Winter 12-9-2003

A user preference perception model using data mining on a Web-based Environment

Chentung Chen

Wei-Shen Tai

Chih-Hao Chu

A User Preference Perception Model Using Data Mining on a Web-based environment

Chen-Tung Chen[1] Wei-Shen Tai[2] Chih-Hao Chu[3]
[1,3] Department of Information Management, Da-Yeh University
[2] Graduate School of Management, Da-Yeh University
112, Shan-Jiau Rd., Da-Tusen, Changhua, Taiwan
chtung@mail.dyu.edu.tw

Abstract

In a competitive environment, how to provide the information and products to meet the requirements of customers and improve the customer satisfaction will be the key criteria to measure a company's competitiveness. Customer Relationship Management (CRM) becomes an important issue in any business market gradually. Using information technology, businesses can achieve their requirements for one to one marketing more efficiently with lower cost, labor and time.

In this paper, we proposed a user preference perception model by using data mining technology on a web-based environment. First, the users' web browse records are aggregated. Second, fuzzy set theory and most sequential pattern mining algorithm are used to infer the users' preference changes in a period. After the test had processed, we use the on-line questionnaire to investigate the customer satisfaction degree from all participators. The results show that the degree of satisfaction was up to 72% for receiving the new information of participants whose preferences had been changed. It indicates that the proposed system can effectively perceive the change of preference for users on a web environment.

Keywords: Preference Perception, Data Mining, Fuzzy Set Theory, Most Sequential Pattern Mining Algorithm (MSPMA)

1. Introduction

Due to the information technology development and the World Wild Web (WWW) prevalence, the traditional business model has a tremendous revolution. As Customer Relationship Management (CRM) concept arises, more and more businesses pay much attention to the interaction between business and customers gradually. Every interaction and modification improves your ability to fit your product to this particular customer [26]. In a real business environment, it is a huge expense in time and labor to implement a one to one marketing solution in the past. Today, information technology can accelerate businesses to improve their interaction with customers more easily and efficiently.

Using the Internet to provide customers' personal products or services, it is an applicable solution to establish a low cost and high performance channel to accelerate interaction between businesses and customers [7]. As the Internet users and information explosively grow, providing users' personal information or services actively that has become an important issue for those Website owners to retain customers' loyalty and satisfaction [31]. In the other hand, analyzing the users' personal searching or reading behavior that is also useful to accelerate businesses to understand the customers' requirements indeed. According to analytic data, the marketing salesman can provide customers one to one promotion activities.

Over these years, data mining and data warehouse become to power tools for businesses to sense the complicated market trend. Data mining and data warehouse retrieve useful knowledge or information from an enormous database, then integrate, accumulate and translate them into a part of the business wisdom [1]. In other words, data mining provides business managers an effective solution to retrieve valuable information from a variety of data sources.

As data mining and data warehouse grow, businesses can analyze customer's consumptive behavior and divide into several different kinds of groups or communities efficiently[30]. In this paper, we propose a user preference perception model using data mining and fuzzy set theory to detect user's preference trends or changes. It is applicable to accelerate businesses to understand what customers' preferences and their changes. Businesses can provide customers more fit and valuable information or services according to their preferences, and improve customer's satisfactions and achieve the aim of CRM.

The organization of this paper is as follows. First, we introduce some literatures include Data Mining, Knowledge Discovery in Database, Web Mining, Fuzzy Set Theory, and Mining Path Traversal Patterns. Second, we proposed a user preference perception model to calculate the degree of user's preferences and obtain the user's browsing patterns. Third, we built a website to demonstrate the feasibility and effectiveness of the proposed model. Finally, we make some conclusions at the end of this paper.

2. Literature Review

2.1. Data Mining and Knowledge Discovery

As information technology grows, businesses can aggregate and store mass data more easily and efficiently. But businesses expend a lot of space, money and labor in maintaining those data and cannot acquire any additional benefit form them in the past. As the number of data and dimensions swift increases, analytic experts are harder to analyze and interpret data than the past. Therefore, Data Mining (DM) and Knowledge Discovery in Database (KDD) arose to support analyst and retrieve valuable information from mass data [14].

Data Mining is one of the most general approaches to reducing data in order to explore, analyze and understand it [15]. In the other views, scholars identify it is one method that can discover and extract implicit previously unknown and potentially valuable information, relationship and knowledge form previous data [4,10,15]. Knowledge discovery in Database (KDD) is the knowledge intensive processes of discovering knowledge that is implicit in large and diverse database [13,22,28]. It is an inherently iterative process of selecting data, preprocessing it, transforming it into a workable form, data mining over it, and interpreting the results [27]. According to the foregoing definition of DM and KDD, we identify the common aim of them is providing an efficient method to retrieve valuable information or knowledge from mass data. By these ways, businesses can acquire valuable information to make important decisions or strategies and enhance their competitiveness.

