Incorporating Pageview Weight into an Association-Rule-Based Web Recommendation System

Liang Yan[1] and Chunping Li[2]

School of Software, Tsinghua University, Beijing, China, 100084
[1] yanliang04@tsinghua.org.cn,
[2] cli@tsinghua.edu.cn

Abstract. Web recommendation systems based on web usage mining try to mine users' behavior patterns from web access logs, and recommend pages to the online user by matching the user's browsing behavior with the mined historical behavior patterns. Recommendation approaches proposed in previous works, however, do not distinguish the importance of different pageviews, and all the visited pages are treated equally whatever their usefulness to the user. We propose to use pageview duration to judge its usefulness to a user, and try to give more consideration to more useful pageviews, in order to better capture the user's information need and recommend pages more useful to the user. In this paper we try to incorporate pageview weight into the *Association Rule* (*AR*) based model and develop a *Weighted Association Rule* (*WAR*) model. Comparative experiment of the two shows a significant improvement in the recommendation effectiveness with the proposed *WAR* model.

Keywords: Web Personalization, Web Usage Mining, Web Recommender System, Weighted Association Rule Mining.

1 Introduction

Web recommendation systems anticipate the information needs of on-line users and provide them with recommendations to facilitate and personalize their navigation. Web usage mining is a popular approach to building such systems. Usage-based web recommendation systems usually consist of an off-line module and an on-line module. The off-line module mines user access patterns from web logs; the on-line module matches the current user's browsing behavior with the mined access patterns to predict the interest of the user, and recommends pages interesting to him/her. Recommended pages are usually in the form of links appended to the end of the requested page. Such a process is also called web personalization. An extensive study of web personalization based on web usage mining can be found in [1].

There is a common deficiency, however, in most previous approaches to web usage mining and personalization: they ignore the difference in the importance of the pageviews in a user session. It is quite probable that not all the pages visited by the user are of interest to him/her. A user might get into a page only to find it is of no value to him/her, causing irrelevant page accesses to be recorded into the log file. Therefore, it is imperfect to use all the visited pages equally to capture user interest and predict user behavior. A better approach would be to filter out uninteresting pages and use only the pages of interest to the user for the personalization process.

A. Sattar and B.H. Kang (Eds.): AI 2006, LNAI 4304, pp. 577 – 586, 2006.

But how do we judge whether the user is interested in a page or not? Although in usage-based recommendation systems we can't expect users to express likes or dislikes explicitly, it is quite reasonable to suppose that if a user is not interested in a page, he/she won't spend much time viewing it and vice versa. As pageview duration can be calculated from web logs, it is a good choice for inferring user interest.

In this paper we use the time length of a pageview to estimate its importance in a transaction, in order to capture the user's interest more precisely. We assign a weight to each pageview in a transaction according to its duration, and perform *Weighted Association Rule* (*WAR*) mining to discover significant page sets. In the on-line recommendation phase, we also take pageview duration into account to better capture the current user's interest and generate more relevant recommendations.

The rest of the paper is organized as follows: Section 2 talks about related work. In Section 3 we introduce our approach of *Weighted Association Rule* mining to discover significant page sets. Section 4 introduces the online recommendation method using pageview duration as an indicator of user interest. Section 5 introduces our experiment and results. Section 6 concludes this paper and provides ideas for future work.

2 Related Work

The overall process of web personalization generally consists of three phases: data preparation and transformation, pattern discovery, and recommendation [1]. Data preparation and pattern discovery are performed offline. The data preparation phase transforms raw web log files into transaction data ready for data mining tasks [3]. Here a transaction contains a list of pageviews along with their weights. The weights can be binary, representing the existence or non-existence of a pageview in the transaction, or they can be some function of the pageview duration. In the pattern discovery phase, various data mining techniques can be applied to the transaction data, such as clustering, association rule mining, and sequential pattern discovery. Only the recommendation phase is performed online. In the recommendation phase, the recommendation engine considers the active user session in conjunction with the discovered patterns to recommend pages that the user is most likely to visit. [1]

