

Web Page Personalization based on Weighted Association Rules

R. Forsati

Department of computer engineering,
Islamic Azad University, Karaj Branch
Karaj, Iran
forsati@kiau.ac.ir

M. R. Meybodi

Department of computer engineering,
AmirKabir University of Technology
Tehran, Iran
mmeybodi@aut.ac.ir

A. Ghari Neiat

Department of computer engineering,
Islamic Azad University, North Tehran
Branch, Tehran, Iran
a_ghariniyat@yahoo.com

Abstract— Web personalization is the process of customizing a web site to the needs of each specific user or set of users, taking advantage of the knowledge acquired through the analysis of the user's navigational behavior. Personalized recommendation by predicting user-browsing behavior using association-mining technology has gained much attention in web personalization research area. However, the resulting association patterns did not perform well in prediction of future browsing patterns due to the low matching rate of the resulting rules and users' browsing behavior. In this paper, we extend the traditional association rule problem by allowing a weight to be associated with each item in a transaction to reflect the interest/intensity of each item within the transaction. In turn, this provides us with an opportunity to associate a weight parameter with each item in a resulting association rule. We assign a significant weight to each page based on the time spent by user on each page and visiting frequency of each page, taking in to account the degree of interest instead of binary weighting. We present new personalized recommendation method base on the proposed weighted association-mining technique. We show, through experimentation on real data set that this approach results in more objective and representative predictions and shows a significant improvement in the recommendation effectiveness in comparison to the traditional association rule approaches.

Keywords- personalization; association rules; data mining

I. INTRODUCTION

In recent years, the massive influx of information onto World Wide Web has facilitated user, not only retrieving information, but also discovering knowledge. However, web users usually suffer from the information overload problem due to the fact of significantly increasing and rapidly expanding growth in amount of information on the web. Web personalization [1] is one of the promising approaches to tackle this problem by adapting the content and structure of websites to the needs of the users by taking advantage of the knowledge acquired from the analysis of the users' access behaviors.

One research area that has recently contributed greatly to this problem is web mining. Most of the systems developed in this field are based on web usage mining [2]. Web usage mining[3] has gained much attention in the literature as a potential approach to fulfill the requirement of web personalization [1,3,5,6,7,8]. The discovered knowledge indicating users' navigational behavior is useful for the system to personalize the web site according to each user's behavior and profile. Web Usage Mining applying common data mining techniques to the discovery of usage patterns

form web data and making recommendations based on them. These systems are mainly concerned with analyzing web usage logs, discovering patterns from this data and making recommendations based on the extracted knowledge [5,7,9]. Unlike traditional recommender systems, which mainly base their decisions on user ratings on different items or other explicit feedbacks provided by the user [12,13] these techniques discover user preferences from their implicit feedbacks, namely the web pages they have visited. In this paper, we focus on the association-mining method, which is a widely used data analysis method in web usage mining [4,6]. Association rule mining has been successfully applied in the pages recommendation system, web page personalization and is easy to scale to large data datasets [14,7,15,8], but how to incorporate page weight into the AR in the pages recommendation system has not been explored in previous studies.

Weighted Association Rule 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. [10] 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. [11] 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 and page recommendation systems a page might have different importance in different sessions.

In the association rule based recommendation and personalization context, none of these researches took the time spent on each page as an implicit rating for that page. In other words, they assumed all the access web pages are equally important to user navigation pattern profile and ignored the difference in the importance of the pages in a user session and used the time spent for viewing pages just for preprocessing phase. In [16] developed a usage model for predictions based on association rule by using the binary weight and exact matching the active user with the rules. The

combination of similarity measurements and association rule has been proposed in [15,18]. They measure the similarity between items upon the rules instead of user sessions to produce recommendations then search the most similar items to the active user. In [17], Yan et al assign a weight to each page in a transaction according to its duration, and used significant itemset graph to improve the recommendation efficiency.

