International Journal of Innovative
Computing, Information and Control
Volume **9**, Number **2**, February **2013**

ICIC International ©2013 ISSN 1349-4198
pp. 503–523

COMBING CUSTOMER PROFILES FOR MEMBERS' RETURN VISIT RATE PREDICTIONS

Chun-Hao Chen[1], Rui-Dong Chiang[1,*], Yi-Hsin Wang[2] and Huan-Chen Chu[1]

[1]Department of Computer Science and Information Engineering
Tamkang University
No. 151, Yingzhuan Rd., Tamsui Dist., New Taipei 25137, Taiwan
chchen@mail.tku.edu.tw; HuanChen.Chu@gmail.com
*Corresponding author: Chiang@cs.tku.edu.tw

[2]Department of Information Management
Chang Gung University of Science and Technology
No. 261, Wen-hwa 1st Rd., Kwei-shan, Taoyuan, Taiwan
yishing@gw.cgust.edu.tw

Received November 2011; revised March 2012

Abstract. *The major profit of portal companies in Taiwan is generated by online advertising and e-commerce. Effective advertising requires predicting how a user responds to an advertisement and then targeting (presenting the advertisements) to reflect users' favor. As customers' preferences may change over time, we take the different types of past behavior patterns of the registered members to capture concept drifts. Then, we combine the click preference index (CPI) and the preference drifts to propose a Behavioral Preference (BP) model, and to predict the members' return visit rates in the specific category of the portal site. The marketers of the portal site can target the registered members with high return visit rates and design corresponding marketing strategies. The experimental results with a real dataset show that our model can effectively predict the registered members' return visit rates.*

Keywords: Behavioral targeting, Customer profile, Concept drift

1. **Introduction.** Portal is a website considered as a linking page among websites. It presents information from diversified sources in a unified way [1]. Therefore, it is incapable of gathering customers' browsing behavior information after the customers' leaving for external websites. Generally speaking, the portal company only has statistical information about customers' browsing clicks for registered members. However, the portal often integrates with e-commerce in Taiwan [2-4]. It provides information and abundant services on nearly all conceivable topics (categories): news, sports, travel, personal finance, entertainment, games, shopping, goods consignment and much more. Therefore, in addition to click information, the portal company in Taiwan also has the information in purchase records.

The online advertising channels such as sponsor search and contextual ads are showing great market potential. Effective advertising requires predicting how a user responds to an advertisement and then targeting (presenting the ad) to reflect the users' favor. One strategy used in increasing the value of advertising is behavioral targeting. Behavioral targeting aims to target the marketing message to a segment of customers which has the highest likelihood of acting upon that message. For example, if we were in the sports marketing business, our target customer base would be the individuals who are enthusiastic about sports. It is directly perceived that people who visit sport related websites or blogs are likely to be our target. Therefore, to analyze historical behaviors to recognize users' preferences and to build customers' profiles is the core of behavioral targeting [5].

In recent years, capturing the variance of customer preferences correctly and constructing customer profiles on the Internet have been getting considerable attention [6,7], especially in the research of Recommenders System [8-11]. Accurately reflecting the preference of customers relies on constructing a precise customer profile [12]. Generally speaking, there are two types of feedback to construct a customer profile: one is the explicit feedback (rating) and the other is the implicit feedback (click, purchase, or tag) [13]. As pointed out by [14], the implicit feedback better reveals the interests of the customer in comparison with the explicit feedback. For example, Jason liked "*Titanic*", so he rated the film 5 points, even though he had only seen it once, while "*Rambo*" which Jason had seen thirty times, was only rated 3 points. Although Jason's explicit feedback shows that he likes "*Titanic*", his past behaviors reveal that he actually liked "*Rambo*" more. Therefore, we facilitated the members' implicit feedback to construct a member's preference model.

Customer preference changes over time. Often the cause of changes is hidden, which makes the learning task complicated. Hidden changes can induce some radical changes, which are known as concept drift. There are two basic types of concept drift which may occur in the real world: (1) sudden concept drift, and (2) gradual concept drift [15]. For example, after a movie, someone suddenly collects soundtracks for a singer who he disliked before. This is a sudden concept drift. Another example is that, when someone graduates from school and is ready for taking up an occupation, his preference gradually focuses on professional knowledge. This is gradual concept drift. As far as we know, to date, there is few research in using the implicit feedback to construct a customer profile for capturing concept drift.