Some studies have considered the approach of using pages interesting to the user for the recommendation process. In [2], Mobasher et al use statistical significance testing to judge whether a page is interesting to a user. Its main idea is: A duration threshold is calculated for each page using the average duration and standard deviation of the visits to the page; if the duration of a pageview is longer than the threshold, that pageview is considered interesting to the user and vice versa. The drawback of such an approach is that it simply divides pageviews into interesting and uninteresting groups, and neglects the difference in the degrees of interest. For one thing, there isn't a clear division between interesting and uninteresting pages; for another, the degrees of interest are probably not the same for all the interesting (and uninteresting) pages. In this work we try to assign a quantitative weight to each pageview, taking into account the degree of interest.

Clustering and collaborative filtering approaches are ready to incorporate both binary and non-binary weights of pageviews, although binary weights are usually used for computing efficiency [1] [2]. *Association Rule* (*AR*) mining and *Sequential Pattern* (*SP*) mining [4] can lead to higher recommendation precision [1], and are easy to scale to large datasets, but how to incorporate pageview weight into the *AR* and the *SP* models has not been explored in previous studies.

Weighted Association Rule (*WAR*) mining allows different weights to be assigned to different items, and is a possible approach to improving the *AR* model in the web personalization process. Cai et al. [5] proposed assigning different weights to items to reflect their different importance. In their framework, two ways are proposed to calculate itemset weight: total weight and average weight. Weighted support of an itemset is defined as the product of the itemset support and the itemset weight. Tao et al. [6] also proposed assigning different weights to items, the itemset/transaction weight is defined as the average weight of the items in the set/transaction, and weighted support of an itemset is the fraction of the weight of the transactions containing the itemset relative to the weight of all transactions. Both models attempt to give greater weights to more important items, facilitating the discovery of important but less frequent itemsets and association rules. However, both models assume a fixed weight for each item, while in the context of web usage mining a pageview might have different importance in different sessions.

3 Incorporating Pageview Weight

3.1 How to Determine Pageview Weight

In this paper we simply use the duration of a pageview as its weight in the transaction. Using more complex functions to determine pageview weights is left for future work. Several reasons validate the simple approach of using pageview duration as its weight. First, it reflects the relative importance of each pageview, because a user will generally spend more time on a more useful page. Second, the rates of most human beings getting information from web pages shouldn't differ greatly. If we assume a similar rate of acquiring information from pages for each user, the time a user spends on a page is proportional to the volume of information useful to him/her, and is a quantitative measure of the usefulness of that pageview to the user.

3.2 Weighted Support of an Itemset

We use the preprocessing techniques discussed in [3] to extract transaction data from web log files. After the preprocessing phase, we get a set of n pageviews, $P = \{p_1, p_2, \ldots, p_n\}$, and a set of m user transactions, $T = \{t_1, t_2, \ldots, t_m\}$, each transaction t is an l-length sequence of ordered pairs: $t = < (p_1, w(p_1)), (p_2, w(p_2)), \ldots, (p_l, w(p_l)) >$, where each $p_i \in P$, and the weight $w(p_i)$ associated with p_i is the duration (in seconds) of pageview p_i in transaction t.

For example, $t = \{(A, 60), (B, 20), (C, 70), (D, 90)\}$ is a transaction which records that the user spent 60 seconds on page A, 20 seconds on page B, 70 seconds on page C and 90 seconds on page D.

Definition-1 Weighted support of an itemset by a transaction: Weighted support *WSP* of an itemset *X* by a transaction *t* is the quantity of *X* contained in *t*, denoted as *wsp*(*X*, *t*):

$$wsp(X, t) = min\{w(p_{x1}), w(p_{x2}), \dots , w(p_{xk})\} , \quad (1)$$

where $w(p_{xi})$ is the weight of pageview p_{xi} in transaction *t*. In simple words, weighted support of an itemset by a transaction implies how much of the itemset is contained in the transaction. For example, $wsp(A, t) = 60$, $wsp(B, t) = 20$, $wsp(AB, t) = 20$, $wsp(ACD, t) = 60$, $wsp(ABCD, t) = 20$.

