Stay-Move Tree for Summarizing Spatiotemporal Trajectories

Eun-Kyeong Kim[1]

[1] University Research Priority Program (URPP) "Dynamics of Healthy Aging" & Department of Geography, University of Zurich, Zürich, Switzerland
{eun-kyeong.kim}@geo.uzh.ch

Abstract. Summarizing spatiotemporal trajectories of a large number of individual objects or events provides insight into collective patterns of phenomena. A well-defined data model can serve as a vehicle for classifying and analyzing data sets efficiently. This paper proposes *the Stay-Move tree (SM tree)* to represent frequency distributions for types of trajectories by introducing concepts of *stay* and *move*. The proposed tree model was applied to analyzing the Korean Household Travel Survey data. The preliminary results show that the proposed SM trees can potentially be employed to compare/classify spatiotemporal trajectories of different groups (e.g., demographic groups or species of animals). The methodology can potentially be useful to summarize big trajectory data observed from both human and natural phenomena.

Keywords: spatiotemporal trajectory, spatiotemporal data model, movement analysis.

1 Introduction

The rise of new sensor and communication technologies increasingly produces massive trajectory datasets and demands methodologies for analyzing big spatiotemporal trajectory data. Summarizing spatiotemporal trajectories can provide insight into collective patterns of individual moving objects or events. Despite of efforts made for summarizing spatiotemporal trajectories [1-6], there is still a scarcity of representation models for huge trajectory data.

A well-defined data model can facilitate classifying and analyzing data sets efficiently. Modeling movement patterns of moving objects has been carried out by [3-6]. Dodge *et al.* (2008) [3] developed a conceptual framework for moving objects' behavior and classification of movement patterns. Hornsby and Li (2009) [4] introduced a typology for spatiotemporal trajectories of a single object. Schneider *et al.* (2013) [5] characterized daily mobility patterns by directed networks of locations. Inoue and Tsukahara (2016) [6] proposed a hierarchical classification method for categorizing stays in movement trajectories by the frequency of stays. However, many existing studies do not consider a temporal dimension [5-6] or focus on obtaining representative paths of clustered trajectories [1-3].

To summarize large trajectory datasets in a concise but comprehensive way, this study develops a methodology, called a *Stay-Move (SM) tree*, to categorize spatiotemporal trajectories into a simplified trajectory type, represent the (relative) frequency of trajectories of each trajectory type, and compare trajectories of different groups. This study proposes a *Stay-Move (SM) model* as a data model for spatiotemporal trajectories, which is described in Section 2. Then, the SM tree is proposed to arrange all types of trajectories that consist of *stay* and *move* elements and represent a (relative) frequency distribution of types of trajectories of an entire dataset (Fig. 2), which is illustrated in Section 3. The preliminary analysis on Korea Household Travel Survey data collected in 2006 in the Seoul Metropolitan area shows that the proposed methodology allows comparing collective patterns of trajectories of different groups.

2 Stay-Move Model

This study employs a *Stay-Move (SM) model* proposed in author's previous work [7], a typology of spatiotemporal trajectories. To simplify spatiotemporal trajectories, this model adopts the

Spatial big data and machine learning in GIScience, Workshop at GIScience 2018, Melbourne, Australia. Editors: Martin Raubal, Shaowen Wang, Mengyu Guo, David Jonietz, Peter Kiefer.

two concepts of *stay*, defined as a behavior of staying at a place for a significant time, and *move*, defined as a behavior of moving from one place to another by a significant distance because they are basic elements of movements of objects and events (Fig. 1b). This model represents a trajectory as an ordered set of stays and moves. The order of places is determined by the order of visits, regardless of absolute locations of each place (Fig. 1). For example, if one stays at home from the beginning of a day, their home becomes the first place. If they move to places where they already visited, the model takes the previously labeled order of places, which is regressive (Fig. 1b). If they move to a new place, the order of the place increases by one, which is non-regressive (Fig. 1c). A trajectory type is represented as a sequence of numbers where the numbers denote identical stays in order of appearance for the observed time period. For instance, a travel of 'home, work, home' in a day is represented as '121', as is 'work, gym, work.'

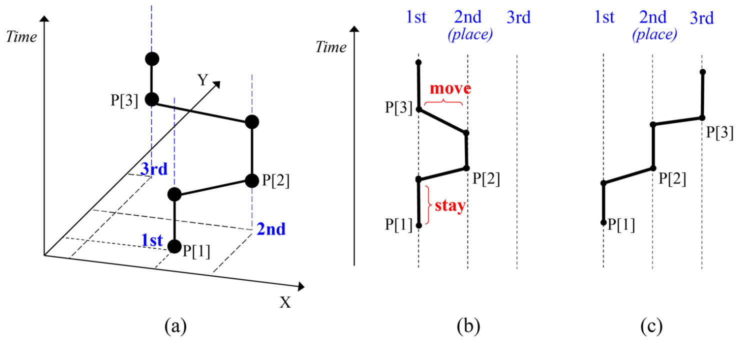

Fig. 1. Concepts of the Stay-Move model. (a) A trajectory in space and time, (b) A regressive type of SM model, and (c) A non-regressive type of SM model.