In this paper, we extend the traditional association rule problem by allowing a weight to be associated with each item in a transaction to reflect the interest of each item within the transaction and develop a novel recommendation algorithm based on proposed weighted association rule mining approach. In the proposed weighted association rule miner, the time spent by each user on each page and visiting frequency of each page are used to assign a quantitative weight to the pages instead of traditional binary weights. The intuition behind this idea is that the time spent on pages [9] and visiting frequency are good implicit interest indicator of a user on those pages. The methodology is like this: first, the weighted association rules of each URL will be extracted from the web log data and similarity between active user sessions will be calculated upon the weighted rules instead of an exact match for finding the best rule. Finally, the recommendation engine will then find the most similar rules to the active user session with the highest weighted confidence by scoring each rule in terms of both its similarity to the active session and its weighted confidence. Proposed weighted measure can enhance the quality of personalized web recommendations that aim to predict which web pages are more likely to be accessed next by current users.

The organization of the paper is as follows: in section 2 we introduce our weighting schema. The proposed weighted association rule and its definitions present in section 3. We represent our page recommendation mechanism in section 4. Experimental result to evaluate proposed method describe in section 5. Section 6 concludes the paper.

II. WEIGHTING SCHEMA

Let $P = \{p_1, p_2, \dots, p_m\}$ denote the set of web pages accessed by users in web server logs after the preprocessing phase [19], each of them is uniquely represented by its associated URL. Also let $T = \{t_1, t_2, \dots, t_n\}$ be the set of user transactions where each $t_i \in T$ is a subset of P. To facilitate the high quality recommendation, we represent each transaction t as an m-dimensional vector over the space of web pages, $t = \langle (p_1, w_1), (p_2, w_2), \dots, (p_m, w_m) \rangle$, where w_i denotes the weight with the i^{th} web page ($1 \leq i \leq m$) visited in a transaction t . The weight w_i in transaction t needs to be appropriately determined to capture a user's interest in i^{th} web page.

The weights can be determined in a number of ways, however in the context of personalization based on clickstream data, the primary sources of data are server access logs. This allows us to choose two types of weights for pages: weights can be binary, representing the existence

or nonexistence of a page access in the transaction or they can be a function of parameters such as duration of the associated page in the user's session to represent the interest of page to a specific visiting user.

Since the recommendation process is based on the behavior of previous users, so the weighting schema must precisely model the user's interest. Recommendation approaches proposed in previous works, however, do not distinguish the importance of different pages and all the visited pages are treated equally whatever their usefulness to the user. They neglect the difference in the importance of the pages and degree of interest in a users' 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. Although in usage-based recommendation systems we can't expect users to express likes or dislikes explicitly, we need a weight measure for approximating the interest degree of a web page to a user.

Inspired by Chan and coworkers [20, 21], we propose a weighting measure which is calculated from web logs to extract the interest of page for the visitor. In our weighting schema, both of time length of a page and visiting frequency of a page are used to estimate its importance in a transaction, in order to capture the user's interest more precisely instead of binary which is typically used in other researches. This approach try to give more consideration to more useful pages, in order to better capturing the user's information need and recommend more useful pages to the user.

Several reasons validate the idea of using pages visit duration as one of the weighting parameters. First, it reflects the relative importance of each page, because a user generally spend more time on a more useful page [9, 17], because if a user is not interested in a page, he/she do not spend much time on viewing the page and usually jumps to another page quickly [22]. However, a quick jump might also occur due to the short length of a web page so the size of a page may affect the actual visiting time. Hence, it is more appropriate to accordingly normalize duration by the length of the web page, that is, the total bytes of the page. The formula of duration is given in Equation (1). Second, the rates of most human beings getting information from web pages should not differ greatly [17]. 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. As page duration can be calculated from web logs, it is a good choice for inferring user interest.

Frequency is the number of times that a page is accessed by different users. It seems natural to assume that web pages with a higher frequency are of stronger interest to users. A parameter that must be considered in the calculating the frequency of a page is the in-degree of that page (e.g. the number of incoming links to the page). It is obvious that a page with large in-degree has more probability to be visited by a user than a page with small one. Specially, in comparing two pages with same visiting rate, the page with small in-

degree is more interesting. The formula of frequency is given in Equation (2).