Definition-2 Weighted support of an itemset (across all transactions): Weighted support *wsp*(*X*) of an itemset *X* across all transactions is defined as follows:

$$wsp(X) = \frac{\sum_{t_i \in T} wsp(X,\ t_i)}{|T| \cdot \overline{w}} , \quad (2)$$

where *T* is the set of all transactions, $\overline{w}$ is the average weight of all the items across all transactions. The numerator is the sum of the weighted support of *X* over all transactions; the denominator is just a normalizing factor making most *wsp* values fall between 0 and 1.

3.3 Weighted Association Rules

In our framework, a weighted association rule has the form of *X* => *Y*, where *X* and *Y* are two itemsets, $X \cap Y = \emptyset$.

Definition-3: Weighted support of the association rule

$$wsp(X => Y) = wsp(X \cup Y) . \quad (3)$$

Definition-4: Confidence of the weighted association rule

$$conf(X => Y) = \frac{wsp(X \cup Y)}{wsp(X)} . \quad (4)$$

3.4 Mining Significant Itemsets

In the traditional association rule mining framework, an itemset is denoted *large* if its *support* is above a predefined minimum support threshold. In our framework, we say an itemset is *significant* if its *weighted support* is above a predefined weighted support threshold. Our approach to mining significant itemsets is based on the Apriori [7] algorithm. Apriori algorithm is an efficient algorithm utilizing the *downward closure property*, that is, any subset of a large itemset is also large. By our definition of weighted support and significant itemsets, there is a property that any subset of a significant itemset is also significant, here called a *weighted downward closure property*. The property always holds because for any itemset *X* and *Y*, if $X \subset Y$, by definition we have $wsp(X, t) \geq wsp(Y, t)$, hence $wsp(X) \geq wsp(Y)$.

4 Recommendation Engine

4.1 Significant Itemset Graph

We use a *Significant Itemset Graph* to improve the recommendation efficiency. Fig. 1 gives an example of the *Significant Itemset Graph*. The idea comes from [8], in which the data structure is called the *Frequent Itemset Graph* because the itemsets stored in it are *frequent* itemsets.

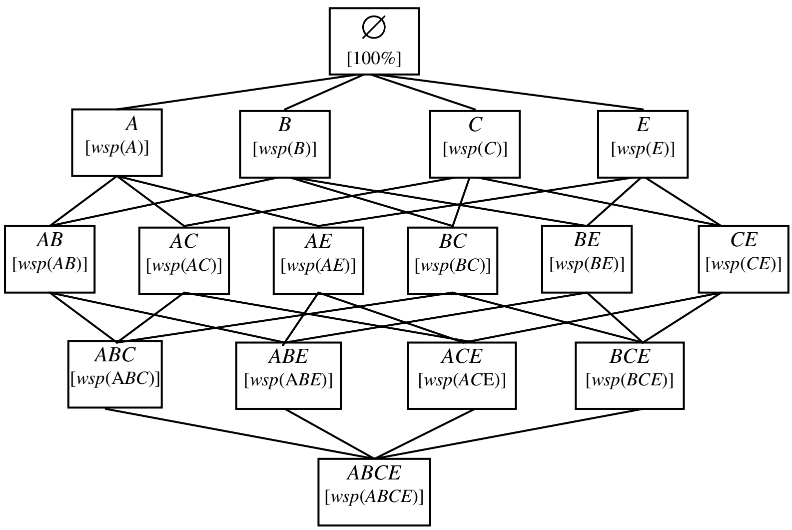

Fig. 1. *Significant Itemset Graph* - An Example. Each node stores a significant itemset along with its weighted support. For a node N containing itemset X, each child node of N corresponds to a significant itemset $X \cup \{p\}$.