2.2. Web mining

The Internet has become one of the most popular information communication media in the world recently. It provides abundant, varied, and interactive information through the Web [4,11,18]. Although users can acquire amounts of information from Internet resources, they also face the serious information overload problem. Thus far, browsers and search engines cannot provide the information to meet users' requirements exactly and solve the information overload problem on the Internet effectively[7,20,21]. Web mining is one of effective tools to analyze users' browsing behavior and provide users the information meet their need respectively [18]. Web mining can be broadly defined as the discovery and analysis of useful information from the WWW [11,24]. Etzioni [12] divided Web mining into three parts as follows:

(1) Resource discovery. Locating unfamiliar documents and services on the Web.
(2) Information extraction. Extracting specific information from newly discovered Web resources automatically.
(3) Generalization. Uncovering general patterns at individual Websites and across multiple sites.

Furthermore, there are three major categories in Web mining area as follows [2,5,11,18,20]:

(1) Web content mining. The discovery of useful information from the Web.
(2) Web structure mining. Trying to discover the model underlying the link structure of the Web.
(3) Web usage mining. Trying to make sense of the data generated by the Web surfer's sessions or behaviors.

Web usage mining obtains logs file, user profiles, registration data, user sessions or transactions, cookies and any data as the results of interactions from web severs or proxy servers [18]. It mines those data to discover valuable information about user's intention. In this paper, we use Web usage mining to discovery the user's browsing patterns and perceive preference changes from his/her cookies.

2.3. Fuzzy Set Theory

A fuzzy set $\tilde{A}$ in a universe of discourse X is characterized by a membership function $\mu_{\tilde{A}}(X)$ which associates with each element x in X a real number in the interval [0,1]. The function value $\mu_{\tilde{A}}(X)$ is termed the grade of membership of x in $\tilde{A}$ [8,32].

Definition 2.1. A fuzzy set $\tilde{A}$ of the universe if discourse of X is convex if and only if for all x_1, x_2 in X,

$$\mu_{\tilde{A}}(\lambda x_1 + (1-\lambda)x_2) \geq \min(\mu_{\tilde{A}}(x_1), \mu_{\tilde{A}}(x_2)), \quad (1)$$

where $\lambda \in [0,1]$.

Definition 2.2. A fuzzy set $\tilde{A}$ of the universe if discourse of X is called a normal fuzzy set implying that $\exists x_i \in X$, $\mu_{\tilde{A}}(x_i) = 1$.

A fuzzy mumber $\tilde{n}$ is a fuzzy subset in the universe of discourse X whose membership function is both convex and normal (shown in Fig. 1).

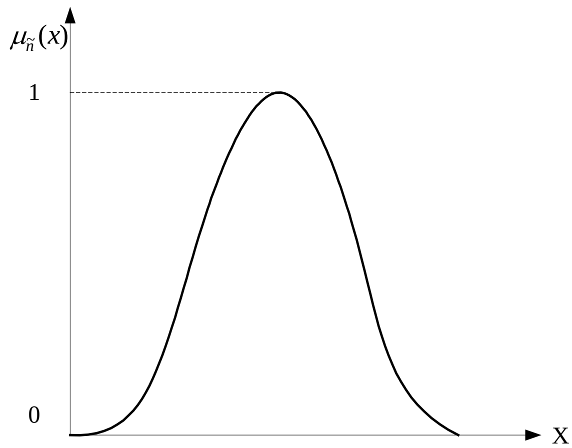

Fig. 1 A fuzzy number $\tilde{n}$.

Definition 2.3. The α-cut of a fuzzy number $\tilde{n}$ is defined as

$$\tilde{n}^{\alpha} = \{x_i : \mu_{\tilde{n}}(x_i) \geq \alpha, x_i \in X\}, \quad (2)$$

where $\alpha \in [0,1]$.

$\tilde{n}^{\alpha}$ is a non-empty bounded closed interval contained in X and it can be denoted by $\tilde{n}^{\alpha} = \left[n_l^{\alpha}, n_u^{\alpha}\right]$, n_l^{α} and n_u^{α} are the lower and upper bounds of the closed interval, respectively. Fig. 2 shows a fuzzy number $\tilde{n}$ with α-cut [17,33].

Up to now, the fuzzy set theory has applied in many areas such as fuzzy logic control [6], fuzzy decision-making [19], fuzzy expert system [16], and fuzzy information retrieve [23] etc. In this paper, we apply the membership function of fuzzy set theory to represent and calculate the degree of user's preferences. It can help us to deal with the relationship between user's browsing time and the length of web page content efficiently.

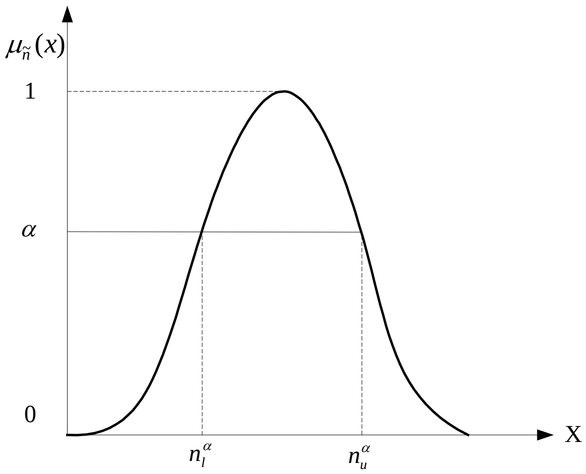

Fig. 2 A fuzzy number $\tilde{n}$ with α-cut.

