

# Mining Individual Behavior Pattern Based on Semantic Knowledge Discovery of Trajectory

Min Ren<sup>1,2,3</sup>, Feng Yang<sup>\*2</sup>, Guangchun Zhou<sup>2</sup> and Haiping Wang<sup>2</sup>

<sup>1</sup>Shandong Normal University, China

<sup>2</sup>Shandong University of Finance and Economics, China

<sup>3</sup>Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology, China

This paper attempts to mine the hidden individual behavior pattern from the raw users' trajectory data. Based on DBSCAN, a novel spatio-temporal data clustering algorithm named Speed-based Clustering Algorithm was put forward to find slow-speed subtrajectories (i.e., stops) of the single trajectory that the user stopped for a longer time. The algorithm used maximal speed and minimal stopping time to compute the stops and introduced the quantile function to estimate the value of the parameter, which showed more effectively and accurately than DBSCAN and certain improved DBSCAN algorithms in the experimental results. In addition, after the stops are connected with POIs that have the characteristic of an information presentation, the paper designed a POI-Behavior Mapping Table to analyze the user's activities according to the stopping time and visiting frequency, on the basis of which the user's daily regular behavior pattern can be mined from the history trajectories. In the end, LBS operators are able to provide intelligent and personalized services so as to achieve precise marketing in terms of the characteristics of the individual behavior.

*Keywords:* spatiotemporal data, semantic trajectory, clustering, knowledge discovery

## 1. Introduction

Due to the rapid development and wide application of such technologies as GPS, GSM, wireless sensors, etc., people can easily collect a large amount of users' real-time location data, which keep a record of their latitude and longitude coordinates at a certain time. And all those locations of the user, sorted in accordance with the time, constitute the user's behavior trajectory. Trajectory data, the real records of a user's

offline activities, often imply rich information of the spatial structure and his behavior, which can reflect his personal intentions, interests and hobbies and behavior patterns. This constitutes the research foundation of Location-based Service (LBS). Therefore, it has gradually become a hot research topic to mine the valuable information contained in individual trajectories.

Human history trajectories recorded the movement processes of their activities that tended to have a certain purpose and often showed a high degree of temporal and spatial regularity [1]. Deep research on trajectory data can contribute to not only analyzing users' activity patterns in a certain area [2], moving patterns [3,4] (such as walking, cycling, driving), etc., but also to determining daily life habits and behavior patterns of the individual and common features of the user groups [5,6], and even to predicting the behavior of the moving object in the following period of time [7,8] and understanding social relations of user groups [9,10], which has great significance for personalized recommendations [11,12] and precision marketing.

Knowledge discovery of users' activities, however, is highly challenging, due to the difficulty in directly finding users' behavior from their history trajectories. This is so mainly because the trajectory is represented by spatial coordinates with timestamps, the latitude and longitude coordinates in a particular reference coordinate system. The specific application usually does not focus on the coordinate values, but on the geographical positions that imply certain se-

\* corresponding author

semantic information and can be associated with the things related to them when mentioned. The current major relevant research analyzes the trajectory coordinates sequence by means of traditional data mining technology. But the spatial location data lack the wealth of useful information found in common databases, such as geographical background knowledge, application semantic knowledge, etc. Thus, the existing research cannot effectively extract implicit knowledge of the trajectory because of the lack of information.

The authors' contributions are as follows. Firstly, a Speed-based Clustering Algorithm was designed to mine a user's stops in a single trajectory accurately, which was not affected by noises and acnodes. Secondly, considering the user's characteristics and his movement environment, a POI-Behavior Mapping Table was created with the help of knowledge database (POI database) and on the basis of statistics of mass questionnaires to extract knowledge from the large-scale trajectory data in order to mine a more clear, accurate and lucid individual behavior pattern. At present, the research that combines knowledge discovery of trajectory and data mining of moving objects is quite limited.

In addition, related works are discussed in Section 2, and the following section defines relevant terms of the research problem. Then, the authors propose a Speed-based Clustering Algorithm and design a POI-Behavior Mapping Table in Section 4. In order to validate the research methods, Section 5 conducts some experiments by using data set Geolife13 and analyzes results in detail. Finally, the authors conclude the research work and discuss the future work in Section 6.

