Research Article

Interesting Activities Discovery for Moving Objects Based on Collaborative Filtering

Guan Yuan, Shixiong Xia, and Yanmei Zhang

School of Computer Science and Technology, China University of Mining and Technology, No. 1, College Road, Xuzhou, Jiangsu 221116, China

Correspondence should be addressed to Shixiong Xia; xiasx@cumt.edu.cn

Received 19 April 2013; Revised 16 June 2013; Accepted 6 July 2013

Academic Editor: Saeed Balochian

Copyright © 2013 Guan Yuan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of location-based service, more and more moving objects can be traced, and a great deal of trajectory data can be collected. Finding and studying the interesting activities of moving objects from these data can help to learn their behavior very well. Therefore, a method of interesting activities discovery based on collaborative filtering is proposed in this paper. First, the interesting degree of the objects' activities is calculated comprehensively. Then, combined with the newly proposed hybrid collaborative filtering, similar objects can be computed and all kinds of interesting activities can be discovered. Finally, potential activities are recommended according to their similar objects. The experimental results show that the method is effective and efficient in finding objects' interesting activities.

1. Introduction

In the real world, with the development of location-based service, the activity of moving objects in some regions can be tracked and their motion trajectory will be recorded by positioning device. In a sense, the trajectory data of an object is its activities in the region. Therefore, the interesting activity discovery can be transferred to finding objects' interesting regions from their historical trajectories. Moving objects' activities are often along with some purposes. Therefore, finding and studying the interesting activities of moving objects can help to understand their behavior well [1, 2]. However, the moving object activities are of many characteristics. These characteristics can be obtained through analyzing the trajectory data. Therefore, it is very important to discover their interesting activities.

There are various kinds of activities for a moving object. However, different activity types are often along with different features, and these features cannot be analyzed from the traced trajectory directly. Therefore, finding interesting activities discovery for moving objects is a challenging task [3], which is mainly due to: (1) the sampled trajectory of moving object is expressed by $\langle(Lat., Lon.), timestamp\rangle$; because of

its triviality, discovering moving objects' behaviors from their historical trajectories directly is very difficult [4]; (2) moving objects' interesting activities are not only known (which can be analyzed from the trajectory explicitly) but also unknown ones (some activities whose objects may be interesting but have not visited before), so it is difficult to find more complete interesting activities of objects synthetically [5].

In order to solve the problems mentioned previously, we propose a work on interesting activities discovery for moving objects based on collaborative filtering (CF, and in the rest of this paper we use CF to abbreviate collaborative filtering). First, we generate the objects' activity sequence and abstract the features of each activity on the basis of our previous work [3, 4]. Second, we build an object-activity-time matrix (OATM) with three dimensions, which are object identification, hot region, and sequence. Third, on the basis of the OATM, we compute object's interest degree to each hot region comprehensively and generate an object-region interest degree matrix (IDM). Fourth, combined with CF, similar objects can be queried and their common interesting activities can be found. Finally, a serial of potential interesting activities can be recommended to the moving objects according to the K -nearest neighbor (KNN) algorithm with

FIGURE 1: Framework of interesting activities discovery for moving objects.

a given threshold. As discussed above, the framework of the proposed work is described in Figure 1.

In summary, the contributions of this paper are as follows.

- (1) A moving object activity model is proposed, by which the redundant and trivial trajectories can be quantized and transformed to the meaningful moving object activity in a simple way.
- (2) The interesting drifting function is firstly introduced in the moving object interesting activity discovery, which can solve the problem of object interesting changes effectively.
- (3) The hybrid CF algorithm is introduced, taking full account on interesting drifting and interesting type, by which objects' interesting activity can be computed comprehensively and their potential activities can be discovered.

The rest of this paper is organized as follows. Section 2 describes the motivation of this paper and gives the related work of the topic. Section 3 presents an objects activities representation method and their interest degree computation method. Section 4 proposes the novel hybrid CF to the discovery of object's interesting activities. Section 5 conducts the experimental results on real data sets. Finally, Section 6 draws conclusions and points out some future works on moving object data mining.

2. The Motivation of the Work

The sampled trajectories are the record set of the moving objects, which can record their history activities with location, time, and other information. The related definitions on trajectory are given in [3, 4], which are also available in this paper. In Figure 2, there are two trajectories of Bob and Jim, and we can see that they share some common passed regions. Therefore, we can infer that the rest regions on their trajectories also have the possibility of mutual sharing by the two persons.

