Research on Data Mining in Social Networks based on User Features

Liyun KONG

Automobile and Information Engineering Department
Guangxi ECO-Engineering Vocational and Technology College
Liuzhou,545004, China

Abstract—this paper first introduces social network data sets, the definition of technology and the current situation and challenges in the research of social network, which introduces the research significance of multi-dimensional social network data sets. Then this paper introduces 3D social data acquisition scheme, including the scheme, the reality of social information acquisition scheme of virtual social information acquisition and self-assessment information acquisition scheme. This paper collected and made three - dimensional social data mobile social network data sets, explored the law of motion of users, from the angle of multiple social data mining to explore the difference and connection between the multi-dimensional social characteristics, the experimental results show that the prediction accuracy of forecasting model has higher in the paper.

Keywords: data mining; social network; data extraction; user characteristics

I. INTRODUCTION

With the rapid development of Internet and computing, social behavior is to be record the data in the form of down and be made into social data set. These social data set have focused on these social networking database on the research of many scholars, the field of sociology, behavioral science, communication science, anthropology, and practitioners in many fields of R & D, production, advertising and marketing. The researchers hope to massive social data mining through different angles, so as to analyze the characteristics and laws of mining the social potential. This article from the diversity of social data point of view, acquisition and produced include three-dimensional social data mobile social network and based on the data set to explore the social characteristics of people.

Online social network data set collect and store the users in the social network virtual interactive information, in accordance with the functional properties of online social networks, mainly divided into the following three categories:

1). Friends SNS. Friends social networks social networks are mapping the real social network, it uses the real social relationships to attract a large number of users to participate in the. International popular dating sites include: Facebook, My Space and Cy world; the domestic popular dating sites include: Renren, QQ space, happy net; in addition, there are Linkedin and Xing for business professionals. Friends social

networking site data set includes: friends, log, albums, state data.

2). Blog SNS. The blog social network social networks provides the most basic function is a blog publishing and between users and between users focus on service, pay close attention to each other acts constitute the class of online social network [5]. The online social network is often directed network, if the user A attention to the user B and user B does not concern the user A, then there exists only to the edges of B from A and absent from B to A side. Blog famous social networking sites include: Twitter, Google Blogger, MicrosoftLive Spaces, Sina Weibo etc.. Blog social network data set includes: friends, comments, topic, label data information.

3).instant communication network. The instant communication network is to rely on the communication of social network instant communication system is constructed. In this communication system, each user has their own contacts list, according to the contact list buddy relationship can form the instant communication of social networks, the representative communication social networks including MSN, QQ, Skype etc.. However, because of the instant communication system mostly depend on the client communicate with chat and does not provide API open platform, so there are few instant communication network data sets.

II. ACQUISITION SCHEME OF SNS

Social information acquisition system includes mobile terminal data acquisition module, data storage module and mobile terminal of cloud based data storage and monitoring module. Compared with other data acquisition device (such as a Imote: Bluetooth data collector, PMTR: pocket mobile data collector), mobile phone data collection with comprehensive, convenient carrying, has the advantages of short development cycle etc.. So this paper choose smart phones as the acquisition terminal scheme of realistic social information collection. In the experiment volunteers carry a Android smartphone. The data acquisition program in the mobile phone through the acquisition of GPS, Bluetooth, Wifi and other data record of the student movement and position information. The motion information is stored at the same time not only upload to the cloud server in the mobile terminal, convenient management to monitor. Acquisition system architecture shown in Figure 1, consists of three modules: mobile terminal data acquisition module, data

storage module for mobile terminal, the cloud data storage and monitoring module.
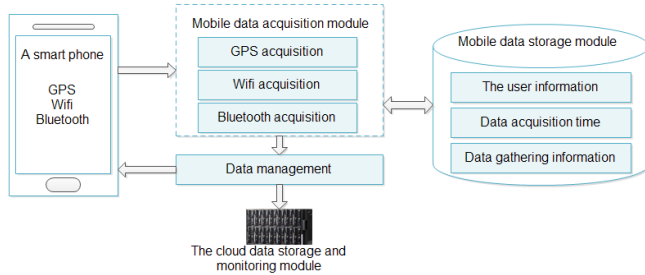

Fig.1 Structure of social information acquisition system

III. SCHEME OF VIRTUAL SOCIAL INFORMATION ACQUISITION

Virtual networking is only the rise in twenty-first Century, but with the Internet develops rapidly, currently the most fiery online social networking platform Renren and Sina micro-blog enrollment has billions user. Considering the experimental subjects are college students, so I chose to virtual social information for college students of virtual social networking platform one one Renren collection.

All open platform use OAuth 2 as the authentication and authorization protocol. OAuth 2 is an open standard, which allows third party applications to access the stored in the service on the web for various data resources in user permission (such as photographs, video, friends list, etc.), and in this process the site without will tell the user account and password information to third party applications. Renren OAuth authentication and authorization process shown in figure 2.

