

Discovering Region of Interest in Location-based Social Networking Services on Road Network

Rong Tan¹ and Wen Si¹

¹College of Information and Computer Science, Shanghai Business School,
201400, Shanghai, P. R. China
tanrong529@gmail.com, siw@sbs.edu.cn

Abstract

Region of Interest (ROI) discovery is among the most common functions in Location-based social networking services (LBSNS). While former researches mainly utilize the accurate location coordinates history, the Road-context-based Active Region Extraction Algorithm (RAREA) proposed in this paper explores the method to extract those regions with road contexts. Experimental results illustrate that by analyzing the characteristics of those road contexts, ROIs are able to be discovered with high efficiency. And our work shows that privacy protection and personalized services can be both achieved in LBSNS.

Keywords: Region of Interest Discovery, Location-based Social Networking Services, Mobile Computing, Location Privacy Protection, Road Network, Social Computing

1. Introduction

With the development of mobile positioning technology and the popularity of smart portable devices, temporal and spatial contextual information has been increasingly utilized to mine relevant data in recent years. For location-based social networking services (LBSNS) [1-3], both relations discovery and content delivery can benefit from the mined information. Moreover, the areas where user frequently takes activities could be discovered [4-6]. Location factors are able to play an important role in establishing new social relationships since the similarity of those active areas in each other's life can be used as a useful reference [7-9]. The closer location could give rise to a greater possibility of sharing a common topic and strong sense of identity. On the other hand, users' preferences could be obtained as well. If the contents pushed by LBSNS are closer to the users' preferences, they are more likely to cause the users' attention so as to realize the relevant purposes [10-11].

At present, coordinates of longitude and latitude as well as the trajectories are two kinds of raw data widely used to mine users' preferences in LBSNS. Coordinates are generally utilized in users' "check-in" activities, namely the coordinates represent the position of the POIs visited by users, such as the restaurant, shopping mall, etc., [12-13]. Through the analysis of the characteristics of the POI objects, some preferences of the user could be obtained. The trajectory is another kind of movement model that has arisen in recent years [14-15]. It is a record of the moving object's location in a certain time and space, and can be represented as a series of positional entities and time entities sequences. Compared with the analysis based on coordinates, the trajectory analysis can obtain more user information, including users' important activity places, habits, behaviors and preferences, etc.

The mining of both coordinates and trajectories requires accurate location and scene information. However, for LBSNS, it is inadvisable to display users' precise locations or to acquire position data for a long period of time since it is likely to cause users' privacy leakage resulting in unpredictable potential dangers. It has become the consensus of the academic field and the industrial circle in recent years that the non-precise location

information is utilized to protect the location privacy in the mobile environment. Furthermore, the acquisition of trajectory information requires recording the users' location data for a long time so that the consumption of portable mobile device resources is large which may affect the normal proceeding of other works. And although the ability of mobile devices to access location data has been greatly enhanced, it is still subject to a number of objective factors. For instance, GPS can only be located outdoors. When users need to switch between different transports, GPS's positioning function will be significantly reduced. Therefore, the application of LBSNS needs not only to adjust the form of location information, but also to put forward relevant mining analysis methods.

This paper mainly describes the method of data mining based on road network. The road context refers to the user's exact location mapped on the road network. By changing the display granularity of the location information, the privacy is protected. This method has been applied to a number of mature LBSNSs. For example, when a user of Sina Weibo which is one of the biggest LBSNS in China wants to shares his location, the position information is displayed in the form of "XX Road, XX District, XX City, XX Province". Although the granularity of information is changed, data mining can also be carried out by utilizing the road and temporal contextual information. It is found that the vast majority of users usually act in limited geographic ranges. Based on this hypothesis, the Road-context-based Active Region Extraction Algorithm (RAREA) is proposed in this paper. The RAREA algorithm assumes that the users' regions of interest are closely related to the road contexts they frequently move around. In the next section, the road context model will be introduced first.

2. Road Context Model

RAREA assumes that the vast majority of people's daily life is regular, and people's preferences are rarely changed in a certain period of time. Therefore, for the vast majority of people, the range of daily activities will be limited to a few relatively small regions. When they use the mobile social networking applications to share road context, there must be certain rules behind it. It is conducive to extracting the set of users' active regions related to these road contexts to understand and discover the laws behind these phenomena. The data statistics in above section have indirectly verified the assumption.

