Computers, Environment and Urban Systems 58 (2016) 97–106

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

A contextualized and personalized model to predict user interest using location-based social networks

Ming Li [a,*], Günther Sagl [b], Lucy Mburu [a], Hongchao Fan [a]

[a] GIScience Research Group, Institute of Geography, Heidelberg University, Germany
[b] Geoinformation and Environmental Monitoring, Engineering & IT, Carinthia University of Applied Sciences, Austria

ARTICLE INFO

Article history:
Received 21 December 2014
Received in revised form 29 March 2016
Accepted 30 March 2016
Available online xxxx

Keywords:
User interest
Context-awareness
Personalization
Prediction
Location-based social networks

ABSTRACT

The accurate determination of user interest in terms of geographic information is essential to numerous mobile applications, such as recommender systems and mobile advertising. User interest is greatly influenced by the usage context and varies across individuals; therefore, a user interest model should incorporate these individual needs and propensities. In this paper, we present an approach to model user interest in a contextualized and personalized manner based on location-based social networks. Multinomial logistic regression is employed to quantify the relationship between user interest and usage context at both the aggregate and individual levels. The proposed approach is tested in a real-world application using Foursquare check-ins issued between February and June 2014 in the three major cities of Chicago, Los Angeles and New York. Results demonstrate the capability of the contextualization process for capturing contextual influences on user interest, and that such influences can be observed at a fine-grained scale at the individual level through the personalization process. The proposed approach therefore enables contextualized and personalized estimation of user interest, thereby contributing useful information to follow-up mobile applications.

1. Introduction

Thanks to the rapid technological advancements in portable devices and telecommunication, citizens are increasingly exploring the world with their mobile devices at hand every day. The trend of computations shifting from the desktop to the mobile environment has enabled geographic information to emerge as a centerpiece to enhance human–computer interaction (HCI). Mobile users unwittingly create and utilize geographic information when they interact with the real world through their mobile devices, and such interaction often involves making decisions depending on the immediate surroundings, nearby venues, facilities, services, etc. Among the great variety of available geographic information, only a small portion is usually of interest to an individual. Presenting all available information regardless of relevance would inevitably undermine the efficiency of decision making (Reichenbacher, 2005; Li, Sun, & Fan, 2015). This makes it crucial for mobile application developers to acquire precise and personalized knowledge of user interest taking into account their contextual conditions.

Two complexities arise in determining what information will be of interest to which users. The first stems primarily from the dynamic nature of mobile user interests. The interests vary significantly and are influenced by the usage context, which also varies from one person to another. The model of user interest should, thus, take this dynamism into account. The second obstacle regards the identification of the data that will represent user interest in a reliable manner. It is very difficult to determine the exact contexts and interests of mobile users, because both aspects are vague and heterogeneous. Fortunately, nowadays' mobile communication behavior, specifically the technical possibility in combination with users' desire to instantly share information with others via internet platforms, provides potential solutions to aforementioned problems. With the emergence of location-based social networks (LBSNs) such as Twitter and Foursquare, individuals are continuously sharing their current location, engagements, opinions about events or objects, etc. Consequently, a plethora of publicly available data is generated from these platforms, and can potentially reflect the behavior of millions of citizens at a remarkable level of detail (Roick & Heuser, 2013).

The Foursquare platform, as an example, is one of the most popular LBSNs worldwide, with more than 55 million users and over 6 billion check-ins (BrandonGaille, 2015). Every time a user checks-in at a place, a unique data entry, i.e., a check-in record, is created in the Foursquare database with attached contextual information such as the time and location. The location information in Foursquare is not only recorded with geographic coordinates, but also characterized with predefined hierarchical categories. This categorization is one of the outstanding features that distinguish Foursquare from other LBSNs (Cramer, Rost, & Holmquist, 2011), and it provides a special insight into the semantics of the locations and, in turn, the user interest. Hence, the check-in data provides a convenient way to learn how user interest relates to

* Corresponding author.
E-mail address: ming.li@geog.uni-heidelberg.de (M. Li).

http://dx.doi.org/10.1016/j.compenvurbsys.2016.03.006
0198-9715/

usage context. The knowledge derived from such data can thereafter be applied in a variety of follow-up applications, such as information retrieval, recommender systems, mobile advertising, etc.

To this end, this paper proposes an approach to uncover the influence of context on user interest at both the aggregate and the individual levels. Two processes are at the core of this approach: the contextualization process applies multinomial logistic regression to all users and quantifies the relationship between user interest and usage context at the aggregate level. The personalization process distinguishes personal differences and quantifies the relationship at the individual level. Together, these two processes have the potential to reveal new insights into user interest and usage context, as well as into the relationships between both.

The rest of the paper is organized as follows. Section 2 presents a brief review of existing studies on context-awareness, prediction and follow-up applications. Section 3 presents the proposed approach with its conceptual framework and main modules. Section 4 tests the approach using Foursquare datasets from three US cities. Finally, Section 5 concludes the work and provides directions for future work.

2. Related work

With computation becoming increasingly pervasive, plenty of research has been conducted for the mobile environment in the recent years. Related research directions include, for example, the mobile usage context and context-aware applications (Section 2.1), various kinds of prediction applications (Section 2.2) and user-interest-based applications (Section 2.3).

