Travel Pattern Analysis from Trajectories Based on Hierarchical Classification of Stays

Ryo Inoue, Motohide Tsukahara

Graduate School of Information Sciences, Tohoku University, 6-6-06 Aramaki Aoba, Aoba, Sendai 980-8579, Japan
Email: {rinoue, tsukahara}@plan.civil.tohoku.ac.jp

Abstract

It has recently become possible to utilize a large amount of detailed trajectories for travel pattern analysis. However, methods proposed in this area have limitations in applying the data taken over a wide area. To overcome these limitations, this paper proposes extraction of travel patterns through hierarchical classifications of stays and travel patterns based on the Huffman coding algorithm. The results of experiments conducted using the proposed method on trajectories in Okinawa, Japan confirm its feasibility for analyzing travel patterns.

1. Introduction

A huge amount of trajectory data has become available with the widespread use of positioning technology. One of its applications is tourist activity analysis, with research focused on extracting typical travel patterns actively underway.

Previous studies in this area first detect and code stays, then analyze travel patterns from stay sequences. Many approaches have been proposed for the latter process. Zheng *et al.* (2007) analyzed transition probabilities between sites, Giannotti *et al.* (2007) mined sequential patterns, and Shoval and Isaacson (2007) and Shoval *et al.* (2015) used sequence alignment. However, the former process has not been sufficiently investigated. Most of the studies conducted simply judged stays by the preset area classification. Giannotti *et al.* (2007) and Zhang *et al.* (2009) extracted stays via density-based analyses; however, no threshold setting criteria were presented.

Previous analyses are limited especially when the study area is wide. However, because places of interest and arrival and departure times vary, the number of travel patterns is significant. This results in difficulty finding similar patterns. Thus, adjusting the resolution of the analysis to reduce patterns is essential.

This study focuses on "stays," which each consists of a visited place and its arrival and departure time. The analysis resolution of a stay may vary from site- to region-basis in space and from minute- to day-basis in time. Analysis methods whose results change according to resolution settings are useless as they cause difficulty interpreting results. We believe that hierarchical classification of travel patterns to spatio-temporal resolution settings is key to solving the problem. Thus, this paper proposes classification of travel patterns based on Huffman coding of stays, and reports on tests of its applicability using trajectory data obtained in Okinawa, Japan.

2. Hierarchical Classification of Stays and Travel Patterns

Huffman coding outputs compact code with average code length close to Shannon entropy: the average information contained in the data. It constructs a binary tree by repeating the aggregation of the two least-frequent data elements. The algorithm can classify stays hierarchically; however, only their frequencies are considered, their similarities are disregarded. The stay classification approach proposed in this paper permits grouping stays according to similarity.

2.1 Stay Classification

Let d_{ij} denote the Euclidian distance between the locations of stays i and j, at_i and dt_i denote the arrival and departure times of stay i, respectively, and α denote a weight parameter between spatial and time difference. The proximity of stays is defined by

$$p_{ij} = \alpha d_{ij} + (1-\alpha)\{|at_i - at_j| + |dt_i - dt_j|\}. \ (0 \le \alpha \le 1) \quad (1)$$

By utilizing the link setting condition of relative neighborhood graph, stays are adjudged to have similarity if the following conditions are satisfied:

$$p_{ij} \le \max\{p_{ik}, p_{jk}\} \quad \forall k \ne i, j. \quad (2)$$

Through the classification of stays, the stay pair with the smallest proximity is grouped first if multiple pairs have the same frequencies, and the similarity of stay groups is evaluated by their elements.

Every node on the tree represents a class, and nodes on a path from the root to leaf nodes represent classes from low to high resolutions.

The classifications are dependent on α. We propose to select α that minimizes the average code length, as it outputs the most compact classification of stays. The average code length in this method is larger than that of Huffman coding; the difference indicates additional information by considering stay similarity.

2.3 Travel Pattern Classification

We classify travel patterns based on stay classification by dividing stay classes individually. Figure 1 indicates an initial state in which the nodes with bold lines signify active classes. Figure 1(a) illustrates a state in which the root node of the stay classification tree is divided into classes #0 and #1, and Figure 1(b) shows a travel pattern classification at this stage: classes with stay #0, #1, and both stays. Figure 2 shows the classification in the next stage. Class #1 that covers a wider range is divided, and travel patterns are hierarchically classified into seven groups.

