Electronic Commerce Research and Applications 17 (2016) 1–11

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

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

Exploring the sequential usage patterns of mobile Internet services based on Markov models

Xiaohang Zhang [a,b,*], Cenyue Wang [a], Zhengren Li [a], Ji Zhu [b], Wenhua Shi [a], Qi Wang [a]

[a] School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1107, USA

ARTICLE INFO

Article history:
Received 14 July 2015
Received in revised form 21 February 2016
Accepted 22 February 2016
Available online 28 February 2016

Keywords:
Mobile internet
Behavior pattern
Multi-state model
Hidden Markov model
Clustering

ABSTRACT

Mobile Internet has developed rapidly, and various types of mobile Internet services have changed people's lifestyles profoundly. Consequently, there is a broad market for mobile Internet service providers. To provide better service and attract users, service providers must understand their users' behavior patterns. This study proposes a framework to model users' mobile online behavior based on a multi-state model and a hidden Markov model; this study also extracts typical sequential behavior patterns through clustering methods. The results of the experiments display several characteristic behavior patterns that can guide service providers in application designing, operating, and marketing.

1. Introduction

With the development of the 3G/4G network and an intelligent terminal industry, the quantity of mobile Internet users is increasing rapidly. According to the 2014 ICT figures released by ITU (ITU 2014), the number of mobile Internet users reached 2.3 billion, with 55% of these users in developing countries that lead mobile-broadband growth. By the end of 2014, China had 649 million Internet users and 557 million mobile Internet users, according to CNNIC survey (CNNIC 2014a). Additionally, the percentage of Chinese users accessing the Internet via mobile grew to 83.4% as of June 2014, for the first time surpassing the percentage of users (80.9%) who access the Internet via PCs (CNNIC 2014b). As the figures show, the mobile Internet market is a new rapidly growing market for online service providers. To seize mobile Internet market share, online service providers, including telecommunication operators and Internet enterprises, are committed to developing and operating various mobile Internet services.

Through mobile Internet, people can send instant messages, watch videos, read online books, order takeout and so on. All types of mobile online behavior have generated a huge volume of behavior-record data, which is stored by telecommunication operators and Internet enterprises. With the development of the mobile Internet market, the behavior data volume will continue to grow rapidly. Knowledge such as users' preferences and habits is covered in those mobile online behavior records. By analyzing user behavior data, online service providers can understand mobile Internet users' behavior patterns. Furthermore, that knowledge can help telecommunication operators plan network resources and design pricing policies; in addition, it can assist Internet service providers in developing marketing strategies and designing new services.

In the mobile Internet user behavior research field, previous researchers have studied network traffic, geospatial dynamics of application usage and single service behavior from several aspects. However, most studies are based on a single service or focus on underlying technology aspects. To assist in business decisions, we need to extract knowledge from all-service records of user behavior and analyze this from a business perspective. Thus, we can develop managerial proposals regarding mobile online user behavior.

This study proposes a novel framework for analyzing the sequential behavior patterns of various mobile Internet services. This framework transforms users' behaviors into state transition processes, train a multi-state model (MSM) and a hidden Markov model (HMM) separately to describe user behavior and understand behavior patterns by clustering users and extracting the typical behavior characteristics of each cluster. Although both of the MSM and HMM originate from Markov model, their results are significantly different (see Section 3) and their combination can improve the understanding of behavior patterns.

* Corresponding author at: School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China.

http://dx.doi.org/10.1016/j.elerap.2016.02.002
1567-4223/

The remainder of this paper is structured as follows. In Section 2, the related literature regarding Internet service usage pattern and model algorithms are thoroughly reviewed. Section 3 presents this study's framework and process. Section 4 describes the data used in this study and the experimental results. The managerial implications of the results are discussed in Section 5, and the contributions and limitations of this study are presented in Section 6.

2. Literature

This study's objective is to propose a novel framework for analyzing the behavior patterns of various mobile Internet services. The literature review is composed of two parts: studies on mobile Internet user behavior and the related models applied in this study. The latter includes a multi-state model, a hidden Markov model, distance computation and clustering. The literature reviews are shown in Table 1.