In current researches, there are many works that studied detecting the hot region from trajectory data [3, 5, 6]. Assuming that the regions are frequently visited by Bob and Jim, and

their geographical places are described in Figure 2. Therefore, we can make the following inferences: Bob generally eats breakfast at home, and then goes to the gym, then does shopping, and returns home after buying some food in a nearby supermarket. At the same time, Jim leaves home and goes to restaurant for his breakfast, and then does bodybuilding, next he goes shopping, in the early evening, he buys food in a nearby supermarket and goes home. These details show that bodybuilding and shopping are common hobbies shared by Bob and Jim. Therefore, these spatially overlapping regions indicate that the two persons have some similar interest activities; that is, they both love sports and shopping. If the two persons are similar enough in their interest activities, we can find one person's interest activities based on those of another person using the advanced hybrid CF [7–9].

With this idea, this paper presents a novel work on interesting activities discovery for moving objects based on hybrid CF.

3. Representation of Moving Object Activity

In our previous work, we have studied the moving pattern of moving objects from their trajectories [3, 4]; thus, we know that once an object stays in a hot region can be viewed as one of his activities. Therefore, we develop an efficient algorithm called DB-HR [3] to find the hot regions where moving objects stay in for a long time. In this section, we assume that objects' activities have already been generated by DB-HR through finding their hot regions. So, we need to formally represent activities for further analysis on the interesting activities finding for moving objects.

Definition 1. Activity (a): an activity means a moving object's once access to a hot region, and it is a four-tuple, denoted as $a = \{O_{id}, HR_{id}, StartTime, EndTime\}$, where O_{id} is the identification of the object, HR_{id} is the hot region, and $StartTime$ and $EndTime$ are the timestamps that the object enters and leaves the hot region.

Using moving object activity, trajectory can be converted into a sequence of activities: $S = \{a_1, a_2, a_3, a_4, \dots, a_n\}$, where S is the activity sequence and a_i is the i th activity in the S . With the activity sequence, the triviality of trajectory can be avoided. Moreover, by calculating the activities in the same region of the activity sequence comprehensively, the quantitative calculation result is the interest degree of moving object for each region.

3.1. Activity Matrix Generation. In order to clearly express the spatiotemporal relationship between objects' activities, in this paper, we establish an object-activity-time matrix (OATM, as shown in Figure 3) to represent the objects' activities and times structurally. Then, we can calculate moving objects' interest degree in the activities by converting the object activity sequence into object activity matrix. In this way, on one hand, it is convenient for later calculation by converting the original multidimensional scattered information into a

FIGURE 2: Relation between moving objects' activities.

FIGURE 3: Matrix of activity hot regions and time frame for moving objects.

unified data; on the other hand, studying more comprehensive interest degree calculation method can intelligently quantify the interest degree of moving object activities, and the numerical value of interest degree can also reflect the characteristics of the object.

Figure 3 gives the matrix of activity regions and time frame for moving objects. It can be seen from the matrix that the moving object may access region HR_2 many times in a given time period. By this way, we can know that the object may be more interested in this region. Therefore, we give the calculation rules of interest degree for object activity afterwards.

3.2. Calculation of Interest Degree for Object Activities. The more activities happened in a certain hot region, the higher interest degree it is for a moving object [5]. In this sight, we make further abstraction on activity types, then give the definition of object activity type (Activity Type, AT) and moving object activity interest degree (Interest Degree, ID):

Definition 2 (activity type (AT)). Activity type refers to the category of the hot regions, denoted as $AT = \{at_i \mid HR_i, O_{id}, TimeLength, Count, LastTime\}$. The attributes represent activity region, object identification, activity duration, number of activities, and the last access time, respectively. Activity type is mainly decided by the categories of hot region. For example, supermarket and shopping centers are both commercial areas, while persons in these two regions share different types of activities. Therefore, using activity type to classify the activities of moving objects can discover interesting activity of object from more general aspects.

Definition 3 (interest degree (ID)). The interest degree of moving objects in related regions is a measurement of how much the object likes the activity. The interest degree of moving object mainly includes three aspects: (1) the activity count that moving objects visited in a given hot region; (2) the activity duration that moving objects spent in the given hot region; (3) the activity time that moving objects spent in a given hot region.