3. Application

3.1 Trajectory Data, Stay Detection, and Minimum Resolution Settings

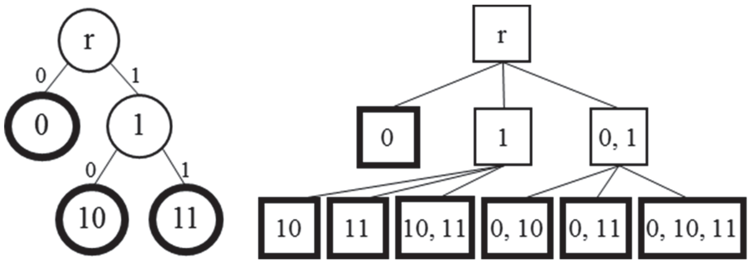
The proposed method was applied to rental car trajectories in Okinawa, Japan. The trajectories consisted of positions at one to five-second intervals observed by onboard GPS devices of 614 tourist groups from August 29 to December 1, 2014 in three to five-day trips (Figure 3).

Stays are defined as states where the tourists stayed in a 100-meter radius circle for more than 15 minutes, and their locations are defined as the centroids of trajectories. Trajectories with less than four stays per day were excluded from farther analysis: 4,167 stays and 823 daily sequences of stays were extracted.

The minimum resolution for analysis was set as follows. Neighbor stay locations were aggregated to analyze visits to the same site by different tourists; the set of locations within 300

(a) Stays. (b) Travel patterns.

Figure 1. Classifications when stays are classified in two.



(a) Stays. (b) Travel patterns.

Figure 2. Classifications when stay #1 is divided into #10 and #11.

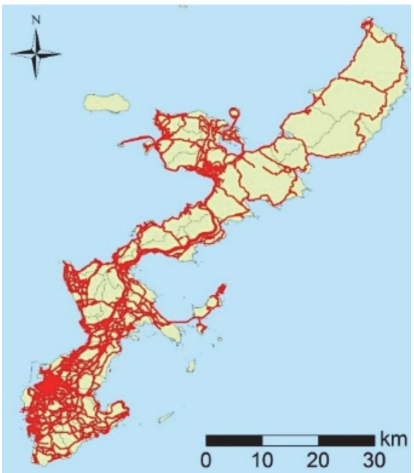


Figure 3. Positions in trajectories.

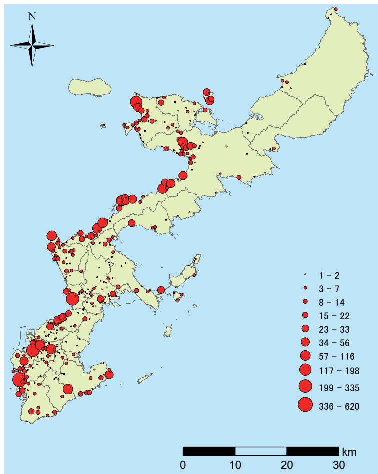


Figure 4. Stay locations and frequencies.

meters were aggregated. Figure 4 shows 399 locations; circles represent locations and their sizes represent visit frequencies. The temporal resolution was set at 10 minutes. On aggregating the same stays, the total number of stays returned was 3,714.

3.2 Hierarchical Classification of Stays and Travel Patterns

α was set to 0.55 to minimize the average code length, by the search within the range from zero to one at 0.05 intervals. When stays were classified in 28 classes, for example, 823 travel sequences classified into 689 patterns, travel patterns became rich in variety. Figure 5 shows the stay classification results and Table 1 shows sample classes. The characteristics of classes are apparent from their locations and arrival and departure times. Further, classes allocated in close proximity to each other on the tree have high similarity. Table 2 shows sub-classes of class “0011” in 156 classifications, a more detailed classification.

