© 2017 Institute of Thermomechanics CAS, v.v.i.

A study on data characteristics of mobile user

Bing Han[2], Yinglan Fang[2,3], Yong Yang[2],
Yongmei Zhang[2]

Abstract. With the development of communication technology, mobile data has become quite large. It provides personalized service to mobile users by extracting mobile user data characteristics. The paper has analyzed mobile user data and established the sparse characteristics model and the dense characteristics model of mobile data based on the sparse density of users' geographical location information. The mobile user data characteristics model is implemented on the server side and the personalized server information is pushed to the client. The mobile user data characteristics are displayed on the client side so that the user can check the individual data characteristics. Mobile users can be recommended to clients based on the design of mobile data sparse characteristics models and dense characteristics models. At the same time, mobile users can select different conditions to check their activity area.

Key words. Mobile users, data characteristic, sparse characteristic, dense characteristic.

1. Introduction

Researchers have paid more attention to the analysis and research on the characteristics of mobile users' data. In order to analyze the characteristics of users, the paper [1] proposed a digital map resource retrieval mode based on Web Services and designed a unified search result fusion algorithm and meta data update algorithm of digital map resources based on Web Services. Paper [2] located data by making mobile users ' phone and gave the extraction steps of mobile users and identification of key mobile users' active areas and classification of mobile users 'active areas. Paper [3] used the resident travel data in Beijing to construct the OD distribution and population distribution cloud evolution of residents. Paper [4] proposed

a new method to control robotic walking support system that is adaptive to user characteristics. Paper [5] compared several representative methods, such as Fab [6], Siteseer [7], and LAW [8] that simulate user preferences and construct reliable and efficient intelligent information search systems. These methods have used the user interest characteristics analysis technology. Paper [9] proposed a scheme by analyzing typical cell phone call data, finding the characteristics of mobile user data and other characteristics of mobile users. According to the characteristics obtained by the mobile user data, it could mine deeper and more valuable information and apply high-level applications. Paper [10] proposed an adaptive approach of user interest analysis, which uses a dynamic update strategy and takes full account of user preferences changing with time and scope.

2. Related theories and technologies

2.1. Mobile user data characteristics classification

User data characteristics will be divided different division types according to different classification criteria. Different classification standards represent different meanings and the characteristics of mobile users obtained are also accordingly different. As mobile devices are free of time and place, mobile data becomes more complex. Mobile users change with time or place, and their mobile data is closely related to time and location. Therefore, the characteristics of mobile user data can be divided into static and dynamic characteristics based on time and place attributes.

Mobile data static characteristics mainly refers to the data characteristics of mobile users themselves, such as the mobile user's name, gender, occupation, ID number and so on. These attributes will not change in the short term and change the current mobile user characteristics. Therefore, mobile user data characteristics can be described based on these attributes. Static characteristics of mobile users can be used as the basis for the study of data features, there is a basic characteristics analysis of mobile data.

2.2. Mobile user data characteristics usage

Compared with the traditional user data, mobile data is more abundant type and larger scale and more value. For mobile data analysis, different mobile user data characteristics have different uses. As the most important use of mobile user data characteristics research personalized services includes personal recommendation, personalized information retrieval and personalized site. Based on the purpose of providing mobile users with personalized products or services, mobile data characteristics can be better studied.

3. Mobile user data characteristics analysis

Mobile users can use mobile devices anytime and anywhere, and mobile data is generated immediately. By analyzing the mobile data, the characteristics of mobile data can be found and the data characteristics of mobile users can be extracted.

3.1. Mobile user data characteristics source

Mobile users in this paper refer to users who hold smart phones. When mobile users use smart phones with a slide-in App, it generated mobile data and analyzed the characteristics of mobile users and explored the preferences of mobile users. The business data generated when a user uses a sliding screen App. When mobile users use the slide App, the App application software automatically collects the information and the information generated by the user operating the smart phone, such as the phone number and gender and date of birth of the user registration App personal information. Personal information generated when a mobile user uses the App, such as the number of sliding screens of the mobile user, the number of slide pictures, the number of times the user clicks the link, and the like. Information is collected by the App application, such as other App information installed in the smart phone: App name, version, etc.

Mobile devices used by mobile users themselves. The information of the smart phone itself can also be used as the data source for mobile user data characteristics analysis. The information about mobile devices is type of mobile device and device serial number and mobile operating system version, etc. Mobile operator information in mobile devices is operator name and operator attribution and carrier code, etc.

