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.

3.2. Procedure

A logger installed on these iPhones ran as a background process during the study period. It did not interrupt usage or require user activity in order to record data. Instead, the logger automatically uploaded usage data to a server on a daily basis. The participants were not given any instructions on how they use their instrumented devices. The logger automatically hashed communications data. Thus, privacy was maintained in the collection of SMS, voice phone, and e-mail data allowing users to interact with their devices more naturally (Tossell et al., 2012). An informed consent document explained exactly what data were collected by the logger and was signed by each participant in this study.

The focus of this study is solely on the web logs collected from the Safari web browser. The custom logger captured user interactions on the web by accessing the history file once per day. Data collected from URL visits included the date and time of each URL visit, the name of the URL visited, and the referring URL. Client-side data such as the buttons selected to access the URLs (e.g., frequency of back arrow pushes) was not collected due to technical restrictions in the Safari application. Similarly, the method used to access the page (manual text entry, link click, bookmark) was not recorded.

One important technical note regards the way Safari logs web history. When a user re-visits a web page, Safari increments a frequency counter but only preserves the timestamp for the most recent visit to that site. Since log data was collected nightly, if a user visits the same page multiple times in a day, timing data is lost for all but the most recent visit. By comparing the count entry to the number of available temporal data, we calculated that roughly 5–6% of the data was missing in timestamps. We have excluded this portion of the data from further analysis. As a result, the data do not represent an entirely veridical representation of the user's browsing history. However, we do not have reason to believe that this biases one model over another.

4. Model overview

A variety of cues can be used to predict which website a user will visit next. These cues can be classified into two groups: *contextual cues* and *historical cues*. Contextual cues reflect the current state of the user and the device. For example, physical location is a contextual cue because a user's browsing habits may differ when he is at work or home. In contrast, historical cues reflect a user's past browsing history. For example, the frequency or recency with which a site is visited would be considered historical cues.

We present results from seven models. Three baseline models leverage only a single cue. These are the Recency, Frequency, and Context models. Three additional models integrate both frequency and recency in making predictions. The Frecency model (Connor et al., 2010) is based on the algorithm implemented in Mozilla Firefox, and the New Frecency model (Ruderman, 2014) is another algorithm proposed by Mozilla. We use these as benchmarks to compare our proposed History model, which borrows an algorithm from cognitive psychology (Anderson et al., 2004) and is described below. Finally we test a History/Context model which augments the History model by incorporating a limited amount of context.

4.1. Proposed model

Determining which site a user will visit next can be framed in terms of Bayesian inference: given a site's prior *visitation history* and the likelihood of visiting that site in the *current context*, how likely is it that the user will visit that site next? The solution to this problem can be determined through Bayes' rule:

$$P(S_i|H_i \cap C) = \frac{P(C|S_i \cap H_i)P(S_i|H_i)}{P(C|H_i)} \tag{1}$$

In Eq. (1), $P(S_i|H_i \cap C)$ represents the probability that a particular website S_i is visited next, given the current context C, and the site's visitation history H_i. Though it is often more convenient to work with the equivalent log odds version of Eq. (1):

$$\underbrace{\ln\frac{P(S_i|H_i \cap C)}{P(\neg S_i|H_i \cap C)}}_{\mathbb{O}} = \underbrace{\ln\frac{P(C|S_i \cap H_i)}{P(C|\neg S_i \cap H_i)}}_{\mathbb{C}} + \underbrace{\ln\frac{P(S_i|H_i)}{P(\neg S_i|H_i)}}_{\mathbb{H}} \tag{2}$$

For convenience, we refer to different portions of this equation as the odds term ($\mathbb{O}$), the context term ($\mathbb{C}$), and the history term ($\mathbb{H}$). Given this formulation, the site which has the highest probability of being visited next is that which has the highest posterior odds. Thus, to find the website S_i that is most likely to be visited next, we can select the maximum a posteriori (MAP) over all i:

$$\arg\max_{S_i} \ln\frac{P(S_i|H_i \cap C)}{P(\neg S_i|H_i \cap C)} \tag{3}$$

The difficulty in this problem lies in the characterization of visitation history (H_i) and the current context (C). For instance, visitation history might be operationalized as frequency of visits to a site, or recency of visits to a site, or some combination of the two. Likewise, the current context might include the page the user is currently on and the current physical location of the device (GPS coordinates). We will consider some of these characterizations in sequence, but first it is useful to define some commonalities between the potential models we will be comparing.

4.2. Model commonalities

Below we describe a set of models used to predict which website a user will visit next. All models were evaluated by analyzing log files and did not intervene with the user's behavior.