As a matter of fact, many unknown factors affect users when they sharing the road context. For example, users' important activity locations are in the vicinity of these roads, locations attract the users are on the roads, or the users have to pass the roads when going to the destinations. Users share the road context for corresponding hidden reasons. Although it is unrealistic to find all hidden reasons and to make a quantitative analysis, appropriate analysis can be carried out in other aspects so as to find the answer.

First of all, based on the road data, the importance levels of each road for users are obviously different. It is the first work to measure the importance level of the road context shared by users. Secondly, the roads with higher importance levels are more likely to generate links with potential active regions. It is also of great importance to measure the internal relation between the road context shared by users and their potential active regions. Only through the comprehensive analysis of the importance of road context and its intrinsic link with the potential active region can the set of users' active regions be extracted.

2.1. Road Context Definiton

Road context shared by users is corresponding to a path in a road network. The road context model is defined as follows:

Definition 2.1. The road network $RN=(V,E,w)$ can be expressed as a directed weighted graph, where V is the set of all set, E is the set of edges between two points. The

weight function $w: E \rightarrow \mathbb{R}^+$ is the length of edge, which is the Euclidean distance between two points.

Definition 2.2. Road context r_i is a path in road network RN and a vertex sequence $r_i = \{v_1, v_2, \dots, v_n\}$. The first vertex r_i of the sequence is the starting point of the road, and the last vertex is its terminal point. The intermediate point is the cross vertex between road r_i and other roads. Any two consecutive vertices form a section of road r_i , which is denoted as $e(v_i, v_{i+1})$. The length of road r_i is $len(r_i) = \sum_{i=1}^{n-1} w(e(v_i, v_{i+1}))$, namely the sum of the weights of the edges in the road position r_i .

2.2. Importance Level of Road Context

The road context shared by users has a different level of importance, which is denoted as $IL(r_i)$. The level is mainly determined by the following three aspects:

A. Road count weight ($w_{count}(r_i)$)

The larger counts of the road context shared by users indicate higher importance level of the road. Thus the corresponding count weights will be greater. Here, the weight of road counts is the count value of the road, namely $w_{count}(r_i) = count(r_i)$.

B. Road length weight ($w_{len}(r_i)$)

It is more difficult to estimate the impact of longer roads on users. On the contrary, it is relatively easy to judge the importance levels of shorter roads. Huqingping Highway is taken as an example. It passes through the Qingpu District of Shanghai, bordering on Yan'an Road Viaduct in the east, and to Pingwang Town in Jiangsu Province in the west. The overall length is over 40 km. If the road context shared by a user is Huqingping Highway, it is difficult to determine the importance level of each section. Therefore, it is conducive to improving the accuracy to appropriately reduce the weight of longer roads, and increase the weight of shorter roads. The weight of road length is the product of count weight and the weight ratio coefficient of road length α , namely $w_{len}(r_i) = \alpha * w_{count}(r_i)$. The ratio coefficient α of the weight of road length is divided into five categories, which is as shown in Table 1.

Table 1. The Weight of Road Length

Road length interval (Unit: km)	Very short (0,2]	Short (2,3]	Ordinary (3,4]	Long (4,5]	Very long (5,∞)
Weight ratio coefficient α	0.3	0.25	0.2	0.15	0.1

C. Road type weight($w_{type}(r_i)$)

Similar to the weight of the road length, the weight of the road type, takes into account different road types. Its importance level to determine the active region will be different. In Table 2, the road is classified according to the Open Street Map [75]. Different types of roads are endowed with different weights. The weight of road type is the product of count weight and the ratio weight coefficient of road type, namely $w_{type}(r_i) = \beta * w_{count}(r_i)$.