## 2. Related Works

### 2.1. Trajectory Data Clustering

The basis of research on trajectory is to determine users' significant areas from the trajectory data by using the clustering algorithm, which has drawn more and more attention. Ashbrook and Starner [7] used K-means clustering algorithm to find important places in trajectories. Its key limitation is the number of clusters that must be given a priori, but this number is often unknown in the problem. At the same time, the

noise and acnode have great influence on the clustering results. The density-based clustering algorithm was used more commonly, compared with others. DBSCAN [14], a representative and popular density-based algorithm, was widely used due to its outstanding advantage that it can discover clusters of arbitrary shape, disregarding the noise. Karli et al. [15] eliminated high speed movement data, and then used DBSCAN algorithm to extract important places from trajectories with slow speed only. But DBSCAN algorithm is very sensitive to the input parameters Eps and MinPts, which represent the radius and the minimum number of points. Therefore, some researchers improved DBSCAN algorithm in many aspects. For example, Zhou et al. [16] put forward a density and join-based clustering algorithm named DJ-Cluster for the extraction of important places from spatio-temporal trajectory data. The main idea was that if the number of a point's neighbors within distance Eps was less than MinPts, it was labeled noise. Otherwise, if any neighbor belonged to an existing cluster, it and its other neighbors joined in and if not, all points formed a new cluster. The method improves the efficiency and saves much memory. Palma et al. [17] proposed a spatio-temporal clustering algorithm based on speed to identify stops of a single trajectory and find interesting places. The method was a variation of DBSCAN, which replaced MinPts with a minimal duration MinTime, defined a new function of the Eps-linear-neighborhood of a point and used the quantile function to calculate the Eps. In addition, OPTICS, another interactive density-based clustering method, was also applied for the cluster of trajectory data. For example, Max [18] designed a time-based OPTICS algorithm with the density on the basis of the spatial and temporal properties of a trajectory to discover stops in it. In addition, many scholars designed other algorithms to analyze the trajectories. For example, Zhao et al. [19] extracted interesting places of the moving objects' trajectories using Minimum Bounding Boxes (MBB), thus the original trajectories can be expressed and then effectively clustered by using a hierarchical clustering algorithm. Kirmse et al. [20] determined users' daily stop locations with the aid of meanshift clustering algorithm. Hu et al. [21] proposed an incremental DPMM-based clustering algorithm for trajectories, which could determine an appropriate number of clusters automatically.

Spatial and temporal trajectory data are massive, multi-dimensional, noisy and incomplete.

Therefore, it is essential to design a clustering algorithm with high performance and good scalability. Certain location devices like GPS are unable to capture moving objects in a sealed space (such as elevators, buildings) which leads to the loss of signals and a blank in the trajectory [15, 22]. For this reason, objects' activities were seemingly not reflected in trajectory effect. Moreover, these activities were often significant for understanding of the characteristics of users' behavior, but they were overlooked in the above-mentioned studies. For users' stops, this paper took into account both dense and disappeared sample points in a trajectory, and at the same time emphasized the integrity of a trajectory when analyzing it, which avoided the interference of the noise data.

## 2.2. User's Behavior Mining

Individual behavior pattern mining is important research into knowledge discovery of trajectory. Currently the related research is few and dispersed. Liao et al. [2] built an Activity Model, based on the Relational Markov Network, to label the user's activities such as staying at home, working, shopping, dining out, visiting, and so on. Hu et al. [23] used Web Ontology Language (OWL) to understand buddies' relationships, locations and activity for automatic context-aware presence management. Yang et al. [22] proposed mining individual life patterns and routines from individual trajectory information and developed a mining framework of life patterns. However, this study only offered the idea and process of trajectory data mining, but did not give the specific and detailed mining method. Kirmse et al. [20] explained the semantic of daily occupied areas of users, combining with the characteristics of time. For example, users stopped longest in their working place during the day while at home in the evening. But they did not associate the occupied area with a significant place. Ying et al. [24] made full use of space information to assist the interpretations of their research results. At first, users' daily occupied places, obtained by clustering, were related to their neighboring points of interest; e.g., the park, school, bank, hotel, and so on, and then sorted according to users' trajectories like "home-school-company-home" to realize the semantic interpretation. Li et al. [25] statistically analyzed hot spots in which the user stayed in accordance with the period to find a user's repetitive periodic behavior. But this

analysis could only deal with a single periodic model of the behavior and failed to solve multiple periodic models and group periodic model, etc.