Related content information recommended by mobile users is recommended pictures and picture links and integral value of pictures, etc. Classified information of the recommended content is category name and category coding and length of recommendation, etc. Recommended sequence information is sequence name and sequence picture details and image duration, etc. Recommended sequence information for mobile devices is mobile phone number and recommended serial number.

3.2. Mobile user data characteristics classification

According to the source of the characteristics of mobile user data, the mobile data in this paper will change with different time or place, It results the mobile data show sparse characteristics and dense characteristics. According to the sparse and dense degree of the mobile user's geographical location information, the mobile data presents sparse characteristics and dense characteristics.

Sparse characteristics. Mobile user data will be generated only after the mobile user is registered, and the data acquisition of the location will be generated slowly. Mobile data is sparse at this time. Sparse characteristic mainly displayed in the data type and quantity of the data. The mobile data is mainly compose of the basic mobile users information. It includes gender and age and etc. The content of the mobile user library information is the types of content and content details and mobile device

information and equipment serial number and device type and operating system version, etc. These mobile data constitute the sparse characteristics of mobile data.

Dense characteristics. The geographic location data of mobile users will increase with time, location and collection accuracy. According to the generation of mobile data to estimate the behavior characteristics of mobile users, the recommendations log is produced and related information by increasing the choice of recommendations and recommending the content of mobile users interest. The dense characteristics are mainly reflected in the location data of mobile users. It includes longitude, dimension, location time, mobile phone number and so on. The content recommendation logs of mobile users include recommendation information and recommended time and recommendation equipment number, etc. The sliding screen information of mobile users includes the number of sliding screen and sliding screen content and number of unlock. Sequence receive log includes receive device number and receive sequence and receive type, etc. Push error log information includes device number and error message and wrong type.

3.3. Mobile user data characteristics model research

Mobile users often display different mobile user access behavior when it access the location. Based on characteristics extracted mobile user data, mobile user can analyze the sparse and dense characteristics in sparse data and data intensive. It can also establish the mobile user data characteristics model and recommend personalized interest to mobile users.

Mobile data presents sparse characteristics as mobile data begins to emerge. At this time, mobile users have less location data and can not effectively personalize the interest points to mobile users. According to the existing data, the sparse characteristics of mobile users is extracted, and the sparse characteristics model is established, and the interest points are personalized to mobile users. In this paper, a sparse characteristics model of mobile data is established based on three sparse characteristics of gender and age and category. According to the model, the meta data stored in the server is first processed, namely the location semantic transformation. The semantic data is classified according to the class of interest. And the interest point category library is stored in the server. According to the sparse characteristics of the extracted mobile data which includes age characteristics and gender characteristics and category similarity characteristics, the characteristics model is calculated. Based on the comprehensive characteristic model calculation formula, the comprehensive model is calculated. According to the results of the comprehensive characteristics model, it select the individual interest point categories or interest points from the interest point category library for mobile users.

In the process of the mobile data accumulating over time or location, the data of mobile users' access location data slowly from sparse to dense. This shows that mobile users have access to a large number of locations, the number of mobile users is increasing. And mobile data is more reflective of users' access behavior and access characteristics. At this time, based on the current data characteristics, three representative characteristics of mobile data density are extracted: geographical location

characteristics, time characteristics and category characteristics. Based on the data density of the extracted mobile data, this paper establishes a mobile data intensive characteristics model.

Previous models only considered the key factors affecting interest recommended time characteristics and characteristics of geographical location and does not take into account the characteristic of the category of the point of interest. Or it has failed to give a comprehensive consideration the characteristics of three kinds of models. The recommended results are more accurate than the recommended results. According to the model, firstly the meta data stored on the computer is processed, and the data of the latitude and longitude position are converted into identifiable position semantics. Then the semantic data is divided according to the industry category, and the interest point category library is saved on the computer. According to the model formula, the model of geographic location characteristics model, time characteristics model and category characteristics model are calculated according to the comprehensive model. Finally, according to the final result of the calculation, it selects the individual interest point or interest point category from the interest point category library and recommends it to mobile users.

4. Design and implementation of characteristics model

4.1. Mobile data sparse characteristics model design

When the mobile data is sparse, the mobile data sparse characteristics model is designed based on the extracted mobile data sparse characteristics. The mobile data sparse characteristics model can solve the personalized recommendation problem when mobile data is sparse. When the mobile data is sparse, the mobile user's position can not be analyzed because the mobile user's basic data and location data are relatively few. Therefore, the mobile user's position data can not be analyzed to predict the mobile user's access behavior. This paper extracts the characteristics of age, gender and category similarity to establish a mobile data sparse characteristics model.

