Expanding Typologies of Tourists' Spatio-temporal Activities Using the Sequence Alignment Method

Junya Kawase[a], and
Fumiko Ito[a]

[a] Department of Urban System Science
Tokyo Metropolitan University, Japan
j.kawase0922@gmail.com

Abstract

Several studies have been conducted on tourists' activities that are suitable for understanding the typologies of these activities as a whole in time and space using the sequence alignment method (SAM). In this study, we expanded this analysis using SAM to clarify the characteristics of tourists' activities in detail. Corresponding data were acquired by conducting a survey among tourists visiting the Ueno Zoo. The data collected from these tourists were then classified based on their preferred route types and the entrance/exit gates used by them. Consequently, we identified some of the typical activities in which they engaged and obtained detailed information on the characteristics and characteristic elements of those activities.

Keywords: sequence alignment method; spatio-temporal activity; global positioning system; time-space path; Kernel density estimation

1 Introduction

A better understanding of the spatio-temporal activities of tourists is beneficial for the tourism industry. In most tourists' activities researches, it is common that statistical analysis is applied to activity data of classified subjects by age, gender or their attribute data. However, tourists with similar attribute data do not necessarily take similar activities. Consequently, we have to understand characteristics of tourists' activities and classify subjects of researches by their characteristics. There have been many studies aimed at understanding tourists' activities by visualisation, typologies, and other methods. However, because of wide variations in their activity patterns, it is difficult to obtain detailed information on the characteristics of tourists' spatio-temporal activities. Tourists are very arbitrary in choosing the sites they visit, the combination and order in which they visit those sites, and the time they spend visiting each site. To solve this difficulty, some researchers have applied the method of sequence alignment analysis to the typologies of tourist activities. The sequence alignment method (SAM) is the basic tool of bioinformatics, and it is the identification of residue-residue correspondence (Lesk 2013). Wilson (1998) illustrated its use in the analysis of daily activity patterns derived from time-use diaries. Shoval and Isaacson (2007) conducted GPS-tracking activity analysis of tourists visiting the Old City of Akko (Israel) and obtained a taxonomic guide tree from which they derived clusters of typical patterns by applying SAM. In a similar fashion, Shoval et al. (2015) conducted typologies of tourists visiting Hong Kong.

In order to perform a detail discussion for marketing or promoting tourists' activities, we have to understand tourists' activities in detail. As above studies indicate, application of SAM to all the subjects' activities is suitable for obtaining the typologies of tourists' activities in the destinations. In other words, they obtained comprehensive sketches of tourists' activities in the above-mentioned destinations. However, such a comprehensive sketches may not be enough for the detail discussion. In addition, this approach may not be suitable for some cases, such as cases in which the starting/ending points of subjects' routes are different to each subject, the cases of subjects' fields of activities are different obviously to each subjects. In such cases, it is natural that subjects having same starting/ending points or fields of activities are considered same clusters. These similarities can be clarified by viewing their GPS logs on map enough without using SAM. Therefore, it is preferable to divide the subjects based on their broad routes categories classified by GPS logs directly. The process of determining their broad routes categories is different depending on the purpose of the study.

In this paper, we will introduce our attempt for resolution of above-mentioned issues. We roughly divided subjects visiting a zoo based on the areas and the entrance/exit gates they passed. For each group of visitors, SAM is applied to their trajectories for typologies of their activities. Although this work seems simple, we consider that it is necessary to remove one by one impediment to the better use of SAM by such a work.

2 Summary of Analysis

We used ClustalTXY—an alignment software package—to calculate the representative Hagerstrand time-space trajectories (Wilson 2008). ClustalTXY software package calculates the distance between two sequences by using an algorithm based on both activity and geographic difference of subject. The software deals with two files that represent one subject's event: Sequence file contains information related to their activity, and Location file contains information related to where this event occurred. Sequence similarities are measured using the weighted sum of the activity mismatch cost and the location mismatch cost.

