

Temporal Rules for Mobile Web Personalization

Martin Halvey

Adaptive Information Cluster
School of Computer Science and
Informatics,
University College Dublin, Ireland
martin.halvey@ucd.ie

Mark T. Keane

Adaptive Information Cluster
School of Computer Science and
Informatics,
University College Dublin, Ireland
mark.keane@ucd.ie

Barry Smyth

Adaptive Information Cluster
School of Computer Science and
Informatics,
University College Dublin, Ireland
barry.smyth@ucd.ie

ABSTRACT

Many systems use past behavior, preferences and environmental factors to attempt to predict user navigation on the Internet. However we believe that many of these models have shortcomings, in that they do not take into account that users may have many different sets of preferences. Here we investigate an environmental factor, namely time, in making predictions about user navigation. We present methods for creating temporal rules that describe user navigation patterns. We also show the benefit of using these rules to predict user navigation and also show the benefits of these models over traditional methods. An analysis is carried out on a sample of usage logs for Wireless Application Protocol (WAP) browsing, and the results of this analysis verify our hypothesis.

Categories and Subject Descriptors

H5.4 Hypertext/Hypermedia, H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous, H 1.2 User/Machine Systems

General Terms

Algorithms, Experimentation, Human Factors

Keywords

WWW, WAP, mobile, temporal models, user modeling.

1. INTRODUCTION

The goal of total personalization for adaptive hypermedia depends on discovering all of the factors that impact user behavior in different contexts. These factors include obvious things like explicitly stated user preferences and less obvious variables like implicit preferences, patterns of past behavior, and the physical environment. The real challenge facing hypermedia is to properly identify these factors, the way in which they influence behavior and the degree of that influence. In previous work, we have identified time as an important environmental factor that impacts user navigation behavior in a mobile-internet portal [2]. However, in this previous work, we made several key knowledge-engineering steps to segment the log data. It would be much better if this knowledge-engineering step could be automated, that is to say if we had a technique that captured the way in which behavior is influenced by time. In this paper, we develop such an automated method for determining temporal rules that describe user navigation. We use these methods to describe user navigation

with respect to URL's. We find that the methods that we describe in this paper learn a greater number of high confidence rules than traditional association rule mining methods.

2. TIME BASED USER PROFILES

The problem of mining association rules can be formalized as follows [1]. Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. Associated with each transaction is a unique identifier, called its TID. We say that a transaction T contains X , a set of some items in I , if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subseteq I$, $Y \subseteq I$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence c if $c\%$ of transactions in D that contain X also contain Y . The rule $X \Rightarrow Y$ has support s in the transaction set D if $s\%$ of transactions in D contain $X \cup Y$.

Our proposal is that each transaction T has associated with it a timestamp t . So for some rules R of the form $X \Rightarrow Y$, the rule could be refined to R' of the form $X \cup t \Rightarrow Y$ such that confidence c of $X \Rightarrow Y$ for transactions that take place in transaction set D during times t is greater than confidence c over all of the transaction set D . Time periods can be determined at the start of the process and can be any period less than the total time period over which the data can be collected, i.e. seconds, years, days or hours could be used provided they are not longer than the total time period.

We are proposing four methods to determine these temporal types of rules. The first is *point better*, for each point in the time period that the confidence c is improved we create a new rule. A point in the time period will depend on how time is being measured, for example days or hours etc. This rule may be refined to include sequences of points, which we call *sequence better*. By sequence we mean a consecutive period of one or time points. The third method is called *point threshold*, this is a refinement of the point better method. For each point in the time period that the confidence c is above a certain threshold we create a new rule. This may require learning some initial rules that are below the threshold value so that the improvement may be above the threshold value. This rule may be refined to include sequences of points, which we call *sequence threshold better*.

Each of these methods requires an existing algorithm to learn an initial set of rules and each of these methods may also require multiple scans of the data set or initial set of rules as can be seen in the example algorithms below. For the remainder of this work the existing algorithm that is used is Apriori [1] and it shall also provide the baseline for all of our experiments that follow.

3. PREDICTING USER NAVIGATION

To evaluate the effectiveness of the techniques that we have outlined in this paper we analyze web logs from a mobile Internet portal of a major European operator. This data, gathered in September 2002, involved 1,168 users and 147,700 individual user sessions. The data set was segmented into four parts, each part is the equivalent of 1 weeks browsing on the WAP portal. In turn each possible combination of 3 weeks was used to learn rules using Apriori. The point threshold and sequence threshold methods were then applied to each of the possible combination of 3 weeks. Each of the rule sets was then used to try and predict user navigation in the 4th week. Here we compare results based on the accuracy of predictions, number of predictions made and number of applicable rules that may be applied for each user. Initially we had hoped that we could compare our techniques with an existing algorithm (i.e. Apriori). However the algorithm did not learn a sufficient number of high confidence rules for a fair comparison to take place. In fact for a majority of users no rules were learned in the 0.6 – 1 confidence range. However in the figures below we can see the results of a comparison between the point threshold method applied to days (PTD) and hours (PTH), and the sequence threshold method for days (STD) and for hours (STH).