4.2. Mobile data intensive characteristics model design

According to the characteristics of data when mobile data is dense, extracting the specific characteristics that can reflect the characteristics of mobile data intensive. These characteristics reflect the behavior habits of mobile users and are especially important for personalized recommendation of mobile users. In this paper, the mobile user's geographical location characteristics, time characteristics and characteristics characteristics are extracted to design a mobile data-intensive characteristics model.

4.3. Mobile user data characteristics comprehensive model design

According to the mobile user data characteristics model, it can get personalized interest points when data is sparse and data intensive. The sparse characteristics comprehensive model solves the problem of personalized recommendation when data is sparse, which combines three characteristics: age, gender, and category. The dense characteristics integrated model has been designed to personalize recommendations when data is dense. It has took into consideration geographic location and time and category.

When the mobile data is sparse, this paper extracts three representative characteristics of age, gender and category similarity, and analyzes the impact of each characteristics on the mobile user access behavior. But the recommended point of interest is not very consistent with the mobile user's access requirements. In order to accurately recommend mobile users to point of interest, this paper designed a comprehensive model which takes into account the mobile user's age and gender and category similarity.

In the case of mobile data density, this paper takes geographic location, time and category three specific characteristics to build a mobile data intensive characteristics model to personalize the points of mobile users interest. To get a personalized recommendation not to mobile users access points of interest, it need to integrate these three methods - geography characteristics, time characteristics, class characteristics. Considering the three factors in the design of the model, the influence of the result of the recommended put forward a comprehensive model. Computation formula is as follows:

$$S(u, li) = fGeo(u, li) + ftim(u, t, li) + fca(u, c, li), \qquad (1)$$

$S(u, li)$ represents the comprehensive score of mobile user u for the non - accessed interest points. $fGeo(u, li)$ represents the geographical location characteristics correlation degree. ftim (u, t, li) represents the time characteristic correlation degree. $fca(u, c, li)$ represents the characteristic correlation. According to the recommended results obtained by the dense characteristic model, k points of interest with the highest degree of relevance are recommended to mobile users.

4.4. Mobile User Data Characteristics Model Implementation

The mobile user data characteristics model is the main module of the system, and its main function is to establish the mobile user data characteristics recommendation model based on the extracted data characteristics. According to different mobile data characteristics models, it selects individual interest points or interest point categories to be recommended to mobile users. This part mainly consists of data characteristics recommendation and display implementation and mobile user trajectory tracking implementation and mobile user activity area distribution.

According to the mobile user data characteristics model, it recommend personalized interest points to mobile users. In order to show the personalized interest point of recommendation to mobile users, this paper develops a characteristics of

mobile user data characteristics recommendation. This function is to display the recommended interest points or interest point categories to facilitate mobile user viewing. Mobile users can view the recommended points of interest and click on the point of interest, It can view the details of the point of interest, such as address, semantics and interest points of the distance between the current position and mobile users. The function select different mobile user characteristics models to recommend personalized interest points to users based on the sparse density of location data of mobile users. The number of location data of the user (156****5648) is 5240, so the user sparse characteristics model is adopted to recommend the personalized interest points. It has shown the personalized point of interest and point of interest category information of the users (156 **** 5648) near the North China University of Technology. Then click on the "South Gate" point of interest, it can see the user has visited the point for 21 times and with 188 meters from the user's current location and other information. The recommended results for mobile users (156 **** 5648) at the current and current time are in keeping with the mobile user's personal interests and reflect the data characteristics of the mobile users.

It has shown the personalized point of interest and point of interest category information of the users (188 **** 8785) near the North China University of Technology at 2 o'clock in the afternoon. Then click on the "North China University of Technology Service Building" point of interest, it can See the user has visited the point of interest for thirty times. The point of interest near 22 meters from North China University of Technology service building. Mobile users (188 ***** 8785) at the current location and the current time of the recommended results are different. Objectively it shows that different mobile users interest in their hobby is different. And it shows the choice of different mobile data characteristics Model, it can get different recommended results.

Mobile user trajectory tracking can track the movement of mobile users to analyze the movement behavior of mobile users, as well as the tracking behavior of mobile users seeking interest points. Mobile user trajectory tracking includes real-time trajectory tracking and historical track view. The real-time trajectory tracking of mobile users can see the current location of the mobile user, as well as the current trajectory. The history trace can look at the track of the selected date based on the date of the selection. When the mobile user clicks "Start", real-time track tracking will start. When the mobile user clicks "stop", then the real-time tracker will stop. When the mobile user switches to the history track interface, it can view the history track.