To sum up, there are few systematic guidance and research schemes for knowledge discovery in trajectory data sets of moving objects. This paper established a POI-Behavior Mapping Table through questionnaires, based on which users' behavior patterns in POIs were analyzed and semantics of the trajectory were interpreted more clearly and accurately.

## 3. Problem Definitions

Trajectory data mining can help us to obtain effective, potentially useful, and understandable information from abundant spatio-temporal data. In this section, some definitions of the relevant research are presented in detail.

**Definition 1. Trajectory:** Trajectory is a discrete sequence based on the time, made up of a series of spatio-temporal sample points; i.e.,  $T = \{p_1, p_2, \dots, p_n\}$ , where  $p_i$  ( $i = 1, \dots, n$ ) is a real-time coordinate point recorded by the GPS device after each sampling time, usually represented by a triple  $(lngt, lat, t)$ , which means the location coordinates ( $lngt$  is the longitude and  $lat$  is the latitude) of a person at time  $t$ .

A trajectory is often obtained with the help of the GPS device. However, when a user is in a relatively sealed space such as an elevator or basement, the device tends to produce erroneous data called noise (sample points in the blue circle) because of inaccurate location, as shown in Figure 1, or it loses parts of data without capturing the mobile object, leading to a blank in a trajectory (sample points in the red oval), as shown in Figure 1. This means that the sampling point disappears for a period of time  $\Delta t$  at time  $t$ , and after the limited time  $\Delta t$ , it will reappear.

Generally, users move fast at certain times when sample points are sparse in the trajectory, while at other times they move at a lower speed, showing relatively dense sample points that have more important research value than the fast moving ones. These dense areas are defined as stops.

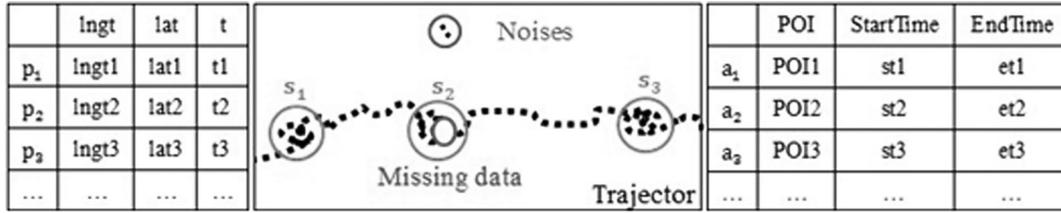


Figure 1. Research framework.

**Definition 2. Stop:** Stop is a triple  $(L, StartTime, EndTime)$ , where  $L = \{p_i, p_{i+1}, \dots, p_m\}$  is a subtrajectory of  $T$ , which is a geometry cluster of trajectory sampling points.  $StartTime = p_i.t$  is the start time of the stop and  $EndTime = p_m.t$  is the end time. Stops of specific applications are valuable only when the duration  $MinTime$  is sufficient, that is

$$StartTime - EndTime \geq MinTime$$

Stops are divided into two types. The first, denoted by Stop I, is a subtrajectory made up of dense sample points, in which the speed of any two contiguous trajectory sampling points is less than the normal walking speed and even approximately equals zero. For example,  $S_1$  and  $S_3$  in Figure 1. The second, named Stop II, is a subtrajectory with lost signals, such as  $S_2$  in Figure 1.

Stops are semantically important parts of a trajectory. A stop often means that a user stayed in a geographical area for a certain amount of time. If a user stayed there for a long period of time, it is reasonable to believe that he was very interested and might have visited or been engaged in an activity. Moreover, a stop often corresponds to a specific significant place, such as park, cinema, supermarket, etc., which is called Point of Interest (POI).

**Definition 3. POI:** POI refers to a specific place, usually denoted by  $(id, name, category, loc)$ , where  $id$  is a unique number, and  $loc$  is its location that can be described by the coordinates (longitude and latitude) of the center position; i.e.,  $loc = (longitude, latitude)$ .

Trajectory has no semantic geographical information, while the most important characteristic of POI is its informative presentation. When a user's activity associates with a specific POI, which increases the space and semantic characteristics of the location, his behavior in the area can be understood.

**Definition 4. Activity:** Activity is a stop of a user in a POI, denoted by a tuple  $(POI, StartTime, EndTime)$ . In the end, a user's trajectory is converted to the sequence of his activities in POIs. An activity sequence is a set of activities in accordance with the chronological order, represented by  $Activities = \{a_1, a_2, \dots, a_n\}$ , where  $a_i$  ( $i = 1, \dots, n$ ) is an activity in the sequence.