We use time spent by a user for viewing a page and frequency of visiting as two very important pieces of information in measuring the user's interest on the page, so we assign a significant weight to each page in a transaction according to these definitions as Equation (3).

$$\text{Duration}(p) = \frac{\text{Total Duration}(p)}{\text{Size}(p)} \quad (1)$$

$$\max_{Q \in T} \left(\frac{\text{Total Duration}(p)}{\text{Size}(p)} \right)$$

$$\text{Frequency}(p) = \frac{\text{Number of visit}(p)}{\sum_{Q \in T} \text{Number of visit}(Q)} * \frac{1}{\text{Indegree}(p)} \quad (2)$$

$$\text{Weight}(p) = \text{Frequency}(p) * \text{Duration}(p) \quad (3)$$

III. WEIGHTED ASSOCIATION RULE BASED RECOMMENDATION MODEL

As mentioned before, we first extend the traditional association rule problem by allowing a weight to be associated with each item in a transaction to reflect interest of each item within the transaction. In turn, this provides us with an opportunity to associate a weight parameter with each item in a resulting association rule, which called a weighted association rule (WAR).

Weighted association rule is useful in some sense. For example, the product, which has higher profit margin, should be paid more attention. Weighted Association Rule (WAR) mining allows different weights to be assigned to different items, and is a possible approach to improving the ARM model in the web personalization process.

In this model greater weights are given to more important items, facilitating the discovery of important but less frequent itemsets and association rules. However, previous models assume a fixed weight for each item, while in the context of web usage mining a page might have different importance in different sessions.

In the following we describe weighted rules with the definition of associated parameters. We extend the Apriori [23, 24] by adapting its parameters based on weighted items. In the next section we employ this algorithm for page recommendation.

Weight Settings:

Given the transformation of user transactions into m -dimensional space as vectors of weights of web pages, $t = \langle (p_1, w_1), (p_2, w_2), \dots, (p_m, w_m) \rangle$ where each $p_i \in P$, the weight w_i associated to page p_i is a non-negative real number to reflect the importance of page p_i in transaction t according to Equation (3). Inspired by Tao[32], we modify the measures exist in Apriori algorithm in the following definitions to reflect the weighting schema.

Definition-1. Weighted item:

Item weight is a value attached to an item (page) representing its significance. We denote it

as $w(p_i) = \text{Weight}(p_i)$, which is calculated using the Equation (3).

Definition-2. Weight of an itemset in a transaction:

Based on the item weight $w(p_i)$, the weight of an itemset X , denoted as $w(X, t)$, can be derived from the weights of its enclosing items. One simple way is to use the minimum weight of the all items in the itemset as the weight of whole itemset as shown in Equation (4).

$$w(X, t) = \begin{cases} \min(w(p_1, p_2, \dots, p_k)) & X \subseteq t \\ 0 & X \not\subseteq t \end{cases} \quad (4)$$

Where k is the number of items in the itemset.

Alternatively, we can use the average weights of its enclosing items as the itemset weight. Our experiments show that the minimum weight has better quality.

Definition-3. Transaction weight:

By assigning a weight to each item and itemset, we also assign a weight to each transaction to be used in the calculation of the support of each itemset. Assigning weight to transactions gives us the possibility to distinguish between different transactions. Usually the higher a transaction weight, the more it contributes to the mining result. One simple way is to calculate the average weights of all items that enclosed in each transaction. The weight of each transaction $w(t_k)$ is calculated as shown in Equation (5).

$$w(t_k) = \frac{\sum_{i=1}^{|t_k|} w(p_i)}{|t_k|} \quad (5)$$

Definition-4. Weighted support of an itemset across all transaction:

We modify the support of an itemset, Weighted support $wsp(X)$ of an itemset X across all transactions is defined as follows:

$$wsp(X) = \frac{\sum_{t \in T} w(t) * w(X, t)}{\bar{w} * \sum_{k=1}^{|T|} w(t_k)} \quad (6)$$

Where \bar{w} is the average weight of all the items across all transactions, and T is the set of all transactions.

