Int. J. Human-Computer Studies 83 (2015) 43–50

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

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

A Bayesian approach to predicting website revisitation on mobile phones ☆

Jeffrey C. Zemla [a,*], Chad C. Tossell [b], Philip Kortum [c], Michael D. Byrne [c,d]

[a] Department of Cognitive, Linguistic, & Psychological Sciences, Brown University, Providence, RI, United States
[b] United States Air Force Academy, Colorado Springs, CO, United States
[c] Department of Psychology, Rice University, Houston, TX, United States
[d] Department of Computer Science, Rice University, Houston, TX, United States

ARTICLE INFO

Article history:
Received 12 June 2014
Received in revised form
3 June 2015
Accepted 4 June 2015
Communicated by Duncan P. Brumby
Available online 12 June 2015

Keywords:
World wide web
Mobile phones
Computational modeling

ABSTRACT

Mobile web browsing is highly recurrent, in that a large proportion of user's page requests are to a small set of websites. Despite this, most mobile browsers do not provide an efficient means for revisiting sites. Although significant research exists on prediction in the personal computer realm, little work has been done in the mobile realm where physical constraints of the device and mobile browsing behaviors are vastly different. The current research proposes a Bayesian model approach, based on a cognitive model of memory retrieval that integrates multiple cues in order to predict the next site a user will visit. These cues include frequency of site visitation, the recency of site visitation, and the context in which specific sites are accessed. The model is assessed using previously collected web logs from 24 iPhone users over the course of one year. Our model outperforms simpler models based on frequency or recency, which are sometimes implemented in desktop browsers. Potential applications of the model are discussed with the objective of increasing browsing efficiency on mobile devices.

1. Introduction

Previous research has revealed strong regularities in the way people retrieve information on the web. Statistical patterns underlie the number of links a user follows within a webpage (Huberman et al., 1998), how often people revisit webpages (Tauscher and Greenberg, 1997), and how people select links on a webpage (Fu and Pirolli, 2007). An understanding of patterns in web browsing behavior can be useful for developing a more efficient browsing experience. Indeed, models of web browsing behavior on personal computers (PCs) have been used to provide more personalized user support for revisits to web pages (Obendorf et al., 2007), optimize caching (Yang and Zhang, 2003), and improve search (White and Drucker, 2007). These improvements may attenuate usability problems such as page loading delays, once labeled as the primary usability problem for the web on the PC (Sears et al., 1997).

Now, smartphones allow users to access the web anywhere without having to retreat to a PC. While many websites are optimized for viewing and interaction on smaller mobile devices, smartphones have noted usability problems with page loading delays (Tossell et al., 2012a; Oulasvirta et al., 2005) similar to those faced by PC users in the 1990s. Even the simplest and most common tasks are significantly less efficient than the same tasks performed on PCs (Tossell et al., 2010).

The goal of the current study is to develop and assess predictive models of web use on iPhones in order to attenuate problematic page loading delays. A logs-based approach is used to collect real usage data from iPhone users "in the wild" over the period of one year. These data are modeled using a technique applied to both human cognitive processes and information foraging that leverages context and usage history. The fit of these models is examined to assess the utility of the predictive technique for providing a more efficient, personalized browsing experience.

2. Background

Early research characterizing web surfing behavior on PCs has found that over half of all page requests were to previously visited pages (Tauscher and Greenberg, 1997; Catledge and Pitkow, 1995; Cockburn and McKenzie, 2001). Despite this recurrent nature, web browsers are not always optimized for accessing previously visited material. While tools such as bookmarks and history lists are designed to facilitate revisitation, these methods are often inefficient. These lists can easily become cluttered or outdated, which may result in low

☆This paper has been recommended for acceptance by Duncan P. Brumby.
* Corresponding author.
E-mail address: jzemla@brown.edu (J.C. Zemla).

http://dx.doi.org/10.1016/j.ijhcs.2015.06.002
1071-5819/

usage (Aula et al., 2005). History lists are used only by a small fraction of users (Tauscher and Greenberg, 1997) and though bookmark usage is more common, many users express frustration at the need to keep them organized (Abrams et al., 1998).

The problems associated with bookmarks and history lists have led users to adopt alternative strategies for revisiting websites, such as e-mailing URLs to themselves or searching Google with appropriate keywords (Aula et al., 2005). One might suspect that with an increase in high-bandwidth internet connections over the past decade, a web search may in fact be an efficient method for revisiting pages. However, even this may be problematic in some circumstances, particularly with inexperienced users. In 2010, for instance, many users voiced their outrage over Facebook's new layout after trying to access the site from a Google search for "facebook login." In fact, what users thought was the "new Facebook" was actually a blog article about Facebook which had topped the list of search results (Melanson, 2010).