## 4. Research Methods

### 4.1. Speed-based Clustering Algorithm

In the real world, individual activity is often purposeful and usually indicated by stops, which helps to understand a user's behavior much better. According to Definition 2, this paper proposes a Speed-based Clustering Algorithm (SBC for short) to find out two kinds of stops in the trajectory.

#### 4.1.1. Pseudo Code

Trajectory data of mobile objects have a strong spatio-temporal correlation. Considering the continuity of the two adjacent trajectory points at a particular time, the algorithm initially started clustering based on the speed to obtain all possible candidate stops, and then merged two adjacent candidate stops according to their time interval in order to avoid dropping without satisfactory conditions due to abnormal data (namely noise) produced by the GPS equipment that located falsely.

This algorithm is determined by density-based algorithm, but integrates some additional different concepts.

**Definition 5. Neighborhood of a point:** Let  $T = \{p_1, p_2, \dots, p_i, \dots, p_n\}$  be a trajectory.  $P_i$  and  $P_{i+1}$  are neighbors, if

$$distance(p_i, p_{i+1}) / (p_{i+1}.t - p_i.t) \leq MaxSpeed \quad (1)$$

Given that the Earth is spherical, the distance between two points  $P_i$  and  $P_{i+1}$  is calculated by the following formulas:

$$distance(p_i, p_{i+1}) = 2r \sin^{-1}(\sqrt{\varphi}) \quad (2)$$

and

$$\varphi = \sin^2 \left( \frac{1}{2}(p_i.lat - p_{i+1}.lat) + \frac{1}{2} \cos(p_i.lat) \cdot \cos(p_{i+1}.lat) \sin^2(p_i.lngt - p_{i+1}.lngt) \right) \quad (3)$$

where  $r$  is the radius of the Earth.

If  $P_i$  and  $P_{i+1}$  are neighbors and  $q$  is the neighbor of  $P_{i+1}$ ,  $q$  is also the neighbor of  $P_i$ . The

neighborhood of a point  $P_i$  is the maximal set of its neighbors (the neighborhood is ordered by time). Each point in a neighborhood is called core point.

**Definition 6. Cluster:** Let  $\{p_k, p_{k+1}, \dots, p_m\}$  be points of a neighborhood. If satisfying  $p_m.t - p_k.t \geq MinTime$ , the neighborhood is a cluster with respect to  $MaxSpeed$  and  $MinTime$ .

**Definition 7. Noise:** Let  $\{p_k, p_{k+1}, \dots, p_m\}$  be a cluster. If  $P_i$  does not belong to this cluster where  $k \leq i \leq m$ , it is defined as a noise.

The pseudo code of this algorithm is shown in Figure 2.

There are four main merits of SBC for clustering trajectory data compared with DBSCAN [15] and CB-SMoT [17]:

```

Input: T // A trajectory
      MaxSpeed // Maximum speed for clustering
      MinTime // Minimum duration for clustering
      MinInterval // Minimum time interval for merging
Output: S // Set of stops
S = { } ; Auxiliary = { } ; L = { } ; n = 0 ; // Initialize
// compute the candidate stops
for i = 1 to |T| do // |T| denotes the number of points in T
  // compute the distance between two consecutive points
  d = distance ( p[i], p[i+1] ) ;
  speed = d / ( p[i+1].t - p[i].t ) ;
  if ( speed < MaxSpeed ) then // record the points into L
    if ( L == { } ) then
      L U= { p[i], p[i+1] } ;
      StartTime = p[i].t ;
    else
      L U= { p[i+1] } ;
      EndTime = p[i+1].t ;
    end if
  else if ( L != { } ) then // record a candidate stop into Auxiliary
    n++;
    Auxiliary U= { ( L, StartTime, EndTime ) } ;
    L = { } ;
  end if
end if
end for
// Merge the contiguous candidate stops
i = 1;
while ( i <= n ) do
  Stop = Auxiliary[i] ;
  while ( Auxiliary[i+1].StartTime - Stop.EndTime < MinInterval ) do
    Stop.L U= Auxiliary[i+1].L ;
    Stop.EndTime = Auxiliary[i+1].EndTime ;
    i++ ;
  end while
  if ( Stop.EndTime - Stop.StartTime > MinTime ) then
    S U= { Stop } ;
  end if;
  i++ ;
end while

```

Figure 2. Pseudo code.