The SM model is useful to simplify complicated spatiotemporal trajectories. Although this model can be applied to any trajectory data, one needs to be cautious in applying the SM model to noisy data. Datasets containing no descriptions (e.g., home, work, stop-over) for each location, e.g., GPS data, need to be preprocessed. A set of location points that constitutes a daily trajectory needs to be discretized into move(s) and stay(s) according to an analyst's definition of moves and stays. If the definitions allow capturing stays and moves at a micro scale of space and time, the number of segments of a SM model may increase to the excessive degree, which is not ideal to summarize trajectory data concisely.

3 Stay-Move Tree

This study proposes a *Stay-Move (SM) tree* to organize all types of trajectories in an SM model by a tree structure. As seen in Fig. 2, each node of a tree represents a specific trajectory type constructed by the SM model. The root node at the top of the tree represents a trajectory without any moves for a period of time (e.g., a day), and its child node represents a trajectory type with one more move, extending from its parent node (Fig. 2). Each parent node branches into its child nodes along with a possible place set that includes previously visited places or a new place. Trajectory types with the same number of moves are located at the same depth of a tree. The height of a tree is determined by the maximum number of moves of a single trajectory in the data.

Here is a more formalized description. Let a list of locations where the n-th object once stayed until the m-th move be $\mathrm{Loc}_m[\mathrm{O}_n] = \{\mathrm{location}[i] \mid i \leq m\}$, and let a sequence of ordinal numbers that substitutes elements of $\mathrm{Loc}_m[\mathrm{O}_n]$ by order of appearance of locations be $\mathrm{LocOrder}_m[\mathrm{O}_n] = \{\mathrm{order}[i] \mid i \leq m\}$. Then, the combination of elements belonging to $\mathrm{LocOrder}_m[\mathrm{O}_n]$ can represent each type of trajectory as illustrated in Fig. 2. In addition, a possible location set can be defined as $\mathrm{PossibleLoc}_m[\mathrm{O}_n] = [\mathrm{Loc}_m[\mathrm{O}_n] - \{\mathrm{location}[m]\}] \cup \{\mathrm{location}[m+1]\}$.

Due to its predefined structure embracing all types of trajectories, the SM tree enables to represent the (relative) frequency distribution of trajectory type (Fig. 2). A (relative) frequency distribution over an SM tree gives insights into the nature of movements including the level of activities and the diversity of visited places. Based on a (relative) frequency distribution, SM trees can be used to compare different sets of trajectories grouped by demographic characteristics (e.g., age group), regions, time periods (e.g., Monday vs. Saturday), or animal species.

4 Classification of Stay-Move Trees

With Stay-Move trees, it is possible to investigate how similar different groups of trajectories are by measuring the similarity between a pair of SM trees constructed from a pair of groups of trajectories. A similarity function measuring the degree of similarity between two SM trees can be defined as follows:

$$100 - \sum_{i=1}^{depth}\sum_{j=1}^{node}\alpha_{ij}\left|P_{ij}[Tree_1] - P_{ij}[Tree_2]\right|$$

$$\alpha_{ij} = \text{weights for } j\text{th node at } i\text{th depth (default} = 1)$$

$$P_{ij} = \text{the relative frequency (i.e., percentage) of } j\text{th node at } i\text{th depth}$$

Similarity functions can be defined based on either the frequency or the relative frequency (i.e., percentage), but the similarity function based on the percentage is better to compare different groups of trajectories because it is normalized by the total number of trajectories.