4.2 Capture the User's Current Interest More Precisely

Previous works [1] [4] [8] go the right way to use only the last visited pages to generate the recommendations because *earlier* portions of a user session are not likely to represent the user's *current* information need. However, it still has another shortcoming: It doesn't eliminate the uninteresting pages. We propose that both page freshness and interestingness should be considered to capture a user's current interest. To give more recently visited pages more consideration, we attenuate the weight of each pageview as it gets obsolete. Pageviews with the greatest adjusted weights in the current user session are used to generate the recommendations. In this paper we use a most simple method to attenuate the weight of each pageview: linear attenuation. Given an active user session $S = \langle (p_1, w_1), (p_2, w_2), \ldots, (p_k, w_k) \rangle$, the attenuation factors and adjusted weights are calculated as follows:

Table 1. Attenuation factors and adjusted weights

Pageview	(p_1, w_1)	(p_2, w_2)	…	(p_{k-1}, w_{k-1})	(p_k, w_k)
Attenuation factor	$\frac{1}{k}$	$\frac{2}{k}$	…	$\frac{k-1}{k}$	$\frac{k}{k}=1$
Adjusted weight	$\frac{1}{k}\,w_1$	$\frac{2}{k}\,w_2$	…	$\frac{k-1}{k}\,w_{k-1}$	w_k

4.3 Recommendation Process

The recommendation process is as follows: Given an active user session S, the recommendation engine first adjusts the weight of each pageview, using the attenuation method discussed above, and then selects the pages with the highest adjusted weights to recommend pages. The recommendation score of page p by page set X, denoted as $rec(p, X)$, is calculated as follows:

$$rec(p, X) = wsp(X, S) \bullet conf(X => p)\ , \tag{5}$$

where $wsp(X, S)$ is the weighted support of page set X by the user session S, representing how much X is contained in S, as defined in Section 3.2, and $conf(X => p)$ is the confidence of weighted association rule $X => p$, defined in Section 3.3. Note that $wsp(X, S)$ is calculated using the adjusted weights. The improvement of this approach is that both the weight of X in S and the confidence of $X => p$ are used to determine the recommendation score, not just the confidence value as is used in previous works.

As a subset of X may support p with a higher recommendation score, we choose the highest recommendation score by all the subsets of X as the recommendation score of p. Table 2 gives an example.

The confidence of $\varnothing => p$ is simply the wsp of p. The weight of $\varnothing$ is set to be the average duration of all pageviews across all transactions (here assumed to be 50 seconds in this example). For the last pageview whose duration could not be calculated from the access log, we also use the average duration to estimate it.

Table 2. Calculation of recommendation scores: Given association rules, and the pageview set $X = \{(A, 100), (B, 200)\}$ selected to generate the recommendation, the highest score of 140 is stored as the recommendation score of page p

Weighted Association Rule	$\varnothing => p$	$A => p$	$B => p$	$AB => p$
Confidence	40%	60%	70%	90%
Recommendation Score	50 • 40% = 20	100 • 60% = 60	200 • 70% = 140	100 • 90% = 90

For a k-item page set X, we use all the 2^k subsets of X to generate the candidate recommendations, and the candidates with the highest recommendation scores are recommended to the user. For computing efficiency, an upper bound of 5 is set for the number of pages used to generate recommendations, as the number of subsets increases exponentially with k.

5 Experimental Evaluation

The recommendation approach based on *Weighted Association Rule* (*WAR*) proposed in this work is compared with the traditional *Association Rule* (*AR*) based approach.

5.1 Test Data Set

The comparative test is performed on the preprocessed and filtered sessionized data from the main DePaul CTI web server (http://www.cs.depaul.edu). The data is based on a random sample of users visiting this site during a 2 week period in April of 2002. The original (unfiltered) data contains a total of 20950 sessions from 5446 users. The filtered data files are produced by filtering out low support pageviews, and eliminating sessions of size 1. The filtered data contains 13745 sessions and 683 pageviews. We treat each session as a transaction. Each transaction contains a sequence of pageviews along with their weights (durations).

