Predicting Navigation Patterns on the Mobile-Internet Using Time of the Week

Martin Halvey Mark T. Keane Barry Smyth

Adaptive Information Cluster,
Department of Computer Science,
UCD,Ireland

martin.halvey@ucd.ie mark.keane@ucd.ie barry.smyth@ucd.ie

ABSTRACT

A predictive analysis of user navigation in the Internet is presented that exploits time-of-the-week data. Specifically, we investigate time as an environmental factor in making predictions about user navigation. An analysis is carried out of a large sample of user, navigation data (over 3.7 million sessions from 0.5 million users) in a mobile-Internet context to determine whether user surfing patterns vary depending on the time of the week on which they occur. We find that the use of time improves the predictive accuracy of navigation models.

Categories and Subject Descriptors: H5.4 Hypertext/Hypermedia, H5.m. Information interfaces and presentation, Miscellaneous, H 1.2 User/Machine Systems

General Terms: Experimentation, Browsing, Navigation, Mobile-Web

Keywords: WWW, WAP, mobile, prediction, user modeling, log file analysis.

1. INTRODUCTION

One of the key challenges in adaptive hypermedia and personalization is to improve user's online environment through the exploitation of past behavior, explicit and implicit preferences and environmental factors. Truly adaptive information systems will need to be able to predict user behavior quite accurately in order to deliver on this Holy Grail. In this paper, we examine navigation in the context of the mobile-Internet with a view to predicting user preferences based on an environmental factor (i.e.. time of the week). Cotter and Smyth [1] have proposed that users should not have just one set of preferences but rather groups of preferences, characterizing their personality in different contexts. The fundamental hypothesis behind the present work is that a given user's Internet navigation patterns during office hours will have different preferences than usage on the weekend. One of the key challenges in employing such ideas is to determine the cuts between people's distinct personas and to determine the key categories that might form the basis for user preferences.

2. Task Context

Surfing the World Wide Web (WWW) involves traversing many hyperlinks that connect vast numbers of documents, and in recent years has lead to the problem of information overload [2]. Such problems can be alleviated by adaptive systems that can predict user surfing patterns [5]. These problems become even more acute in using the mobile Internet; specifically, in this case in using Wireless Application Protocol (WAP) telephones to navigate the mobile-Web. For browsing purposes, mobile internet users also have very limited input capabilities, featuring numeric keypads with minimal text entry, unlike the mouse and keyboard options available to PC users. This means that navigation is typically carried out by scrolling through hierarchies of menus, rather than performing a search engine search and direct clicking to a relevant page. Furthermore, at present, WAP users face slow download times and incremental billing costs [4]. So, getting to the right information quickly is important to these users.

Smyth [5] has already shown that when menu-hierarchies are re-ordered to meet users preferences, based on previous navigation patterns, that WAP users find what they want more readily and, interestingly, spend more time on line as a result. However, this work never looked at whether specific environmental factors – like time of the week – could be used to predict different sets of user preferences. Intuitively, users' navigation on the weekend when they have leisure days should differ from their navigation on weekdays when they are at work. Following this intuition, our hypothesis was that users will surf differently during different time periods, and that these differences can be used to make predictions about user navigation. If people's surfing behavior can be predicted accurately then a given URL in a menu-hierarchy could be aggressively promoted to the top of the hierarchy delivering the user's desired page instantly.

To test this hypothesis, we analyzed a large data set from a mobile Internet portal. This data, gathered in September 2002, involved over 421,000 users and 3.7M individual user sessions using WAP phones to access a mobile Internet portal operated by a major European mobile service provider. Halvey et al [3] have previously shown that navigation patterns in this data set follow the same universal law of surfing that has been found in the WWW. By dividing the data set into daytime, evening and weekend sessions we endeavored to determine whether distinct navigation patterns arose and, if found, whether these patterns could be accurately predicted using Markov models. These time periods were chosen as they closely resembled time periods in the pricing plan of the mobile provider and also provided strong boundaries between work and leisure time for example.

3. Predicting User Navigation Using Markov Models

To investigate this issue we constructed k-order Markov models for the evening, weekend and daytime data sets, as well as models for all of the data combined. Each of these models was then used

WWW 2005, May 10--14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

to predict the next URL clicked in the path given the current path for each of the three data subsets and for values of k between 1 and 7. In each WAP menu the user has approximately seven selections (including going back to the previous page) from which they can choose, therefore the chance level baseline is 0.167. However for these experiments a fully connected graph was assumed, to take into account instances where users used a bookmark or typed in a URL in the mid-session for example. As these browsing patterns are limited to one portal, theoretically users have a one in 256 choice, accuracy of approximately 0.004.

