Future Generation Computer Systems 37 (2014) 429–437

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Analysis of user behaviors by mining large network data sets

Zhenhua Wang [a], Lai Tu [a,*], Zhe Guo [b], Laurence T. Yang [c], Benxiong Huang [a]

[a] *Department of EIE., Huazhong University of Science and Technology, Luoyu Rd. 1037, Wuhan, China*

[b] *Controller Technology Development Department, Shanghai Institute of Huawei Technologies Co.Ltd, XinJinQiao Rd. 2222, Pudong New District, Shanghai, China*

[c] *School of Computer Science and Technology, Huazhong University of Science and Technology, Luoyu Rd. 1037, Wuhan, China*

HIGHLIGHTS

- We process 16-week-long CDR data of one million users with fuzzy clustering.
- We find some relations between ARPU level and users behavior patterns.
- We prove that the mobility of users is related to ARPU and communication behavior.
- Results indicate that the top ARPU level users are the most "lonely" ones.

ARTICLE INFO

Article history:
Received 29 March 2013
Received in revised form
16 January 2014
Accepted 17 February 2014
Available online 26 February 2014

Keywords:
User behavior
Data mining
Customer segmentation
Fuzzy *c*-means clustering

ABSTRACT

Understanding the intelligence of human behaviors by mining petabytes of network data represents the tendency in social behaviors research and shows great significance on Internet application designing and service expansion. Meanwhile, the running mobile networks that generate huge data can be the best social sensor for these studies. This paper investigates a practical case of mobile network aided social sensing which uncovers some features of users' behaviors in mobile networks by intelligently processing the big data. The paper studies the users' behaviors with regard to communication, movement, and consumption based on large user data sets. The main contribution of the study is some findings on the relations among these behavior features. We find that the users' calling behaviors are different despite their monthly expenditures being similar, though different consumption level users may have similar communication behaviors. We also find that statistically users with the higher mobility contribute more ARPU than those with lower mobility. Additionally, we also find that the top consumption level users are the most "lonely" ones by exploring the movement clustering patterns of users. These findings are significant to instruct marketing strategies for telecommunication industry.

1. Introduction

Although technologies have been pushing our world and society to a smarter one, human behaviors in the society still keep some inherent features and complexities that are hard to explain. Understanding the intelligence of human behaviors in the real world has great significance in practical application, such as mobile network deployment, traffic engineering, urban planning and service recommendation. While studies on human behaviors were not new in social science, quantitative analyses were not common due to the lack of source of data. Thanks to the computers and networks as they can now give plenty of computational ways of collecting and analyzing data for social studies, which used to depend on surveys in traditional methodology. Thus in the new era of "Big Data", never before have researchers had the opportunity to mine such a wealth of information that promises to provide insights about the complex behaviors of human societies [1,2]. One goal of these researches is to quantitatively uncover the inherent feature of human behaviors and track how our behaviors evolve by mining petabytes of network data.

The studies on human behaviors can be traced since 19th century [3]. The discipline covers psychology, sociology, anthropology, etc., which study different aspects of the nature of human intelligence. However, due to lack of ways of measurement and analysis of large scale of data, the studies mostly focused on individuals or a certain small group or rough qualitative estimation of social behaviors. This was kept almost unchanged until the emergence of modern computing science and network technology. Recent

* Corresponding author. Tel.: +86 18086484600.
E-mail addresses: tulai.net@gmail.com, tulai@hust.edu.cn (L. Tu).

http://dx.doi.org/10.1016/j.future.2014.02.015
0167-739X/

advances in computing and network science have driven the studies on social behaviors to a newly high stage. The universal usage of computational devices leaves huge amount of data that are tightly related with human behaviors. Scientists were able to use computer data, and network data to sense and analyze human behaviors in the society. During this period, a lot of theoretical and practical achievements in social behaviors studies arose. The prevalence of social network in the Internet and the progress in complex network research are best examples for this [4,5]. Meanwhile, with the development and widespread of smartphones and mobile network, they began to show their dominance in sensing human behaviors over the traditional way of mining Internet social media. Therefore, in this paper, we investigate a practical case of mobile network aided social sensing. We study the users' behaviors with regard to communication, movement, and consumption based on mining a large set of mobile user data. The findings may be significant to instruct marketing strategies for the service providers to increase their revenues and lead them to success through Customer Relationship Management (CRM) [6].

When discussing the communication behaviors, movement behaviors and consumption behaviors of mobile users, we aim at three questions:

1. Does each ARPU (Average Revenue Per User) level imply similar communication behavior?
2. Does each ARPU level imply similar mobility level? And what is the relationship between a user's consumption capacity and mobility?
3. Do people with similar mobility patterns have similar communication behaviors?

