Mining Top-K Relevant Stay Regions
from Historical Trajectories

Yung-Hsiang Lin[1], Chien-Hsiang Lai[1], and Po-Ruey Lei[2(✉)]

[1] Department of Computer Science, National Chiao Tung University,
Hsinchu, Taiwan, ROC
{shiang1095,likekaito}@gmail.com
[2] Department of Electrical Engineering,
ROC Naval Academy, Kaohsiung, Taiwan, ROC
cnabarry@gmail.com

Abstract. With increasingly prevalent mobile positioning devices, such as GPS loggers, smart phones, and GPS navigation devices, a huge amount of trajectories data is collected. Users are able to obtain the various location-based services by uploading their trajectories. In this paper, we address that a user's movement behavior is able to discover by their similar shape trajectories and resulted in some regions frequently stay in common, called relevant stay regions. Once a set of stay regions discovered, we can predict the next region where the user intends to go and provide location-based information of the next stay in advance, such as traffic status, targeted advertises, sightseeing recommendations, and so on. Prior works have elaborated on discovering stay region from the whole crowd trajectories and then exploring the relations between the regions to describe the movement patterns for location prediction. However, the trajectories pass the same region may not have the similar movement behavior. Thus, we propose a framework to discover stay regions relevant to the specific movement behavior and then applied in location prediction, called Region Modeling and Mobility Prediction. The proposed framework includes two modules: region modeling and mobility prediction. In the region modeling module, we develop shape clustering method to group the similar trajectories from historical data and then explore the stay region model from trajectory clusters. Based on the discovered region model, the mobility prediction module provide a cluster selection algorithm and several prediction strategies to generate the top-k relevant stay regions. Experiments results on real datasets demonstrate the effectiveness and accuracy of our proposed model on detecting next stay region, comparing with other baseline methods.

1 Introduction

While the mobile positioning devices become prevalent, a tremendous amount of trajectory data is generated. Users can upload their trajectories or check-in data to location-aware web provider(e.g., @trip, Foursquare and Facebook) and obtain the various location-based services such as tourism recommendations and

store advertising in their daily life. In general, trajectory data is a sequence of GPS points and these sequential GPS points record the users' true movement. If some trajectories frequently and repeatedly appear in a user's historical trajectory data, we are able to suppose that the user may have a certain movement behavior. However, due to the uncertainty of GPS position collected, it is difficult to discover that the trajectories are completely repeated. For example, in Fig. 1, there are three historical trajectories of one user in different days with different colors. These trajectories are not completely the same but pass through some regions in common. More specifically, those trajectories have similar shape and all of them have stayed in some specific regions together(Region R1, R2, and R3). This observation shows that the user usually takes similar routes if they have visited the similar set of stay regions. Thus, we claim that a user's movement behavior can be discovered by clustering the historical trajectories with similar shape and then resulted in some stay regions where the user frequently visited. The sets of stay regions which discovered from the similar trajectories are relevant to a user's movement behavior and called relevant stay regions. Once a set of stay regions discovered, lots of location-aware information and applications could be provided to user, such as inferring regions for sightseeing and tourism recommendations, distributing coupons of stores near by stay location, estimating the traffic status on the way to destination, and even predicting the possible next stay for navigation system to set the destination automatically. In this paper, we focus on the problem of mining stay regions from historical trajectories and applied in location prediction problem.

Fig. 1. An example of the similar-shape historical trajectories and passed some stay regions in common.

Given a set of trajectories, prior works have studied the location prediction problem in which given the users current location, the problem is to predict the next location or the location at a specific time. In [7,10], the authors first map the user's historical trajectories into the regions by grid-cell system and estimate

the transition probability between the mapped regions. Then, according to the user's current location to predict the future movement based on the probability model. The authors in [5,9] construct a decision tree according to trajectories and use tree structure to denote the stay locations for prediction. These existing research works focus on discovering stay region from the whole crowd trajectories and then exploring the relations between the regions to describe the movement patterns for location prediction. However, the trajectories stay in the same region may not have the similar movement behavior. For instance, a user usually stays a restaurant for lunch in the daytime and then go back to work. In the night-time, the user sometimes exercises at the gym in the same region and then go back home. Both trajectories stay in the same region but those are two different movement behaviors. In this paper, we first discover the user's movement behavior by trajectory clustering and then explore the stay regions from each group of trajectories. Based on those sets of stay regions from the similar movement behavior, the next stay region prediction is able to be improved. The advantage of predicting the future location by the stay regions generated from the similar movement behavior is that the regions where the user frequently stay are the same even if the user detours his/her route for traffic or other instance.