At each site transition, the models predict the top four sites that the user is most likely to visit next (i.e., the four sites with the largest posterior odds). A site transition is defined as moving from one domain or sub-domain to another. As such, transitions within a site (e.g., navigating from one Wikipedia page to another) are ignored by the model: no prediction is made, and the frequency count for that site is not increased. Intra-site navigation is ignored specifically because no click effort is saved by a predictive model when the user is searching for a link on the current page. Additionally, a user's current site (the site currently loaded in the browser) is always excluded from the list of predictions. If the current site has a posterior odds in the top four, it is replaced by the next most likely site such that four predictions are always made.

Prediction accuracy is defined as the percentage of time that one of the four predicted sites was in fact the next site visited. Prediction accuracy is calculated individually for each user and then averaged across users to calculate model accuracy. Later, we examine how model performance is affected by the number of predictions made, e.g., predicting two sites instead of four.

5. Results

5.1. Overview

An overview is presented in Fig. 1 and details regarding each model are reported in the sections below.

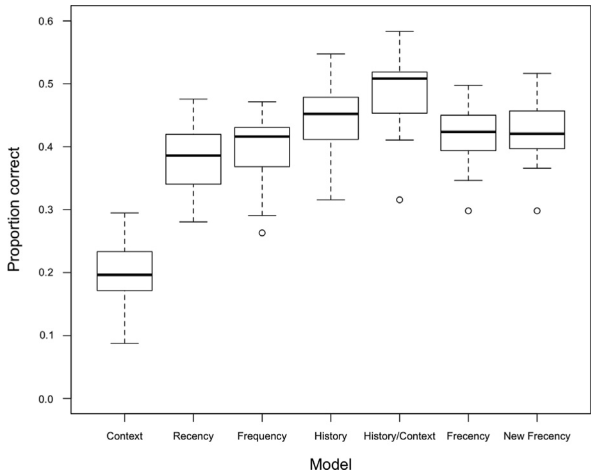

Fig. 1. Each box plot represents the distribution of predictive accuracies across 24 subjects for each model type. For each model, four predictions are generated ($N=4$).

We tested three baselines which incorporate only a single cue (Recency, Frequency, and Context). We found that the most recently visited site (context model) is the least predictive cue, whereas frequency and recency provide substantially better predictive accuracy. We also test three models which integrate both frequency and recency (History, Freceny, and New Frecency). Our proposed model, the History model, statistically outperforms the other two and achieves 44.4% accuracy. Finally, we test a model which incorporates frequency, recency, and context, and found that this model is able to successfully boost performance, achieving 48.9% accuracy.

In total, 42,724 unique pages (i.e., full URLs) were recorded in the log. The number of unique pages per user ranged from 112 to 4948 (median: 1511). 58,021 page transitions were logged, ranging from 120 to 5701 per user (median: 1854). This results in roughly an average of 26% revisitation rate per user.

4724 unique sites (i.e., domains) were recorded in the log. The number of unique sites per user ranged from 37 to 671 (median: 235). The number of inter-site transitions totaled 26,345, ranging from 59 to 2934 per user (median: 939). This results in a site revisitation rate of over 80% per user on average.

Users relied on search heavily, issuing over 17,500 searches across the entire study period. 56% of browsing sessions (defined as activity between when a browser is opened until it is closed) consisted of at least one search.

5.2. Frequency model

One of the simplest models we can test is a frequency model, in which the four most frequently visited sites are predicted each time the user visits a new site. Context is ignored, and thus the model is simplified to:

$$\mathbb{O} = \mathbb{H} \tag{4}$$

Here, our characterization of $\mathbb{H}$ assumes that *only* frequency information is important, and thus any other information related to a site's visitation history, including temporal information, is ignored.

Overall, this model does a decent job of predicting the next site the user will visit. On average, 39.6% the sites visited by the user are correctly predicted by the model.

5.3. Recency model

Another baseline model that we can test is a recency model. In this case, the four most recently visited sites are predicted. This characterization ignores frequency information and context. In this case, however, our characterization of $\mathbb{H}$ now reflects the recency of visits rather than the frequency of visits.

The accuracy of this model is slightly worse than the frequency model, predicting 38.2% of site transitions. A paired t-test comparing the accuracy of recency and frequency models for all 24 subjects suggests that the frequency model may be favorable to the recency model, however a two-tailed test does not quite reach significance, $t(23)=1.88, p=.07$.

5.4. Context model

In addition to historical cues such as recency and frequency, mobile browsing behavior may be influenced by a variety of contextual cues. Possible contextual cues include the user's current location, time of day, or the most recently-visited site. We focus on only one contextual cue which we felt would be most predictive: the most recently-visited site. Previous research has indicated that this cue is a reliable indicator of website revisitation (Pitkow and Pirolli, 1999; Su et al., 2000).