We explore the above three issues in this paper. We notify that the detailed information about human mobility across a large population can be collected by mobile operators that record the closest base station when a call is generated. Herein, we use 16-week-long (from September to December, 2009) calling records of about one million users in a metropolis in China as our data set to conduct research. Each record includes the serving base station's ID, the start and end times of each phone call (outgoing and incoming calls are distinguished), as well as the monthly billing information of each user. For preserving the users' privacy, the record of each phone call is anonymous.

We apply both the value-based and behavior-based segmentation methods to investigate the difference between the users who have different or similar ARPU. That is, we firstly divide the users based on their ARPU into different groups as preliminary division. Then, we study their behaviors by using fuzzy c-means (FCM) clustering. The contributions of our paper are:

- Verify that the same ARPU level users have different behavior patterns, while different ARPU level users may have similar communication behaviors;
- Prove that the users' mobility levels have relation with their ARPUs and communication behaviors. People who make less calling or like making calls at night have less mobility than others;
- Find that the top ARPU level users are the most "lonely" ones, which imply that they are willing to move alone and do not like being with others.

The rest of paper is organized as follows. Section 2 briefly reviews related work. Section 3 introduces a communication behavior indicator and a user segmentation algorithm based on the FCM that are used to study the questions proposed by us. In Section 4, we investigate user behavior patterns and extract salient characteristics that indicate the relationships among users' consumption capacity, communication time, mobility, and locations. Finally, we further discuss our results and conclude the paper in Section 5.

2. Related work

Telecommunication industry has been developing rapidly since early 1990s. With increased market competition and years of development, the number of mobile users is becoming saturated. Notably, the proportions of mobile users are even more than 100% in some metropolises. This implies that some users subscribe more than one mobile number. Such a situation forces mobile operators to turn their efforts from increasing the number of subscribers to retaining the existing ones. As a result ensuring the quality of mobile services becomes the most important factor to enhance competitiveness. Typically, the ISP's are in search of data that possess key information about their users and try to distinguish users and in-depth understand the needs of different user groups.

To know the relationship between customer value and behavior, an appropriate method for customer segmentation is critical in CRM. The customer value, in most instances, only reflects the consuming capacity of a user. Segmentation based on it cannot distinguish the difference of user behaviors, which, however, contains useful and valuable information about the characteristics, preferences and desires of users.

Customer segmentation methods are classified in terms of segmentation dimensions and the purposes of segmentation. Generally, there are four kinds of customer segmentation methods [7]: demographics segmentation [8], lifestyle segmentation [9–12], behavior segmentation [13] and benefit segmentation [14]. Demographic segmentation treats geography as an important dimension. However, the globalization of markets and the rapid development of information technology weaken the relevance between customer and geographic characteristics. Lazer firstly proposed the method to identify and segment customers based on their lifestyle [11], which includes: Activity, Interests and Opinion (AIO) [9,10,12]. However, this kind of segmentation is hard to do since it is generally impossible to get customers' comprehensive lifestyle data in practice. Behavior segmentation classifies customers by analyzing their behavior patterns [13]. Supported by information technology, we can handle a large amount of data to get useful results with this method. But simply using behavior segmentation cannot disclose customer value or benefit, which is mostly concerned by the operators. There are many ways to calculate the value of customers. The most popular one is segmenting users based on ARPU. The famous pyramidal model typically divides the users based on the ARPU into 3 clusters: high, medium, and low ARPU, respectively. In this paper, we choose user behavior as an indicator to give an in-depth insight of the relationship between customer value and behaviors by considering the weakness and advance of the existing segmentation methods.

CRM is a broadly recognized, widely implemented strategy to manage and nurture a company's interactions with its users and sales. Its overall goal is to find, attract, and win new users, retain existing ones and entice former users back, and reduce the costs of marketing and customer services [6]. In order to achieve successful customer relationships, data mining (DM) is generally applied to understand the characteristics and desires of users [15]. The DM is a process of finding hidden patterns, associations, rules and statistically significant structures in large databases [16]. Clustering is a DM technique with applications in the areas of data exploration, segmentation, targeted marketing, and cross-selling [17,16]. With clustering, the CRM aims to segment users into discrete groups that share similar characteristics, such as age, gender, interests, and consumption habits. However, we notify that DM methods in the previous work seldom cluster users based on both their behaviors and ARPU.

Recently, analysis of user behaviors is becoming a popular approach to understand users. A number of applications ranging from city planning to resource management in mobile communications rely on the understanding of human behaviors. For example, Lopez-Paris et al. found that pedestrians usually walk along straight pavements in cities. Based on this, they proposed an approach to automatically agent navigation in realtime applications [18]. Eagle and Pentland found that different MIT staff's mobility patterns are different: higher entropy of junior staff indicates