In this paper, we propose a framework to discover stay regions relevant to the specific movement behavior and then applied in location prediction, called Region Modeling and Mobility Prediction. Specifically, the framework includes two modules: region modeling module and mobility prediction module. In region modeling module, given a use's individual trajectory data, we develop the shape clustering method to discover the user's movement behavior by grouping the trajectories with the similar movement shape. Based on the trajectory clusters, a stay regions model is discovered for each trajectory cluster. In mobility prediction, given the starting location and current time, top-k stay regions relevant to the user's movement behavior will be provided. We design an algorithm to select trajectory cluster, i.e. the specific movement behavior, by considering the user's starting location and current time. Then, several prediction strategies are proposed to generate top-k stay regions those relevant to the user's specific movement behavior. We evaluate the performance of our proposed framework by real-world dataset generated by mobile users in Taiwan. In addition, we also compare with the other existing approaches. The extensive experiments results demonstrate the effectiveness of our framework and the accuracy of next stay region prediction ia able to be improved.

The contributions of this paper are summarized as follows:

1. We address that a user's movement behavior can be discovered by clustering the historical trajectories with similar shape and then resulted in some regions where the user frequently stayed. Such a set of stay regions is relevant to a specific movement behavior, called relevant stay region.
2. We propose a framework, called Region Modeling and Mobility Prediction, to discover movement behavior by proposed trajectory shape clustering and predict top-k relevant stay regions where a user intends to go.

3. Extensive experiments are conducted on real datasets to evaluate our proposed framework. The results demonstrate that our framework is more effective and accuracy than existing works.

The rest of this paper is organized as follow: Sect. 2 introduces the current research works of the location and destination prediction. Section 3 states the problem and gives an overview of our proposed framework. Section 4 presents the shape clustering and region modeling on historical trajectory. Section 5 describes mobility prediction module for top-k stay regions generation. Performance studies are presented in Sect. 6. Finally, Sect. 7 concludes this paper.

2   Related Work

There are many research works discussing the problem of location prediction, they only focus on discovering the frequent region from the whole crowd trajectories and then exploring the relations between the regions to describe the movement patterns for location prediction. However, the trajectories stay the same region may not have the similar movement behavior. A stay region is supposed to be relevant to a movement behavior, in other words, the stay regions should be generated from similar trajectories. The route to next stay region may be detoured due to traffic jams and other instances. Furthermore, the next region where the user frequently move to may not be changed according to the user's movement behavior. Therefore, those relevant stay regions can be adopted to improve the accuracy of location prediction. In this section, we first discuss some research works on the location prediction and then introduce the research works related to mining stay regions.

A number of location prediction techniques have been proposed in data mining literature. Markov model has been widely applied in predicting destinations for a specific individual as well [1,10]. In [10], the author uses a Markov model to offline prepare the probabilities needed to efficiently compute the posterior probability for any given query trajectory online. Some existing works use the external information to predict destination [6,12], these external information such as the distributions of different districts (ground cover), of traveling time, of trajectories length, the accident reports, road condition, and driving habits often enhance the prediction accuracy. Even context information such as time-of-day, day-of-week, and velocity has been incorporated as the features in training the Bayesian network model for prediction [4]. Chen et al. [2] used a tree structure to represent the historical movement patterns and then matched the current partial trajectory by stepping down the tree. Trajectory pattern [5,9] are first explored and apply on location prediction. There are some existing works focus on extracting the stay regions for semantics mining. The authors in [11] proposed the concept of stay point detection to discover the stay regions. Considering the both spatial and temporal information, the authors in [8] propose a sequential clustering method to extract the stay regions.

