 885–899.

Asakura, Y., Hato, E., 2004. Tracking survey for individual travel behaviour using mobile communication instruments. Transportation Research C 12 (3/4), 273–291.

Asakura, Y., Iryo, T., 2005. Tracking individual travel behaviour using mobile phone without information from base stations. Journal of Advanced Transportation 39 (1), 105–116.

D'Este, G.M., Zito, R., Taylor, M.A.P., 1999. Using GPS to measure traffic system performance. Journal of Computer-Aided Civil and Infrastructure Engineering 14, 273–283.

Doherty, S.T., Noel, N., Gosselin, M-L., Sirois, C., Ueno, M., 2001. Moving beyond observed outcomes: integrating global positioning systems and interactive computer-based travel behavior surveys. Transportation Research Circular, 449–466.

Draijer, G., Kalfs, N., Perdok, J., 2000. Global positioning system as data collection method for travel research. Transportation Research Record 1719, 147–153.

Forrest, T.L., Pearson, D.F., 2005. Comparison of trip determination methods in household travel surveys enhanced by a global positioning system. Transportation Research Record 1917, 63–71.

Hato, E., 2006. Development of MoALs (mobile activity loggers supported by GPS-phones) for travel behavior analysis. Paper presented at Transportation Research Board 85th Annual Meeting.

Itsubo, S., Hato, E., 2006. Effectiveness of household travel survey using gps-equipped cell phones and web diary: comparative study with paper-based travel survey. Paper presented at Transportation Research Board 85th Annual Meeting.

Johnson, S.C., 1967. Hierarchical clustering schemes. Psychometrika 32 (3), 241–254.

Kracht, M., 2004. Tracking and interviewing individuals with GPS and GSM technology on mobile electronic devices. Paper Presented at the 7th International Conference on Travel Survey Methods, Costa Rica.

Marchal, F., Hackney, J., Axhausen, K.W., 2005. Efficient map matching of large global positioning system data sets: tests on speed-monitoring experiment in Zürich. Transportation Research Record 1935, 93–100.

Murakami, E., Wagner, D.P., 1999. Can using global positioning system (GPS) improve trip reporting? Transportation Research C 7, 149–165.

Ohmori, N., Nakazato, M., Sasaki, K., Nishii, K., Harata, N., 2006. Activity diary surveys using GPS mobile phones and PDA. Paper Presented at Transportation Research Board 85th Annual Meeting.

Wermuth, M., Sommer, C., Kreitz, M., 2003. Impact of new technologies in travel surveys. In: Stopher, P., Jones, P. (Eds.), Transport Survey Quality and Innovation. Pergamon Press, pp. 465–469.

Wolf, J., Guensler, R., Bachman, W., 2001. Elimination of the travel diary: experiment to derive trip purpose from global positioning system travel data. Transportation Research Record 1768, 125–134.

Wolf, J., 2004. Applications of new technologies in travel surveys. Paper Presented at the 7th International Conference on Travel Survey Methods, Costa Rica.

Zito, R., D'Este, G.M., Taylor, M.A.P., 1995. Global positioning systems in the time domain: how useful a tool for intelligent vehicle-highway systems? Transportation Research C. 3 (4), 193–209.