Definition-5. Weighted confidence of the weighted association rule:

We define the weighted confidence of association rule for weighted rules as follows:

$$wconf(X \Rightarrow Y) = \frac{wsp(X \cup Y)}{wsp(X)} \quad (7)$$

Definition-6. Weighted rules:

For each rule, besides the weighted confidence and weighted support, we also add the weight of each page. The result of weighted association rule mining conceptually described as follows:

$r = \langle (p_1, p_2, \dots, p_k), (q_{k+1}, q_{k+2}, \dots, q_{k+m}), (w_1, w_2, \dots, w_{k+m}), \delta, \alpha \rangle \in R$, where $(p_1, p_2, \dots, p_k), (q_{k+1}, q_2, \dots, q_{k+m})$ present the body and head of the weighted rule respectively, w_i represent the weight of i^{th} page in the rule, δ represent the weighted

support and α represent the weighted confidence of the rule.

IV. A RECOMMENDATION ENGINE USING WEIGHTED ASSOCIATION RULES

The methods based on association rule mining to compute a recommendation set for the current user session, use a sliding window to control the number of session pages to be matched against the association rules [14]. So, maintaining a history depth may be important in the recommendation service to provide reasonable suggestions. In the following, we present our mechanisms for this purpose.

A. Modify user's current session

Maintaining a history depth may be important because most users navigate several paths leading to independent pieces of information within a session. Previous works [11, 14, 25] use a fixed-size sliding window over the current active session to capture the current user's history depth and generate the recommendations.

The sliding window of size n and go the right way over the active session allows only the last n visited pages to influence the recommendation value in the recommendation set because most of users go back and forth while navigating a site to find the desired information, and it may not be appropriate to use earlier portions of the user session to represent the user's current information need. However, this method does not distinguish the importance of different pages, and all the n last visited pages are treated equally whatever their usefulness to the user. A better approach would be to filter out uninteresting pages and use only the pages of interest to the user for the personalization process.

Another parameter can also be used to associate an additional measure of significance with each page in the user's active session is weight of page. Although it seems that the recently visited pages by user are more appropriate to be used for the recommendation, but in many cases the user have a burst behavior. He navigates between pages to find an interesting page and spent much of his time on that page and then repeats this process. So, the place of a page in the user session is not the only parameter influencing the selection of predictor pages. Hence we consider the freshness of a page and its weight simultaneously to choose the predictor pages.

In contrast to using a sliding window to preserve only the most recent session information for the matching work, inspired by Yan and Li [17], we propose a measure for approximating the user's current interest and filter out uninteresting pages by using a most simple method to capture the weight of interest of each page. We formulate the freshness of a page and its weight simultaneously to signify pages in user's current session as follows. First, the session is weighted as done for transactions. This guarantees that the time spent by user on each page and the frequency of page is reflected in the weight of each page. To apply the freshness of each page to its significance, we define the following parameter for each page:

$$Fresh(p_i) = \frac{i}{|w|} \quad i=1, 2, \dots, |w| \quad (8)$$

Where $|w|$ is the size of sliding window and i is the place of page in the sliding window where 1 is assigned to the first visited page. In this Equation the last page is the freshest page. Also, the weighted vector should be normalized to effectively reflect the impact of freshness. The weight of each page is normalized as follows:

$$W_{\text{normalized}}(p_i) = \frac{w(p_i)}{\sum_{j=1}^n w(p_j)} \quad (9)$$

Therefore, in the weight measure we devised, fresh and $W_{\text{normalized}}$ are valued equally. We use the harmonic mean of $W_{\text{normalized}}$ and fresh to represent the interest degree of a web page to a user in the session. Equation (10) guarantees that Interest of a page is high only when $W_{\text{normalized}}$ and fresh are both high.

$$Interest(p_i) = \frac{2 * Fresh(p_i) * W_{\text{normalized}}(p_i)}{Fresh(p_i) + W_{\text{normalized}}(p_i)} \quad (10)$$

For example let

$S = \langle (A, 30), (B, 20), (C, 5), (D, 5), (E, 4), (F, 10) \rangle$ is an active user session after calculating the weight of each page according to Equation (3). Fig. 1 shows the comparison between our method and traditional sliding window. As we set the length of slide window to 3, the traditional method use the 3 latest pages from current session by choosing the page set $X = \{D, E, F\}$ but our method chooses the set $X = \{A, B, F\}$.