3   Framework Overview

A user's movement behavior can be discovered by clustering the historical trajectories and then resulted in some regions where the user frequently stayed. Such a set of stay regions relevant to the movement behavior, called *relevant stay region*. We claim that the users' stay regions will be relevant to their movement behavior. Based on the discovered sets of relevant stay regions, the prediction on next stay region can be improved. Thus, we propose a framework to discover movement behavior and generate relevant stay regions and then applied in location prediction, called Region Modeling and Mobility Prediction.

Figure 2 shows the proposed framework which is comprised of two components: Region Modeling Module and Mobility Prediction Module. Since there are two subtasks in the region modeling module, exploring movement behavior and detecting the relevant stay regions. We first propose a trajectory clustering method called shape-clustering to explore the movement behavior by grouping the similar-shape trajectories. Based on the trajectory clusters, the sets of relevant stay regions are discovered. Specifically, a set of stay regions which is relevant to specific movement behavior is discovered from each trajectory cluster. For detecting the relevant stay regions of each trajectory cluster, we detect stay points from each trajectory and then adopt a Share Nearest Neighbor clustering(SNN-clustering) to cluster the stay points as stay regions. Finally, this module generates a relevant region prediction model for next stay region prediction. In mobility prediction module, we design an algorithm to select trajectory cluster by considering the user's starting location and current time, i.e. determining the specific movement behavior by the user's current movement. Then, several prediction strategies are proposed to generate top-k stay regions those relevant to the user's specific movement behavior.
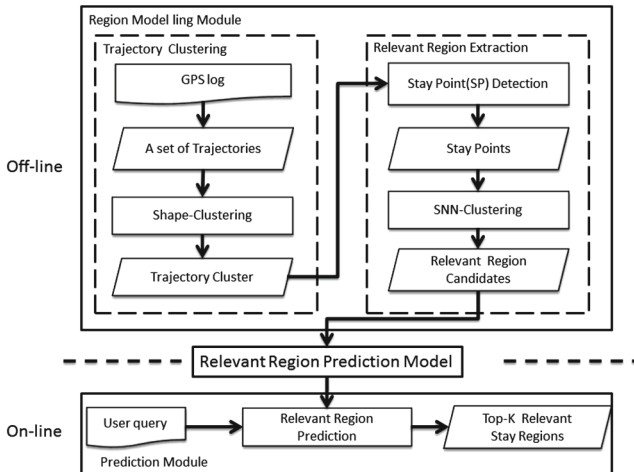

Fig. 2. Framework

4   Movement Behavior Discovery and Region Modeling

In this section, we will discuss how to discover the relevant region model for next region prediction. This includes two subproblems, namely, discovering movement behavior from trajectories and mining relevant stay region by these behaviors. To achieve the goal, we first propose Shape-Clustering to explore the group of the similar-shape trajectories. Then, based on the discovered trajectory clusters, we develop a method to extract the relevant stay region from these trajectory clusters.

4.1   Shape-Clustering

The first step in our framework is to find the movement behavior by clustering the trajectories with the similar movement. More specifically, we attempt to cluster the similar-shape trajectories so as to be able to represent the user's movement behavior. In existing work, a shape-based pattern detection method has been used to detect streaming time series data [3]. We adopt the concept on trajectory data to find out the similar-shape trajectory. The similar-shape trajectories imply that these trajectories have the similar movement behavior. As shown in Fig. 3, there are two trajectories T_i and T_j denoted as sequential GPS points. Then, we define $segment_{i,n}$ as a nth trajectory segment in T_i in the defined time interval. To avoid the problem of trajectory segmentation [8], each segment has a overlap to smooth the segmentation. We calculate the average position(the two red points) in each segment according to latitude and longitude of GPS points in a segment. The distance d is calculated and defined as distance measure between the two average position(the two red points).
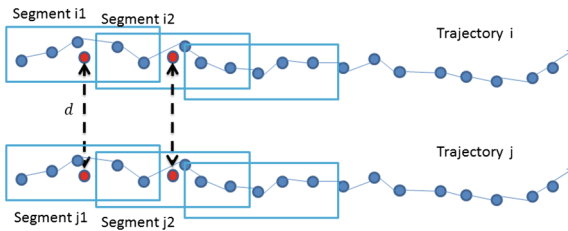

Fig. 3. An example of shape clustering

The shape clustering algorithm is shown in Fig. 4. If the distance d less than the distance threshold δ, these two segments are considered as similar segments. In addition, we define a similar counter mentioned as $SimSeg$ in the algorithm to count number of similar segments. Thus, the value of similar counter can be considered as similarity between two trajectories. Then, each two trajectories whose similarity is higher than similarity threshold $SimThres$ should be group into the same cluster. Otherwise, the trajectory form a new cluster independently.