We perform 10-fold cross-validation on the CTI data set. In each of the 10 iterations, the data set is divided into training (90%) and evaluation (10%) data sets. The training set is used to generate the *AR* (*frequent* itemsets) and *WAR* (*significant* itemsets) models, and the evaluation set is used to test the recommendation effectiveness of the *AR* and *WAR* based approach.

5.2 Evaluation Methodology

To compare the *AR* based model and the *WAR* based model fairly, we select the same number of frequent/significant k-itemsets for the *AR* and *WAR* based models. Our evaluation methodology is as follows: Each transaction t in the evaluation set is divided into two halves; pages in the first half are used to generate the recommendation set R_t, and pages in the second half, denoted as $Eval_t$, are used to evaluate the generated recommendations. For the *AR* based recommendation approach, we adopt the *active session window*, i.e. the last visited $|win|$ pages [1] [4] [8], to generate recommendations. For the *WAR* based model, we use pages which are both fresh and significant in the first half to generate recommendations, as described in Section 4.2. To guarantee an appropriate minimum session size, we only use transactions containing at least 4 pageviews.

We propose two evaluation metrics: *Weighted Coverage* (*WC*) and *Weighted Precision* (*WP*), to evaluate the recommendation effectiveness. For a transaction t in the evaluation data set, *Weighted Coverage* and *Weighted Precision* are defined as:

$$WC_t = \frac{\sum_{R_t \cap Eval_t} w(p_i)}{\sum_{Eval_t} w(p_i)} \ , \tag{6}$$

$$WP_t = \frac{\sum_{R_t \cap Eval_t} w(p_i)}{|R_t| \cdot \overline{w}} \ , \tag{7}$$

where $w(p_i)$ is the weight of pageview p_i in transaction t; $|R_t|$ is the size of the recommendation set, and $\overline{w}$ is the average pageview weight across all transactions.

For example, if the evaluation half of transaction t is $Eval_t = \{(A, 80), (B, 20)\}$, it is evident that recommending page A captures more of the user's information need than recommending page B. In previous studies, however, recommending either page will get a coverage score of 50%. In our framework, weighted coverage of recommending page A is 80%, and that of recommending page B is 20%.

Still with the evaluation half $Eval_t = \{(A, 80), (B, 20)\}$, a recommendation set $R_1 = \{A, C, D, E\}$ is more precise than $R_2 = \{B, C, D, E\}$, although both have only 1 page needed by the user. Assuming the average pageview weight $\overline{w} = 50$, weighted precision of R_1 is $\frac{80}{4 \cdot 50} = 40\%$, and weighted precision of R_2 is $\frac{20}{4 \cdot 50} = 10\%$.

The overall weighted coverage and weighted precision scores are calculated as the *weighted* average of WC_t and WP_t scores for each transaction, where $w_t = \sum_{Eval_t} w(p_i)$ is the weight of $Eval_t$, and $Eval$ is the evaluation (10%) data set:

$$Weighted\ Coverage = \frac{\sum_{Eval} WC_t \cdot w_t}{\sum_{Eval} w_t}, \tag{8}$$

$$Weighted\ Precision = \frac{\sum_{Eval} WP_t \cdot w_t}{\sum_{Eval} w_t}, \tag{9}$$

The reason to use *weighted* average is that we should give greater weight to a score associated with a more important session. For example, a weighted coverage score of 80% may be meaningless, if $Eval_t = \{(A, 8), (B, 2)\}$ with page A being recommended, as the user is probably not interested in either page, spending less than 10 seconds on each. Another weighted coverage score of 60% may be more meaningful, if $Eval_t = \{(A, 60), (B, 40)\}$ with page A being recommended, because this evaluation set contains pages more useful to the user.

The overall weighted coverage and weighted precision scores for each of the 10 rounds are averaged to calculate the final weighted coverage and weighted precision scores.