2.4. Mining Path Traversal Patterns

In a distributed information-providing environment where documents or objects are lined together, users are apt to traverse objects back and forth in accordance with the links and icons provided [10]. Therefore, extracting the patterns from user's log data in such environment is referred to as mining path traversal patterns. It facilitates web-site owners or managers to understand what access paths are users interested in and relationships between web pages. These information or relationships will be the important references for website owners or managers to improve the website structure or content. For the sake of extracting valuable user's access patterns from log data, there are several procedures is described as follows [9].

(1) Determining maximal forward references from the original log data. Using algorithms to convert the original sequence of log data into a set of traversal subsequences. Each traversal subsequence represents a maximal forward reference from the starting point of a user access.

(2) Determining the large reference sequences. Using algorithms to determine the frequent traversal patterns and the termed large reference sequences from the maximal forward references. A large reference will appear a sufficient number of times in the database.

(3) Determining the maximal reference sequences. After large reference sequences are determined, maximal reference sequences can then be obtained in straightforward manner. A maximal reference sequence is a large reference sequence that is not contained in any other maximal reference sequences.

3. User Preference Perception Model

In this paper, we analyze log data from web servers to obtain user's surfing tracks and browsing time. Then both of fuzzy set theory and maximal reference sequences are applied to perceive the changes of user's preference in our model.

3.1. Converting log data

Obtaining user's surfing paths from the log data of web servers and convert them into user's traversal subsequence sets by user's number respectively. The traversal subsequence is a set of web-page nodes and sorted by the time-stamp in a determinate period, e.g. T_1={Y, B, H, C, T, I, H, J} and T_2={Y, H, P, T, I, K, H, J}.

3.2. Eliminating the uninterested nodes

Eliminating the node whose browsing time exceeds the maximum threshold value from the user traversal subsequence set. In this paper, the maximum threshold value is denoted by $t_{\max}$. When the number of users visit the web page, we can compute the average of browsing time (t_{avg}). If a visitor's browsing time exceeds the average of browsing time in this web page, the degree of the visitor's preference is regarded as 1. It means that this visitor shall interest in this web page completely. Contrary to the foregoing situation, a visitor's browsing time more short means the visitor has much less interests about this web page. Therefore, the degree of the visitor's preference that depends on the browsing time is represented by the membership function of fuzzy set theory (shown in Fig. 3). The degree of preference (P) is calculated as follows.

$$P_{ki} = \begin{cases} 0, \; t_{\max} < t_{ki} \\ 1, \; t_{avg} \le t_{ki} \le t_{\max} \\ \dfrac{t_{ki}}{t_{avg}}, \; 0 \le t_{ki} < t_{avg} \end{cases} \tag{3}$$

where, t_{ki} is the browsing time of the kth visitor in the ith web page, P_{ki} is the degree of the kth visitor's preference in the ith web page.

According to the membership function of the preference degree, α-cut is applied to divide the web pages into two categories as follows.

(1) If $P_{ki} \geqq \alpha$, we think that the kth visitor has

interests about the *i*th web page. We reserve this web page in database to analyze the preference of this visitor at the rest of procedures.

(2) If $P_{ki} < \alpha$, we think that the *k*th visitor has no interests about the *i*th web page and eliminate this web page in our model.

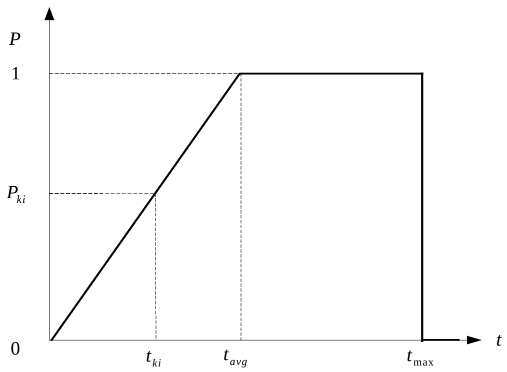

Fig. 3 The membership function of the preference degree.

3.3. Extracting browsing patterns

After the foregoing procedures, the interesting web pages of all users will be stored in database D_T. In this paper, each web page belongs to a category. Then, Full Scan (FS) algorithm, essentially utilizes the concept of DHP (Direct Hashing and Pruning), is applied to solve the discrepancy between traversal patterns and association rules [25].

For example, a user's traversal subsequence set is shown in Table 1. Recall that L_k represents the set of all large *k*-references and C_k is a set of candidate *k*-references. The procedures of the FS are described as follows.

Step 1. Candidate 1-reference sequences and their support value (appearance frequency) are listed in C_1 (shown in Table 2).