Algorithm 1 Shape-Clustering Algorithm

Input: A trajectory $T_i = p_1p_2p_3...p_j...p_{len_i}$, A trajectory $T_j = p_1p_2p_3...p_j...p_{len_j}$, similarity threshold $SimThres$, distance threshold δ

Output: Trajectory clusters

$SimSeg = 0$
$TotalSeg = 0$
$TS_i \leftarrow$ timestamp of p_1 in T_i
$TS_j \leftarrow$ timestamp of p_1 in T_j
while $TS_i \leq$ timestamp of p_{len_i} and $TS_j \leq$ timestamp of p_{len_j} **do**
 $Avgpoint_i \leftarrow$ average of points in segments$[TS_i, TS_i + interval]$
 $Avgpoint_j \leftarrow$ average of points in segments$[TS_i, TS_i + interval]$
 $TotalSeg \leftarrow TotalSeg + 1$
 if distance between $Avgpoint_i$ and $Avgpoint_j \leq \delta$ **then**
  $SimSeg \leftarrow SimSeg + 1$
 end if
 $TS_i \leftarrow$ **SegmentSmoothing**(TS_i)
 $TS_j \leftarrow$ **SegmentSmoothing**(TS_i)
end while
if $SimSeg / TotalSeg \geq SimThres$ **then**
 put T_i and T_j into same trajectory cluster
end if

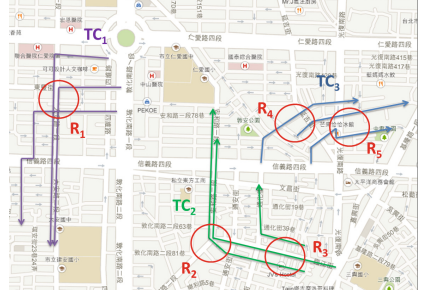**Fig. 4.** Shape clustering algorithm

Fig. 5. An example of region modeling

4.2 Relevant Region Extraction

According to the result of shape-clustering, the second step is to detect possible relevant stay regions for each trajectory cluster. We adopt the conventional approach to extract stay region from trajectories [8,11]. We first detect the stay points from each trajectories. A stay point is detected when the consecutive points of a examined point do not exceed the predefined distance threshold during the specified period of time threshold. Then, the clustering method is applied to group the stay points those are close enough. The cluster of stay points is able to represent a region where the user frequently stays. A stay region is a summary of a set of similar stay points from different trajectories. To define the similarity between stay points and discover the stay regions, we adopt SNN (shared nearest neighbor)density-based clustering. When applying SNN density based clustering to discover stay regions, we constrain the searching range of nearest neighbors is a radius D_h around the examined nodes. We define a stay point is in a stay region if each stay point of which contains at least $MinSR$ number of neighbors in the distance radius D_h. The points without $MinSR$ nearest neighbors are viewed as non-stay points and discarded. All the connected components in the resulting graph are clusters finally. These clusters can be considered as stay region candidates where an object often stay for certain activities. After generating the relevant stay region candidates, to avoid the region formed by traffic jam, we check whether the region is on the road exactly or not. If stay regions are located on road, the stay regions are removed from candidates since the regions may be formed by traffic jam. For an example in Fig. 5, the result of this module are three trajectory clusters(TC_1, TC_2 and TC_3) and the trajectories in the same cluster have similar shape with each other. In each cluster, there are some relevant stay region candidates($R_1, R_2, ..., R_5$) generated by stay points detection and SNN-clustering.