The objects' interests are always changing and evolving constantly; in this paper, we introduce interest drift function [9] to solve this problem. On the basis of interest drift function, we give the calculation formula of interest degree for moving object. For example, the interest degree of object O in region HR_i is computed as follows:

$$ID_{HR_i}^O = \left(\frac{\sum_{a, hr \in HR_i} C_{HR_i}}{\sum_{a, o_{id} \in O} C_{at}^O} + \frac{\sum_{a, hr \in HR_i} TL_{HR_i}}{\sum_{a, o_{id} \in O} TL_{at}^O} \right) \times h(t). \quad (1)$$

In which, $\sum C_{HR_i}$ is the total activities that the object O stayed in the hot region HR_i . $\sum C_{at}^O$ represents object O 's total activities of in the data set. $\sum TL_{HR_i}$ denotes the total duration of the object O in the HR_i . $\sum TL_{at}^O$ represents the total duration of the object O in the data set. $h(t)$ is the interesting drift function, and it is the function of time t . Through the function, the retention degree of the object activity in the interesting region HR_i can be calculated given the object interesting drift function $h(t)$ as follows:

$$h(t) = m \times \left(\frac{t - t_{\min}}{t_{\max} - t_{\min}} \right)^2 + 1 - m. \quad (2)$$

In (2), $t_{\min}$ is the earliest time that the moving object access to related hot region, and $t_{\max}$ is the most recent statistical time. Let DS be the set of statistical data, then $t_{\min} = DS.BeginTime$, and $t_{\max} = DS.EndTime$ ($t_{\min} \leq t \leq t_{\max}$). m is the coefficient of interests drifting, namely, the speed of object's interest drift, $m \in [0, 1]$. When $m = 0$, no interests drift happens, and when $m = 1$, the fully nonlinear interest drift happens; the function value of interest drift is between 0 and 1. According to the definition of interest drift coefficient [10], each object's interest drift coefficient is not the same. In this regard, this paper uses parameter m to denote different interest drift.

In order to make it more flexible to calculate the object's interest degree to a certain activity, we need to introduce

two factors to adjust the importance of visit count and visit time. Therefore, we define the weighting vector of the interest degree, $W = \{W_C, W_{TL}\}$. W_C is the weight of the visit count, and W_{TL} is the weight of visit time. The interest degree of object activity can be adjusted by setting different numerical ratios. Therefore, we give the interest degree formula as follows:

$$ID_{HR_i}^O = \left(W_C \times \frac{\sum_{\forall a, hr \in HR_i} C_{HR_i}}{\sum_{\forall a, o_{id} \in O} C_{at}^o} + W_{TL} \times \frac{\sum_{\forall a, hr \in HR_i} TL_{HR_i}}{\sum_{\forall a, o_{id} \in O} TL_{at}^o} \right) \times h(t). \quad (3)$$

By setting the weighting vector of the interest degree, the interest degree of an object in a particular region can be calculated by a more comprehensive and more flexible way. By using (3), the matrix in Figure 3 can be converted into matrix of -hot regions objects, shown in Figure 4. By this way, the activities of objects in their interesting regions are converted from multidimensional sampled data into a single numerical data, and it is convenient for further calculation.

In practical application, due to the randomness of object activity and the longer time interval, the matrixes in Figures 3 and 4 are often very sparse. In fact, the object activities are all of specific purposes; for example, the objects no matter how long they stay in a supermarket or in a shopping center, they all do shopping. Therefore, if we give full consideration into the type of activity, it may increase the accuracy of the interesting activities discovery, and at the same time, it can reduce a great deal of redundancy. Suppose that HR_1 , HR_2 , HR_3 , and HR_4 which are interesting regions of O_1 , HR_1 , and HR_2 belong to the same type T_1 , while HR_2 and HR_3 belong to the same type T_2 , then the moving object in the two kinds of activity types has a comprehensive interest degree, expressed as follows:

$$\overline{ID}_{T_m}^O = \frac{\sum_{HR_i \in T_m} (\sum_{\forall a, hr \in HR_i} C_{HR_i} \times ID_{HR_i}^O)}{\sum_{HR_i \in T_m} \sum_{\forall a, hr \in HR_i} C_{HR_i}}. \quad (4)$$

The object's total interest degree to a certain activity type is the average weighted values of its interest degree to each single activity, which belongs to the same activity type. This value can synthetically reflect the comprehensive level of interest degree that the moving objects visit to a certain type of regions. The interest degree of activity type-objects can be calculated by (4).