Table 2. Road Type Weight

Type	Primary	Secondary	Tertiary	Trunk	Highway	Residential
Weight ratio coefficient β	0.15	0.2	0.25	0.05	0.05	0.3

To sum up, the importance level of road context is equal to the sum of road count weight, road length weight and road type weight. It is defined as the following formula:

$$\begin{aligned}
 IL(r_i) &= w_{count}(r_i) + w_{len}(r_i) + w_{type}(r_i) \\
 &= (1 + \alpha + \beta) * count(r_i)
 \end{aligned} \tag{1}$$

2.3. Influence Force of Road Context

In order to measure the close degree of the intrinsic link between the road context and the potential active regions, the concept of influence force of road context is proposed. It is assumed that each piece of road context will have an influence force on any other location in the space around it. The sum of all influence forces of each location reflects the possibility that it could become an active region. The location closer to the road context is more likely to be affected by the influence force, and the possibility that it could become the active region is greater. As the distance increases, the influence will gradually decrease. It shows that the relation between the location and the potential active region gradually decreases. This process can be considered as a typical distance attenuation phenomenon. In the existing distance attenuation model, the exponential function, power function and Gauss function can reflect this phenomenon. The exponential function is selected as the distance attenuation model to calculate the influence level of road context. Moreover, in the road network, it is obviously inappropriate to utilize Euclidean distance as the distance measure in distance attenuation model. Therefore, the shortest path calculation method is utilized here, which is defined as follows:

Definition 2.3. It is assumed that there is a point $u \in V$ in road network $RN = (V, E, w)$. Its shortest path to the other point $v \in V$ is $SP(u, v)$. If the distance between u and v remember is denoted as $dist(u, v)$, $dist(u, v) = len(SP(u, v))$.

Definition 2.4. The average distance between any point $p \in V$ in space and any road r_i is denoted as $avgDist(p, r_i)$, which is equal to the average value of the shortest path between point p and all vertices $v_j \in r_i$ in $r_i = \{v_1, v_2, \dots, v_n\}$. It is defined by formula 2 as follows:

$$avgDist(p, r_i) = \frac{1}{n} \sum_{j=1}^n len(SP(p, v_j)) \tag{2}$$

If there is a point $q \notin V$ in space, q' is the point when q is mapped to its nearest edge $e_i \in E$. Average distance from point q to any road r_i is the sum of the distance $Dist(q', p)$ between point q' and its nearest point and the average distance between point p and road r_i , which is defined by formula 3 as follows:

$$avgDist(q, r_i) = dist(p, q') + \frac{1}{n} \sum_{j=1}^n len(SP(p, v_j)) \tag{3}$$

Combined with the importance level of road context, the calculation method of the influence level of the road context is defined as follows:

Definition 2.5. Any point p in the space is subject to the influence force of road context $r_i \in R'$, (R' is the set of road context objects shared by users) which is denoted as $IF_{r_i}(p)$. Its size is defined by formula 4 as follows:

$$IF_{r_i}(p) = IL(r_i) * \exp(-\lambda * avgDist(p, r_i)) \tag{4}$$

The sum of the influence force of the road context shared by all users at point p is:

$$IF_{R'}(p) = \sum_{r_i \in R'} IF_{r_i}(p) \tag{5}$$

It is observed that the influence force is proportional to the importance level of the road context and inversely proportional to the average distance between point p and r_i . As for the distance attenuation function $\exp(-\lambda * avgDist(p, r_i))$, parameter λ is the experience parameter, the general range of values is in the interval of [0.001,0.002]. Greater influence force of a point indicates that its location is representative in users' daily activities, and it is more closely related to users' potential active regions, vice versa. Therefore, if these representative points are clustered by a certain method, the corresponding user active regions could be extracted.

3. Road-Context-Based Active Region Extraction Algorithm

Active region extraction is divided into three steps. 1) Division of geographical space based on grid. 2) Extraction of active regions in each effective grid through relevant algorithms. 3) Combination of the results so as to get the final set of users' active regions.

3.1. Grid Division

Before extracting the set of user active regions, the whole geographical space is divided into a number of first level grid regions, which are denoted as $SG_{m \times n} = \{G(i, j) | 1 \leq i \leq m, 1 \leq j \leq n\}$. Depending on the difference in the city of the objects and the actual situation of the shared road location, the unit length of the grid area can range from 0.02 to 0.05 latitude and longitude coordinates. Here, the interval based on the latitude and longitude coordinates takes into account that the road network is generally stored in latitude and longitude format in spatial database.

Afterwards, each first level grid region G is divided into $k \times k$ sub-grids, which are denoted as $G_{k \times k} = \{g(i, j) | 1 \leq i \leq k, 1 \leq j \leq k\}$. The unit length of each sub-grid is 0.001 latitude and longitude coordinates interval, and the actual corresponding length is about 110 meters. The unit length can be used to achieve better calculation accuracy and reduce the amount of calculation.