To establish the effect of context, we first look at a model which incorporates context but ignores prior history. That is:

$$\mathbb{O} = \mathbb{C} \tag{5}$$

In this case, $\mathbb{C}$ can be computed using the user's visitation history up until the current point. Formally, $\mathbb{C}$ is equal to the log odds of arriving at S_i from C as opposed to arriving at $\neg S_i$ from C.

Overall, the context model performs relatively poorly, achieving an average of only 19.9% accuracy. This demonstrates that this contextual cue is likely to account for less of the variance than history when predicting site visitation. However, if the successful predictions made by the context model are different than (or only partially overlap with) the successful predictions made by historical cues, it is possible that we can increase prediction accuracy by examining the full model established in Eq. (2).

5.5. History model

A better characterization of $\mathbb{H}$ should take into account not only frequency of visitations, but also the recency of each visit. The form that this function should take is not obvious; however, this problem has been addressed before within the information retrieval literature. Anderson and Schooler (1991) noted two important characteristics present in many informational systems. First, the odds that a particular piece of information will be needed in the future decays as a power function of the time since it last appeared. Second, the odds that a particular piece of information will be needed in the future increases as a power function of how many times it has appeared in the past. These patterns have been observed in a variety of systems such as library book loans (Burrell, 1980), e-mail content (Anderson and Schooler, 1991), and human memory (Anderson et al., 2004). These two observations can be captured in a single equation where the odds of needing a piece of information is represented as a linear summation of odds from all previous presentations:

$$\mathbb{O} = \ln \sum_{j=1}^{n} t(S_{i_j})^{-d} \tag{6}$$

The right-hand side of Eq. (6) captures the aforementioned observations. The odds of needing a particular item S_i increases with the number of observations n, as designated by the summation. However, the contribution of each observation decays as a power function since the time elapsed since that observation, $t(S_{ij})$. The parameter d is a free parameter that represents the rate of decay, and later we examine how model performance changes as a function of this decay rate. For an initial analysis, we set d to .5, as this decay rate has been used often to model a wide range of phenomenon related to the accessibility of human memory (Anderson, 2007).

Pitkow (1997) has previously shown that both the frequency and recency of web page revisitation can be characterized by power distributions. However, while Pitkow characterizes frequency and recency of revisitation as following power distributions, the fit of a model which integrates these observations (i.e., Eq. (6)) was not assessed.

In our model, the log odds of visiting each site are recomputed using Eq. (6) after each visitation. The time $t(S_{ij})$ is calculated as the amount of time that has passed since that particular visit. Thus, the computed odds reflect the sum of all temporally discounted visits to that site. For now, context is still ignored. Once again, only the characterization of $\mathbb{H}$ has changed.

This model achieves an average prediction accuracy of 44.4%, which is statistically better than the frequency model, $t(23) = 8.27, p < .001$. In fact, the prediction accuracy of the History model is equal to or better than the prediction accuracy of the frequency and recency models for all 24 subjects in the dataset.

As seen for a sample user in Fig. 2, the distribution of posterior log odds in this model roughly follows a power distribution. It is apparent that only a few sites have very high odds, whereas there is a long tail of sites with low odds.

5.6. History/context model

Now that we have characterized both visitation history ($\mathbb{H}$) and context ($\mathbb{C}$), we are in a position to test the History/Context model using Eq. (2). Though the History model achieves good predictive accuracy, the addition of context may improve it further. In particular, those sites in the long tail of the History model (Fig. 2) are unlikely to be predicted by the history model, yet may be predicted by the context model. A model that incorporates both historical and contextual cues could compensate for this.

One complication is that the numerator of the context term of Eq. (2) is sometimes zero, because a user has never transitioned from C to S_i. This presents a problem in the current model, where the log of each term must be computed to integrate the context and history terms, since the log of zero is undefined. To remedy this, a Laplace correction was used by adding .01 to the numerator and denominator. Thus if a site S_i has never been visited from C, it is implicitly assumed that the base rate probability of this transition is $.01/(N+.01)$ where N is the number of prior visits to S_i. Note that in the context model, a Laplace correction can be avoided by simply not taking the logarithm, which does not affect rank ordering. However in the History/Context model it is necessary to take the log in order to integrate the history and context terms of the model.

The History/Context model results in an average prediction accuracy of 48.9%, which is statistically better than the History model, $t(23) = 6.43, p < .001$. For 22 of 24 users, the History/Context model performs equal to or better than all other models that were tested.