Through the comprehensive calculation for interest degree, the interest degree matrix of hot regions-objects (Figure 4) and the interest degree matrix of activity type (Figure 5) are generated. Afterwards we will use hybrid CF to calculate the similarity degree of moving object based on their interesting matrixes and find their interesting activities.

4. Interesting Activity Discovery for Moving Objects

To better describe the work proposed in this paper, notation of basic symbols is given in Table 1.

TABLE 1: Parameters notation.

Parameter	Description
S_o	The activity sequence of moving objects
$Len(S)$	The length of the activities sequence
$T(S)$	Time span of activity sequence
$HRset$	The set of hot regions
S_i^o	The i th activity of object O in the sequence S
$Len(HRset)$	The total count of activities in the $HRset$

Moving objects

→

		O_0	O_1	O_2	$\cdots$	O_i	$\cdots$	O_m
Activities	HR_0	ID_0^0	ID_0^1	ID_0^2	$\cdots$	ID_0^i	$\cdots$	ID_0^m
	HR_1	ID_1^0	ID_1^1	ID_1^2	$\cdots$	ID_1^i	$\cdots$	ID_1^m
	HR_2	ID_2^0	ID_2^1	ID_2^2	$\cdots$	ID_2^i	$\cdots$	ID_2^m
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
	HR_n	ID_n^0	ID_n^1	ID_n^2	$\cdots$	ID_n^i	$\cdots$	ID_n^m

FIGURE 4: The interest degree matrix of hot regions-objects.

4.1. Interesting Activities Discovery. In this section, we use hybrid CF method to find the potential interesting activities for moving objects. Similarity computation is the key of collaborative filtering method. The cosine similarity is often viewed as the standard similarity function for collaborative filtering algorithm [7, 11], and it uses the angle between two vectors to represent their similarity. The function is shown as follows:

$$\text{Sim}(O_a, O_b) = \frac{\sum_{i \in I} ID_i^a \times ID_i^b}{\sqrt{\sum_{i \in I} (ID_i^a)^2} \sqrt{\sum_{i \in I} (ID_i^b)^2}}. \quad (5)$$

Formula (5) is the standard cosine similarity function. ID_i^a represents the interest degree of moving objects O_a in hot region HR_i . In order to avoid the differences in moving objects while visiting different hot regions, we need to improve the cosine similarity function. Pearson [11] proposed the modified cosine similarity function on the basis of the standard one as follows:

$$\text{Sim}(O_a, O_b) = \frac{\sum_{i \in I} ((ID_i^a - \overline{ID}_a) \times (ID_i^b - \overline{ID}_b))}{\sqrt{\sum_{i \in I} (ID_i^a - \overline{ID}_a)^2} \sqrt{\sum_{i \in I} (ID_i^b - \overline{ID}_b)^2}}. \quad (6)$$

Among them, $\overline{ID}_a$ and $\overline{ID}_b$ are the average interest degrees of moving objects O_a and O_b in all hot regions. The adjusted cosine similarity function improves the interest degree deviation of different objects by introducing the average interest degree. We use O_a and O_b commonly visited regions as their interest activities ($HR_a \cap HR_b$), denoted as HR' . The value of $\text{Sim}(O_a, O_b) \in [0, 1]$, and the bigger $\text{Sim}(O_a, O_b)$ is, the higher similarity between O_a and O_b is. Through the similar function

	Moving objects						
	O_0	O_1	O_2	$\dots$	O_i	$\dots$	O_m
Activity type	T_0	$\overline{ID}_0^0$	$\overline{ID}_0^1$	$\overline{ID}_0^2$	$\dots$	$\overline{ID}_0^i$	$\dots$
	T_1	$\overline{ID}_1^0$	$\overline{ID}_1^1$	$\overline{ID}_1^2$	$\dots$	$\overline{ID}_1^i$	$\dots$
	T_2	$\overline{ID}_2^0$	$\overline{ID}_2^1$	$\overline{ID}_2^2$	$\dots$	$\overline{ID}_2^i$	$\dots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	T_n	$\overline{ID}_n^0$	$\overline{ID}_n^1$	$\overline{ID}_n^2$	$\dots$	$\overline{ID}_n^i$	$\dots$

FIGURE 5: The interest degree matrix of activity type-objects.

