International Journal of Database Theory and Application
Vol.9, No.4 (2016), pp. 135-142
http://dx.doi.org/10.14257/ijdta.2016.9.4.12

Precision Advertising Based on the Scene of Trajectory Mining

Yun Wu, Feng Yang*, Min Ren and Le Han

Shandong University of Finance and Economics, China
wuyunjy@163.com

Abstract

The value of offline data is continuously realized due to the rapid development of GPS, GSM and Wireless Sensor Technology. The Mobile Internet has become more common and with the advancement of technology, people have become more dependent on the mobile intelligent terminal. The Advertising model is influenced subtly by changes to the audiences' lives and online patterns. Location Based Social Networking Services promote the precision advertising model into being, which is based on mobile internet. This paper used the methods of trajectory mining and POI clustering to create activity mapping table based on the users' location track data, then established user behavior feature model and built an advertising platform to implement precision advertising.

Keywords*: Data, Location Based Social Networking Service, Mobile internet, Precision advertisement*

1. Introduction

The Location Based Social Networking Service [1], which is based on the SNS (Social Networking Service), improves people's interactive modes on the internet platform and shortens the effective distance between people by means of LBS (Location-based Service). Meanwhile it brings people's online activities more in tune with their lives, in turn serving to improve their lives. With the rapid development and extensive use of GPS, GSM and Wireless Sensor Technology, people can more easily collect a large amount of real-time position data that records their latitude and longitude coordinates in a certain period of time. As smartphones and mobile devices usage spreads, the uses of the two factors in precision advertisement: people and position, gradually become more prominent.

In the commercial era, the advertisers are perfecting the advertisement which is constantly driven by the goal of profit maximization in order to achieve the precision injection. The netizens' Internet-enabled devices are gradually transferred from computers to smart mobile devices. It is not enough to analyze the user's interests just by relying on their browsing history, but more is needed to combine the user's location (offline data) to explore the trajectory model and excavate a richer life interest. Rao and Minakakis once said that a great business opportunity [2] is contained in the application of the LBS----the service of precision advertising delivery based on the analysis of the location in 2003.

This paper, based on the user's trajectories, adopts an improved trajectory analysis algorithm to extract the stay points, then scenarize the trajectory, analyze the user's behavior characteristics to create and improve the activity mapping table and user behavior database. We further explore the precision advertising based on the user's geographical location with the aid of the existing advertising technology and the network application platform. The GPS trajectory data research is performed under the premise of user's awareness and authorization and without invading privacy.

ISSN: 2005-4270 IJDTA

2. Research Status

In the study of GPS trajectory extraction for the users, D. Ashbrook extracted the stay points from the original data and adopted time as a measure [3]. Yu Zheng brought the method of stay point using distance factor based on the time [4-5]. Ran Xue separated the beginning and the end of the trajectory and then combined them according the conditions found by extracting the stay point in the middle [6]. However, all above algorithms are inaccurate as they all ignored the situations where the speed may change and the track point may be lost in the non-static state. The research of user tracks is used more for interests mining, hot spots finding, and traffic management applications, crime tracking and tourism recommendations.

Currently, internet advertising generally uses content targeting, IP exclusion, category exclusion, keywords negation and other techniques for advertising. Directional advertisement analyzes the user's characteristics of surfing online and finds the interests by obtaining the cookie and the website to advertise it to the people who need it. The results of researches of C. Wang *et al.* show that advertisements which aim at users' interests can improve click rates and the user's experience [7]. P. Chatterjee *et al.* also agreed that the more targeted advertisement got, the more effective it will be [8]. At present, the research of accurate advertising is based on the method of directional advertising although most are based on the online data. Without further study of users' interests, the huge user information has not been used and the full effects of advertisement injecting for maximum profits have not been realized.

The cooperative relation between the user's accurate activity based on the geographic position track and the precision advertisement has not been deeply studied. Therefore it is a valuable subject that lies in the online lives changing from computers to mobile intelligent terminals and the development of social communication style and mobile localization technology.

3. The Mechanism of Precision Advertising Based on the Scene of Trajectory Mining

Figure 1 shows the system of precision advertising based on the scene of trajectory mining flow chart. The behavior of the user's location contains abundant semantics and we can find the user's interests and information such as the hot area combining with the real life scenes of location. The processing of trajectory scenarios changes the relationship between advertising and consumers from "one-to-many" to "one-to-one".