Albeit the page A is visited first by user but as it has a large weight than D, E and F so it is the more interested for user and included in our window in contrast to the traditional method that escapes it.

Current Session	A	30	B	20	C	5	D	5	E	4	F	10
Fresh	A	1/6	B	2/6	C	3/6	D	4/6	E	5/6	F	6/6
Normalized Weight	A	30/74	B	20/74	C	5/74	D	5/74	E	4/74	F	10/74
Interest	A	0.22	B	0.29	C	0.10	D	0.09	E	0.094	F	0.23

Traditional Slide Window Scheme for Current Session D E F

Proposed Slide Window Scheme for Current Session A B F

Figure 1. Comparison between our method and traditional sliding window

B. Recommendation Mechanism

We developed a usage model for predictions based on weighted association rule. There are two phases in our system. First, the weighted association rules of each URL will be extracted from the web log data, the rules produced is representing the behavior of user's navigation on the web site. Secondly, the recommendation engine will search the top-n most similar weighted rules to the active user session before generating recommendation for the user. During the second phase instead of exact match between the active user and rules, we use a similarity measure for finding the most similar rules.

Each of the weighted association rules $r = \langle (p_1, p_2, \dots, p_k), (q_{k+1}, q_{k+2}, \dots, q_{k+m}), (w_1, w_2, \dots, w_{k+m}), \delta, \alpha \rangle \in R$ obtained in the mining stage described in the previous section, are represented as a set of page-weight pairs. This will allow for both the active session and the association

rules to be treated as m-dimensional vectors over the space of page in the site. Thus, given a weighted association rule r , we can represent the left-hand side of the each rule r_L as a vector: $r_L = \{w_1, w_2, \dots, w_m\}$, where

$$w_i = \begin{cases} \text{weight}(p_i, r_{L_i}), & \text{if } p_i \in r_L \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

Similarly, the current user session is also represented as a vector $S = \{s_1, s_2, \dots, s_m\}$ where s_i is a significance weight associated with the corresponding page reference, if the user has accessed p_i in this session, and $s_i = 0$, otherwise.

Then we compute the matching score between association rules that capture relationships among page based on their co-occurrence in navigational patterns of users and the current active session. The matching score between them is defined as:

$$\text{Dissimilarity}(S, r_L) = \sum_{i: r_{L_i} > 0} \frac{2 * (w(s_i) - w(r_{L_i}))^2}{w(s_i) + w(r_{L_i})} \quad (12)$$

$$\text{Match Score}(S, r_L) = 1 - \frac{1}{4} \sqrt{\frac{\text{Dissimilarity}(S, r_L)}{\sum_{i: r_{L_i} > 0} 1}} \quad (13)$$

S and r_L represent the active user and left hand side of weighted association rule, respectively.

As the algorithm tries to find rules that are similar to the active user session, the similarity measure between a rule and the active session is dependent on the magnitude of the left-hand side of the rule. Association rules might have multiple items on the right hand side of the rules but, due to the nature of the prediction problem in this paper recommendations are independent of one another and users will select only one of several recommendations so we only use rules that have singleton right-hand sides.

The recommendation engine is the online component of a usage-based personalization system in order to determine which items are to be recommended, a recommendation score is computed for each page p_i . Two factors are used in determining this recommendation score: the overall matching score of the active session to the weighted rules as a whole, and the weighted confidence of the rule. The recommendation scores for the active user are computed by multiplying these factors. Given the weighted association rule and active session S , a recommendation scores for the active session, is computed as follows:

$$\text{Rec}(S, X \Rightarrow p) = \text{Match Score}(S, X) * \text{wconf}(X \Rightarrow p) \quad (14)$$

The improvement of this approach is that instead of exact match between the active user and association rules, both of the similarity between rules and current session and the weighted confidence of $X \Rightarrow p$ are used to determine the recommendation score, not just the confidence value as is used in previous works [7, 14]. We choose the highest recommendation score as the recommendation to the active session.