Step 2. After scanning through D_T, those *k*-references in C_k with count exceeding the threshold become L_k. Thus, FS gets L_1 and makes a hash table to count the number of occurrences of each 2-reference.

Step 3. Staring with $k = 2$, FS generates C_k based on the hash table count obtained in the previous pass and determines the set of large k-references L_k, and makes a hash table to determine the candidate (k + 1)-references. If L_k is nonempty, the iteration continues for the next pass (shown in Table 3 and 4).

In mining association rules, a set of candidate references C_k can be generated from joining L_{k-1} with itself, denoted by L_{k-1}* L_{k-1}. If we set the threshold equal to 3, the results of extracting browsing patterns are listed in Table 5.

Table 1 The original traversal sequences.

Traversal Sequences	Item Set
T_1	*<Y,B,H,C,T,I,H,J>*
T_2	*<Y,H,P,T,I,K,H,L,J>*
T_3	*<P,H,I,T,Y,J,H,P>*

Table 2 C_1 To L_1

C_1			L_1	
Item Set	Support		Item Set	Support
<T>	3		*<P>*	3
<B>	1		*<Y>*	3
<C>	1		*<J>*	3
<H>	6		*<I>*	3
<I>	3		*<H>*	6
<J>	3		*<T>*	3
<Y>	3			
<P>	3			
<K>	1			
<L>	1			

Table 3 C_2 To L_2

C_2			L_2	
Item Set	Support		Item Set	Support
<TH>	3		*<TH>*	3
<TI>	2		*<TJ>*	3
<TJ>	3		*<HT>*	3
<TY>	1		*<HI>*	3
<TP>	1		*<HJ>*	5
<HT>	3		*<HP>*	3
<HI>	3		*<IH>*	3
<HJ>	5		*<IJ>*	3
<HY>	1		*<YH>*	5
<HP>	3		*<YJ>*	3
<IT>	1		*<PH>*	3
<IH>	3			
<IJ>	3			
<IY>	1			
<IP>	1			
<JT>	0			
<JH>	1			
<JI>	0			
<JY>	0			
<JP>	1			
<YT>	2			
<YH>	5			
<YI>	2			
<YJ>	3			
<YP>	2			
<PT>	2			
<PH>	3			
<PI>	2			
<PJ>	2			
<PY>	1			

Table 4 C_3 To L_3

C_3			L_3	
Item Set	Support		Item Set	Support
<THT>	0		*<HTJ>*	3
<THI>	0		*<HTH>*	3
<THJ>	2		*<HIH>*	3
<THP>	1		*<HIJ>*	3
<HTH>	3		*<YHJ>*	4
<HTJ>	3			
<HIH>	3			
<HIJ>	3			
<HPH>	1			
<IHT>	0			
<IHI>	0			
<IHJ>	2			
<IHP>	1			
<YHT>	2			
<YHI>	2			
<YHJ>	4			
<YHP>	2			
<PHT>	1			
<PHI>	1			
<PHJ>	2			
<PHP>	2			

Table 5 User browsing patterns.

User Browsing Patterns	Support
<P,H>	3
<H,P>	3
<Y,H,J>	4
<Y,I,J>	3
<H,I,H>	3
<H,T,J>	3
<H,T,H>	3

3.4. Perceiving user's preferences

According to user browsing patterns and the degree of preference in each category, changes of a user's preferences will be perceived in the proposed model. In this study, each browsing pattern is regard as a document and its web pages that belong to the same category are regarded as a same term. The vector and the weighted value are applied to represent the degree of preference in different categories. The degree of preference in each category is calculated as follows.

$$P_i = (w_{i1}, w_{i2} \cdots, w_{ik}) \tag{4}$$

where, the P_i is the ith user's preference degree in each category, w_{ij} is the ith user's preference degree in the jth category.(j=1, 2,…,k)

If a category appears in the almost of browsing patterns, it's the degree of importance should be less than the other categories appears in little patterns. Categories that are less important for browsing patterns should be assigned lower weights, whereas categories that are more important in a browsing pattern should be assigned higher weights. According to this concept, we use TFIDF (Term Frequency – Inverse Document Frequency) [29] to adjust the degree of preference in each category from a browsing pattern. The formula is described as follows.

$$w_j^* = \sqrt{\frac{f_j}{n} \times \log\left(\frac{p_n}{p_j} + 1\right)} \tag{5}$$

where, w_j^* is the degree of preference in the jth category, f_j is the appearance frequency of the jth category, n is the total number of appearance frequency from each category in all browsing patterns, p_n is the total number of browsing patterns, p_j is the number of browsing patterns that contain the jth category.