In the paper, we would like to predict members' return visit rates (the possibility of each registered member who may click (browse) the websites at least once or purchase at least one product with respect to a specific category of the portal) within a given time frame. The main contributions in this study are listed as follows:

1. The first one is that the proposed approach takes the different types of past return visit patterns as basis to capture the members' concept drift over time.
2. The second one is that the proposed approach combines the click preference index (CPI) and the preference drifts to propose a Behavioral Preference (BP) model to predict the members' return visit rates.
3. The third one is that the proposed approach provides a basis for the marketers of portal sites to design corresponding marketing strategies.

The experimental results with a real dataset show that the model can achieve our goal and can predict members' return visit rates accurately.

2. **Related Works.**

There are two types of feedback to construct a customer profile: explicit feedback and implicit feedback. Since our research is based on implicit feedback, we review the related literature here in some detail.

Several studies have applied customer implicit feedback for a product into a customer profile for predicting customer product preferences. These implicit feedback have included: user purchase patterns, web page visits, and web browsing paths [16]. In addition, [17] has proposed that customers of online stores go through four main shopping steps: product impression, click-through patterns, basket placement, and purchase. These shopping steps are the fundamental elements in building a customer profile. These researchers used customer behavioral data, including click frequency, "add to cart" frequency, and purchase frequency to construct customer profiles [18-21]. Among them, [19] proposed the following approach to build customer profiles that is depicted in Equation (1):

$$P_{i,j} = \frac{P^c_{i,j} - \min(P^c_{i,j})}{\max(P^c_{i,j}) - \min(P^c_{i,j})} + \frac{P^b_{i,j} - \min(P^b_{i,j})}{\max(P^b_{i,j}) - \min(P^b_{i,j})} + \frac{P^p_{i,j} - \min(P^b_{i,j})}{\max(P^p_{i,j}) - \min(P^p_{i,j})} \tag{1}$$

In Equation (1), $P_{i,j}$ is the degree of preference customer i for a product j; $P^c_{i,j}$ is the number of times of customer i clicks on product j within a given time period; $P^b_{i,j}$ is the number of times customer i adds product j to the basket within the given time period; $P^p_{i,j}$ is the number of times customer i purchases product j within the given time period. Equation (1) shows that the customer preference fraction $P_{i,j}$ is the sum of those three normalized values, $P^c_{i,j}$, $P^b_{i,j}$, $P^p_{i,j}$. It ranges between 0 and 3. The larger the customer's $P_{i,j}$ is, the more the customer likes the product. Through experiments, they confirmed that their method can predict customer product preferences.

Table 1 gives an example of the customer preference model based on Equation (1). It is clear that $\max(P^c_{i,j})$ is 10, $\min(P^c_{i,j})$ is 4, $\max(P^b_{i,j})$ is 8, $\min(P^b_{i,j})$ is 2, $\max(P^p_{i,j})$ is 8, and $\min(P^p_{i,j})$ is 0. $P_{i,j}$ of customer A is $(10-4)/(10-4)+(2-2)/(8-2)+(0-0)/(8-0) = 1$. Similarly, $P_{i,j}$ of customer B and C are 0.92 and 2, respectively.

TABLE 1. An example of the customer preference model

	$P^c_{i,j}$	$P^b_{i,j}$	$P^p_{i,j}$	$P_{i,j}$
Customer A	10	2	0	1
Customer B	6	4	2	0.92
Customer C	4	8	8	2

In general, the model above is used in the Recommenders system. If we directly apply it to the return visit rate prediction, it will lead to an inaccurate in prediction because the customers' past behavior patterns are not considered. A detailed discussion of this will follow in the next section.