- (i) The speed *MaxSpeed* between two consecutive sampling points is used for clustering instead of the density of sampling points, mainly because many sampling points in a trajectory are lost easily in some area where sampling points are relatively sparse while the speed is still slow due to the location equipment and the activity types of moving objects.
- (ii) The minimal time *MinTime* replaces a minimal amount of points *MinPts* in a neighborhood, which can avoid problems such as the absence of points because of some equipment failure and time differences in the sample rate.
- (iii) The algorithm uses quantile function to calculate the relative parameter *MaxSpeed* related to the mean and standard deviations instead of an absolute, fixed value given manually.
- (iv) Two neighboring clusters are merged according to the given threshold *MinInterval*, which can reduce the loss of some stops that are split into many clusters due to noises of error location, but the durations of some clusters are less than *MinTime*.

#### 4.1.2. Parameters Estimation

This algorithm includes four parameters that are the trajectory data  $T$ , maximum speed threshold *MaxSpeed*, the minimum stopping time threshold *MinTime*, and the minimum time interval threshold *MinInterval*, respectively.

*MaxSpeed* represents the maximal speed between two consecutive points, which is used to identify the lower speed parts of a trajectory, corresponding to POI. If a moving object uses different vehicles in a trajectory, its speed obviously changes. For example, when a user is on a motor vehicle, his average speed might reach 30-50 m/s in urban areas, but it might fall to 0-5 m/s in traffic jams. And when he walks the normal walking speed is usually 2-3 m/s. Therefore, if the *MinSpeed* value is set to be 3 m/s, traffic jam can become a stop by mistake. We calculate this parameter according to [17]. For a trajectory  $T$ , we calculate the speed between two consecutive sample points, and then we determine the mean and the standard deviation of these speeds. Through the two parameters a

Gauss curve can be drawn accordingly. On the basis of the curve, the value of *MaxSpeed* is able to be determined by using the quantile function:  $[0, 1] \rightarrow R$ , an inverse cumulative distribution function. The quantile function is defined as:

$$F^{-1}(p, \mu, \sigma) = \mu + \sigma\sqrt{2}f^{-1}(2p - 1) \quad (4)$$

where

$$f^{-1}(x) = \sum_{k=0}^{\infty} \frac{c_k}{2k+1} \left( \frac{\sqrt{\pi}}{2} x \right)^{2k+1} \quad (5)$$

$$c_k = \sum_{m=0}^{k-1} \frac{c_m c_{k-1-m}}{(m+1)(2m+1)} \quad (6)$$

and  $c_0 = 1$ .

According to the formulas above, the *MaxSpeed* value is lower than the mean, usually in the interval of  $[0, \mu - \sigma]$ .

*MinTime* is not an absolute time setting. But in general, if an object stayed in an area for more than 30 minutes [9, 17], it can be inferred that there are related activities of the object in the area.

*MinInterval* can be set according to the sampling interval.

#### 4.2. POI-Behavior Mapping Table

The research on users' behavior focused on the semantic that represented the matching of trajectories and POIs and on a certain amount of stopping time. Users' behavior is closely related to the classification of POIs that are based on map data of LBS operators and quite different from each other. In the previous research [26], we used a two-stage classification and divided POIs into 15 categories and 150 subcategories, according to the BAIDU Map.

In order to understand individual behavior patterns and analyze the semantics of the trajectory, a POI-Behavior Mapping Table was established which defined the durations of possible behavior in each subcategory POI; i.e., the possible behavior, the minimal and maximal duration of the behavior obtained through the questionnaires and actual investigations. Parts of POI-Behavior Mapping Table are showed in Table 1.

POI Category	POI Sub-category	Behavior	Minimal Duration	Maximal Duration
Catering	Chinese Restaurant	delivery	15 minutes	45 minutes
		eating	1 hours	4 hours
		working	3 hours	10 hours
	Chinese Fast Food Restaurant	take-out	5 minutes	30 minutes
		delivery	15 minutes	45 minutes
		eating	30 minutes	1 hours
		working	3 hours	10 hours
...	...	...	...	
Leisure & Entertainment	Cinema	watching films	1 hours	3 hours
		working	4 hours	8 hours
	...	...	...	...

Table 1. Parts of POI-Behavior mapping table.