For example, if the ith user's initial preferences are both of H and T categories, his/her preferences are denoted by $P_i = (w_H, w_P, w_Y, w_I, w_J, w_T) = (1,0,0,0,0,1)$. After a determinate period, if the system gets user's browsing patterns are as same as table 5. The degree of preference in the H category is calculated as follows.

$$w_H^* = \sqrt{\frac{9}{19} \times \log\left(\frac{7}{7} + 1\right)} = 0.377$$

According to the results of preference measurement, user's preferences are denoted by P_i = (0.377, 0.262, 0.212, 0.262, 0.726, 0.262). When the threshold of the degree of preference is 0.3, we filter out those preferences whose degree of preference are lower than the threshold and retain those preferences whose degree of preference are equal or higher than threshold. Thus, user's new preferences are denoted by $P_i = (w_H, w_P, w_Y, w_I, w_J, w_T) = (1,0,0,0,1,0)$. Comparing the discrepancy between initial and new preferences, the change of user's preferences are perceived easily and obviously.

4. System design and experiments

In this paper, a website for providing personal news is built to demonstrate the feasibility and effectiveness of the proposed model. The system architecture and experiments are elaborated as follows.

4.1. System architecture and functions

In this study, a website for providing personal information is built in a personal computer with a Pentium 4 processor and Linux operation system. The architecture of this proposed system is shown in Fig. 4. The major components of this system and their functions are described as follows.

(1) Web server log database. When a registered member logged in the website, all of his/her surfing paths and browsing time are recorded and stored in the server log database.

(2) Converting log data. The system will convert log data into user's traversal subsequence sets by user's

registered number respectively.

(3) Eliminating user uninterested nodes. In this stage, the system depends on threshold limit value of browsing time and α-cut to eliminate those web pages whose browsing time are excess long or the degree of preference are excess low.

(4) Extracting browsing patterns. FS will extract applicable browsing patterns from user's traversal subsequence sets.

(5) Perceiving user's preferences. After a determinate period, user's new preferences are perceived in this component. The system will provide user news according to his/her new preferences.

(6) Providing general or personal news. User can select any ways that meet his/her requirements for providing news. If the function of general news providing is selected, website will provide all categories of news in news database by date. Otherwise, website will provide user personal news by his/her preferences in user profile database.

(7) News search engine. users can search news that they needs manually; even both of date and the type of category are avail conditions in the news search proceeding.

(8) User profile database. When a user registered in the web site, all of information, including name, password, email, initial preferences and the others data are stored in user profile database.

(9) News database. The news from several news providers are stored and sorted by their category attribution. In periodical, the system will update news database every four hours.

4.2. Experiments and results

In order to verify the feasibility and effectiveness of the preference perception model we proposed, a website is built and experiments are implemented at Da-Yeh University in Taiwan. The results of the experiment sufficient indicate this proposed model is effective in perceiving the change of user preferences.

4.2.1. Experimental procedure

In this experiment, the participants are students at Da-Yeh University. The topical categories are divided into "Stock and Security", "Society and Culture", "International", "Home and Living", "Mainland China", "Business and Economy", "Regional", "Government", "Entertainment", "Science and Computer", "Recreation and Sports". The experimental procedure is described as follows (shown in Fig.5).

Step 1. Member registration. All participants filled out their profiles like name, password, email address and preferences in the registration form. (shown in Fig.6).

Step 2. Providing member personal news. When a member login the website, he/she will see the personal news by category according to his/her preferences from user profile database (shown in Fig.7).