Historical track view has shown the user (188 **** 8785) in the history of the track view. When it select the view date on March 1 at 2017, it can see the historical track from North China University Technology Fifth Teaching Building to the North Industrial University student apartment four. It can be seen from the above tracking of mobile users that the mobile users' trajectories are different from the historical ones because of the different time and location characteristics of the mobile users. This also shows that the mobile users have different data characteristics.

Mobile user activity area distribution can view user activity area to analyze mobile user activity area preference. Mobile user activity area distribution can be

used to view the activity area during the user's time period according to the start time and end time. Users can also select activity area semantics to view user activity areas that belong to this semantic. User activity preferences can be analyzed by the user activity areas.

It has shown the activity area distributes thermal force map information for the user (188 **** 8785) from February 17 to April 21 in 2017. It can be seen from the figure that the user's active area in this time period is mainly concentrated near north china university technology and university Students apartments. It indicates that the user often operates in these two areas. The user's activities in the north china university technology has larger area and more frequent activities. So the figure shows the user's thermal map points is larger and has darker color. It also reflects the mobile user's activities in these two areas data characteristics are different.

5. Conclusion

The study of this paper extracts the characteristics of mobile users related to data and designs a mobile user data characteristics model. When the location data of mobile users are sparse, the basic personal information and geographical location data of mobile users are analyzed. And the mobile users' age and gender and access categories are extracted. And the sparse characteristics model of mobile data is designed. This model can be used to solve the data sparse, mobile user points of interest personalized recommendations. When the location data of mobile users change with time or location and sparsely densely populated, the mobile users' location data and converted location semantic information are used to extract the geographical location characteristics. These data characteristics designed a mobile data intensive characteristics model. This model can realize personalized recommendation function of mobile user's point of interest when data is dense. Using design of mobile data sparse characteristics model and intensive model, it combines with geographical position record number threshold and chooses a model of mobile user personal recommendation point of interest.

References

[1] LIU W. B., HU T. T., JIANG F., WANG J.: *Heterogeneous Digital Map Resource Retrieval Algorithm Based on WebServices.* Geospatial Information *14* (2016), No. 12, 61–65.

[2] CHEN J., HU B., ZUO X. Q., LE Y.: *User characteristics of mobile phone positioning data.* Journal of wuhan university information science edition *39* (2014), No. 6, 734–738.

[3] HU Z. W., DENG X. Y., GUO J. F., WEN H. M.: *Characteristics of Resident Trip Demand Based on Mobile Phone Location Data.* China Intelligent Transport Annual Meeting (2013), 889–897.

[4] OSCAR C. J., YASUHISA H., KAZUHIRO K.: *A New Control Approach for a Robotic WalkingSupport System in Adapting User Characteristics.* IEEE Transactions on Systems Man & Cybernetics Part C *36* (2006), No. 6, 725–733.

[5] FRAGOUDIS D., ET AL.: *User Modeling in Information Discovery: An overview.* Pro-

ceedings of Advanced Course OR Artificial Intelligence, ACA1999, Greece, July (1999), 410–419.

[6] BALABANOVIE M., SHOHARN Y.: *Learning Information Retrieval Agents: Experiments with Automated Browsing.* In: Proceedings of the AAAI Spring Symposium Series on Information Gathering from Heterogeneous, Distributed Environments March (1995), 13–18.

[7] RUCKER J., POLANCO M. J.: *Siteseer: personalized navigation for the Web.* Communications of the Acm *40* (1997), No. 3, 73–76.

[8] EDWARDS P., BAYER D., GREEN, C. L. AND PAYNE, T. R.: *Experience with Learning Agents which ManageInternet-Based Information.* AAAI Spring Symposium on Machine Learning in Information Access, Stanford, March (1996), 25–27.

[9] S. R. SUBRAMANYA: *Analysis of Mobile Phone Call Data to Determine User Characteristics and to Enhance User Experience.* International Journal of Computer Applications (2012), NCRTC No. 9, 9–13.

[10] HOCHUL J., TAEHWAN K., JOONGMIN C.: *Adaptive User Profiling for Personalized Information Retrieval.* Third 2008 International Conferenceon Convergence and Hybrid Information Technology (2008), 836–841.