3.2. Active Region Extraction Algorithm based on Road Context

It is the most important to determine the region range to extract the active regions. In order to calculate the range of active regions, the concept of grid density is introduced. For each sub-grid, the size of the influence force of the road context is its density, which is defined as follows:

Definition 3.6. For a grid $g(i, j) \in G_{n \times n}$, where $1 \leq i \leq n, 1 \leq j \leq n$, its four vertices are assumed as $V' = \{v_{tl}, v_{lb}, v_{rb}, v_{rt}\}$, and its $density(g(i, j))$ is the average value of the influence force on the four vertices of the road context, which is defined by formula (6):

$$density(g(i, j)) = \frac{1}{4} (IF_{R'}(v_{tl}) + IF_{R'}(v_{lb}) + IF_{R'}(v_{rb}) + IF_{R'}(v_{rt})) \quad (6)$$

Afterwards, the Road-based Active Region Extraction algorithm (RAREA) is described. Firstly, RAREA calculates the threshold value δ , and its function is to find the grid regions with larger density. Formula 7 defines its calculation method:

$$\delta = (Max(density(g)) + Min(density(g))) / 2 \quad (7)$$

For all $g(i, j) \in G_{n \times n}$, if its $density(g(i, j))$ is greater than the threshold value δ , and it is a maxima, it will be added into a maximum priority queue Q . In the meantime, the sub-grid density smaller than the threshold value δ is set as zero. The maxima refer to the density of the sub-grid is not less than the density of the other sub-grid in eight directions. After that, all elements in Q are conducted with the following operations successively:

1. Dequeue of the element with the greatest priority (namely the grid object with the largest density in the current queue), which is assumed as $g'(i, j)$. If it has not been accessed, it will be added to the Visited list to implement the next step. Otherwise, the first step is roll-back until Q is empty
2. Add the grid neighbors along the four different directions (up, down, left, right) of $g'(i, j)$ of successively. The average density of the grid regions after the addition is calculated. If the average density value drops rapidly after adding a new network neighbor, the addition of new grid is cancelled. In the meantime, the calculation in this direction is cancelled. Take the left direction as an example (similar to the calculation mode in other directions), if $avgDensity(G') - avgDensity(G'') > \theta$, the calculation in this direction will be cancelled.

Where, G' and G'' represents the sets of sub-grids before and after adding new grid neighbors in this direction. $avgDensity(G')$ and $avgDensity(G'')$ are reckoned into the average density values respectively. θ is the descendent factor of average density values. Based on θ , the grid neighbors with the closest density value can be found around the grid with the highest density value. Therefore, the grid region with the highest density value can be formed so as to find the road range with the most significant influence force. Formula 8 defines the default calculation method of the descendent factor θ :

$$\theta = \frac{\delta}{\max(density(g(i, j))) - \min(density(g(i, j)))} \quad (8)$$

3. When the calculation is completed in the four directions, four sub-grids in which the calculations are stopped in each direction are recorded, and assumed as $g_{left}(i, j-l)$, $g_{right}(i, j+r)$, $g_{top}(i-t, j)$ and $g_{bottom}(i+b, j)$ respectively. Therefore, the minimum bounding rectangle (MBR) covering these four grids is a candidate active region, which is added into the set of candidate active regions (CAR).
4. Finally, after obtaining the set of CAR, it is needed to carry out the intersection judgment of rectangles. The intersection of rectangles in the plane includes two kinds of cases. As for the judgment of the intersection between each two candidate active regions, if there is an intersection, they will be merged. The new active region after the merger is the minimum bounding rectangle covering the two active regions. The intersection judgment is repeated until none of the regions within CAR intersect with each other. The returned CAR is the set AR of active regions in this grid region.