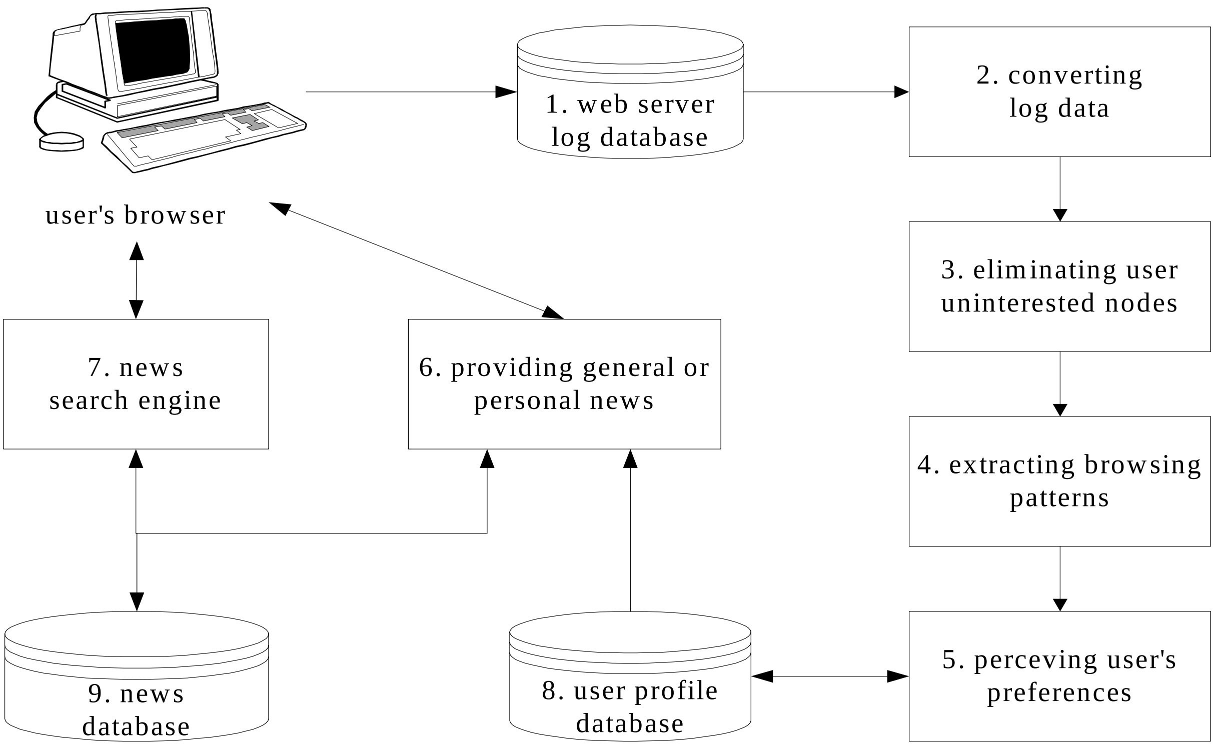

Fig. 4 The system architecture of the website.

Step 3. Recording member’s browsing behavior. The cookies in the member’s browser are used to record all of surfing web pages, browsing time and timestamp. Till the member access the next web page, the cookies will be sent to the web server and stored in web server log database.

Step 4. Perceiving member’s preferences. According to the log data in three weeks, user’s preferences are analyzed based on the user perception model we proposed. Due to the average of adult reading speed is 200 words per minute approximately [3], the average of reading time is applied to filter out those web pages that average of browsing time are excess long. In addition, the α is set 0.5 and threshold of traversal sequences is set 9. Thus, both of user browsing patterns and the degree of preference in each category are obtained from log data.

Step 5. Providing member personal news. The website selects several news from those categories that meet user’s preferences randomly. Members can see those selected news by category.

Step 6. Satisfaction survey. In each bottom of news web page, members were asked to fill the questionnaire of satisfaction. The entire members and individual degree of satisfaction are calculated by the formula of satisfaction measurement.

4.2.2. Results analysis

During the experiment for the proposed preference perception model, there are 16176 pieces of news and 8087 log records are retrieved. Using the preference perception model, 576 avail maximal reference sequences and 314 browsing patterns are extracted and the changes of preference in 23 participants are perceived (shown in Table 6).

The results of the degree of satisfaction measurement for all participants are shown in Table 7. According to the results of satisfaction measurement, the average of the degree of satisfaction from member 3, 10 and 21 (their preferences had not changed) is 0.637. In the other side, the average of the degree of satisfaction from the other members (their preferences had changed) is 0.724. Thus, we can discover the degree of satisfaction for all members that preferences had been changed is higher than the others with no changes. In other words, the almost members can satisfy the results of the preference perception model we proposed.

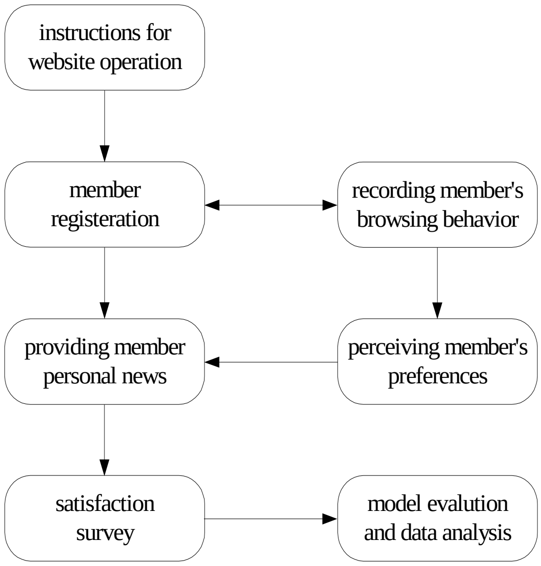

Fig. 5 A flowchart for experimental processes.

Fig. 6 Fill out form of the member profile.

Fig. 7 Personal news and category title.

5. Conclusions

In this paper, we proposed a user preference perception model using data mining on a web-based environment. After aggregating the user's web browse records, fuzzy set theory and most sequential pattern mining algorithm are applied to perceive the changes of user's preferences in a period. According to the user preference perception model, a web site for providing personal news is built to demonstrate the foregoing model we proposed. As the result of the degree of satisfaction survey from all participators, the degree of satisfaction are up to 72 % from the users whose preference had been changed in the test period.

Because of the convenience and prevalence the Internet provides, more and more businesses provide information, service and product through it. If businesses can provide customer suitable service or products, they will improve not only competitiveness but also customer satisfaction and loyalty. The user preference perception model we proposed would accelerate businesses to provide more suitable information, services and products that meet the customer's requirements and make both of them enjoy the benefit through a web-based environment.

6. Acknowledgement

The authors gratefully acknowledge the financial support of the National Science Council, Taiwan, under project numbers NSC 91-2213-E-212-016 and NSC 91-2745-P-212-001.