The results of the daytime model are shown in Figure 1, these results clearly illustrate the benefits of using time of the week. For reason of space we are only showing one of the graphs, the other results display the same characteristics. First, the predictive accuracy of the models for the different time periods (weekend, day, evening) are all better than that for the merged data that does not distinguish time periods; indeed, its performance is quite poor at 0.19-0.31. Second, the partitioned data sets are the best predictors of navigation in their corresponding time periods; that is, the weekend model best predicts navigation for weekend hits, the evening model best predicts navigation for evening hits, and so on. Third, the predictive accuracy of these models is quite high (0.5-0.6 in the best k-order models).

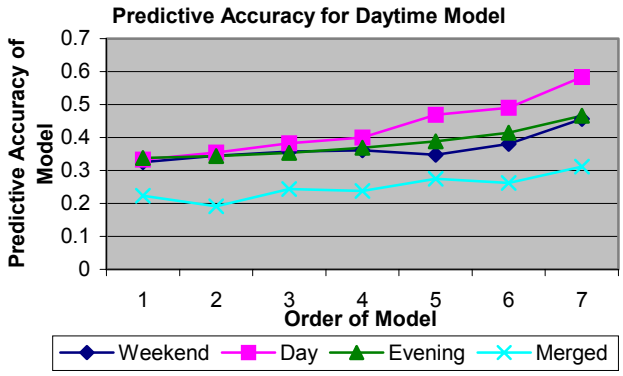

Figure 1: The predictive accuracy for each of the four models for different daytime periods with respect to k.

We also verified the benefit of such predictive model by using them in Smyth & Cotter [5] empirically verified Click-Distance Model. This model allows us to measure the benefits, to users of different types of personalization, in reduced navigation time. Using the click distance model the average "effort" involved in a user navigating to and selecting a link is 3.5. This is because the majority of the menus have 6 possible selections, resulting in a maximum click distance of 7 (6 scrolls plus 1 select) and a minimum of 1(one select) this averages 3.5. If we promote a URL and the user selects this URL this results in a click distance of 1(one select). If however we promote a URL and the user does not select the promoted URL, this can result in a click distance of 3.5 (if the promoted URL is higher in the menu than the selected URL normally) which we call the "best case" scenario, or the average click distance can increase to 4.5 (if the promoted URL is lower in the menu than the selected URL normally increasing the effort to the user) which we call the worst-case scenario. Table 1 illustrates the difference between best case and worst-case scenarios for a particular time period; these values are simulated using the accuracies in figure 1.

Table 1: Best Case and Worst Case Click Distance for k = 7 for daytime data using daytime model and all data model.

Time	Best Case	Worst Case	Average
All Data	2.75	3.45	3.5
Daytime	2	2.2	3.5

As can be seen from Table 1 using a predictive model corresponding to a time period reduces the click distance significantly more, in particular in the worst-case scenario.

4. Conclusions and Future Work

We have presented our hypothesis, which is that user navigation patterns in the Internet change depending on the time of day; users can have different personas at different times of the day. We have shown that the predictive accuracy for Markov models for the specific time periods are far more accurate than models based on undifferentiated data. We have also shown that these models can usefully reduce navigation times when they are used to re-order menus in a WAP portal. Therefore, it looks time time-of-day segmentations are a useful instrument for better personalization in adaptive information systems. In our future work, we are examining methods for automatically segmenting data to learn the best cuts to support this type of personalization. This segmentation of the data could quite easily be used in conjunction with some other predictive model, for example the ClixSmart navigator [5] to make more accurate predictions about user navigation and adapting portal structure to the needs of users.

5. ACKNOWLEDGMENTS

This material is based on works supported by the Science Foundation Ireland under Grant No. 03/IN.3/I361.

6. REFERENCES

[1] Cotter, P, & Smyth, B. (2000). " PTV: Intelligent Personalised TV Guides". Proceedings of the 12th Innovative Applications of Artificial Intelligence (IAAI-2000) Conference. AAAI Press

[2] Maes, P. Agents that reduce work and information overload. Communications of the ACM, 37(7):30-40, 1994.

[3] Halvey M., Keane M.T. & Smyth B., "Mobile Web Surfing is the same as Web Surfing", Communications of the ACM, 2005. (In Press).

[4] Ramsay, M., Nielsen, J. Nielsen Report, "WAP Usability Deja Vu : 1994 All Over Again", (2000).

[5] Smyth, B. and Cotter, P. (2002) "Personalized Adaptive Navigation for Mobile Portals". Proceedings of the 15th European Conference on Artificial Intelligence - Prestigious Applications of Intelligent Systems. Lyons, France