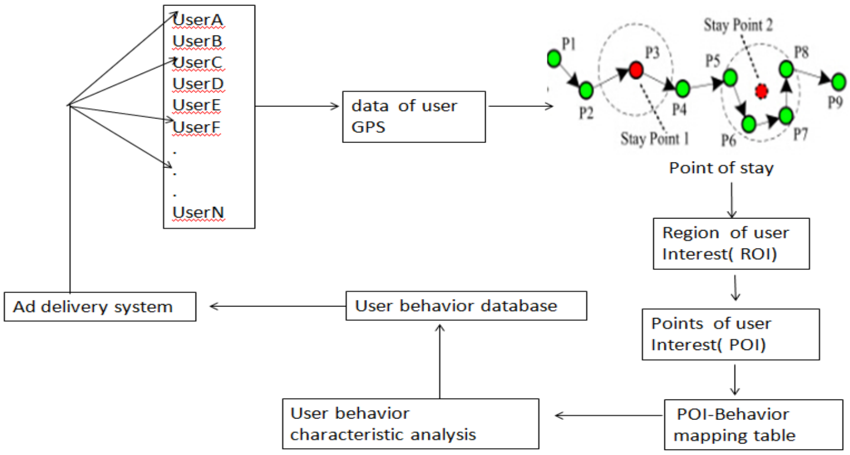

Figure 1. The System of Precision Advertising Based on the Scene of Trajectory Mining

3.1 The Extraction and Clustering of Stay Points

Usually each GPS point includes longitude, dimension and a time stamp, expressed as triples $p_i=(x_i,y_i,t_i)$. A series of chronological GPS points form a GPS trajectory, expressed as $Traj=p_1\text{--}p_2\text{--}\ldots\text{--}p_n$ (meet $p_i.T<p_{i+1}.T$).

The distance between two GPS point is calculated with the Haversine formula [9]. As GPS points are large enough, first the stay points are extracted from the trajectories using the stay point detection algorithms a large number of scholars put forward. In consideration of emergencies or a change of the velocity in a circular region, we propose a new stay point extraction algorithm.

The principle of this algorithm:

(i) We automatically set the first and the last point of a trajectory as stay points,

(ii) Starting from the second point of a trajectory, within a certain region the times of the speed of two adjacent points faster than the maximum speed of human walking is less than the speed threshold Speed Threth. The distance between the user' s first GPS record point and other points is less than the distance threshold dist Treh. At the same time, the time interval between the user' s first GPS record point and the last point is greater than the time threshold time Threth. In this way the three conditions are satisfied and the average of the points records to a stay point;

(iii) Merge the point that we get in step i and the stay point that we get in step ii.

(iv) Calculate the distance between the first point and its close stay point and the distance between the last point and its close stay point. If the distance is less than the combined threshold d (distance threshold), we merge them. Conversely we set as a single stay point.

After we obtain the stay points, we use a clustering algorithm to make the stay points regionalized, and get the user's interest area. The most frequently used clustering algorithms are DBSCAN [10], DENCLUE [11], OPTICS [12] and so on. In order to more accurately cluster the users' stops, we use the clustering algorithm called Clustering by fast search and find of density peaks [13] proposed by Alex Rodriguez and Alessandro Laio. The density of the clustering center itself is greater than its neighbors surrounding it. At the same time the distance between one clustering center and the other clustering center is relatively larger.

The principle of this algorithm:

(i) Calculate the distance d_{ij} between any two points p_i and p_j ;

(ii) The formula of density of the point is

$$\rho = \sum_{j \in I_S\{i\}} \aleph(d_{ij} - d_c) \tag{1}$$

and

$$\aleph(x) = \begin{cases} 1, x < 0 \\ 0, x \geq 0 \end{cases} \tag{2}$$

Where: d_c is truncated distance, according to the characteristics of the data to be set. According to this formula, the density of the points p_i values can be understood as the number of points whose distance from the point in the P to p_i is less than d_c.

(iii) The distance σ_i

$\{q_i\}$ is the descending order index of $\{\rho_i\}$,namely

$\rho_{q1} \geq \rho_{q2} \geq \ldots \geq \rho_{qN}$

so

$$\delta_{qi} = \begin{cases} \min\{d_{q_i q_j}\}, i \geq 2 \text{ and } j < i \\ \max\{\delta_{qj}\}, i = 1 \text{ and } j \geq 2 \end{cases} \tag{3}$$

(iv) In a range the point with larger ρandσ value is the clustering center and around the point and area defines the interest area.