Table 6 The results of perceiving user preference.

No.	The weight in each categories	Initial preferences	New preferences
1	4(0.516); 6(0.317); 9(0.256)	(00010100100)	(00010100100)
2	6(0.623); 11(0.299)	(00000111011)	(00000100001)
3	10(0.549)	(00000000010)	(00000000010)
4	10(0.549)	(01110000110)	(00000000010)
5	4(0.202); 9(0.444); 11(0.325)	(00111001111)	(00010000101)
6	2(0.303); 4(0.214); 7(0.214); 9(0.399); 11(0.358)	(00110000101)	(01010010101)
7	10(0.549);	(00000001110)	(00000000010)
8	4(0.400); 7(0.221); 9(0.355)	(01010000100)	(00010010100)
9	2(0.277); 4(0.283); 7(0.238); 9(0.184); 10(0.324)	(00010000011)	(01010010010)
10	9(0.367); 11(0.408)	(00000000101)	(00000000101)
11	2(0.364); 4(0.448); 9(0.364)	(01011010010)	(01010000100)
12	3(0.549)	(01111011110)	(00100000000)
13	4(0.362); 7(0.384); 9(0.265)	(00010010110)	(00010010100)
14	2(0.253); 3(0.263); 4(0.272); 7(0.358); 9(0.233)	(01100001111)	(01110010100)
15	9(0.549)	(01110011110)	(00000000100)
16	2(0.316); 4(0.255); 7(0.400)	(01111000010)	(01010010000)
17	2(0.139); 4(0.421); 9(0.328); 11(0.230)	(00000000001)	(00010000101)
18	9(0.549)	(11000000100)	(00000000100)
19	2(0.304); 4(0.219); 9(0.431)	(00000000100)	(01010000100)
20	7(0.321); 9(0.472)	(01010010110)	(00000010100)
21	4(0.398); 7(0.331); 9(0.241)	(00010010100)	(00010010100)
22	4(0.466); 9(0.289)	(10010100110)	(00010000100)
23	11(0.549)	(00000000101)	(00000000001)
24	2(0.214); 9(0.354); 10(0.153); 11(0.397)	(01000000011)	(01000000101)
25	4(0.309); 9(0.491)	(00010000000)	(00010000100)
26	2(0.374); 3(0.349); 6(0.272); 8(0.410); 9(0.167)	(00100101010)	(01100101000)

Table 7 The results of satisfaction measurement.

No.	The degree of satisfaction in each category	Average
1	6(0.867); 9(0.667); 4(0.8)	0.778
2	11(0.8); 6(0.6)	0.700
3	10(0.667)	0.667
4	10(0.933)	0.933
5	4(0.667); 9(0.6); 11(0.8)	0.689
6	9(0.8); 4(0.6); 2(0.533); 7(0.6); 11(0.8)	0.667
7	10(0.933)	0.933
8	4(0.667); 9(0.467); 7(0.6)	0.578
9	4(0.733); 10(0.933); 2(0.8); 7(0.733)	0.800
10	11(0.8); 9(0.667)	0.733
11	2(0.8); 9(0.733); 4(0.533)	0.689
12	3(0.733)	0.733
13	4(0.733); 9(0.6); 7(0.867)	0.733
14	2(0.6); 7(0.467); 9(0.8); 4(0.733); 3(0.2)	0.560
15	4(0.533); 9(0.667)	0.600
16	2(0.933); 7(0.8); 4(0.933)	0.889
17	9(0.533); 4(0.733); 11(0.867)	0.711
18	9(0.667)	0.667
19	4(0.6); 2(0.267); 9(0.533)	0.467
20	9(0.667); 7(0.933)	0.800
21	4(0.667); 7(0.4); 9(0.467)	0.511
22	9(0.6); 4(0.667)	0.633
23	11(0.6);	0.600
24	11(1); 9(1); 2(1)	1.000
25	4(0.733); 9(0.667)	0.700
26	6(0.667); 3(1); 2(0.6); 8(0.933)	0.800

References

[1] Berry, Michael J. A. & Linoff, G., Data Mining Techniques: For Marketing, Sales, and Customer Support, Wiley, 1997.

[2] Borges, J. & Levene, M., "Data Mining of User Navigation Patterns," *Lecture Notes In Computer Science archive Revised Papers from the International Workshop on Web Usage Analysis and User Profiling*, Springer-Verlag, 1999, 92-111.

[3] Celce-Murcia, M., "Reading for Academic Purpose: Guidelines for the ESL/EFL teacher," *Teaching English as a second or foreign language* (the 3rd ed.), Boston: Heinle, 2001, 187-204.

[4] Chakrabarti, S., "Survey articles: Data mining for hypertext: a tutorial survey," *ACM SIGKDD Explorations Newsletter*, 2000, 1(2), 1-11.

[5] Chakrabarti, S., Dom, B.E., Kumar, S.R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D. & Kleinberg, J., "Mining the Web's link structure," *Computer*, 1999, 32(8), 60-67.