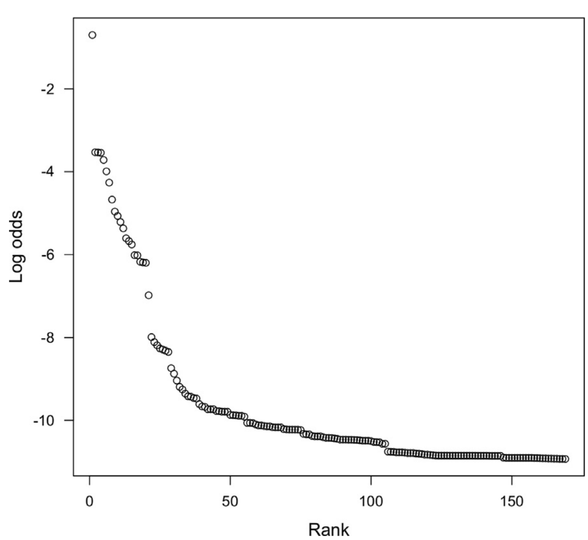

Fig. 2. The distribution of posterior log odds for each site is shown for a sample participant. Each circle represents a site in the user's data log. Only a few sites have very high odds because they have been accessed both recently and frequently. Odds shown are those calculated at the very end of the dataset.

5.7. Frecency model

While the History model appears to perform well on our data set, comparisons to the Recency and Frequency models should be interpreted cautiously. Several modern web browsers include algorithms for ranking web sites that are not strictly recency- or frequency-based. One such algorithm is the Frecency algorithm (Connor et al., 2010), implemented in Mozilla's Firefox.

As with the History model, a site's score is represented by the temporally discounted sum of all visits to that site. However instead of using a power function, the algorithm uses a step-function to assign weights based on the recency of each visit. Specifically, visits are separated into one of the five discrete bins which determine its weight (see Table 1).

In Mozilla's implementation, each visit's weight is multiplied by a bonus score that is determined by how the site was accessed; e.g., a URL that was typed receives a bonus of 2, whereas URLs that are clicked as links receive a bonus of 1.2. A site's value is determined by the sum over all visits:

$$Score_i = \sum_{j=1}^{n} bonus_j * weight_j \qquad (7)$$

This model shares many similarities to the History model. Notably, it integrates both recency and frequency, and decays in a sub-exponential manner. Unfortunately, our logs did not contain enough information to compute the bonus score for each visit, as we did not record method of access. Instead, we tested a simplified version of the model in which a constant bonus score was used for all visits, as has been suggested elsewhere (Fitchett and Cockburn, 2012).

The Frecency algorithm results in an average prediction accuracy of 42.1%. This algorithm is directly comparable to the History model in that it incorporates recency and frequency, but not context. The History model achieved higher prediction accuracy, and outperformed the Frecency model for 22 of 24 subjects, $t(23) = 6.65, p < .001$.

5.8. New frecency model

Finally, we test an alternative Frecency algorithm proposed by a member of the Mozilla Firefox team (Ruderman, 2014). The proposed algorithm abandons the binning structure in favor of a continuous

Table 1
Frecency bins.

Time	Weight
0–4 days	100
4–14 days	70
14–31 days	50
31–90 days	30
90+ days	10

exponential decay model. While an exponential decay function may have some computational advantages, the choice is at odds with previous literature showing that website revisitation is better characterized by power-law decay (Pitkow, 1997; Dezsö et al., 2006), as are similar information retrieval systems (Burrell, 1980; Anderson and Schooler, 1991; Adamic and Huberman, 2002).

The relationship is governed by the following equation:

$$Score_i = \sum_{j=1}^{n} bonus_j * e^{-\lambda t(S_{ij})} \tag{8}$$

where

$$\lambda = \frac{\ln 2}{30 \text{ days}} \tag{9}$$

Again, we did not have enough information to calculate the bonus score for each visitation and thus weighted each visitation equally. This algorithm results in an average prediction accuracy of 42.4%, slightly higher than the original Frecency algorithm. However, the History model outperforms the New Frecency model, achieving higher prediction accuracy for 22 of 24 subjects.

6. Summary

Seven different models for predicting website revisitation were tested and compared. The prediction accuracies of each model are shown in Fig. 1. A limited context model, which takes into account only site to site transitions, fares relatively poorly, whereas models that incorporate historical cues offer better predictive accuracy. Three models incorporate frequency and recency. Our History model outperforms reduced versions of the Frecency model used in Firefox, as well as a potential replacement for Frecency. The History/Context model is the only model to incorporate all three suggested cues, and achieves the highest level of predictive accuracy while maintaining relatively low variance among 24 subjects. This suggests that the model can perform robustly across users regardless of individual differences in web surfing behavior.

6.1. Reduced assumptions

Up until now, we have only examined the prediction accuracy of each model under the assumption that the top four sites are counted as successful predictions. Fig. 3 shows how model performance changes as a function of the number of predictions made by the model. By definition, increasing the number of predictions increases model accuracy. However in all models, this increase appears to be non-linear. Each successive prediction offers only a modest increase in predictive accuracy.