5   Mobility Prediction

After extracting the relevant stay regions from trajectory clusters, a relevant stay region model for mobility prediction is constructed. Given a user's current location and time, the task of mobility prediction is to select the best trajectory cluster and determines the score of relevant stay region candidates respectively. Thus, the two-stage prediction module is proposed, including trajectory cluster selection and prediction strategy. Finally, according to the user's current location and time, the top-k relevant stay regions will be generated for next stay region prediction.

Due to a trajectory cluster is summarized from the similar-shape trajectories, we address that the trajectory cluster is able to imply a user's movement behavior. The next stay region prediction can be improved by estimate the user's next stay region on his/her movement behavior. By considering a user's current location and time, we propose a method to select the best trajectory cluster for prediction. Generally, movement behavior is usually relevant to time or location of a stay region. Thus, the average time and location of relevant stay region candidates in the trajectory cluster are determined by the score of the trajectory cluster. The average time and location of relevant stay region candidates are defined as Eqs. 1 and 2. We formulate the scoring function of a trajectory cluster as shown in the Eq. 3. The scoring value of the trajectory cluster which is closer to zero shows that the trajectory cluster has the better matching according to the user's current behavior. The trajectory cluster is selected as the candidate for generating the relevant stay regions.

$$TC_i.time = \frac{\sum_{j=1}^{n} SR_j.time}{n}, \forall SR_j \ in \ Traj \in TC_i \tag{1}$$

$$TC_i.loc = \frac{\sum_{j=1}^{n} SR_j.loc}{n}, \forall SR_j \ in \ Traj \in TC_i \tag{2}$$

$$score_{TC_i} = \frac{\frac{|TC_i.time-q.time|}{MAX(|TC.time-q.time|)} + \frac{d(TC_i.loc,q.loc)}{MAX(d(TC.loc,q.loc))}}{2} \tag{3}$$

Based on selected trajectory cluster, we evaluate each relevant stay region candidate which belong to the selected cluster and generate the top-k relevant stay region by their evaluated score. For scoring the potential of a relevant stay region where the user may move to, we design several prediction strategies by considering the user's current location and time. Three prediction strategies are proposed: (1) *Near Time First*, (2) *Near Location First* and (3) *High Frequency First*. *Near Time First*(Eq. 4) means that if the historical stay time on the stay region is closer to current time, the score is higher.

$$score_{NT,i} = \frac{1}{log_2(\Delta Time + 2)} \tag{4}$$

In the same manner, *Near Location First*(Eq. 5) denotes that the distance between stay region and current location, the score is higher.

$$score_{NL,i} = \frac{1}{log_2(\Delta Distance + 2)} \tag{5}$$

The last one is $High\ Frequency\ First$(Eq. 6), the more times the user has been stay region, the score is also higher.

$$score_{Freq,i} = \frac{Frequency_i}{Frequency_{max}} \qquad (6)$$

Moreover, we consider that different users may have good effect by using different score functions, so the weighted average of these three score functions will be adopted. As the Eq. 7, the combination of score functions can be set different α,β and γ for different user. In general, the next stay location often near the current location and the time at next stay location often near the current time, so α and β can be set larger than γ. Otherwise, if the user often stay some specific location many times so that we can use $High\ Frequency\ First$ to predict the next stay region very well, then the γ can be set larger. After this section, we evaluate the effect of these three criterions and the combination of score functions in experiment section.

$$score_{Comb,i} = \frac{\alpha \times score_{NT,i} + \beta \times score_{NL,i} + \gamma \times score_{Freq,i}}{\alpha + \beta + \gamma} \qquad (7)$$

6   Experimental Results

6.1   Dataset Description

For this study, we use a real-world trajectory datasets: trajectory data of trips from a website called @trip (http://www.a-trip.com/). @trip is a platform let users can upload their travel logs or check-in data and share these data for other users. We extracted a trajectory dataset which consists of 1,243 users, 14,039 trajectories and 13,192,283 GPS points. In order to test the effectiveness under various scenarios, the experiments are conducted on 10 selected users with different movement behaviors. Their stay regions relevant to different movement behavior are labeled as the groundtruth for evaluation.