Revisiting websites on a smartphone is typically less efficient than on a PC due to the nature of smartphone usability. Although many web sites provide mobile-optimized versions of their site, navigation can still be difficult. Limited screen real estate means users must scroll and zoom to find information, and often encounter slow page loading delays over mobile networks. While there is still a significant revisitation rate on mobile phones, reliance on browser features such as bookmarks is uncommon (Tossell et al., 2012a), perhaps because the user interface for accessing bookmarks is more cumbersome than on a PC. Again, users often rely on Google searches to revisit sites—even for those sites which a user visits most frequently (Tossell et al., 2012a). With considerably longer page loading delays over mobile networks, this strategy is much less efficient than on a PC. Other strategies, such as directly typing in a URL, are also considerably slower on mobile phones (Sauro, 2010). It appears that there are currently very few options for users to quickly access previously visited material on a smartphone. This is especially problematic given that mobile phones are sometimes used in short intervals compared to desktop computers (Cui and Roto, 2008), and brief interactions account for up to a third of smartphone sessions (Oulasvirta et al., 2012).

Mobile research has sought to mitigate these usability deficiencies through the exploitation of cross-device web usage and context. For example, Kane et al. (2009) found 75% of sites visited on users PCs are also accessed on their associated mobile device. History lists on users PCs can be shared with their mobile devices for quicker retrieval on-the-go. A few lines of research have focused on the use of context to enhance application usage (see Baldauf et al., 2007; Chen and Kotz, 2000 for reviews). Some of these examples are rather mundane: exploiting temporal context (e.g., time of day) to prompt a reminder from the calendar application. Other examples are more novel: leveraging location-based information and activity recognition to generate recommended points of interest, tourist info, and train schedules on an application (de Pessemier et al., 2014). However, context is often overlooked in models of web site revisitation.

We propose a Bayesian method to integrating multiple cues in order to predict the next site a user will visit. This approach is based on a cognitive model of memory retrieval (Anderson and Schooler, 1991; Anderson and Lebiere, 1998) that has since been applied to other information retrieval systems (e.g., Stanley and Byrne, 2013). Pitkow (1997) has previously noted the relevance of this model to information retrieval on the web. The current research extends this work by formally testing this model on a real-world dataset, while also incorporating additional contextual cues (Anderson et al., 2004).

Tauscher and Greenberg (1997) proposed several methods for ordering history lists on a PC, including simple methods such as ordering by frequency or recency of page visits. These methods were evaluated on their ability to predict future website revisitation, and thus help users quickly access previously visited material. These methods provide a useful benchmark to assess more complex models of website revisitation.

Other models of website revisitation have been proposed in the literature (e.g., Fitchett and Cockburn, 2012), and several modern web browsers already include an algorithm for predicting revisitation. For example, Mozilla Firefox uses an algorithm called *frecency*, a portmanteau of frequency and recency, which offers page suggestions on new tabs and URL predictions when users type in the address bar (Connor et al., 2010). An additional algorithm has been proposed by the Mozilla team to replace the frecency algorithm (Ruderman, 2014). We test both of these models on our data set to offer representative comparisons to our approach.

Unlike many previous studies, the current research focuses on predicting revisitation to *sites* (i.e., domains and sub-domains) rather than individual *pages*. This approach has been found to be more suitable given the increasingly ephemeral nature of the web. Indeed, Weinreich et al. (2008) examined PC-based web logs and found that a large proportion of web events were accessing dynamic pages and web applications. In the latter, URLs are often not informative at later time points: URLs are frequently generated dynamically, the content of these pages change frequently over time, and revisits may redirect to a home page or log-in screen. In the mobile space, the amount of visits to dynamic pages and web applications is exacerbated (Tossell et al., 2012a). Search has become more fundamental to mobile web usage relative to browsing (Church et al., 2007). Tossell (2012) found that search activities consumed over 30% of all web usage. As opposed to PC searches, mobile searching is often triggered by contextual factors (Teevan et al., 2011). The revisitation rate to pages has decreased significantly over time (Zhang and Zhao, 2011), from 58% in 1997 (Tauscher and Greenberg, 1997) to 46% for desktop usage ten years later (Obendorf et al., 2007) to 25–35% on mobile phones (Kane et al., 2009; Tossell et al., 2012a). Conversely, site revisitation rates have increased, from 70% for desktop browsing (Obendorf et al., 2007) to as high as 90% for smartphone browsing (Tossell et al., 2012a). Based on these findings, researchers have recommended pointing users to top-level sites to access dynamic content within those sites.

3. Methods

A field study was conducted as part of a larger evaluation of Internet use on smartphones. Data were collected using the LiveLab software (Shepard et al., 2010). More complete details about the current dataset are available in Tossell et al. (2012a). A short review of the methodology is reported below for convenience. An anonymized copy of the dataset is available at URL http://livelab.recg.rice.edu/.[1]

3.1. Participants

Twenty-four students (14 males, 10 females; mean age 19.2 years old) were recruited to participate in a longitudinal study on smartphone usability. Participants were not paid, but instead given an iPhone 3GS running iOS 3.1.3 to use as their primary phone. Participants were allowed to keep the phone at the completion of the study, which lasted one year.

[1] Due to a technical error, one user's data contains 12 fewer URLs than the dataset used in this study.