One additional assumption we made concerning the History and History/Context models is an arbitrary setting of the decay weight (d). Though we chose a value consistent with what has been used in previous research, this value is likely to vary depending on the domain of interest. Fig. 4 demonstrates that prediction accuracy is marginally affected by the choice of decay rate, though accuracy remains high across a range of possible decay values.

7. Discussion

The History/Context model reviewed here provides a simple and well-established Bayesian framework for integrating contextual and historical cues to predict website revisitation. To account for historical cues, we implemented an algorithm used in cognitive psychology (Anderson and Schooler, 1991) and demonstrate its effectiveness in predicting website revisitation.

7.1. Additional cues

Of the models we tested, only a single contextual cue is tested: the most recently visited site. It is very likely that additional cues that we did not test include could improve the fit of the model. For instance, Rahmati et al. (2012) highlight the importance of contextual relevance of web browsing. Additional cues such as time of the day (morning, afternoon, evening), day of the week (weekend, weekday), and physical location may also influence web browsing behavior.

Mozilla's Frecency algorithm includes information regarding method of accessing a URL (e.g., typed, bookmark, click, etc.) which seems particularly well suited for reducing click effort. Additional mobile cues might include the most recent voice call or most recent app used.

In our model, we explored only a limited amount of cues to establish the effectiveness of integrating history and frequency using a specific algorithm, and incorporated a single contextual cue to validate the potential for easily including additional cues. Incorporating more cues in a similar manner is likely to produce modest increases in predictive accuracy at the cost of model

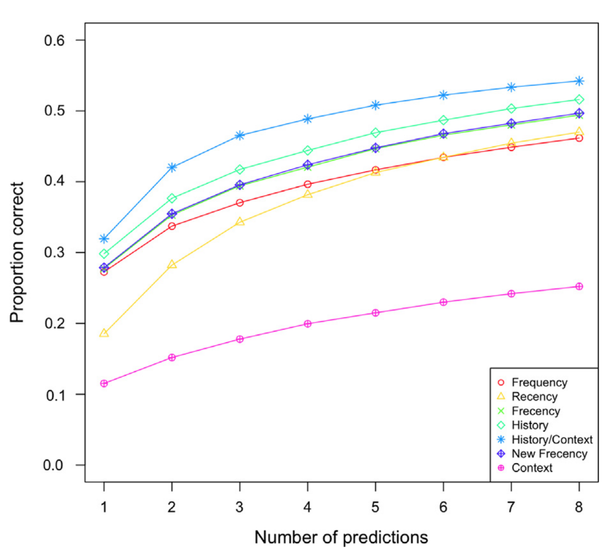

Fig. 3. The accuracy of each model type is depicted as a function of the number of predictions, N.

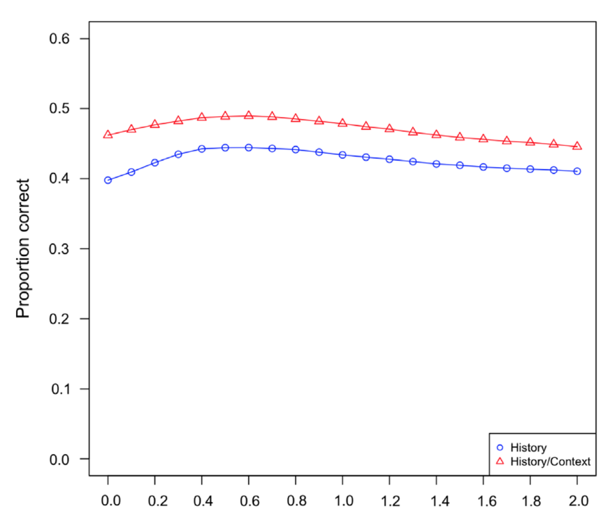

Fig. 4. The decay rate parameter (d) affects the prediction accuracy of the History and History/Context models. However, both models perform robustly across a range of decay rates.

complexity. The challenge in incorporating additional cues lies in discretizing them in a way that allows us for computing conditional probabilities (i.e., binning).

The History/Context model reviewed here performs well in spite of the fact that it has only two parameters: a decay rate and a Laplace correction. In our initial analysis both of these parameters were fixed, though subsequent analysis shows model performance is relatively stable across different decay rates. Model performance is also unlikely to vary significantly with such a small Laplace correction value. Additional parameters might be considered, such as a weighting parameter for context and history; currently these cues are weighted equally (as suggested by a naive Bayes model). If these parameters are computed for each user individually, the model's performance may improve by accounting for individual differences in web surfing behavior.