	Moving objects						
	O_0	O_1	O_2	$\dots$	O_i	$\dots$	O_m
Moving objects	O_0	1					
	O_1	S_{10}	1				
	O_2	S_{20}	S_{21}	1			
	$\vdots$	$\vdots$	$\vdots$	$\dots$	1		
	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	1
	O_m	S_{m0}	S_{m1}	S_{m2}	$\dots$	S_{mi}	$\dots$

FIGURE 6: Similarity matrix of moving objects.

Formula (6), we can get an object similarity matrix (shown as Figure 6), which contains object similarity between each other. For the similarity is symmetrical, the similarity matrix can be showed as a low triangular matrix.

There are two major tasks in finding object interesting regions: (1) $\forall HR_i, \exists HR_j \rightarrow O_a$ has never accessed ($HR_i \notin S_{O_a}, HR_{set}$); then we will predict the interest degree of object O_a in HR_i ; (2) finding N ($N > 1$) regions that object O_a most likely to visit from the hot regions which it never visited, as the potential interest regions of object. The similar matrix is symmetrical, and we can easily determine the most similar objects. Here, we use K to denote the count of similar objects.

By calculating the similarity between moving objects, the selection of k similar moving objects for unknown hot region HR_i based on hybrid CF is proposed (the K moving objects which are most similar to O_a). We define these similar objects as $S(O_a)$ and $|S(O_a)| \leq k$; then we can forecast their interest degree in HR_i , expressed by $\overline{ID}_j^a$, and is shown as follows:

$$\overline{ID}_j^a = \overline{ID}_a + \frac{\sum_{O_x \in S(O_a)} \text{sim}(O_a, O_x) \times (\overline{ID}_j^x - \overline{ID}_x)}{\sum_{O_x \in S(O_a)} \text{sim}(O_a, O_x)}. \quad (7)$$

In (7), $\overline{ID}_a$ and $\overline{ID}_x$ are the average interest degrees of O_a and O_x to other hot regions; $\overline{ID}_x^a$ is the predictive interest degree of O_a to HR_x . In the process of interesting activity discovery, we generally select k most similar neighbors in the similarity matrix, use the k nearest neighbors to recommend predictive interesting activity, and give the predictive interest degree.

4.2. Flow of the Proposed Algorithm. Therefore, based on the above analysis, we give the approach of interesting activity discovery for moving objects based on hybrid CF, shortly for approach of interesting activity discovery (IAD). Through IAD algorithm, potential interesting activities can be discovered for objects. The pseudocode of IAD algorithm is shown in Algorithm 1.

Algorithm IAD requires 3 input parameters, which are the recommended object O_{id} , the number of the nearest neighbor k , and the threshold of predictive interest degree ID_{th} . This algorithm mainly includes the following steps. Lines 01–04 are mainly used to load sequence of moving object activity and extract related activities; then calculate the interest degree of object-activity and the similarity between objects. The calculated data is stored in the matrix, which is convenient for further analysis. Lines 05–07 are mainly used to predict the potential interesting activities based on the object similarity matrix and the object interest degree matrix. Also, the predictive interest degree for potential activities is given. Since the algorithm only calculates the activities' interest degree for moving objects and discoveries of objects' interesting activities, therefore, we just need to update the calculated interest degree. The worst case of the time complexity is $O(m^2 + n^2)$, where m is the number of moving object and n is the total of hot regions. After the first calculation, the time complexity of the algorithm is a constant $O(m + n)$, where m is the time consumption of finding similar neighbor and n is the time of recommending the potential interesting activities of objects.

5. Experiments and Analysis

To validate the method proposed in this paper, we develop a module of interesting activities discovery for moving objects in our trajectory data mining system (TrajMiner). The data set is the GPS trajectory data [12], abbr. GeoLife. The metadata of the trajectory is stored in the text files. TrajMiner firstly transformed the GeoLife metadata into the standard trajectory data by noise reduction, data cleaning, and trajectory reconstruction. Then, it stores the trajectory data into the SQLServer 2008. In this experiment, we select 2983 trajectories of 50 moving objects, which include 6037452 sampling points, to analyze the accuracy and efficiency of the proposed method.