Pseudocode of RAREA is as follows:

Algorithm Road-based Active Region Extraction

Input:

$G_{n \times n} = \{g(i, j) | 1 \leq i \leq n, 1 \leq j \leq n\}$: Set of grids, each grid has the density value

$density(g(i, j))$

Output:

AR : Set of active regions extracted

Algorithm:

- 1: $Q = \emptyset$; //A max priority queue whose member is grid and the priority of each element is the density value
 - 2: $Visited = \emptyset$; //A list which records the visited grids
 - 3: $CAR_G = AR_G = \emptyset$; //Lists which record the candidate active regions and active regions respectively in grid G
-

```

4: Calculate the threshold  $\delta$ ; // formula 4-7
5: Calculate the descendent factor  $\theta$ ; // formula 4-8
6: for each ( $g(i, j) \in G$ ) do
7:   if  $density(g(i, j)) \geq \delta$  and  $density(g(i, j))$  is a maxima then
8:     add  $g(i, j)$  to  $Q$ ;
9:   else
10:    set  $g(i, j) = 0$ 
11:   end if
12: end for
13: while  $Q$  is not empty do
14:    $q = Q.pop() = g(i, j)$  is first element of  $Q$ ;
15:   if  $q \notin Visited$  then
16:     add  $q$  to  $Visited$ ;
17:     for each  $direction \in \{left, right, top, bottom\}$  do
18:       while  $(avgDensity(G') - avgDensity(G'')) \leq \theta$  do
19:         set  $G' = G''$ ;
20:         add new neighbor grid  $g$  into  $G''$ ;
21:       end while
22:       mark left_index, right_index, bottom_index and top_index;
23:     end for
24:      $mBox = MBR(left\_index, right\_index, bottom\_index, top\_index)$ ;
25:     add  $mBox$  to  $CAR_G$ ;
26:   end if
27: end while
28: while foundIntersection do
29:   set foundInsection = false;
30:   for each pair ( $CAR_G[i], CAR_G[j]$ ) in  $CAR_G$  do
31:     if  $CAR_G[i]$  intersects with  $CAR_G[j]$  then
32:       set foundInsection = true;
33:        $CAR_G[i] = MBR(CAR_G[i], CAR_G[j])$ ;
34:       remove  $CAR_G[j]$  from  $CAR_G$ ;
35:     end if
36:   end for
37: end while
38:  $AR_G = CAR_G$ ;

```

6. Experiment and Analysis

In order to verify the accuracy and efficiency of the active region extraction algorithm based on the road context and the social relation recommendation algorithm based on active region, a series of experiments are designed. For the active region extraction algorithm based on the road context, it is the basis of the social relation recommendation algorithm based on active regions. The strengths and weaknesses of its efficiency have a direct impact on the accuracy of the recommendation. Therefore, the overall efficiency of the algorithm, the impact of the grid region division and the accuracy of extracting the active regional are assessed. As for the social relation recommendation algorithm based on active region, it is assessed in the aspect of overall efficiency of the algorithm.

Experimental environment and data

The experimental computer running the algorithm is a DELL Optiplex host. Its hardware configuration is as follows: CPU Core 2 Duo 2GHz, RAM 2048MB. All

algorithms are written in Java. Background relation database system is Postgre9.1, and GIS database is Postgis1.5. All road data sets in the experiment are collected from Chinese road data set provided by the open street map organization. A total of 352234 road data records are collected. Road context is cited from the data of 20 Sina Weibo users, including a total of 745 road data records.

Overall performance

Firstly, the overall performance of the road-context-based active region extraction algorithm (RAREA) is assessed. The default method of grid division is adopted. That is, the unit length of each first level grid region is 0.03 latitude and longitude coordinate interval. The unit length of each sub-grid under the first level grid region is 0.01 latitude and longitude coordinate interval. The correspondingly actual lengths are about 3000 meters and 100 meters. 100 active sub-grids (sub-grid with road context) are randomly extracted among those collected among the data of 20 Sina Weibo users. 5,10,15 and 20 pieces of road context are adopted to test the efficiency of RAREA respectively, which are averaged after the calculation. Figure 5, shows the execution time of RAREA.