3.2POI-Active Mapping Table

After the stay point clustering, we can get the user's activities in a particular area and their stay time. Then exploring the "area users stayed matching with natural geographical area and stayed a certain period of time", we can infer that stay area is the user's interest point and also infer the potential user activities. In order to more accurately analyze likely social activities that users partake in at the interest point, we combine the user's location with stay time to make a user trajectory scenario. Then we do practical survey statistics to get the time that the general public does all kinds of activities accounting for general items such as dinner, outings, KTV, gym, *etc*. And we establish point of interest- activity mapping table (PAM).

A POI-Active Mapping Table is created according to area classification that the map data provider shows. Generally the classification of map data has two levels, the large category does not have more than 20 kinds and the small category has hundreds of class segments. Through a lot of questionnaires, we obtain the average time of public daily life activities recording the data in a POI-Active Mapping Table as shown in Table 1.

Table 1. Poi-Active Mapping Table

POI Category	POI Sub-category	Behavior	Minimal Duration	Maximal Duration	Average time
Catering	Chinese Restaurant	Eating&Drinking	30min	4h	0.999h
	Western Restaurant		30min	4h	1.18h
	Chinese Fast Food		15min	3h	0.781h
	Café		15min	3h	1.267h
	Tea house		30min	4h	1.398h
	…		…	…	…

Leisure & Entertainment	Cinema KTV Video Game Room	Seeing films&Playing	1h 1h 30min	4h 8h 4h	1.89h 2.179h 1.421h
Scenic spot	Park Memorial Museum Tourist attractions	Playing&Visiting	30min 15min 30min	4h 4h 5h	1.825h 1.404h 2.301h

3.3Analysis of Users' Behavior Characteristics

Deep investigation of the trajectory not only contributes to the analysis of the methods of travel (such as walking, biking, driving) of users within a certain areas [14], but also determines the daily life hobbies, individual behavior patterns and common characteristics of user groups [15]. It can even predict the moving targets' behavior at a following time [16].

In daily life, people's behavior and activities represent a certain sense, which can reflect personal preferences, needs or what someone is about to do. According to the analysis of user events, combed with the user's location and the stay time, the behavior feature model of users can be built after creating POI--activities mapping table. The processed trajectory data of users can be used on the feature model to match the user behavior. The association rules will be found by correlation analysis of a users' interest and another interest activities and time series analysis is used to analyze the behavioral pattern of user. For example, the Hugetable system of Wuhan Borqs is a good platform for user behavior analysis. Obtaining user behavior information [17] is the key factor for us to advertise and find the target audience. Also user behavioral characteristics analysis is a necessary prerequisite to achieving accurate delivery of advertising.

3.4 Advertisement System

After the function and stability requirements analysis, we design the system architecture consisting of the data model level, system control layer and view layer. Based on the users' behavior database, using internet technology, program development technology and the traits of mobile intelligent terminal APP, we develop a mobile application advertising system consisting of the service and the clients [18]. Advertisers can place ads on the platform through client APP, and the server-side provide support to maintain the normal operation of the platform. When placing ads on the advertising platform, the similarity analysis between user behavior and advertising should be carried out first. Then the overall similarity score between user interest and advertising is calculated to find the audience's best matching with the advertising theme and finally present the ads in front of the audience [20].

4. Experimental Analysis

The GeoLife GPS Trajectories Data Set that The Microsoft Asia research institute opened to the outside world was used for the experimental data in this paper [19] [21]. The data set from 178 users involved in Geolife programs collected over four years. The trajectory of the GPS data set was made of a series of points that contain latitude and longitude with a timestamp, and the data set contains 18465 paths, more than the total length of 100000 meters. The GPS data once got an interval of 1 to 5 seconds from the GPS receiver or GPS phone records [22]. Most of the data from the data set was collected in Beijing and Figure 2 shows the distribution of the data set in Beijing. In the diagram the brighter the area the more points in the area. The thresholds of speed, time and distance were respectively set as 1.5 m/s, 30 min, and 200 m in this experiment, and we selected the data focused on a certain user as the experimental data. The picture Figure3 shows the different number of stay points obtained by different extraction algorithms. The “time” represents the stay point extraction algorithm using time as judgment criteria, and the stay points number is 39. Although this method can ensure accuracy, it can only be used to extract still points, and the stay point semantics gained is too monotonous. The “t-d” shows a time and distance criteria of residence extraction algorithm. The number of stay points is 2776. Despite obtaining more points by this method, the accuracy is lower. The “speed” extraction algorithm adds speed as the criteria in the second method, and the number of points it gave was 1922. This approach ignores the starting points, which also represent a certain user information, leading to the uncomprehensive extraction points. The “my-tds” extraction algorithm is used in this paper and the number of stay points is between the second and third methods, which can ensure a sufficient amount of information without losing accuracy.