We selected the Ueno Zoo (Taito, Tokyo) for our study because it is well suited for the application of SAM to GPS tracking of tourists' activity surveys. Zoo visitors' spatio-temporal activities are limited in time and space to the zoo area. Fig. 1 shows the map of Ueno Zoo. Ueno Zoo has a 14.2 ha site area, two entrance gates and three exit gates. The site is partitioned into two gardens (East Garden and West Garden). A bridge and a small monorail line connect the two gardens. All visitors travel around the site on foot except for movements between the two gardens. We divided the zoo site into thirty zones based on their spatial connections and functions. Each of the zones was assigned a code with two alphabets (Fig. 1).

We conducted the survey on 30th June 2013 at the Ueno Zoo to obtain GPS data of visitors. At the Main Gate, we distributed small GPS loggers to visitors who agreed to participate in our survey. When the visitors left the zoo from one of the three exit gates, we collected the loggers from them, and asked them to answer questionnaires to obtain their attribute data, such as age, gender, and accompanying persons. We obtained 113 valid sets of GPS logs and visitors' attribute data.

Using the obtained GPS data, we made a Sequence file and a Location file for each of the subjects. We replaced the subjects' location with a code representing a zone once every minute. Thus, we created a Sequence file. If applicable logs did not exist or were error logs, we substituted gaps (-) in place of the error logs. In this approach, the Sequence files contain geographical information of activities of the experiment subjects. Thus, we inputted the elapsed time from the opening of zoo to the Location files. They were expected to represent the temporal mismatch cost. This approach is similar to the approach used by Yabe (2010).

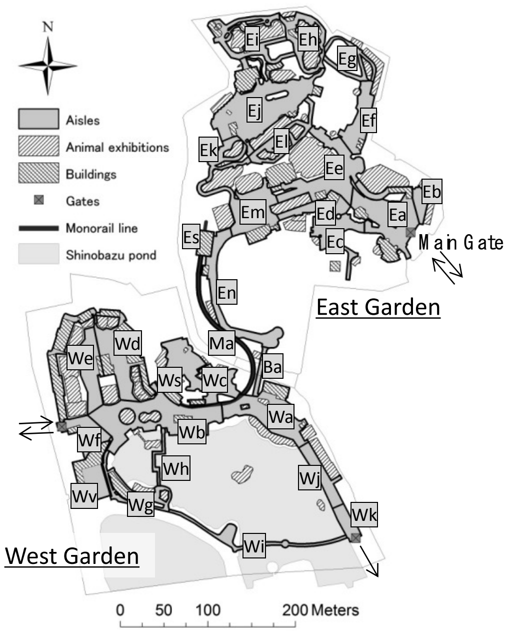

Fig. 1. A Map of Ueno Zoo

3 Analysis and Results

3.1 Preliminary Analysis

We conducted typologies of all subjects' trajectories to clarify their overall picture as a preliminary analysis for confirming points needing improvement of existing approach. Consequently, we obtained a cladogram and assigned nomenclature to each six clusters. We tried to interpret the details of typical routes of each cluster. However, these descriptions didn't represent adequately the characteristics of each subject' activity included into each cluster. Each cluster contained many types of routes practically. Although there were incomplete typologies, we could understand some characteristics of visitors' activities as overall picture. We found that the clusters have characteristics by three points:

(i) broad routes (clear moving as represented by movement between the two gardens or gates used by them)
(ii) whether stayed for tens of minutes somewhere or not
(iii) which zone they stayed for tens of minutes

These results are useful to some extent, but it is roughly that descriptions detailed characteristics of visitors' activities of Ueno Zoo. Therefore, we considered that an approach for clarifying the more detailed characteristics of subjects' activities is required.

3.2 Main Analysis and Result

For clarifying the more detailed characteristics of subjects' activities, we tried to expand the analysis. We divided the 113 subjects into three groups based on their broad routes taken while traversing the zoo as a whole. They are as follows:

(i) East Garden → West Garden Group (46 subjects)
(ii) East Garden → West Garden → East Garden Group (54 subjects)
(iii) other route Group (13 subjects, excluded from this analysis)

Subjects in Group (i) departed from the Main Gate, went around the East Garden, moved to the West Garden, and exited the zoo using one of the two gates in the West Garden. Subjects in Group (ii) went back to the East Garden and exited the zoo using the Main Gate. These two routes accounted for the majority.