5.1. Accuracy Analysis. In this part, we adopt precision and recall functions to validate the accuracy of IAD. The precision and recall functions are given as follows:

$$\begin{aligned} \text{Precision} &= \frac{\sum_i TP_i}{\sum_i (TP_i + FP_i)} \times 100\%, \\ \text{Recall} &= \frac{\sum_i FN_i}{\sum_i (TP_i + FN_i)} \times 100\%, \end{aligned} \quad (8)$$

where TP_i denotes true positive, which means that the object has visited HR_i and the algorithm was truly recommended to him. FP_i denotes false positive, which means that the object did not visit HR_i , but it was truly recommended to him. FN_i is

Input: O_{id} , k , ID_{th}

Output: the latent activity set of O_{id}

01: Abstract all the activity sequences and stored them in the S_{set} ;

02: Build object-visit-matrix M according to the S_{set} ;

03: Compute object-region interest degree from M , and build the interest degree matrix M_{id} ;

04: Search similar objects using hybrid CF algorithm, and generate similar matrix M_s according to M_{id} ;

05: Find the k most nearest neighbors to O_{id} from M_s ;

06: On the basis of the k most nearest neighbors, recommend the latent activities for O_{id} ;

07: If the O_{id} 's interesting degree to the recommended activities is less than ID_{th} , then remove the recommended activity, otherwise, recommend them to O_{id} .

End.

ALGORITHM 1: Interesting activity discovery (IAD).

TABLE 2: Data volume of the six stages.

Phase	Count of objects	Count of trajectories	Count of sampling points
1	50	1080	1,000,170
2		2091	2,004,283
3		2319	3,002,400
4		2672	4,004,758
5		2870	5,001,380
6		2983	6,037,452

false negative, which means that the object have visited HR_i , while it was not recommended to him. Therefore, the higher precision is, the better our method is, and the lower recall is, the better our method is. Figure 7 shows the accuracy of our proposed method.

In Figure 7, P and R are the precision and recall of the ID_{th} value noted in the corresponding bracket. We analyze the accuracy in two situations, one is associated with interesting drift (Figure 7(a)), and the other is without interesting drift (Figure 7(b)). From the two figures, we can see that with the increase of K and ID_{th} , the precision increases slightly, and the recall decreases slowly. The precision associated with interesting drift seems smooth (with $m = 0.5$), while the precision without interesting drift is not very stable.

5.2. Efficiency Analysis. In order to validate the efficiency of the proposed algorithm, in this section, we adopt the incremental validation method for IAD verification. First, we select a subset of GeoLife for static calculation of interesting activity. The subset includes 1000170 sampling points from 1080 tracks of 50 objects, and 104 hot regions can be discovered by the DB-HR [3] from trajectories. Then, we increase the number of training data on the basis of the first phase gradually from 1000 to nearly 3000 trajectories of the 6 stages. For each training stage, we analyze the efficiency of the algorithm. Table 2 gives the number of training data in different stages.

In order to evaluate the effectiveness of IAD algorithm, we introduce mean absolute error (MAE) [11, 13] as the efficiency measurement. MAE is the most commonly used

efficiency measurement of recommendation because it is easy to understand and it can be intuitive, convenient, and accurate to evaluate the quality of algorithm. MAE is the interest degree difference between the predictive activities and real activities. Therefore, the smaller MAE is the higher efficiency the algorithm is. In order to compare the experimental results, this section improves (4) by removing the interest shift function $h(t)$. Getting the simple interest degree of the object $\vec{ID}_{HR_i}^O$, both of the weights are set to 0.5:

$$\vec{ID}_{HR_i}^O = 0.5 \times \frac{\sum_{\forall a, hr \in HR_i} C_{HR_i}}{\sum_{\forall a, o_{id} \in O} C_{at}^o} + 0.5 \times \frac{\sum_{\forall a, hr \in HR_i} TL_{HR_i}}{\sum_{\forall a, o_{id} \in O} TL_{at}^o}. \quad (9)$$