In order to better understand the user's activities and analyze the user's interests, the clustering results are mapped to baidu map, then POI activities matching. Table 2 by average activity time from the questionnaire survey, we analyze real user consumption time, and statistics user consumption behavior in the POI subclass.

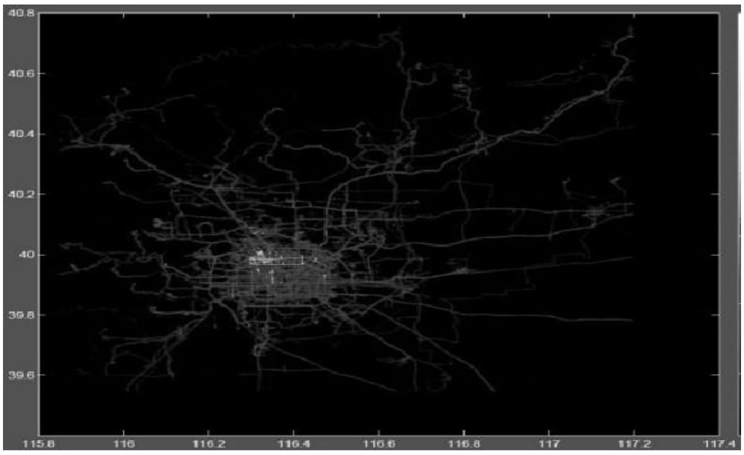

Figure 2. The Distribution of Data Set in Beijing

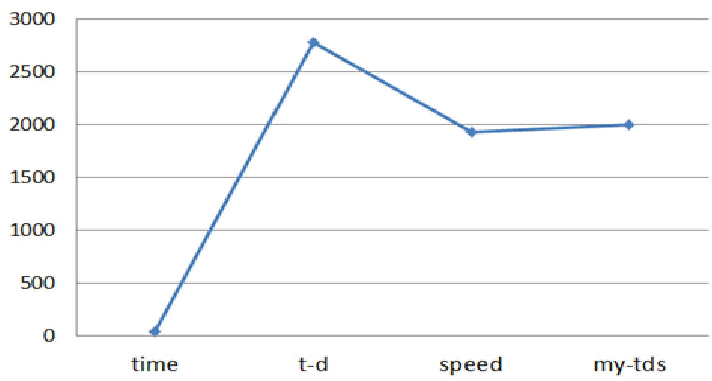

Figure 3. The Number of Stay Points

Table 2. Clusting and Matching

POI Category	POI Sub-category	Visit Times	Average Duration (h)
Catering	Chinese Restaurant	44	1.8290
	Western Restaurant	5	1.7181
Leisure & Entertainment	Cinema	10	2.8668
	KTV	6	3.9063
Shopping	Shopping Center	16	2.1583
	Supermarket	25	1.0396

5. Conclusions

Personal characteristics and life-loving activities can be determined from the raw trajectory data and by analyzing the user activity over an extended time period to define the user's daily behavior patterns. With cloud computing and the arrival of the big data era, it is essential to more effectively integrate and analyze more deeply the users trajectory data to obtain more knowledge and intelligence information and hence, increase user participation and improve the user experience. Therefore, we can provide users with more humane and more effective location-based services to help businesses improve the potential consumer intentions survey. In addition, we can provide more potential customers with personalized information and afford advertisers a more favorable precise advertising delivery platform, leading to a win-win situation between users and advertisers.

Acknowledgement

This research is supported by National Natural Science Foundation of China (NSFC) Grant No. 71172046 and MOE (Ministry of Education in China) Youth Fund Project of Humanities and Social Sciences (No. 13YJC630162).

