Expert Systems with Applications 42 (2015) 2764–2771

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A privacy-aware feature selection method for solving the personalization–privacy paradox in mobile wellness healthcare services

Namyeon Lee [a], Ohbyung Kwon [b,*]

[a] Sungkyul University, Anyang, Gyunggi-Do, Republic of Korea
[b] School of Management, Kyung Hee University, 1 Hoegi-Ding, Dondaemun-Gu, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Available online 21 November 2014

Keywords:
Privacy decision-making
Feature selection
Personalization–privacy paradox
Privacy calculus
Mobile wellness healthcare services

ABSTRACT

Despite the vast number of available opportunities, customers' privacy concerns can inhibit their acceptance of mobile wellness healthcare services. Personalization of these services may go some way toward alleviating these concerns. The essence of personalization in mobile wellness healthcare services is feature selection because users may be concerned about disclosing private information, although it may be useful in provision of personalized services. Therefore, an optimal feature selection method is needed which considers both these privacy concerns and the quality of personalization. Such safe and accurate data collection would facilitate understanding of customers' preferences while safeguarding their privacy. In this study, a privacy-aware feature selection method is proposed based on the personalization–privacy paradox, and this paradox is explored in the context of wellness healthcare services with consideration of the personal characteristics of customers using these services.

1. Introduction

Recent advances in mobile technologies have presented decision-makers (e.g., marketing managers) with a new form of mobile healthcare services, mobile wellness services (MWS), which refers to healthcare services provided via mobile devices, including reliable access to and transmission of medical information, location management, and support for patient mobility (Varshney, 2007; Wu, Li, & Fu, 2011). In order to define MWS, we first need to provide a definition of wellness. In previous studies, the definition of wellness varies considerably (Hattie, Myers, & Sweeney, 2004; Krout, 2007; Ryan & Deci, 2001; Travis & Ryan, 1988). According to the definition of Zender and Olshansky (2009), wellness is an individual's subjective experience of overall life satisfaction in relation to physical, mental, emotional, spiritual, social, economic, occupational, and environmental dimensions. In other words, wellness does not mean only physical health, but also psychological health and many other kinds of health as well. Based on this definition of wellness and the work of researchers on design in information systems (Chatterjee, Chakraborty, Sarker, Sarker, & Lau, 2009), we define MWS as the usage to varying degrees of mobile technologies to increase physical and psychological wellness and provide healthcare services across locational, temporal, and contextual boundaries. For example, if a user feels stress or depression, MWS may provide online music streaming or recommend therapy shops in the user's vicinity to decrease that stress or depression based on the user's current context (e.g., activity level, body temperature, heartbeat) or information about user preferences (e.g., favorite genre of music). These private data may be obtained by sensors in mobile devices or third-party tools such as Nike+.

While traditional healthcare services focus on the patient, wellness healthcare services also focus on individuals other than the patient, such as caregivers or other interested parties. Therefore, the target market of MWS is wider than that for traditional healthcare services. In addition, MWS aims to prevent illnesses rather than to treat diseases, as compared with traditional healthcare services. Moreover, in the process of service selection, MWS offers another possible option, whereas previously, patients have had little choice but to use traditional healthcare services.

To ensure quality and customer satisfaction, the designers of MWS try to understand the current living context and preferences of their customers by utilizing one or more location tracking technologies, including GPS, cellular networks, wireless LANs, and RFID. Moreover, users must build a personal profile by providing information such as age, gender, address, and medical history. In addition, individual questions may be posed and self-reported health status information may be requested. Such information is

* Corresponding author. Tel.: +82 2 961 2148; fax: +82 2 961 0515.
E-mail addresses: nylee@sungkyul.ac.kr (N. Lee), obkwon@khu.ac.kr (O. Kwon).

http://dx.doi.org/10.1016/j.eswa.2014.11.031
0957-4174/

often very useful for service deliverers to understand their customers' preferences.

However, in soliciting information from customers, a paradox arises between personalization and privacy because increasing the quality of personalized services inevitably requires provision of personal data. Obtaining and reusing user data for the purposes of personalization in healthcare settings raises privacy issues. Hence, a trade-off exists between personalization and privacy in service design. Individuals may anticipate possible loss of privacy when they are requested to provide personal information. They may prefer to retain control over their personal information, its accessibility to others, and the manner in which it is revealed (Stein, 2002). Privacy is defined as "the ability of the individual to control the terms under which personal information is acquired and used" (Westin, 1967, p.7). Correspondingly, privacy control in the context of MWS means the capability of MWS to optimize the degree of personalization in an autonomous manner. In other words, the system may autonomously measure the level of usage based on a user's personal information such as his or her profile, living context, and preferences. This does not mean that MWS does not work if a user does not want to provide her or his personal information. In such cases, MWS provides general, non-personalized services.