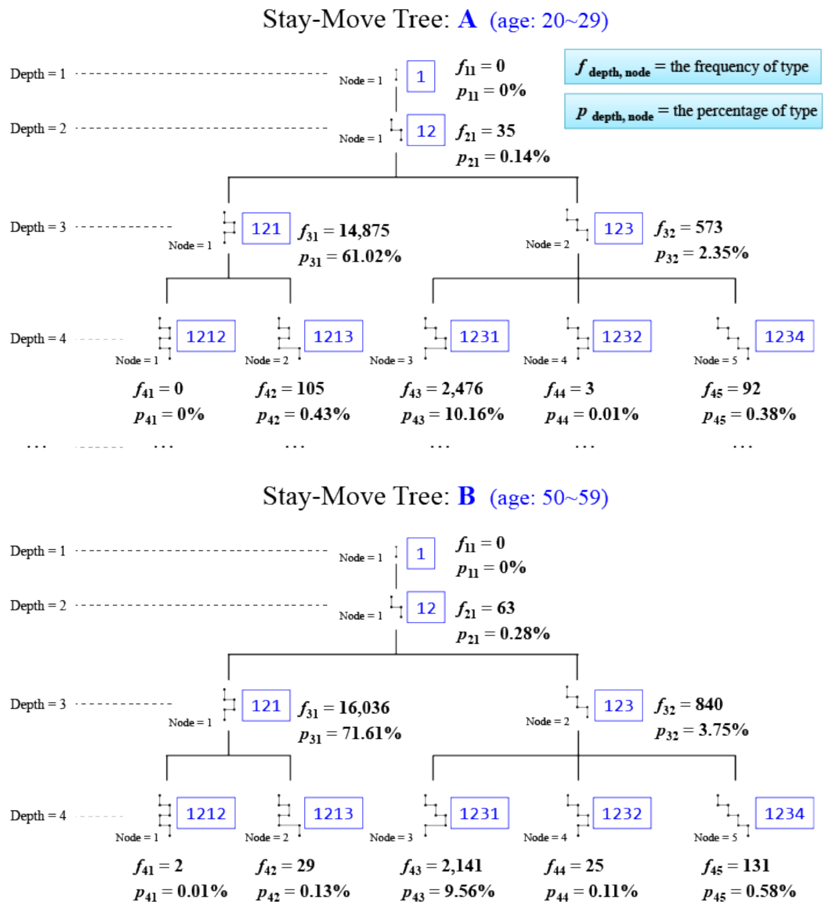

Fig. 2. The frequency distribution (f_{ij}) and the relative frequency distribution (P_{ij}) of daily trajectory type of a Stay-Move tree for two groups of people (A: age of 20-29; B: age of 50-59). Nodes beyond the depth of 4 are not represented in this figure for both trees, A and B, due to the limited space.

5 Preliminary Result

To show the utility of the proposed methodology, I conducted preliminary analysis on Korea Household Travel Survey data collected in 2006 in the Seoul Metropolitan area. The data contains about 100k daily trajectories and about 300k trips. Each trip has attributes including administrative area codes, time, and place type of departure and arrival as well as mode of transportation. To keep SM trees in a manageable size, this analysis does not include stop-over places but only destination places.

The frequency distributions and relative frequency (percentage) distributions of trajectory types of SM trees were constructed for age groups in the Seoul Metropolitan area (Fig. 2), and then, the similarity of each pair of SM trees was calculated (Table 1). The similarity matrix shown in Table 1 reveals that age groups of 10-19 and 20-29 have different patterns in relative frequency (percentage) distributions of trajectory types from the rest of age groups. A close look into the percentage of each type enabled a better understanding of those patterns. Age groups of 10-19 and 20-29 show the higher percentage of two types of trajectories: '12131' (e.g., home–school–home–extracurricular activities–home) and '1231' (e.g., home–school–extracurricular activities–home) than other age groups. While age groups of 40-69 have similar relative frequency distributions each other, the age group of 30-39 are different from them.

In the future, I plan to extend the data analysis to other big trajectory data including large GPS data.

Table 1. Similarity of relative frequency (percentage) distributions of different age groups

	Age Group						
	10~19	20~29	30~39	40~49	50~59	60~69	70~119
10~19	-	-	-	-	-	-	-
20~29	84.4%						
30~39	48.8%	56.0%					
40~49	52.3%	57.6%	89.2%				
50~59	55.7%	60.4%	82.9%	91.5%			
60~69	55.2%	59.9%	87.1%	94.6%	93.7%		
70~119	52.6%	59.1%	91.2%	93.1%	87.9%	91.8%	

Legend
100.0%
40.0%

6 References

1. Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, *19*(6), 639-668.
2. Sinha, G. & Mark, D. M. (2005). Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographical Systems*,*7*(1), 115-136.
3. Dodge, S., Weibel, R., & Lautenschütz, A.-K. 2008. "Towards a taxonomy of movement patterns". *Information Visualization* **7**: 240–252.
4. Hornsby, K. & N. Li. 2009. "Modeling spatiotemporal paths for single moving objects". In *Research Trends in Geographic Information Science*, Ed. Gerhard Navratil: 151-167. New York: Springer, Dordrecht, Berlin 2009.
5. Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, *10*(84), 20130246.
6. Inoue, R., & Tsukahara, M. (2016). Travel pattern analysis from trajectories based on hierarchical classification of stays. In *International Conference on GIScience Short Paper Proceedings* (Vol. 1, No. 1).
7. Kim, E.-K. (2009). *3D sequential object clustering for spatio-temporal data analysis*. (master's thesis, Kyung Hee University, 2009).