We set the predictive interest degree of activities for object as $\{P_1, P_2, \dots, P_n\}$. While the actual interest degree set between object activity is $\{R_1, R_2, \dots, R_n\}$, then according to the definition of MAE, the mean absolute deviation MAE is calculated as follows:

$$MAE = \frac{\sum_{i=1}^N |P_i - R_i|}{N}. \quad (10)$$

In the experiment, we set $k \in [3, 10]$ and $ID_{th} \in [0.5, 1]$ to guide the interesting activities discovery. First, we make a comment on the average similarity of each stage under the same value of k . Then, we compare the impact of different values of k on MAE, trying to find out the relationship between the number of the nearest neighbors and the efficiency of recommendation, and the effect of similarity on MEA under different values of k . The contrast relation between k and MAE in different incremental stages is shown in Figure 8. As being the distinguished degree of results when $k = 3, 5, 8$, in Figure 8(b), the predicted MAE is relatively high. As shown in Figure 8, when $k = 8$, the average of similarity between objects compared to the stability number is changed relatively little. It shows that in this case the similarity between objects is more consistent, and increased activity is similar in the neighborhood with the increase in the number. Therefore, it will achieve good effect to recommend among interesting activities for object using the number of neighbor with $k = 8$. For Figure 8(b), the recommendation of interesting activities depends on the size

FIGURE 7: Accuracy analysis of the proposed method.

FIGURE 8: The MAE values with different phases and K.

of training data set; the larger the number of the training data is, the better effect is and the more accurate results are. As can be seen from Figure 8, with the increase of data, the accuracy of interesting activities discovery tends to be stable. In this case, it can achieve more ideal effect with data continuing to increase.

5.3. Effectiveness Analysis. In order to validate the effectiveness of the proposed algorithm, in this section, we compare our algorithm with traditional CFs and KNN-CF [11] (which is a recent proposed advanced CF). Here, we denote UBCF (user-based CF) and IBCF (item-based CF) as traditional

CF algorithms. In this section, another evaluation index coverage [13] is introduced to validate the effectiveness of the compared algorithms. Coverage is a widely used effectiveness evaluation index to check the coverage rate of a recommendation system, the calculation is shown as follows:

$$\text{Coverage} = \frac{\sum_{o_{id} \in O} |P_{o_{id}} \cap I_{o_{id}}|}{\sum_{o_{id} \in O} I_{o_{id}}}. \quad (11)$$

In the formula above, $P_{o_{id}}$ is the activity set that the system recommends to the O_{id} , and $I_{o_{id}}$ is the interesting activity set discovered by the system. $P_{o_{id}} \cap I_{o_{id}}$ is the coverage of

FIGURE 9: Effectiveness analysis with two indexes.

the system recommends and objects' interesting activities. Therefore, the higher coverage is, the more effective algorithm is.

In this experiment, we validate the effectiveness of our hybrid CF from two aspects. (1) the convergence of the compared algorithm with the relation of MAE; (2) the effectiveness of the proposed algorithm with the relation of coverage.

We can read from Figure 9 that our algorithm is better than traditional CFs and recent proposed advanced KNN-CF. In Figure 9(a), the 4 algorithms perform nearly at the same level. When the total count of moving objects is 10 or more, the MAEs of hybrid CF and KNN-CF decrease rapidly and the two algorithms become smooth when total objects are more than 30. Since no strategies are used in traditional CFs, both UBCF and IBCF perform not so well as ours and KNN-CF. The best MAE value of ours is 0.09 when total objects reach 30, while the best MAE value of KNN-CF is 0.13 when total objects reach 37. In Figure 9(b), we can see that our hybrid CF reaching the stable state is faster than the other 3 algorithms. In both experiments of Figure 9, the difference between 4 algorithms is very little when total objects are very few. As shown in Figure 9(b), the best coverage of our hybrid CF reaches 99.8% when the count of objects is more than 40, and the best coverage of KNN-CF reaches 91.3% when the count of objects is more than 35. Since activity type and interesting drifting mechanism is adopted in our algorithm, interesting hot regions which objects visit frequently will have more weights, and this value plays very important role in the object interesting activity discovery. Therefore, our hybrid CF is better than KNN-CF and traditional CFs, such as UBCF and IBCF.

As discussed in [14, 15], the performance of CF algorithm is often influenced by three factors: (1) the data sparseness; (2) the scalability of the work; (3) the original evaluation. In fact, the mentioned three factors are well conquered in our work. Firstly, we use activity type to merge the sparse object's interesting into once class or type. Secondly, through the activity type, a great many activities happened in hot regions with the same type can also be merged, and so, the performance can be well improved. Thirdly, for a hot region

never visited by anyone, it also can be inferred to the certain type and recommended to objects who like it.