7.2. Storage and performance concerns

One concern is that the History and History/Context models leverage a user's entire web history to generate predictions. This may not always be possible for a variety of reasons, including performance, storage, or privacy concerns. Indeed, performance appears to be a factor in designing the Frecency algorithm, presumably for performing online prediction in Firefox's address bar (Ruderman, 2014). One possible alteration to the History/Context model is to use an approximation to Eq. (6) that has been used in the cognitive science literature (Anderson et al., 1998; Anderson, 1993), which requires keeping track of the total number of visits to a particular site but only a timestamp for the most distant visit. This approximation works by assuming visitations are evenly spaced from the first visitation to the current time. This approximation significantly reduces the computational cost, and also reduces the size of the log file. However, this approximation is occasionally inadequate for some domains (Sims and Gray, 2004; Stanley and Byrne, 2014). Another approximation exists which can trade-off fidelity of the original equation with the computational ease of its approximation with a free parameter (Petrov, 2006). For the current data set, the total number of site presentations was sufficiently small that performance and storage issues were not a concern.

Another alternative when timestamps are not available is to use only the temporal ordering of the sites; that is, measure time using the discrete number of sites since a particular visitation. Our results show that this change reduces the prediction accuracy of the models modestly, to 42.9% for the History model and 47.1% for the History/Context Model.

7.3. Possible implementations

A predictive model can be implemented in a variety of ways. One useful technique may be to provide users with a springboard—a list of sites or thumbnails that the user is most likely to visit next—upon opening the browser or a new tab. This technique is used in some PC browsers, such as Google Chrome. Teevan et al. (2009) have shown that providing visual snippets of previously visited sites can efficiently support revisitation to web pages. Other use cases may include ordering websites in the URL entry box when the user prepares to manually type a URL; predictive ad placement, wherein ads are chosen based on where the user is most likely to visit next; or pre-caching pages which have a high probability of being visited, thus improving page load times (such as in Lymberopoulos et al., 2012).

7.4. Limitations

Our choice in methodology has several limitations. In relying strictly on web logs, user behavior may already be biased in certain ways based on the infrastructure of the phone and browser. For instance, Safari has an autocomplete feature that saves users effort when typing in a URL; the list and ordering of suggestions may actually affect which site the user visits next. Additionally, many popular websites provide a native app which can be used in place of the website. Thus, for example, a user may no longer wish to visit http://www.facebook.com once she installs the Facebook app. Although these limitations are inherent, the web logs do reflect a subject's natural usage of the browser. That is, while features such as autocomplete may bias user behavior, these features are unlikely to be removed. Web logs provide a good characterization of how users navigate the web when taking into account these biases. Nonetheless, it is important to actively re-assess the model's predictive performance after making substantive changes to the user interface or other components that may alter the user's browsing behavior.

Our results may also appear optimistically biased, due to our choice to predict site revisitation rather than page revisitation. Our choice was motivated largely by the observation that site revisitation on mobile phones is very high and page revisitation is fairly low. The growing adoption of web applications with dynamic URLs may have affected this. In general, an algorithm which only makes page-to-page predictions would fail to make any prediction at all when the active page is being visited for the first time. Consider the situation in which a user frequently transitions from booking a flight to booking a hotel. The URL for a flight's booking confirmation is unique and does not have much future importance. Likewise, the details of a particular hotel booked in the past are not particularly important. Yet the association between the flight booking website and the hotel booking website is captured by an algorithm which predicts site-to-site transitions. Conversely, there are other situations in which pages are more important than sites. For instance, a user may frequently navigate between specific pages in a technical manual. Here a model that predicts only site-to-site transitions will fail.

An intriguing possibility is to integrate site and page predictions by determining which pages are fairly static in nature (e.g., Wikipedia pages), and future research should explore methods for integrating site and page prediction. Alternatively, there may be other use cases in which site-to-site transitions are uniquely practical or valuable. For example, a tree-like navigation structure in which potential sites are suggested to a user first followed by page suggestions after a user makes an initial selection allows for completely keyboard-less navigation so long as the user is revisiting a previous page. In another use case, predictive ad placement may benefit more from site-to-site predictions when advertisers want users re-directed to a top level page.

8. Conclusion

Navigating the web on mobile phones presents unique challenges which lead to an inadequate browsing experience. Mobile users visit fewer sites overall compared to desktop users, yet they revisit sites with great regularity. These patterns suggest that perhaps predicting web site revisitation can improve the user experience. A Bayesian algorithm borrowed from the memory retrieval literature shows promising results in being able to predict a significant portion of revisitation on mobile phones.

Acknowledgments

This work was supported in part by a National Science Foundation grant NSF IIS/HCC 0803556. The authors would also like to thank three anonymous reviewers for critical feedback that greatly improved the manuscript.