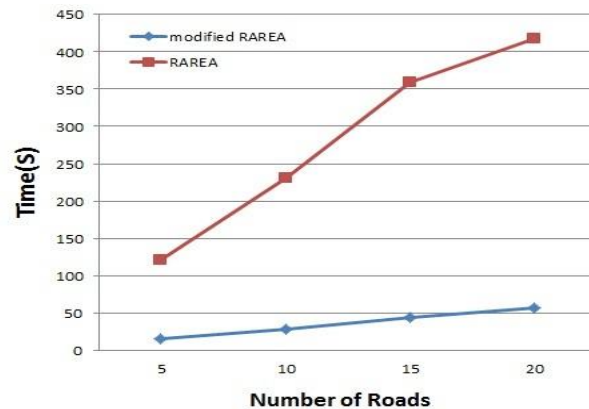


Figure 5. Efficiency of RAREA

As shown in the figure, execution time of RAREA is relatively long, and the efficiency is low. This is for that when the influence force of road context is calculated, the RAREA adopts the classic Dijkstra shortest path algorithm. Each time will spend about 0.002 to 0.003 seconds. There are about 31×31 grid divisions in the sub-grid regions. For any road location information, at least 961 times of the shortest path algorithm needs to be calculated, namely the least time consumption is 1.922 seconds. Therefore, the running time of RAREA is influenced by the number of roads and the number of points in the road. Moreover, RAREA utilizes the average shortest distance when calculating the influence force of the road context. In order to improve the efficiency of the RAREA, it is adjusted to utilize the distance of the central points of the road. That is, the shortest path is calculated for once, which can effectively reduce the number of calculation of the shortest path. The adjusted RAREA is named as modified RAREA. Although the running time of the modified RAREA has been significantly improved compared with the RAREA, it still needs relatively long running time.

Since the vertices of any sub-grid are fixed after the grid division, the shortest paths to any roads are also fixed. Therefore, the shortest distance between all vertices of the sub-grids and all road data can be calculated, and the results are stored. The calculation of the influence force of the road context only needs to endow the results of pretreatment with different weights, which can reduce the running time of RAREA to the maximum extent. The running efficiency of the RAREA after preprocessing is shown in Figure 6. It can be

seen that the efficiency of RAREA has been greatly improved since it omits the time of calculating the shortest path.

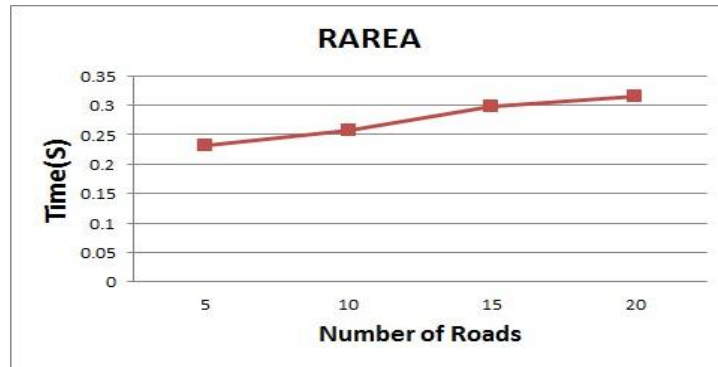


Figure 6. Efficiency of RAREA after Preprocessing

Influence and accuracy of grid division

The default unit length of the first-level grid region is 0.03 latitude and longitude coordinate interval. The default unit length of each sub-grid is 0.03 latitude and longitude coordinate interval. The unit lengths of each first-level grid region are adjusted, which are in the unit intervals of 0.01, 0.03, 0.06 and 0.1. RAREA is utilized to extract the set of active regions of 20 Sina Weibo users, and the changes in the size of the active region set is shown as Figure 7.

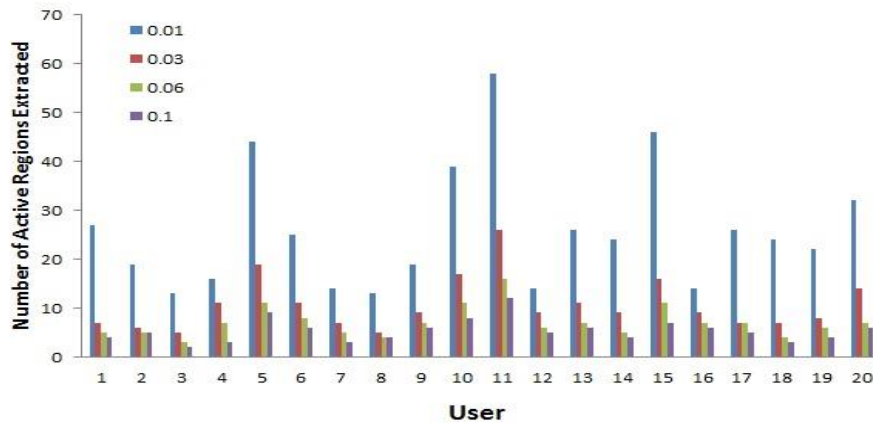


Figure 7. Number of Active Regions Extracted under Different Grid Unit Lengths

It can be seen that the number of the extracted active regions generally decrease with the increase of the unit length of the first-level grid. This is for that if a region with dense road context is segmented during the division of the first-level grid, it is likely to generate effective active regions in different grid division, which results in the errors of the calculation results. Therefore, the division of the first-level grid is likely to affect the final extraction results of RAREA. However, as for the grid with longer unit length, the complexity of the processing and the cost of algorithm will be significantly increased.