## 5. Experiments and Results

The experimental dataset chosen for calculation in this paper is GeoLife Trajectories Vision 1.3, collected by Microsoft Research Asia, including 18,670 trajectories of 182 users in 11,129 effective days, and with the total distance up to 1,292,951 kilometers. Altogether 2,255,871 trajectory sampling points of 31 users living in Beijing urban areas and the surrounding areas were compiled over the course of 325 days (with decimal coordinates of north latitude  $39.47^{\circ} \sim 40.79^{\circ}$  and east longitude  $115.94^{\circ} \sim$

$117.33^{\circ}$ ). Their trajectories are shown in Figure 3, from which we can see that users' activities are mainly concentrated in the main urban districts, with the most intensive amount of sample points being in Haidian district where most volunteers of GeoLife project live or work. The thick lines imply the routes that users walked along more often. Trajectory ID, latitude, longitude and sampling time are selected in our experiments for data analysis.

In Table 2, the clustering result of using SBC on the experimental dataset in this paper is compared with those using DBSCAN and CB-

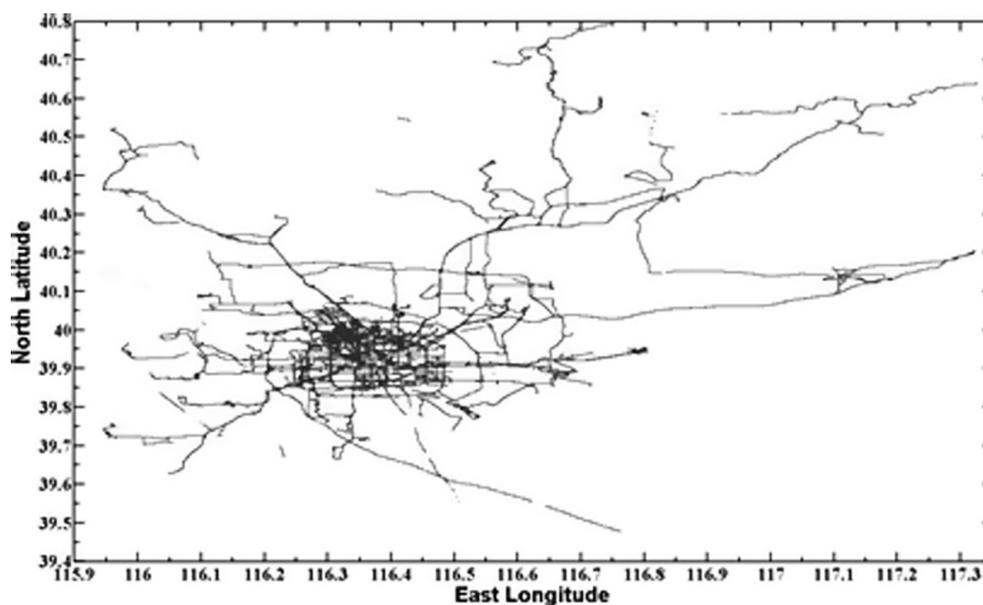


Figure 3. Trajectory distribution of experimental dataset.

SmoT, which illustrates that the clustering effect of SBC is better than the other two algorithms, for it can obtain more Stop I, and especially more Stop II that includes the absent sampling points.

Algorithm	Parameter Setting	Stop I	Stop II	Total
DBSCAN	$Eps = 200$ m $MinPts = 150$	556	0	556
CB-SMoT	$area = 400$ m $MinTime = 0.25$ h $MaxTol = 20$	724	37	761
SBC	$MinTime = 0.25$ h $MinInterval = 3$ h	731	72	803

Table 2. Clustering results.

In order to better understand users' consumption behavior, in this paper clustering results were mapped to Baidu Map and finally 428 POIs were matched. Table 3 describes the statistics of consumer behavior in some sub-categories of "food" and "leisure" POI categories, such as the number of matched POIs of different sub-categories, users' average durations in POIs for the consumption and users' total visit times.

After comprehensively investigating a user's behavior over many days, it could be eventually concluded that in the past month he had visited Wal-Mart 20% of the days. Moreover, in 90% of the visiting days, he arrived between 9:00 to 9:30 and stayed for about 1.5 hours, and in

50% of the visiting days, he took this route: 52 Zhichun Road (Home) → Yonghe King (Chinese Fast Food Restaurant) → Wal-Mart (Supermarket).

Therefore, local merchants can infer consuming places that users often visited and their consumption trends according to their behavior pattern, and at specific times, purposefully emphasize advertisings for products or services users are interested in given their location and behavior in order to improve the precision of the marketing.