References

Abrams, D., Baecker, R., Chignell, M., 1998. Information archiving with bookmarks: personal web space construction and organization. In: Proceedings of the SIGCHI Conference on Human factors in Computing Systems, pp. 41–48.

Adamic, L.A., Huberman, B.A., 2002. Zipfs law and the internet. Glottometrics 3 (1), 143–150.

Anderson, J.R., 1993. Rules of the Mind. Psychology Press, New York, NY.

Anderson, J.R., 2007. How can the Human Mind Occur in the Physical Universe?. Oxford University Press, New York, NY.

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y., 2004. An integrated theory of the mind. Psychol. Rev. 111 (4), 1036–1060.

Anderson, J.R., Bothell, D., Lebiere, C., Matessa, M., 1998. An integrated theory of list memory. J. Mem. Lang. 38 (4), 341–380.

Anderson, J.R., Lebiere, C.J., 1998. The Atomic Components of Thought. Erlbaum, Mahwah, NJ.

Anderson, J.R., Schooler, L.J., 1991. Reflections of the environment in memory. Psychol. Sci. 2 (6), 396–408.

Aula, A., Jhaveri, N., Käki, M., 2005. Information search and re-access strategies of experienced web users. In: Proceedings of the 14th International World Wide Web Conference, pp. 583–592.

Baldauf, M., Dustdar, S., Rosenberg, F., 2007. A survey on context-aware systems. Int. J. Ad Hoc Ubiquitous Comput. 2 (4), 263–277.

Burrell, Q., 1980. A simple stochastic model for library loans. J. Doc. 36 (2), 115–132.

Catledge, L.D., Pitkow, J.E., 1995. Characterizing browsing strategies in the World-Wide Web. Comput. Netw. ISDN Syst. 27 6, 1065–1073.

Chen, G., Kotz, D., et al., 2000. A Survey of Context-aware Mobile Computing Research. Technical Report TR2000-381. Department of Computer Science, Dartmouth College.

Church, K., Smyth, B., Cotter, P., Bradley, K., 2007. Mobile information access: a study of emerging search behavior on the mobile internet. ACM Trans. Web (TWEB) 1 (1), 1–38.

Cockburn, A., McKenzie, B., 2001. What do web users do? An empirical analysis of web use. Int. J. Hum. Comput. Stud. 54 (6), 903–922.

Connor, M., Spitzer, S., Ayala, D., Lee, E., 2010. Frecency Algorithm. 〈https://developer.mozilla.org/en-US/docs/Mozilla/Tech/Places/Frecency_algorithm〉 (Retrieved February 2, 2015).

Cui, Y., Roto, V., 2008. How people use the web on mobile devices. In: Proceedings of the 17th International Conference on World Wide Web. ACM, Beijing, China, pp. 905–914.

de Pessemier, T., Dooms, S., Martens, L., 2014. Context-aware recommendations through context and activity recognition in a mobile environment. Multimed. Tools Appl. 72 (3), 2925–2948.

Dezsö, Z., Almaas, E., Lukács, A., Rácz, B., Szakadát, I., Barabási, A.-L., 2006. Dynamics of information access on the web. Phys. Rev. E 73 (6), 066132.

Fitchett, S., Cockburn, A., 2012. Accessrank: predicting what users will do next. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Austin, TX, pp. 2239–2242.

Fu, W.-T., Pirolli, P., 2007. SNIF-ACT: a cognitive model of user navigation on the world wide web. Hum. Comput. Interact. 22, 355–412.

Huberman, B.A., Pirolli, P.L., Pitkow, J.E., Lukose, R.M., 1998. Strong regularities in world wide web surfing. Science 280 (5360), 95–97.

Kane, S.K., Karlson, A.K., Meyers, B.R., Johns, P., Jacobs, A., Smith, G., 2009. Exploring cross-device web use on pcs and mobile devices, Human-Computer Interaction —INTERACT 2009. Springer, Uppsala, Sweden, pp. 722–735.

Lymberopoulos, D., Riva, O., Strauss, K., Mittal, A., Ntoulas, A., 2012. Pocketweb: instant web browsing for mobile devices. In: Proceedings of the 17th International Conference on Architectural Support for Programming Languages and Operating Systems, pp. 1–12.

Melanson, M., 2010. How google failed its users and gave birth to an Internet meme. 〈http://readwrite.com/2010/02/11/how_google_failed_internet_meme〉 (Retrieved May 15, 2013).

Obendorf, H., Weinreich, H., Herder, E., Mayer, M., 2007. Web page revisitation revisited: implications of a long-term click-stream study of browser usage. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 597–606.

Oulasvirta, A., Rattenbury, T., Ma, L., Raita, E., 2012. Habits make smartphone use more pervasive. Pers. Ubiquitous Comput. 16 (1), 105–114.