In this paper, Renren SDK was developed based on Android software, software access all network server interface through the Oauth certification, grasping the social information, including user ID, user name, ID friends, friend name, access number, log number, number of state, the number of comments.
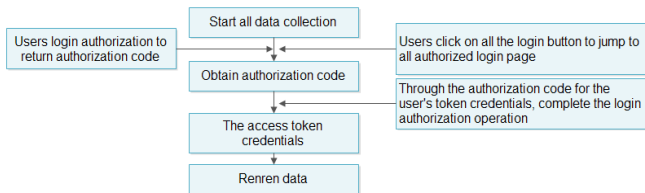

Fig.2 OAuth certification flow chart of Renren

IV. ANALYSIS OF SNS STATISTICS

The movement position data acquisition program will periodically collected volunteers, including GPS data, Bluetooth and Wifi data. Due to the movement of social features of this kind of data reflects the volunteers in real life, so it is called a realistic social data. In addition to this kind of data, this paper was compiled all network information collection program based on Renren SDK, we centralized and acquire everyone nets data of volunteers at the end of 2012.12. All network data reflects the virtual interactive behavior of volunteers in the virtual social networking sites,

and thus are called virtual social data. In the middle of the data acquisition experiment, all the participating volunteers filled out the questionnaire of social relations and the big five personality questionnaire, because these two kinds of social data represent the volunteers subjective self-assessment, and thus are called self-rating social data.

The reality social statistics information see table 1. Internal Bluetooth records each other scanning to Bluetooth between volunteers meet data. Daily active users, see figure 3. From the chart we can see most of the volunteers are actively involved in the experiment. The collected data of GPS and Bluetooth data figure distribution in day time 4. GPS curve reflects the volunteer open data acquisition software of GPS data acquisition time distribution, Bluetooth curve reflects the volunteer encounter time distribution. From this picture can be found in the GPS data mainly to the day time period of 24 points in 9 points, and the distribution of all words, hourly GPS data collected accounted for the total amount of data from 5% to 6%. Bluetooth data has two peaks, respectively is 11 points and 15 points. This is because the volunteers in the two time periods to class. The volunteers from a department, each other has many similar courses, resulting in more meet. At the same time can be found in the morning meeting is higher than that in the afternoon meeting from the figure, this shows that volunteers must be with more same class in the morning.

Table.1 The statistics information of social data

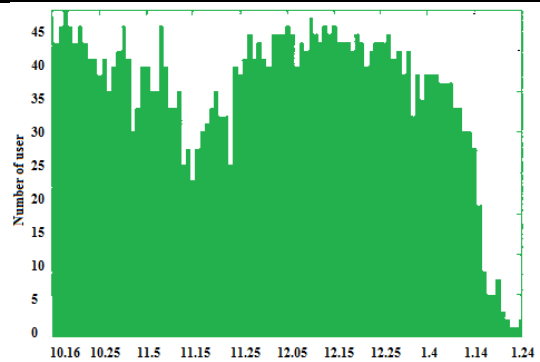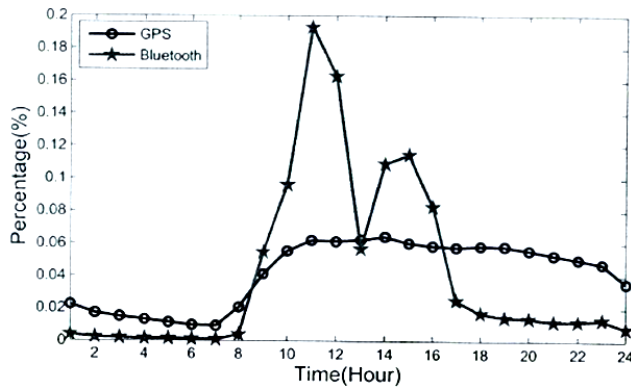|	
Collected on the number of GPS	1591114
Collected Wi-Fi MAC address	88445
Collected to the number of independent Wi-Fi MAC address	5248
Bluetooth records were collected	310937
Internal Bluetooth records collected	170278

Fig.3 Statistics of daily active user

Fig.4 Distribution of GPS, Bluetooth data in a day

V. CONCLUSIONS

This chapter is analyzed social characteristics experiment of mobile user based on social data set. According to the data distribution of the motion user, divided into five geographical position of community. Starting from the five location community perspective, in-depth analysis of the characteristics of mobile characteristics, meet and community characteristics of users. The experiment found that users encounter characteristic of whole obeys a power-law distribution, but has different characteristics in different meet location community. In addition, mobile data users reflects the user has obvious regularity in the location of mobile community, and members of the community the geographical location is also constitute a relatively stable

REFERENCES

[1] Chen Chongle. Improved algorithm for mining association rules in the application log analysis of network user access [J]. Shanghai Normal University, 2007

[2] Zhao Guohua. Research and application of Web data mining based on XML [J]. Tianjin University of Science and Technology, 2010

[3] Ou Jianwen, Dong Shoubin, Cai Bin. Topic information extraction from template web pages [J]. Journal of Tsinghua University: Natural Science Edition, 2005, 45 (1): 1743-1747.

[4] Han Jiawei, Campbell. Data mining: concepts and techniques [M].Beijing: Mechanical Industry Press, I, 2001:, 232-233.