Thus, we conducted typologies of their routes to clarify the detailed characteristics by SAM for above mentioned two groups. Consequently, we obtained two cladograms and assigned nomenclature the clusters. We tried to interpret the details of typical routes of each cluster by using time-space path maps and Kernel density estimation maps. The time-space path maps represent subjects' movement by lines that increase in height by one meter for every elapsed minute. Kernel density estimation maps represent the hot spots from their GPS logs.

Group (i) has seven clusters (Cluster 1 to 7). Cluster 1 is the most typical type of Group (i) (Main Gate → Ea → Ee → Ef → Eg → Eh → Ei → Ej → Ek → El → Ee → Em → West Garden), and Cluster 2, 3, 4 and 5 are considered as the derivatives of Cluster 1. Cluster 6 contained a halt for tens of minutes at Shinobazu pond terrace ("Wb" in Fig.1). Both Cluster 1 and Cluster 7 moved on the north side of East Garden, but the direction are opposite. For Group (i), the direction of route in north side of East Garden and a halt for tens of minutes at Shinobazu pond terrace are the most characteristic elements of their activities.

Group (ii) has five clusters (Cluster 8 to 12). Cluster 8, 9 and 10 are typical types of Group (ii), where they halted for tens of minutes can be considered as characteristic elements of their activities. The typical route of these clusters is through Main Gate → Ea → Ee → Ef → Eg →Eh → Ei →Ej → Ek → El → Ee (or direct to Em), moving to the north side of West Garden later, and then back to East Garden. Cluster 8 halted for tens of minutes in Ej zone. Cluster 10 halted for tens of minutes in Wb zone. Cluster 9 did not halt for tens of minutes anywhere. Cluster 12 contains many routes; however, we can point out they went around south side of West Garden characteristically.

With a few exceptions, subjects of Group (i) didn't halt for tens of minute at any zones other than Wb unlike subjects of Group (ii). There may be relationship between routes of visitors of Ueno Zoo and when they take a rest. Thus, we believe beneficial results could be obtained by conducting an additional study focused on this issue. For visitors of Ueno Zoo, north side of East Garden is also one of the important elements. It is an important management point for the Ueno Zoo.

4 Conclusion and Future Work

We tried to expand the analysis by SAM to clarify the detailed characteristics of tourists' activities. We divided subjects visiting the zoo based on the areas and the entrance/exit gate they used and conducted topologies of their routes by using SAM. We divided subjects into three groups based on their broad routes clarified by their GPS logs, and conducted typologies of their activities.

As a result, we observed some typical patterns in their activities and the detailed characteristics of the patterns, and clarified the characteristic elements of the visitors' activities. The result of this study will contribute to the management of the zoo. As Ueno Zoo has a footfall of ten thousand or more visitors per day, these results is beneficial for improvement of institution for mitigating congestion.

However, this method still requests several improvements. In the first, we have to consider more carefully the influence of long halt time in one zone on the typologies of tourist activities by SAM. Halting for tens of minutes in one zone is represented by continuous same code (if he/she stayed for ten minutes in Ea zone, Sequence file contained codes "EaEaEaEaEaEaEaEaEaEa"). Such a continuous same code may cause unfairly high prioritization in calculating mismatch cost. Secondly, we should discuss about the size of zones. Although, in this study and previous studies used SAM, the size of zones differs one by one, this should be discussed in relation to time resolution issue. Further analysis is required to examine the above-mentioned issues.

We will expand the number of test subjects for further test of the results in this study. In addition, we will conduct similar surveys in other target area for further development of this study.

References

Lesk, A. (2013). *Introduction to bioinformatics.* Oxford University Press.

Shoval, N., & Isaacson, M. (2007). Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American geographers* 97(2); 282-297.

Shoval, N., McKercher, B., Birenboim, A., & Ng, E. (2015). The application of a sequence alignment method to the creation of typologies of tourist activity in time and space. *Environment and Planning B: Planning and Design* 42(1); 76-94.

Yabe, N. (2010). Application of Sequence Alignment Analysis to GPS Tracking Data. Proceedings of the 19th Annual Meeting of GIS Association of Japan; CD-ROM (in Japanese).

Wilson, W. C. (1998). Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A* 30(6); 1017-1038.

Wilson, C. (2008). Activity patterns in space and time: calculating representative Hagerstrand trajectories. *Transportation* 35(4): 485-499.