Oulasvirta, A., Tamminen, S., Roto, V., Kuorelahti, J., 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 919–928.

Petrov, A.A., 2006. Computationally efficient approximation of the base-level learning equation in act-r. In: Proceedings of the Seventh International Conference on Cognitive Modeling. Citeseer, pp. 391–392.

Pitkow, J.E., 1997. Characterizing World Wide Web Ecologies Ph.D. thesis.

Pitkow, J.E., Pirolli, P., 1999. Mining longest repeating subsequences to predict World Wide Web surfing. In: Proceedings of 2nd USENIX Symposium on Internet Technologies and Systems.

Rahmati, A., Shepard, C., Tossell, C., Zhong, L., Kortum, P., 2012. Practical Context Awareness: Measuring and Utilizing the Context Dependency of Mobile Usage. Technical Report, Rice University.

Ruderman, J., 2014. New Frecency Algorithm. 〈https://wiki.mozilla.org/User:Jesse/NewFrecency?title=User:Jesse/NewFrecency〉 (Retrieved February 2, 2015).

Sauro, J., 2010. iPhone vs. Desktop: Which is faster? 〈http://www.measuringusability.com/iphone-desktop.php〉 (Retrieved May 16, 2014).

Sears, A., Jacko, J.A., Borella, M.S., 1997. Internet delay effects: how users perceive quality, organization, and ease of use of information. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 353–354.

Shepard, C., Rahmati, A., Tossell, C., Zhong, L., Kortum, P., 2010. LiveLab: measuring wireless networks and smartphone users in the field. ACM Sigmetrics Perform. Eval. Rev. 38, 15–20.

Sims, C.R., Gray, W.D., 2004. Episodic versus semantic memory: an exploration of models of memory decay in the serial attention paradigm. In: Proceedings of the International Conference on Cognitive Modeling, pp. 279–284.

Stanley, C., Byrne, M.D., 2013. Predicting tags for StackOverflow flow posts. In: Proceedings of the 12th International Conference on Cognitive Modeling.

Stanley, C., Byrne, M.D., 2014. Using large datasets to examine the performance space of ACT-R's base-level learning mechanism. Talk presented at the Midwest Cognitive Science Conference.

Su, Z., Yang, Q., Lu, Y., Zhang, H., 2000. Whatnext: a prediction system for web requests using n-gram sequence models. In: Proceedings of the 1st International Conference on Web Information Systems Engineering. vol. 1, pp. 214–221.

Tauscher, L., Greenberg, S., 1997. How people revisit web pages: empirical findings and implications for the design of history systems. Int. J. Hum. Comput. Stud. 47 (1), 97–137.

Tauscher, L., Greenberg, S., 1997. Revisitation patterns in world wide web navigation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 399–406.

Teevan, J., Cutrell, E., Fisher, D., Drucker, S.M., Ramos, G., André, P., Hu, C., 2009. Visual snippets: summarizing web pages for search and revisitation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Boston, MA, pp. 2023–2032.

Teevan, J., Karlson, A., Amini, S., Bernheim, A.J.B., and Krumm, J. 2011. Understanding the importance of location, time, and people in mobile local search behavior. In: Proceedings of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 77–80.

Tossell, C., Kortum, P.T., Shepard, C.W., Rahmati, A., Zhong, L., 2010. Assessing the performance of common tasks on handheld mobile computers. In: Proceedings of the 54th Human Factors and Ergonomics Society Annual Meeting, pp. 542–546.

Tossell, C.C., 2012. An empirical analysis of internet use on smartphones: characterizing visit patterns and user differences. Unpublished Doctoral Dissertation. Rice University.

Tossell, C.C., Kortum, P., Rahmati, A., Shepard, C., Zhong, L., 2012a. Characterizing web use on smartphones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2769–2778.

Tossell, C.C., Kortum, P., Shepard, C.W., Rahmati, A., Zhong, L., 2012. Getting real: a naturalistic methodology for using smartphones to collect mediated communications. Adv. Hum. Comput. Interact. 2012, 1–10.

Weinreich, H., Obendorf, H., Herder, E., Mayer, M., 2008. Not quite the average: an empirical study of web use. ACM Trans. Web 2 (1), 1–26.

White, R.W., Drucker, S.M., 2007. Investigating behavioral variability in web search. In: Proceedings of the 16th International World Wide Web Conference, pp. 21–30.

Yang, Q., Zhang, H.H., 2003. Web-log mining for predictive web caching. IEEE Trans. Knowl. Data Eng. 15 (4), 1050–1053.

Zhang, H., Zhao, S., 2011. Measuring web page revisitation in tabbed browsing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1831–1834.