6. Conclusions

In this research, a novel work on interesting activity discovery for moving object is presented. Firstly, a moving object activity model is proposed to store objects' activities and their attributes. Secondly, by introducing activity type and interesting drifting mechanism, the hybrid CF is proposed to compute objects interesting activities comprehensively. Thirdly, based on newly proposed hybrid CF, objects which are similar to each other can be found, and their potential interesting regions can be predicted. We have conducted numerous experiments from several aspects on real dataset and demonstrated the accuracy, efficiency, and effectiveness of our hybrid CF. In the future study, we can address moving objects' interesting routes discovery based on the work proposed in this paper.

Acknowledgment

This work was supported by the Fundamental Research Funds for the Central Universities, China (under Grant no. 2013QNA25).

References

- [1] Z. H. Li, B. L. Ding, J. W. Han, R. Kays, and P. Nye, "Mining periodic behaviors for moving objects," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*, pp. 1099–1108, New York, NY, USA, July 2010.
- [2] H. P. Cao, N. Mamoulis, and D. W. Cheung, "Discovery of periodic patterns in spatiotemporal sequences," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 4, pp. 453–467, 2007.
- [3] G. Yuan, S. X. Xia, L. Zhang, M. Zhu, and C. Ji, "Asynchronous periodic patterns discovery for moving objects," *Journal of Convergence Information Technology*, vol. 7, no. 9, pp. 286–294, 2012.

- [4] G. Yuan, S. X. Xia, L. Zhang, Y. Zhou, and C. Ji, "An efficient trajectory-clustering algorithm based on index tree," *Transactions of the Institute of Measurement and Control*, vol. 32, no. 7, pp. 850–861, 2012.
- [5] Y. Zheng and X. Xie, "Learning location correlation from GPS trajectories," in *Proceedings of the 11th IEEE International Conference on Mobile Data Management (MDM '10)*, pp. 27–32, Kansas, Mo, USA, May 2010.
- [6] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, "A clustering-based approach for discovering interesting places in trajectories," in *Proceedings of the 23rd Annual ACM Symposium on Applied Computing (SAC '08)*, pp. 863–868, Fortaleza, Brazil, March 2008.
- [7] C.-G. Huang, J. Yin, J. Wang, Y.-B. Liu, and J.-H. Wang, "Uncertain neighbors' collaborative filtering recommendation algorithm," *Jisuanji Xuebao/Chinese Journal of Computers*, vol. 33, no. 8, pp. 1369–1377, 2010.
- [8] H. Wu, Y.-J. Wang, Z. Wang, X.-L. Wang, and S.-Z. Du, "Two-phase collaborative filtering algorithm based on co-clustering," *Ruan Jian Xue Bao/Journal of Software*, vol. 21, no. 5, pp. 1042–1054, 2010.
- [9] Q. N. Li, Y. Zheng, X. Xie, Y. K. Chen, W. Y. Liu, and Y. Ma, "Mining user similarity based on location history," in *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS '08)*, pp. 34–43, Irvine, Calif, USA, November 2008.
- [10] G. Yuan, S. X. Xia, L. Zhang, Y. E. Li, and J. Y. Xiao, "Web users clustering based on interest degree," *Microelectronics & Computer*, vol. 25, no. 9, pp. 112–114, 2008.
- [11] X. Luo, Y.-X. Ouyang, Z. Xiong, and M. Yuan, "The effect of similarity support in k-nearest-neighborhood based collaborative filtering," *Jisuanji Xuebao/Chinese Journal of Computers*, vol. 33, no. 8, pp. 1437–1445, 2010.
- [12] S. Karli and Y. Saygin, "Mining periodic patterns in spatio-temporal sequences at different time granularities," *Intelligent Data Analysis*, vol. 13, no. 2, pp. 301–335, 2009.
- [13] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5–53, 2004.
- [14] H.-L. Xu, X. Wu, X.-D. Li, and B.-P. Yan, "Comparison study of internet recommendation system," *Ruan Jian Xue Bao/Journal of Software*, vol. 20, no. 2, pp. 350–362, 2009.
- [15] X. Y. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, vol. 2009, Article ID 421425, 19 pages, 2009.