## 6. Conclusion

Aiming to determine individual behavior characteristics and living habits from users' raw trajectory data, the Speed-based Clustering Algorithm, an improved DBSCAN, was designed to extract stops in a single trajectory without the influence of noises and acnodes. Next, based on the authors' preliminary research, a POI-Behavior Mapping Table was built through the questionnaires and actual investigations so as to recognize the activities in the POIs users being engaged. With the analysis of users' activities over a long period of time, daily regular behavior pattern of users can be determined, which will help merchants to enhance their understanding of the potential consumer trends, promote personalized information for various potential consumers, and improve the precision of

POI Category	POI Sub-category	Number	Average Duration (h)	Visit Times
Catering	Chinese Restaurant	39	1.8290	44
	Chinese Fast Food Restaurant	9	0.6189	19
	Western Restaurant	2	1.7181	2
	Western Fast Food Restaurant	6	0.8048	17
	Leisure Catering	4	1.0814	5
Leisure & Entertainment	Cinema	9	2.8668	10
	Tourist Resort	10	2.3657	13
	Library	3	2.2823	3
	KTV	6	3.9063	7
Shopping	Shopping Center	7	2.1583	16
	Supermarket	12	1.0396	22
Total		107	1.8792	158

Table 3. Statistic of consumer behavior in some POI sub-categories.

marketing. In their future work, the authors will conduct experiments with the trajectory dataset collected by them and verify the accuracy of the method. Furthermore, trajectories of user groups will be chosen as the research object to analyze group behavior characteristics.

## Acknowledgements

This research is supported by National Natural Science Foundation of China (NSFC) Grant No. 71172046, MOE (Ministry of Education in China) Youth Fund Project of Humanities and Social Sciences (No. 13YJC630162), Natural Science Foundation of Shandong Province (No. ZR2011GL018), Project of Shandong Province Higher Educational Science and Technology Program (No. J11LG11) and Social Sciences Youth Project of Shandong Province (No. 13DGLJ04).

## References

- [1] M. C. GONZÁLEZ, C. A. HIDALGO, A-L. BARABÁSI, Understanding Individual Human Mobility Patterns, *Nature*, 2008.453(7196): 779–782.
- [2] L. LIAO, D. J. PATTERSON, D. FOX, H. KAUTZ, Building Personal Maps from GPS Data, *Annals of the New York Academy of Sciences*, 2006.1093(12): 249–256.
- [3] F. CALABRESE, M. DIAO, G. DI LORENZO, J. FERREIRA, C. RATTI, Understanding Individual Mobility Patterns from Urban Sensing Data: A Mobile Phone Trace Example, *Transportation Research Part C*, 2013.26: 301–313.
- [4] K. ZHENG, Y. ZHENG, N. J. YUAN, S. SHANG, On Discovery of Gathering Patterns from Trajectories, *Proceedings of the 29th IEEE International Conference on Data Engineering*, 2013: 242–253.
- [5] Y. ZHENG, L. ZHANG, X. XIE, W-Y. MA, Mining Correlation between Locations Using Human Location History, *Proceedings of the 17th ACM SIGSPATIAL Conference on Advances in Geographical Information Systems*, 2009: 472–475.
- [6] Y. ZHENG, X. XIE, Learning Location Correlation from GPS Trajectories, *Proceedings of the 11th Int. Conf. Mobile Data Management*, 2010: 27–32.
- [7] D. ASHBROOK, T. STARNER, Using GPS to Learn Significant Locations and Predict Movement across Multiple Users, *Personal and Ubiquitous Computing*, 2003.7(5):275–286.
- [8] J. KRUMM, E. HORVITZ, Predestination: Inferring Destinations from Partial Trajectories, *Proceedings of the 8th ACM Conference on Ubiquitous Computing*, 2006: 243–260.
- [9] X. XIAO, Y. ZHENG, Q. LUO, X. XIE, Finding Similar Users Using Category-Based Location History, *Proceedings of the 18th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*, 2010: 442–445.
- [10] Y. ZHENG, L. ZHANG, Z. MA, X. XIE, W-Y. MA, Recommending Friends and Locations Based on Individual Location History, *ACM Transaction on the Web*, 2011.5(1): 1–44.
- [11] M. YE, P. YIN, W-C. LEE, Location Recommendation for Location-Based Social Networks, *Proceedings of the 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010: 458–461.
- [12] J. BAO, Y. ZHENG, M. F. MOKBEL, Location-based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data, *Proceedings of the 20th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2012: 199–208.
- [13] <http://research.microsoft.com/en-us/projects/geolife/default.aspx>
- [14] M. ESTER, H-P. KRIEGEL, J. SANDER, X. XU, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 1996: 226–231.
- [15] S. KARLIS, Y. SAYGIN, Mining Periodic Patterns in Spatio-Temporal Sequences at Different Time Granularities, *Journal of Intelligent Data Analysis*, 2009.13(2): 301–335.
- [16] C. ZHOU, D. FRANKOWSKI, P. LUDFORD, S. SHEKHAR, L. TERVEEN, Discovering Personal Gazetteers: an Interactive Clustering Approach, *Proceedings of the 12th ACM International Symposium on Advances in Geographic Information Systems*, 2004: 266–273.
- [17] A. T. PALMA, V. BOGORNY, B. KUIJPERS, L. O. ALVARES, A Clustering-Based Approach for Discovering Interesting Places in Trajectories, *Proceedings of the 2008 ACM Symposium on Applied Computing*, 2008: 863–868.
- [18] M. ZIMMERMANN, T. KIRSTE, M. SPILIOPOULOU, Finding Stops in Error-Prone Trajectories of Moving Objects with Time-Based Clustering, *Proceedings of the 2009 International Conference on Intelligent Interactive Assistance and Mobile Multimedia Computing*, 2009: 275–286.
- [19] Z. XIULI, X. WEIXIANG, Clustering Spatio-Temporal Trajectories Based on Compression of Interesting Places, *Journal of Beijing Jiaotong University*, 2011.35(3): 53–57.
- [20] A. KIRMSE, T. UDESHI, P. BELLVER, J. SHUMA, Extracting Patterns from Location History, *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2011: 397–400.