6.2   Performance Evaluation

In this section, we first the evaluate the effectiveness of shape clustering on movement behavior discovery. Then, in order to show the improvement on prediction accuracy, we conduct the experiments compared with other existing methods.

Effectiveness of Shape Clustering. The shape-clustering method play a important role in our framework. We suppose that a user usually takes similar routes if they have visited the similar set of stay regions. Thus, a user's movement behavior can be discovered by clustering the historical trajectories with similar shape and then resulted in some stay regions where the user frequently visited. Such a set of relevant stay regions is able to apply on improving location prediction problem. In order to show the improvement on prediction accuracy, we compare the proposed framework using shape-clustering and without
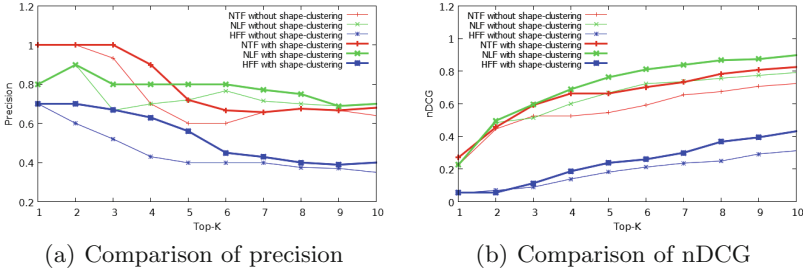
(a) Comparison of precision   (b) Comparison of nDCG

Fig. 6. Effectiveness of shape clustering

using shape-clustering by evaluating their location prediction and top-k relevant region ranking. Figure 6(a) shows the precision of next stay region prediction. The precision of location prediction is improved. In Fig. 6(b), the accuracy of top-k relevant region ranking is evaluated by nDCG. The thick line denotes the framework using shape-clustering and thin line denotes the framework without using shape-clustering. The sorting result of proposed framework with shape-clustering has higher accuracy. We can observe the precision and nDCG are improved by using shape-clustering obviously. The experimental results shows the shape clustering is able to discover the movement behavior effectively and generate the top-k relevant stay regions precisely.

Evaluation of Next Region Prediction. We next test the accuracy of the next region prediction. Given a user's current location and time, the next region prediction is to forecast the next region where the user may possibly stay. We evaluate the prediction accuracy by the proposed framework with various scoring functions, called High Frequency First(HFF), Near Location First(NLF), Near Time First(NTF), and Combination(Combine). Additionally, we compare the performance with the existing methods [9,10]. In [10], the authors predict the destination by computing the posterior probability for any given query sub-trajectory(Sub-Trajectory Synthesis). In [9], they explore the trajectory patterns and construct the decision tree to predict the next stay location(Trajectory Pattern). To compare with them, we use average distance error and precision as measurements for the performance of next region prediction. In Fig. 7(a) and (b), although HFF can not get better effect, NLF and Combine methods have better precision and nDCG value than comparison targets.

Figure 8(a) shows the prediction results represented by average distance error. The experimental result shows that both Combine method and NLF have lower distance error than Sub-Trajectory Synthesis method and prediction by using Trajectory Pattern. We also evaluate the prediction of next location by precision. In Fig. 8(b), we can figure out the propose Combine method has higher precision even the predicting region is getting distance. Furthermore, in Fig. 8(b), the proposed methods NLF and Combine have higher precision than others.
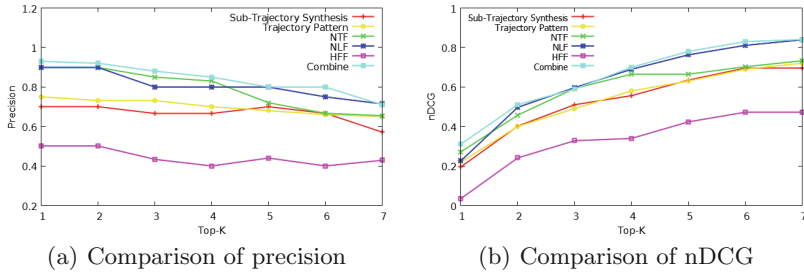
(a) Comparison of precision