5.3 Experimental Results

We vary the number of recommended pages from 4 to 14. Experimental results show that the *WAR* based model increases weighted coverage by 10% ~ 20% and weighted precision by 20% ~ 30%. We also vary $|win|$ from 3 to 5 and do not observe significant impact on the weighted coverage and weighted precision scores.

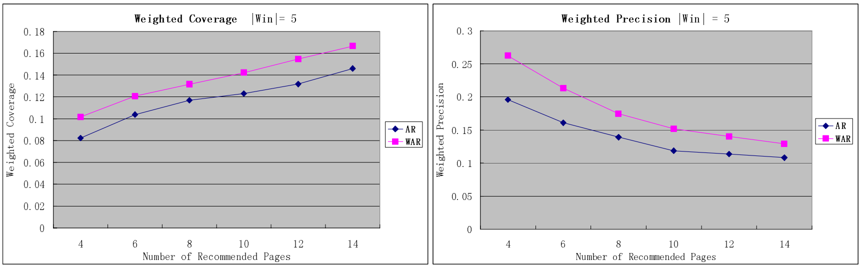

Fig. 2. Weighted coverage and weighted precision of the *AR* and *WAR* based models

6 Conclusion and Future Work

Most usage based page recommendation systems proposed in previous studies do not distinguish the importance of different pageviews in a transaction, and pages interesting to the user and those otherwise are equally used to capture user behavior patterns and generate recommendations. In this work, we use pageview duration to judge the importance of a pageview to a user, and try to give more consideration to pages which are more useful to the user, in order to capture the user's information need more precisely and recommend pages more useful to the user.

We have made a preliminary attempt to incorporate pageview weight into the association rule based recommendation system, and proposed a *Weighted Association Rule* model as an improvement to the traditional *Association Rule* model. Comparative experiments were performed to test the recommendation effectiveness of the traditional *AR* based model and the proposed *WAR* based model. Experimental results have shown that the *WAR* based model could significantly improve the recommendation effectiveness. The reason behind the improvement is that by taking the difference of the interestingness of the visited pages into consideration, our approach can discover and predict user interest more precisely.

The *WAR* based model can easily be extended to the (*contiguous*) *sequential pattern* model, which is a direction for future work. What is the optimal approach to determine pageview weight also needs further study.

Acknowledgment. This work was supported by Chinese 973 Research Project under grant No. 2004CB719401.

References

[1] B. Mobasher: Web Usage Mining and Personalization, In *Practical Handbook of Internet Computing*. Munindar P. Singh (ed.), CRC Press, 2005.

[2] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa: Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data. In *Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWP01)*, Seattle, WA, August 2001.

[3] R. Cooley, B. Mobasher, and J. Srivastava: Data Preparation for Mining World Wide Web Browsing Patterns. In *Journal of Knowledge and Information Systems*, 1(1): 5–32, 1999.

[4] M. Nakagawa and B. Mobasher: A hybrid web personalization model based on site connectivity. In *The Fifth International WEBKDD Workshop: Webmining as a Premise to Effective and Intelligent Web Applications*, pages 59 ~ 70, 2003.
[5] C.H. Cai, A.W.C. Fu, C.H. Cheng, W.W. Kwong: Mining association rules with weighted items. In *Database Engineering and Applications Symposium, 1998. Proceedings. IDEAS'98. International*, pages 68 ~ 77, July 1998.
[6] F. Tao, F. Murtagh, M. Farid: Weighted Association Rule Mining using Weighted Support and Significance Framework. In *Proceedings of the 9th SIGKDD Conference*, 2003.
[7] R.Agrawal and R.Srikant: Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94)*, Santiago, Chile, 1994, pages 487-499.
[8] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa: Effective personalization based on association rule discovery from web usage data. In *Proceedings of the 3rd ACM Workshop on Web Information and Data Management (WIDM01)*, Atlanta, Georgia, November 2001.