V. EXPERIMENTAL EVALUATION

In this section we present a set of experiments that we performed for evaluating the impact of our proposed

technique on the prediction process. As our evaluation data set we used the web logs of the DePaul University CTI Web server [26], based on a random sample of users visiting the site for a 2 week period during April 2002. This dataset contains 13745 distinct user sessions of length more than 1 and 683 distinct pages. We split the data sets in two non-overlapping time windows to form training and a test data set. 70% of the data set (9745 sessions) was used as the training set and the remaining was used to test the system. For our evaluation we presented each user session to the system, and recorded the recommendations it made after seeing each page the user had visited.

A. Evaluation Metrics

In order to evaluate the recommendation effectiveness for our method, we measured the performance of proposed method using 2 different standard measures, namely, Precision, Coverage [6]. Recommendation precision measures the ratio of correct recommendations, where correct recommendations are the ones that appear in the remaining of the user session. For each visit session after considering each page p the system generates a set of recommendations $R(p)$. To compute the Precision, $R(p)$ is compared with the rest of the session $T(p)$ as follows:

$$\text{Precision} = \frac{T(p) \cap R(p)}{R(p)} \quad (15)$$

Recommendation coverage on the other hand shows the ratio of the pages in the user session that the system is able to predict before the user visits them:

$$\text{Coverage} = \frac{T(p) \cap R(p)}{T(p)} \quad (16)$$

B. Experimental Results

In all experiments we measured both Precision and Coverage of recommendations against varying number of recommended pages. We used number of recommended pages varying from 1 to 8 to measure the precision and coverage of algorithms. We used a fixed window size of 3 on recommendation history. In Fig.2 and 3 we observed our system performance in comparison with association rules, which is commonly known as one of the most successful approaches in web mining based recommender systems [7].

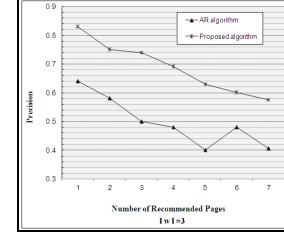


Figure 2. Precision of the AR and proposed algorithm

It can be concluded that our approach is capable of making web recommendation more accurately and effectively against the conventional method. By combining similarity between rules and active user and confidence of the weighted rules, the recommendation engine has selected

only the most relevant pages. Therefore, it increases the effectiveness of the recommendation engine.

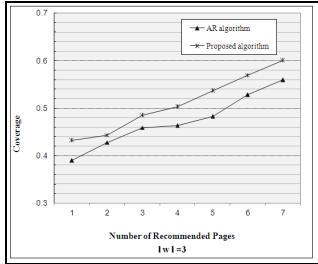


Figure 3. Coverage of the AR and proposed algorithm

VI. CONCLUSION

This paper proposes a new web page personalization system based on the proposed Weighted Association Rule (WAR) model. We extend the association rule mining by assigning a significant weight to the pages based on time spent by each user on each page and visiting frequency of each page. The proposed weighting measure can be used to judge the importance of a page 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. System performance was evaluated under different settings and in comparison with traditional Association Rule based model. The experimental results show that our method is better in precision and coverage rates than the conventional association rule based recommendation, the WAR based model could significantly improve the recommendation effectiveness. As to the recommendation policies, the results show that our sliding window policy is significantly better than the traditional window-sliding.