References

[1] China LBSNS (mobile positioning social services) market prospects and Future Trends Research Report from 2014 to 2018, **(2015)**.
[2] R. Minakakis, “Evolution of mobile location-based services”, Communication of the ACM, vol. 46, no. 12, **(2003)**.
[3] C. Wang, P. Zhang, R. Choiand and M. D. Eredita, “Understanding consumers attitude toward advertising Eighth Amerieaseonf on Information System, **(2002)**.
[4] P. Chatterjee, D. L. H. Whlan and T. P. Novak, “Modeling the eliek stream: Im Plieations for web-based advertising efforts”, Marketing Science, vol. 22, no. 4, **(2003)** , pp. 520-541.
[5] D. Maier, “The complexity of some problems on subsequences and super sequences”, J.ACM, vol. 25, no. 2, **(1978)**, pp. 322-336.
[6] A. Rodriguez and A. Laio, “Clustering by fast search and find of density peaks”, Science, vol. 344, no. 1492, **(2014)**.
[7] L. Liao, D. J. Patterson, D. Fox and H. Kautz, “Building Personal Maps from GPS Data. Annals of the New York Academy of Sciences”, vol. 1093, no. 12, **(2006)**, pp. 249-256.
[8] Y. Zheng, L. Zhang, X. Xie and 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)**, pp. 472-475.
[9] D. Ashbrook and T. Starner, “Using GPS to Learn Significant Locations and Predict Movement across Multiple Users”, Personal and Ubiquitous Computing, vol. 7, no. 5, **(2003)**, pp. 275-286.
[10] L. Jaewan, M. A. Romeo and M. D. BOBBY, “Location-aware agent using Data Mining for the Distributed Location-based Services”, Computational Science and Its Applications Iccsa, **(2006)**, pp. 867-876.

[11] V. Murdoek, M. Ciaramita and V. Plachouras, "A Noisy Channel Approach to Contextual Advertising", Proceedings of the lst International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD.07), **(2007)**.
[12] D. Ashbrook and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users", Personal and Ubiquitous Computing, vol. 7, no. 5, **(2003)**, pp. 275-286.
[13] Y. Ye, Y. Zheng, Y. Chen, J. Feng and X. Xie, "Mining Individual Life Pattern Based on Location History", In proceedings of the International Conference on Mobile Data Management, **(2009)**, pp. 1-10.
[14] Y. Zheng, L. Zhang, X. Xie and W. Y. Ma, "Mining Correlation Between Locations Using Human Location History", In Proceedings of ACM SIGSPATIAL Conference on Advances in Geographical Information Systems, **(2009)**, pp. 352-361
[15] M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A Density-Based algorithm for Discovering Clusters in Large Spatial Databases", in KDD'96, Portland, OR, **(1996)**, pp. 118-127.
[16] A. Hinneburg and D. Keim, "An Efficient Approach to Clustering Large Multimedia Databases with Noise", New York: Proceedings of the 4th ACM SIGKDD, **(1999)**, pp. 58-65.
[17] M. Ankerst, M. M. Breunig, H. P. Kriegel and J. Sander, "OPTICS: Ordering Points to Identify the Clustering Structure", In Proceeding ACM SIGMOD Int'l Conf. on Management of Data, **(1999)**, pp. 49-60.
[18] Y. Zheng, L. Zhang, X. Xie and W. Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories", In Proceedings of International conference on World Wild Web, **(2009)**, pp. 791-800.
[19] L. D. Wolin and P. Horgaonkar, "Web advertising: Gender differences in beliefs, attitudes and behavior", Internet Research, no. 13, **(2003)**, pp. 30-33.
[20] Y. Zheng, X. Xie and W. Y. Ma, "GeoLife: A Collaborative Social Networking Service among User, location and trajectory", Invited paper, in IEEE Data Engineering Bulletin, vol. 33, no. 2, **(2010)**, pp. 32-40.
[21] Y. Zheng, Q. Li, Y. Chen, X. Xie and W. Y. Ma, "Understanding Mobility Based on GPS Data", In Proceedings of ACM conference on Ubiquitous Computing, **(2008)**, pp. 312-321.
[22] R. Xue, "Construction and comparison of user mobility attributes based on geographic location services", Master Thesis of Shandong University, **(2013)**.