In the context of MWS, privacy is a major consideration in individual decision-making about use of the service. However, privacy concerns have not been included in studies of the factors critical to the success of these services (Chatterjee et al., 2009). Selection of features for personalization of services affects both the quality of service and privacy. Various methods may be used to select features in model construction.

The purpose of this study is to propose a privacy-aware feature selection method to cope with the personalization–privacy paradox in mobile wellness services based on the privacy calculus model. This privacy-aware feature selection method minimizes total costs incurred as a result of privacy concerns and service inaccuracy. The cost incurred as a result of privacy concerns is used as a proxy for privacy loss, and the cost of service inaccuracy is used as a proxy for service quality. In our experiment, we build a parsimonious function for the purpose of estimating these costs. Then, by varying the disclosure level of each feature, we derive an optimal set of features which minimizes total cost.

The rest of this paper is organized as follows. The theoretical background is described in Section 2. The proposed methodology is presented in Section 3. The experimental results are provided in Section 4. We conclude in Section 5.

2. Theoretical background

2.1. Privacy calculus model

Privacy calculus theory is a method of cost–benefit tradeoff analysis that accounts for inhibitors and drivers that simultaneously influence the decision of whether or not to disclose information (Culnan & Bies, 2003; Dinev, 2006). During information exchange, consumers' privacy behaviors are driven by privacy calculus (Dinev, 2006). The privacy calculus model illustrates individuals' information disclosure behavior, which has been the subject of much interesting research, including studies based on utility maximization theory from economics (Rust, Kannan, & Peng, 2006) and social contract theory (Milne, 1993; Phelps, Nowak, & Ferrell, 2000) and social exchange theory (Andrade, Kaltcheva, & Weitz, 2002) from social psychology. Conjoint with utility maximization theory, the privacy calculus model explains that individual users perform cost–benefit analyses to decide whether or not to disclose personal information. Furthermore, drawing on social exchange theory (Houston, 1987), Culnan and Bies (2003) introduced the concept of "second exchange" to explain privacy calculus as a utilitarian exchange whereby personal information is given in return for value such as higher quality of services.

Although the concept of privacy calculus is intuitively appealing, few studies have considered the value of privacy calculus for privacy-aware information systems in the context of MWS. Mobile healthcare systems contain not only medical, but also personal information. Users have a vested interest in where their health data is stored and who can view their medical records (Meingast, Roosta, & Sastry, 2006). Misuse of such personal data causes privacy concerns that may prevent them from gaining the full benefit of the healthcare service (Al Ameen, Liu, & Kwak, 2012). About the impact of privacy concerns on service usage, researchers agree that consent to use healthcare data must be obtained. Accordingly, mobile healthcare services such as HealthGear (Oliver & Flores-Mangas, 2006), MobiHealth (http://www.mobihealth.org/), Ubicom (http://www.ubimon.net/), CodeBlue (http://fiji.eecs.harvard.edu/CodeBlue), and eWatch (Maurer, Rowe, Smailagic, & Siewiorek, 2006) have very strong security control. Appropriate education about privacy control must be provided in order for mobile healthcare service to be accepted by patients. In additions, MWS systems must be flexible enough to adjust or compromise to some extent as necessary (Al Ameen et al., 2012). For healthy people using MWS who are not patients, motivation to provide personal data to service providers is lower because they are not dealing with severe diseases or symptoms. Therefore, for the success of mobile wellness care services, the privacy calculus model may be useful to establish a well-designed privacy management mechanism that addresses the needs of both patients and healthy users.

2.2. Feature selection problem

Feature selection is very important for improving the performance of classification systems in many areas such as finance and marketing (Petr, 2013), image processing (Chu, Hsu, Chou, Bandettini, & Lin, 2012), and speech recognition (Flynn, 2012). In general, feature selection algorithms can be categorized into two types: filter and wrapper. The main distinction between the two is that filter algorithms select the feature subset before applying any classification algorithms. Using statistical properties of the various features, the filter eliminates less important features from the subset. For the filter approach, the p-test is often used for feature ranking to evaluate the power of each feature based on training data (Yang, 2013). A higher p-score indicates less overlap between the positive and negative samples in terms of statistical distribution, which means that a feature is able to provide a less ambiguous signal. On the other hand, wrapper feature selection algorithms select feature subsets according to the accuracy of the training data. The classification model can then be learned and tested using the training data (Guyon & Elisseeff, 2003).

The relationship between the feature selection method and the performance of classification algorithms has been examined in prior studies. The results implied that appropriate methods of feature selection improved classification accuracy regardless of the sample size (Chu et al., 2012), and that the informativeness of features in a given data set also affected the performance of classification algorithms. Moreover, regardless of feature selection method, larger sample sizes yielded better performance in terms of accuracy. In general, as the amount of training data increased, the difference in accuracy with and without feature selection was reduced.

However, prior studies have not considered how the privacy concerns associated with each feature affect the performance of