REFERENCES

- [1] M. Eirinaki, M. Vazirgiannis, "Web mining for Web personalization", ACM Transactions on Internet Technology, 3(1), 2003, pp. 1–27.
- [2] J. Srivastava, R. Cooley, M. Deshpande, P. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", SIGKDD Explorations, 1(2), 2000, pp.2-23.
- [3] R. Cooley, B. Mobasher, J. Srivastava, "Web mining: Information and pattern discovery on the World Wide Web, Proceedings of IEEE international conference tools with AI, 1997, pp. 558–567.
- [4] F. H. Wang, S. M.Thao, "A study on personalized Web browsing recommendation based on data mining and collaborative filtering technology", Proceedings of national computer symposium, Taiwan, 2003, pp. 18–25.
- [5] X. Fu, J. Budzik, K. Hammond, "Mining navigation history for recommendation", In Proceedings of the fifth international conference on intelligent user interfaces, 2000, pp. 106–112.
- [6] M. Gery, H. Haddad, "Evaluation of Web usage mining approaches for user's next request prediction", Proceedings of the fifth ACM international workshop on Web information and data management, 2003, pp. 74–81.
- [7] B. Mobasher, R. Cooley, J. Srivastava, "Automatic Personalization based on Web Usage Mining", Communications of the ACM, 43 (8), 2000, pp. 142-151.
- [8] M. D. Mulvenna, S. S. Anand, A. G. Buchner, "Personalization on the net using Web mining", Communications of the ACM, 43(8), 2000, pp.123–125.
- [9] C. Shahabi, A. Zarkesh, J. Abidi, V. Shah, "Knowledge discovery from user's Web-page navigation", In Proceedings of the 7th IEEE Intl, Workshop on Research Issues in Data Engineering, 1997.
- [10] 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, Proceedings IDEAS'98, July 1998, pp. 68 – 77.
- [11] F. Tao, F. Murtagh, M. Farid," Weighted Association Rule Mining using Weighted Support and Significance Framework", In Proceedings of the 9th SIGKDD Conference, 2003.
- [12] M. Deshpande, G. Karypis," Item-based top-N recommendation algorithms", ACM Transactions on Information Systems (TOIS), 2004.
- [13] J. Herlocker, J. Konstan, A. Brodersen, J. Riedel, "An Algorithmic Framework for Performing Collaborative Filtering", Proceedings of 200 Conference on Research and development in Information Retrieval, 2000.
- [14] B. Mobasher, H. Dai, T. Luo, 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.
- [15] A. Demiriz, "Enhancing Product Recommender Systems on Sparse Binary Data", Data Mining and Knowledge Discovery, 2003.
- [16] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Analysis of Recommender Algorithms for E-Commerce", In Proceedings of the 2nd ACM E-Commerce Conference, Minneapolis, 2000.
- [17] Y. Liang, L. Chunping, " Incorporating Pageview Weight into an Association-Rule-Based Web Recommendation System", Springer-Verlag Berlin Heidelberg, AI 2006, LNAI 4304, 2006, pp. 577 – 586.
- [18] A. A. Manan, M. Hanafi, "Web Page Recommendation Model for Web Personalization", Springer-Verlag Berlin Heidelberg, KES 2004, LNAI 3214, 2004, pp. 587–593.
- [19] R. Cooley, B. Mobasher, J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", In Journal of Knowledge and Information Systems, 1(1), 1999, pp. 5–32.
- [20] P.K. Chan, "A non-invasive learning approach to building web user profiles", in: Workshop on Web usage analysis and user profiling, Fifth International Conference on Knowledge Discovery and Data Mining, San Diego, 1999.
- [21] S. Dumais, T. Joachims, K. Bharat, A. Weigend, "implicit measures of user interests and preferences", 2003 workshop report: ACM SIGIR Forum, Fall 2003.
- [22] M. Morita, Y. Shinoda, "Information filtering based on user behavior analysis and best match text retrieval", in: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Springer-Verlag New York, Inc., Dublin, Ireland, 1994, pp. 272–281.
- [23] R. Agrawal, T. Imielinski, A. Swami, "Mining association between sets of items in massive database", International Proceedings of the ACM-SIGMOD International Conference on Management of Data, 1993, pp. 207–216.
- [24] R. Agrawal, 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,pp. 487-499.
- [25] M. Nakagawa, 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, 2003, pp. 59 – 70.
- [26] <http://maya.cs.depaul.edu/~classes/ect584/data/cti-data.zip>.