[6] Chen, C.S., "Design of stable fuzzy control systems using Lyapunov's method in fuzzy hypercubes," *Fuzzy Sets and Systems*, 2003, 139(1), 95-110.

[7] Chen, C.T. & Tai, W.S., "An information push-delivery system design for personal information service on the Internet," *Information Processing and Management*, 2003, 39(6), 873-888.

[8] Chen, C.T., "A fuzzy approach to select the location of the distribution center," *Fuzzy Sets and Systems*, 2001, 118(1), 65-73.

[9] Chen, M.S., Han, J. & Yu, P.S., "Data mining: an overview from a database perspective," *IEEE Transactions on Knowledge and Data Engineering*, 1996, 8(6), 866-883.

[10] Chen, M.S., Park J.S. & Yu, P.S., "Efficient data mining for path traversal patterns," *IEEE Transactions on Knowledge and Data Engineering*, 1998, 10(2), 209-221.

[11] Cooley, R., Mobasher, B. & Srivastava, J., "Web mining: information and pattern discovery on the World Wide Web," *Ninth IEEE International Conference on Tools with Artificial Intelligence Proceedings*, 1997, 558-567.

[12] Etzioni, O., "The World-Wide Web: quagmire or gold mine?" *Communications of the ACM*, 1996, 39(11), 65-68.

[13] Fayyad U.M., Piatetsky-Shapiro, G. & Smyth, P. "From data mining to knowledge discovery: an overview," *Advances in knowledge discovery and data mining*, American Association for Artificial Intelligence, 1996, 1–34.

[14] Fayyad,U. & Uthurusamy R., "Data mining and knowledge discovery in databases," *Communications of the ACM*, 1996, 39 (11), 24-26.

[15] Fayyad,U. & Uthurusamy R., "Evolving data mining into solutions for insights: Evolving data into mining solutions for insights," *Communications of the ACM*, 2002, 45(8), 28-31.

[16] Hall, L.O., "Rule chaining in fuzzy expert systems," *IEEE Transactions on Fuzzy Systems*, 2001, 9(6), 822-828.

[17] Kaufmann, A. & Gupta,M.M. *Introduction to Fuzzy Arithmetic: Theory and Applications*, New York: Van Nostrand Reinhold, 1985.

[18] Kosala, R. & Blockeel, H., "Web mining research: a survey," *ACM SIGKDD Explorations Newsletter archive*, 2000, 2(1), 1-15.

[19] Lee, H.S., "Optimal consensus of fuzzy opinions under group decision making environment," *Fuzzy Sets and Systems*, 2002, 132(3), 303-315.

[20] Madria, S.K., Raymond, C., Bhowmick, S. & Mohania, M., "Association rules for Web data mining in WHOWEDA," *International Conference on. Digital Libraries: Research and Practice*, 2000, 227-233.

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

[22] Markellos, K., Perdikuri, K., Markellou, P., Sirmakessis, S., Mayritsakis, G. & Tsakalidis, A., "Knowledge Discovery in Pattern Databases," *Proceedings of the eleventh international conference on Information and knowledge management*, 2002, 672-674.

[23] Miyamoto, S., "Information clustering based on fuzzy multisets," *Information Processing and Management*, 2003, 39(2), 195-213.

[24] Pal, S.K., Talwar, V. & Mitra, P., "Web mining in soft computing framework: relevance, state of the art and future directions," *IEEE Transactions on Neural Networks*, 2002, 13(5), 1163-1177.

[25] Park, J.S., Chen, M.S. & Yu, P.S., "An effective hash-based algorithm for mining association rules," *Proceedings of the 1995 ACM SIGMOD international conference on Management of data*, 1995, 175–186.

[26] Peppers, D., Rogers, M. & Dorf, B.," Is Your Company Ready for One-to-One Marketing?" *Harvard Business Review*, 1999, 77(1), 151-160.

[27] Phillips, J. & Buchanan, B.G., "Ontology-guided knowledge discovery in databases," *Proceedings of the international conference on Knowledge capture*, 2001, 123-130.

[28] Piatetsky-Shapiro, G., "Knowledge discovery in databases: 10 years after," *ACM SIGKDD Explorations Newsletter*, 2000 , 1(2), 59-61.

[29] Salton, G., "Another look at automatic text-retrieval systems," *Communications of the ACM*, 1986, 29(7), 648-656.

[30] Song, H.S., Kim, J.K. & Kim, S.H., "Mining the change of customer behavior in an internet shopping mall," *Expert Systems with Applications*, 2001, 21(3), 157-168.

[31] Yen, Benjamin P.C. & Kong, Robin C.W., "Personalization of information access for electronic catalogs on the web," *Electronic Commerce Research and Applications*, 2002, 1(1), 20-40.

[32] Zadeh, L.A., "Fuzzy sets," *Inform. and Control*, 1965, 8, 338-353.

[33] Zimmermann, H.J., *Fuzzy Set Theory and its Applications* (the 2nd ed.), Boston: Kluwer Academic Publishers, 1991.