(b) Comparison of nDCG

Fig. 7. Prediction accuracy comparison

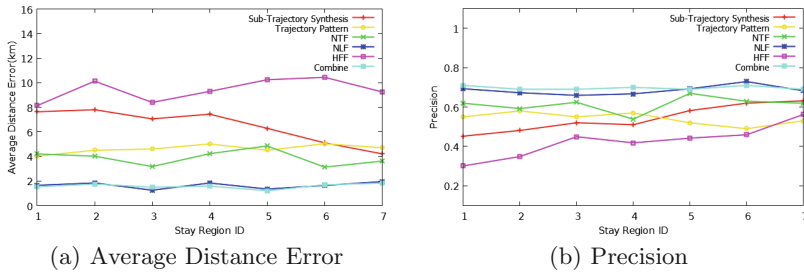

(a) Average Distance Error

(b) Precision

Fig. 8. Average distance error and precision for next stay region prediction

7 Conclusion

In this paper, we address that a user's movement behavior can be discovered by clustering the historical trajectories with similar shape and then resulted in relevant stay regions where the user frequently visited. We propose a framework, Region Modeling and Mobility Prediction, to generate top-k relevant stay regions from movement behavior and apply on location prediction problem. In region modeling module, the proposed shape clustering method explores the user's movement behavior by grouping the trajectories with the similar movement shape. Based on the trajectory clusters, a stay regions model is discovered for each trajectory cluster. In mobility prediction, given the starting location and current time, top-k stay regions relevant to the user's movement behavior will be provided for predicting next region. Experiments based on real datasets have shown that proposed framework is able to explore the movement behavior effectively and generate top-k relevant stay regions accurately. Furthermore, the ability to predict next region has advanced than the prediction methods in the literature.

References

1. Alvarez-Garcia, J., Ortega, J., Gonzalez-Abril, L., Velasco, F.: Trip destination prediction based on past gps log using a hidden markov model. Expert Syst. Appl. **37**(12), 8166–8171 (2010)
2. Chen, L., Lv, M., Chen, G.: A system for destination and future route prediction based on trajectory mining. Pervasive Mob. Comput. **6**(6), 657–676 (2010)
3. Chen, Y., Nascimento, M.A., Ooi, B.C., Tung, A.K.: Spade: on shape-based pattern detection in streaming time series. In: Proceedings of the IEEE 23rd International Conference on Data Engineering, pp. 786–795 (2007)
4. Gogate, V., Dechter, R., Bidyuk, B., Rindt, C., Marca, J.: Modeling transportation routines using hybrid dynamic mixed networks, pp. 116–131 (2005)
5. Jeung, H., Liu, Q., Shen, H.T., Zhou, X.: A hybrid prediction model for moving objects. In: Proceedings of the IEEE 24th International Conference on Data Engineering, pp. 70–79 (2008)
6. Krumm, J., Horvitz, E.: Predestination: Where do you want to go today? Computer **40**(4), 105–107 (2007)
7. Lei, P.R., Shen, T.J., Peng, W.C., Su, I.J.: Exploring spatial-temporal trajectory model for location prediction. In: Proceedings of the 12th IEEE International Conference on Mobile Data Management, pp. 58–67 (2011)
8. Lu, C.-T., Lei, P.-R., Peng, W.-C., Su, I.-J.: A framework of mining semantic regions from trajectories. In: Yu, J.X., Kim, M.H., Unland, R. (eds.) DASFAA 2011, Part I. LNCS, vol. 6587, pp. 193–207. Springer, Heidelberg (2011)
9. Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F.: Wherenext: a location predictor on trajectory pattern mining. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 637–646 (2009)
10. Xue, A.Y., Zhang, R., Zheng, Y., Xie, X., Huang, J., Xu, Z.: Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In; Proceedings of the 29th IEEE International Conference on Data Engineering, pp. 254–265 (2013)
11. Zheng, V.W., Zheng, Y., Xie, X., Yang, Q.: Collaborative location and activity recommendations with gps history data. In: Proceedings of the 19th International Conference on World Wide Web, pp. 1029–1038 (2010)
12. Ziebart, B.D., Maas, A.L., Dey, A.K., Bagnell, J.A.: Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 322–331 (2008)