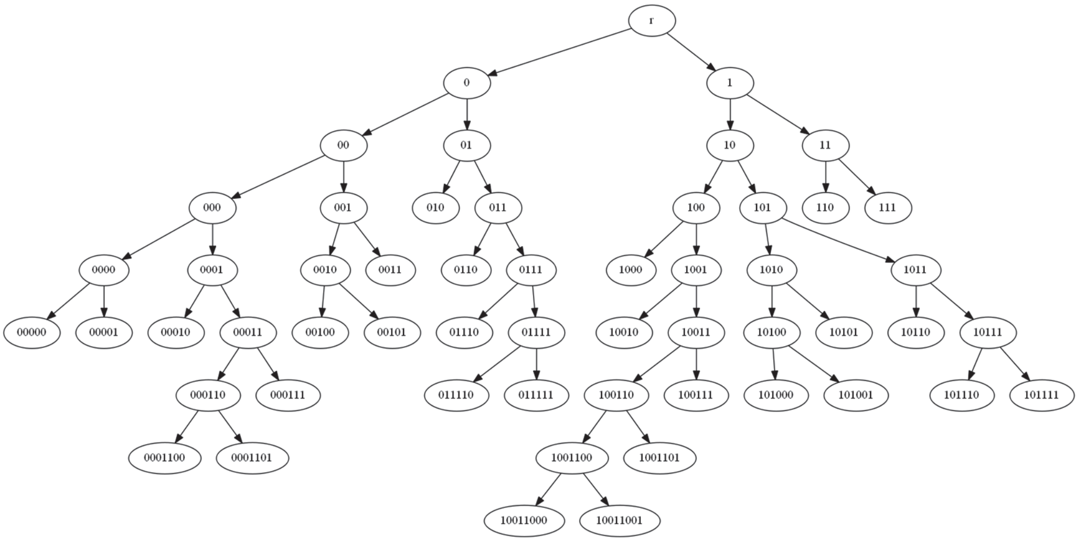


Figure 5. Binary tree of 28 classifications.

Table 1. Stay classes of 28 classifications.

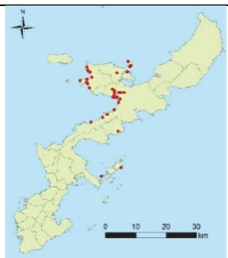
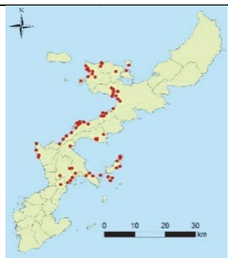
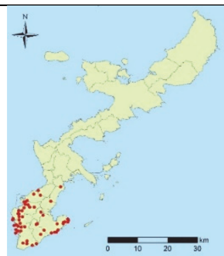
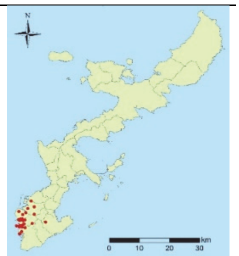
Class	00100	0011	010	0110
Locations				
Frequencies	101	215	414	229
Arrival time: Mean (*SD*)	10:57 (3:24)	16:06 (1:30)	13:48 (2:27)	11:18 (0:50)
Departure time: Mean (*SD*)	12:26 (2:23)	17:05 (1:40)	15:16 (2:20)	12:02 (0:52)

Table 2. Stay classes of 156 classifications.

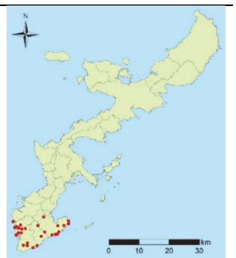
Class	0100	0101
Locations		
Frequencies	210	204
Arrival time: Mean (*SD*)	15:50 (1:28)	11:41 (1:09)
Departure time: Mean (*SD*)	17:18 (1:13)	13:10 (0:55)

Table 3. Most frequent travel patterns at stay classifications in 28 classes.

Travel patterns	Frequencies
010, 0110, 111	8
010, 111	6
0110, 111	5
010, 110, 111	5
010, 0110	5

Table 4. Most frequent travel patterns at stay classifications in 156 classes.

Travel patterns	Frequencies
0110, 1111	3
0101, 0110	3
0100, 0101, 0110	2
0100, 0101, 111011	2
0100, 0101, 111001	2

Tables 3 and 4 show several travel pattern classes, which correspond to the 28 and 156 stay classifications, respectively. Classes {“010,” “0110”} are divided into {“0101,” “0110”} and {“0100,” “0101,” “0110”} as stay “010” is divided into “0100” and “0101.” This confirms that the proposed method can classify travel patterns hierarchically; however, as the travel patterns are diverse, the frequent pattern extraction requires more trajectories.

4. Results

This paper focused on stay classification in travel pattern analyses, and proposed a travel pattern analysis approach from trajectories based on Huffman coding. Experimental results confirm that the proposed method can classify travel patterns; however, the available data were insufficient to discover typical travel patterns. Thus, further analysis with more trajectories is necessary to confirm its effectiveness.

Acknowledgements

Trajectories were provided by the Japan Road Traffic Information Center. This study was supported by JSPS KAKENHI (grants 24241053 and 25249069).

References

Giannotti F, Nanni M, Pinelli F, and Pedreschi D, 2007, Trajectory pattern mining. In: *Proceedings of the 13th ACM SIGKD International Conference on Knowledge Discovery and Data Mining*, 330–339.

Shoval N and Isaacson M, 2007, Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American Geographers*, 97:282–297.

Shoval N, McKercher B, Birenboim A, and 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*, 42(1):76–94.

Zheng Y, Zheng L, Xie X, and Wei M, 2009, Mining interesting locations and travel sequence from GPS trajectories. In: *Proceedings of the International Conference on World Wide Web*, 791–800.