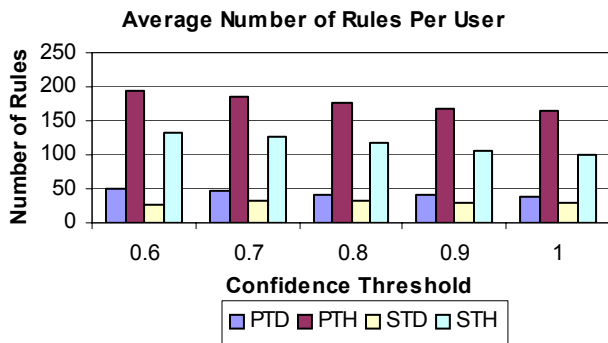


Figure 1: Average number of rules learned for each user for the four methods.

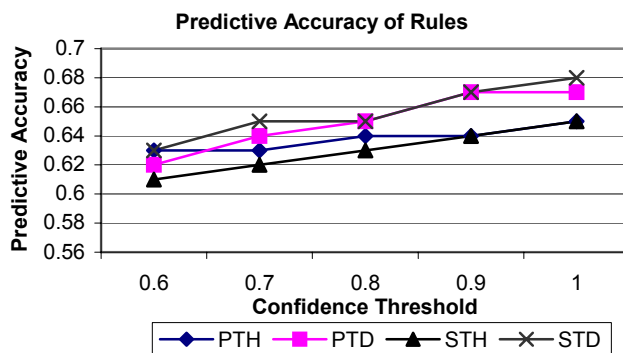


Figure 2: Predictive Accuracy for each user for each of the four methods.

Figure 1 shows the average number of rules for each user. As expected the PTH method learns the highest number of rules for each user. We believe this is because has much more fine grained time points than the methods using days. Also the sequence methods greatly reduce the number of rules that can be applied for

each user. This implies that users do have particular pattern of navigation over a period of time, e.g. a weekend period with respect to days or an evening period with respect to hours. We also analysed the number of times a rule could be applied and also looked at the average number of rules learned and that could be applied for categories of web pages, for reasons of space we do not show these graphs. In general when days are used as the time periods the number of times that a rule was applied was greater than the number of rules that were learned, indicating that a user would have the same navigation pattern at multiple times during the day. However, in the cases where hours are used as the time points the number of times that rules could be applied is generally less than the number of rules learned. Figure 2, shows the predictive accuracy of the rules. The predictive accuracy of the rules is in the range of 60%-70%. This is surprising given that many of the rules have high confidence values. In general we found that the rules learned using days, as time periods are more accurate than those using hours. We drew the conclusion that days are better time points than hours, which appear to be to fine grained. In general for each user less rules were learned using days as time points, the rules are applied more often and the rules are more accurate for predicting user navigation.

4. CONCLUSIONS AND FUTURE WORK

User navigation patterns in the Internet are time dependent; users have different needs and goals at different times. As such we have presented a number of methods for learning rules that can be applied during particular time periods, during which users in general have different goals. We have proven this by analyzing a log file of WAP usage. We have analyzed our methods with respect to pages that users have visited and found that a greater number rules with high confidence and high support values were discovered. Finally we used our methods to try and predict user navigation in a WAP portal by using a number of weeks worth of navigation to predict subsequent weeks navigation. We found that for each user the average number of rules found, and the number of times that those rules could be applied to be greater for the temporal rules that were learned using our methods than those learned using a traditional association rule mining algorithm (i.e. Apriori). Subsequently as high confidence rules were applied more often we have shown that these methods can potentially be of great benefit to users who navigate on mobile devices and aid their experience. Integrating this knowledge with some existing techniques as has been outlined can benefit users who access the Internet on all manner of devices.

5. ACKNOWLEDGMENTS

This research was supported by the Science Foundation Ireland under Grant No. 03/IN.3/I361 to the second and third authors.

6. REFERENCES

- [1] Agrawal R. & Srikant R. Fast Algorithms for Mining Association Rules in Large Databases, In Proceedings of the 20th International Conference on Very Large Databases, VLDB 1994, P.P. 487 – 499.
- [2] Halvey M., Keane M.T. & Smyth B. Predicting Navigation Patterns on the Mobile-Internet Using Time of the Week, WWW (Special Interest Tracks and Posters) 2005, P.P. 958 – 959.