2.1. Studies on mobile Internet user behavior

There are some studies on mobile Internet user behaviors. Cheng and Sun (2012) used message, entertainment and micropayment services to segment customers with an improved segmentation model, the TFM (time, frequency, money) model. Wu and Chou (2011) developed a soft clustering method that uses a latent mixed class membership clustering approach to classify online customers based on their purchasing data across categories. Bose and Chen (2010) selected Internet usage, revenue, services and user categories as the research indicators to cluster customers.

Zhao et al. (2013) exposed four aspects of the characteristics of service visit styles using a weighted user-service bipartite network model. Liu et al. (2012) studied mobile Internet streaming services from the server perspective and showed the great heterogeneity from traditional Internet streaming services in terms of hardware and software differences in mobile devices, different characteristics of mobile videos, and different user access patterns from those in traditional Internet streaming services. Keralapura et al. (2010) studied mobile user browsing behavior by investigating behavior patterns among mobile users based on real mobile network data collected from a large 3G CSP in North America; they also proposed and developed a scalable co-clustering methodology using a novel hourglass model. Ghosh et al. (2011) examined the characteristics of traffic observed at a large number of public Wi-Fi hot-spots deployed in two large metropolitan cities in terms of arrival counts, temporal variations, connection durations, and byte counts and categorized the different venues into certain business types. Chen et al. (2012) studied user movement behavior patterns in terms of the problem of mining matching mobile access patterns based on joining mobile, movement location, and dwell time in a timestamp and service request. Shafiq et al. (2012) provided a fine-grained characterization of the geospatial dynamics of application usage in a 3G cellular data network. Zhao constructed an empirical model for Web browsing on the mobile Internet, which showed that the size of the main object and the size of the embedded object exhibited a Pareto distribution; in addition, the number of embedded objects and the session duration fitted the Weibull distribution well, and the embedded object Inter-arrival time as well as reading time followed Lognormal distribution (Zhao et al. 2011). Lu et al. (2012) proposed a framework to combine user motion patterns and mobile purchase behaviors.

From the above discussion, previous mobile Internet user behavior studies consist of network traffic, geospatial dynamics of application usage and single service behavior. In the mobile Internet data set, however, user behaviors across various mobile Internet services are recorded. For telecommunication operators and Internet enterprises, studies on user behavior of various mobile Internet services and the interrelations among services, which are based on all-service data, are more helpful for understanding mobile Internet user behavior as a whole and assist in making business decisions. In the previous mobile Internet customer segmentation studies, clustering methods were often directly applied to the customers' usage attributes, such as the

Table 1
Literature review.

Components	Contents/Methods	References
Mobile Internet user behavior	Online shopping behavior	Wu and Chou (2011), Hong and Kim (2012)
	Online ticketing behavior	Seret et al. (2014)
	Online Entertainment and Micropayment	Cheng and Sun (2012)
	Internet usage behavior	Bose and Chen (2010)
	Weighted user-service bipartite network	Zhao et al. (2013)
	Heterogeneity from traditional Internet streaming services	Li et al. (2012)
	Scalable co-clustering methodology using a novel hourglass model	Keralapura et al. (2010)
	Characteristics of mobile traffic	Ghosh et al. (2011)
	Mobile access patterns	Chen et al. (2012)
	Fine-grained characterization of the geospatial dynamics of application usage	Shafiq et al. (2012)
	Modeling Web browsing on the mobile Internet	Zhao et al. (2011)
	User motion and purchase behaviors	Lu et al. (2012)
Multi-state model	Review of MSM	Meira-Machado et al. (2008)
	Continuous-time Markov chains	Cox and Miller (1977)
	Applications of MSM in social science	Haan (2010), Kalbfleisch and Lawless (1985)
Hidden Markov model	Introduction to hidden Markov models	Ghahramani (2001)
	Applications of HMM	Och and Ney (2004), Stanke and Waack (2003), and Duong et al. (2005)
Distance between probability density functions	Euclidean space	Cha, 2007
	Bray–Curtis distance	Bray and Curtis, 1957
	Kullback–Leibler divergence	Kullback and Leibler, 1951
	Bhattacharyya distance	Bhattacharyya, 1946
Distance between hidden Markov models	Arbitrary observation densities	Juang and Rabiner, 1985
	Model match	Rabiner, 1989
	BP metric	Panuccio et al., 2002
Clustering methods	K-means	Lloyd, 1982
	k-Medoids	Ng and Han, 2002
	Hierarchical clustering	Johnson, 1967