6. Conclusion

Road context is a pattern manifesting location context that maps the user's location to the nearest road. It could reduce the granularity of information so as to achieve the purpose of protecting the privacy of users' locations. In view of the situation that the

traditional location context mining technology does not support the road context, this chapter proposes an algorithm that could extract users' active regions based on the road context. The set of users' active regions could be extracted by measuring the importance levels of different road context and establishing models to analyze the intrinsic link between the road context and its potential active regions.

Acknowledgements

This research was sponsored by Shanghai Natural Science Fund (No.14ZR1429800, 15ZR1430000) and Young University Teachers Training Plan of Shanghai Municipality under Grant (No.SXY12003).

References

- [1] B. R. Karki, A. Hamalainen and J. Porras, "Social networking on mobile environment", In: Proc. ACM/IFIP/USENIX Middleware Conf. Companion, (2008), pp. 93–94.
- [2] J. Z. Gu, L. He, J. Yang and Z. Lu, "Location aware mobile cooperation design and system", In: Int'l Journal of Signal Processing, Image Processing and Pattern Recognition, 2009, 2(4):49-60.
- [3] N. Kayastha, D. Niyato, P. Wang and E. Hossain, "Applications, Architectures, and Protocol Design Issues for Mobile Social Networks", A Survey. In: Proc. of the IEEE, vol. 99, no. 12, (2011) December.
- [4] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar and L. Terveen, "Discovering personally meaningful places: An interactive clustering approach", In: ACM Trans. Inf. Syst., vol. 25, (2007), pp. 12.
- [5] J. Hightower, S. Consolvo, A. LaMarca, I. Smith and J. Hughes, "Learning and Recognizing the Places We Go", In: Proc. Ubicomp. Tokyo, Japan, (2005).
- [6] J. Cranshaw, E. Toch, J. Hong, A. Kittur and N. Sadeh, "Bridging the Gap Between Physical Location and Online Social Networks", In: 12th International Conference on Ubiquitous Computing. Ubicomp (2010).
- [7] Y. Zheng, L. Zhang, Z. Ma, X. Xie and W. Y. Ma, "Recommending friends and locations based on individual location history", In ACM Transaction on the Web, ACM, (2011).
- [8] E. Cho, S. A. Myers and J. Leskovec, "Friendship and mobility: User movement in location-based social networks", In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), (2011).
- [9] D. Wang, D. Pedreschi, C. Song, F. Giannotti and A. L. Barabási, "Human mobility, social ties, and link prediction", 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'11), (2011).
- [10] Y. Zheng and X. Xie, "Learning travel recommendations from user-generated GPS traces", ACM Transaction on Intelligent Systems and Technology, ACM, (2011) January.
- [11] H. Yoon, Y. Zheng, X. Xie and W. Woo, "Social Itinerary Recommendation from User-generated Digital Trails", Personal and Ubiquitous Computing, Springer Verlag, (2011).
- [12] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar and L. Terveen, "Discovering personal gazetteers: An interactive clustering approach", ACM GIS, (2004).
- [13] V. W. Zheng, Y. Zheng, X. Xie and Q. Yang, "Collaborative Location and Activity Recommendations With GPS History Data", ACM WWW, (2010).
- [14] Y. Zheng, L. Zhang, X. Xie and W. Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories", ACM WWW '09 , (2009), pp. 791–800.
- [15] F. Giannotti, M. Nanni, F. Pinelli and D. Pedreschi, "Trajectory pattern mining", KDD, (2007), pp. 330-339.

Authors



Rong Tan, Dr. Tan is an instructor in department of information and computer science of Shanghai Business School. His research area includes Spatial-Temporal Database, NoSQL and Computer Networks.