- [21] W. HU, X. LI, G. TIAN, S. MAYBANK, Z. ZHANG, An Incremental DPMM-Based Method for Trajectory Clustering, Modeling, and Retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.35(5): 1051–1065.
- [22] Y. YE, Y. ZHENG, Y. CHEN, J. FENG, X. XIE, Mining Individual Life Pattern Based on Location History, *Proceedings of the 10th International Conference on Mobile Data Management*, 2009: 1–10.
- [23] D. H. HU, C-L. WANG, GPS-Based Location Extraction and Presence Management for Mobile Instant Messenger, *Proceedings of the International Federation for Information Processing*, 2007: 309–320.
- [24] J. JIA-CHING YING, W-C. LEE, T-C. WENG, V. S. TSENG, Semantic Trajectory Mining for Location Prediction, *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2011: 34–43.
- [25] Z. LI, B. DING, J. HAN, R. KAYS, P. NYE, Mining Periodic Behaviors for Moving Objects, *Proceedings of the 16th ACM International Conference on Knowledge Discovery and Data Mining*, 2010: 1099–1108.
- [26] M. REN, G. ZHOU, G. GU, H. WANG, Statistical Analysis of User Check-in Behavior and Its Application in O2O Marketing, *Pakistan Journal of Statistics*, 2013.29(6): 973–988.

Received: December, 2014  
 Revised: April, 2015  
 Accepted: April, 2015

Contact addresses:

Min Ren  
 School of Information Science and Engineering  
 Shandong Normal University  
 Jinan, 250014  
 China  
 e-mail: rm\_sd@163.com

Feng Yang  
 School of Computer Science & Technology  
 Shandong University of Finance and Economics  
 Jinan, 250014  
 China  
 e-mail: sdkeylab@163.com

Guangchun Zhou  
 School of Computer Science & Technology  
 Shandong University of Finance and Economics  
 Jinan, 250014  
 China

Haiping Wang  
 School of Computer Science & Technology  
 Shandong University of Finance and Economics  
 Jinan, 250014  
 China

---

MIN REN is a Phd student at Shandong Normal University and an associate professor at Shandong University of Finance and Economics. Her research interests are location service and data mining.

---



---

FENG YANG is an associate professor and Master tutor at Shandong University of Finance and Economics. His research interests are electronic commerce and trajectory mining.

---



---

GUANGCHUN ZHOU is a lecturer at Shandong University of Finance and Economics. His research interest is social network.

---



---

HAIPING WANG is an associate professor and Master tutor at Shandong University of Finance and Economics. Her research interest is marketing.

---