2.1. Context, context-awareness and contextualization: definitions and applications

The notion of context originates from linguistics (Garcia, Duranti, & Goodwin, 1993), and has in the past been considered and defined from many perspectives, depending on the application domains. Among the most frequently cited definitions, we can mention for example the one by Abowd et al. (1999: 306), who define context as “any information that can be used to characterize the situation of an entity”. In the domain of mobile computing, usage context typically includes location, time, user characteristics (activity, mood, etc.), device characteristics (ability, connection, etc.) and information about surroundings (weather, services, etc.) (Nivala & Sarjakoski, 2003; Reichenbacher, 2004). Other than the enumeration of context elements, context is also represented by ontologies (e.g., Wang, Zhang, Gu, & Pung, 2004) or with visual interfaces (Tomaszewski & MacEachren, 2012) to adapt to different application scenarios.

Mobile applications usually need to understand and react to usage contexts in order to better meet the needs of mobile users. This ability is called context-awareness and the process to achieve this ability is contextualization. Many context-aware mobile applications have in the past years been designed for different purposes, contextualized to different types of contexts, and implemented in different ways. Since the creation of mobile context-aware computing by Abowd et al. (1997) with the proposal of Cyberguide, mobile guides have been among the most common context-aware applications. Based on some pre-defined rules, these guides deliver contextualized information in the form of texts (Cheverst, Davies, Mitchell, Friday, & Efstratiou, 2000), images (Lim, 2012), audio data (Chittaro & Burigat, 2005) or maps (Zipf, 2003), and serve in various settings, such as cities (Carlsson, Walden, & Yang, 2008), exhibitions (Oppermann & Specht, 2000) or museums (Ghiani, Paternò, Santoro, & Spano, 2009). Some researchers proposed to adapt not only the information but also the visualization of this information to the context, which led to the emergence of the mobile cartography research field (Reichenbacher, 2004).

The contextualization process used in these applications is typically based on context information collected with sensors (Sagl, Resch, & Blaschke, 2015) and a set of predefined rules in the form of key-value pairs, tagged encoding, ontologies, etc. (see the literature reviews conducted by Chen & Kotz, 2000; Bettini et al., 2010). The rules often involve a complex procedure, including preliminary design, survey activity and follow-up maintenance. In this work, we use a regression model instead of predefined rules for the contextualization process. In this manner, the additional work requiring manual intervention can be replaced by statistical approaches. Moreover, the regression model can provide additional and quantified knowledge regarding how context can influence mobile users, which can benefit many related studies on user behavior.

2.2. Predictions for mobile users: targets and algorithms

As noted by Rudin (2012), prediction lies at the heart of almost every scientific discipline, and it is a key topic in machine learning and statistics. From a purely academic point of view, predictions can assist in gaining knowledge since they allow constructing dynamic models that can directly be tested against the set of previous states. Beyond research, predicting is highly useful in a variety of practical situations. In the literature, one can find a large body of prediction research, with different targets and algorithms, and based on different data sources. For example, LBSN datasets have frequently been used in the recent years to predict social ties and links (Cho, Myers, & Leskovec, 2011; Wang, Pedreschi, Song, Giannotti, & Barabasi, 2011), user behavior and mobility (Do & Gatica-Perez, 2012; Preoţiuc-Pietro & Cohn, 2013), user activity (Bart, Zhang, & Hussain, 2013), users' whereabouts (Steiger, Westerholt, Resch, & Zipf, 2015), users' next visiting places (Gambs, Killijian, & del Prado Cortez, 2012; Noulas & Scellato, 2012), etc. Various algorithms have been used for these applications, such as conditional models (Do & Gatica-Perez, 2012), random walk (Noulas, Scellato, Lathia, & Mascolo, 2012), decision tree (Noulas & Scellato, 2012) and Markov chains (Gambs et al., 2012).

In contrast with previous research, we herein intend to predict user interest in terms of geographic information given their usage context using multinomial logistic regression. We use LBSN datasets, and more specifically the Foursquare check-in datasets, to derive user interest, because this knowledge is essential for many follow-up applications that are introduced in the next subsection.

2.3. Follow-up applications based on user interest

User interest has been extensively studied for a long time within the context of web search and information retrieval, as far back as the era of the traditional web (Claypool, Brown, Le, & Waseda, 2001; Qiu & Cho, 2006). In the era of mobile web, the demand to fulfill user requirements has become even more urgent. In order to support the decision-making process while avoiding unnecessary distractions, many studies (e.g., Da Costa Pereira, Dragoni, & Pasi, 2012; Leung, Lee, & Lee, 2013) have illustrated the need to deliver the information according to the level of priority of user interest.

Knowledge about user interest is useful in a number of ways. The most common way is to rely on user interests to measure user similarity (Liu, Chen, Xiong, Ding, & Chen, 2012; Gao, Dong, & Fu, 2015), whereby similarity between users is a key component of *collaborative filtering*, a powerful tool to enhance information retrieval. User interest is also essential to recommender systems. Based on the understanding of user interest, researchers have designed and implemented various kinds of personalized systems to recommend venues (Bao, Zheng, & Mokbel, 2012; Noulas et al., 2012; Liu, Liu, Aberer, & Miao, 2013), routes (Kurashima, Iwata, Irie, & Fujimura, 2013) and friends (Chu, Wu, Wang, Chen, & Chen, 2013) to mobile users. Mobile advertising is another application field where knowledge about user interest is critical. Researchers have attempted to design personalized mobile advertising to meet the needs of potential customers (Chen & Hsieh, 2012) and