3. **Problem Statements.** In our study, since our data does not contain members' basket placement information, we borrowed the method of [19] to construct a click preference index (CPI) for the members' return visit rate predictions. It is shown in Equation (2).

$$CPI_{i,j} = \frac{C_{i,j}}{Max(C_j)} + \frac{P_{i,j}}{Max(P_j)} \tag{2}$$

where $C_{i,j}$ represents the total number of clicks (not including purchase clicks) for a member i within a given time period j; $P_{i,j}$ represents the total number of purchase records for member i within a given time period j; $Max(C_j)$ and $Max(P_j)$ are the maximum values for the number of clicks and purchases within the given month j. $CPI_{i,j}$ is normalized to avoid a single extreme value skewing the click preference index model for member i. $\frac{C_{i,j}}{Max(C_j)}$ and $\frac{P_{i,j}}{Max(P_j)}$ are respectively limited between 0 and 1. We use the member past implicit feedback in the previous months to predict the return visit rate of the next month. Basically, we expect that the higher CPI will imply the higher return visit rate.

In this investigation, we apply each different CPI_{j-1} to observe its return visit rate in $Month_j$. The return visit rate for CPI_{j-1} is defined in Equation (3).

$$\begin{aligned}&\text{the return visit rate for the same } CPI_{j-1}\\ &= \frac{\text{the total number of return customers in } Month_j \text{ for the same } CPI_{j-1}}{\text{the total number of customers in } Month_{j-1} \text{ for the same } CPI_{j-1}}\end{aligned} \tag{3}$$

Taking data in the music service category for example, we assume that the dataset covers four months (from April to July), and we used the one month data (in June) and the three month data (from April to June) respectively as a training set to compute each member's CPI for the music service category. Then, we observed the return visit rate of the members.

One Month Data:

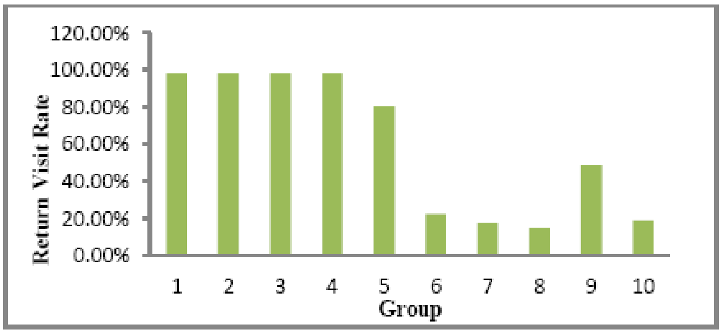


FIGURE 1. Return visit rate in July for CPI_6

Using the data in June as training data, we calculated CPI_6 for the members who browsed for the music service category in June, then we observed the return visit rates of those in July. However, for the members who did not return visit for the music service category in June, their CPI_6 should be zero. Therefore, these members will never be selected as the target members.

As shown in Figure 1, the x-axis represents all the member groups. For brevity, we ranked CPI from high to low and divided the members into ten groups by number of people, with Group 1 having the highest CPI and Group 10 having the lowest CPI. The y-axis signifies the return visit rate in July of each group. In Figure 1, we observe that the trend of CPI_6 curve does not conform to which we expected because Groups 9 and Group 10 do not conform to the standard that a higher CPI corresponds to a higher the return visit rate. As a result, for selecting target members, if the priority option of a website marketer is in accordance with the group sequence, 1-10, it may make some mistakes and incur some unnecessary costs for promotion.

Table 2 presents the respective return statuses for two different types of members with the same CPI in Group 9. Type (1) is the members who revisited the music service category only in June, and type (2) is the members who continuously revisited the music service category from April to June. The result shows that type (1) and type (2) have different return visit rates in July. In other words, the problem of CPI could not differentiate the members with different past behaviors.

Besides, if the members did not visit the music service category in June, their CPI_6 would be zero. We cannot distinguish the return visit rate with CPI_6 for those, even if they have visited in April or May. The result reflected in Table 3 indicates that the different past behaviors may imply different return visit rates.

Now, we extended the training data period from one month to three months, i.e., from April to June. Thus, we computed CPI CPI_{4-6} and observed the members' return visit

TABLE 2. The return status for different types of members with identical CPI_6 in Group 9

	April	May	June	Return visit rate of July
(1)	X	X	O	18.8%
(2)	O	O	O	90.6%

TABLE 3. The return status in July for different types of members with $CPI_6 = 0$

	April	May	June	Return visit rate of July
(1)	O	O	X	13.92%
(2)	X	O	X	7.68%
(3)	O	X	X	5.04%

Three Month Data:

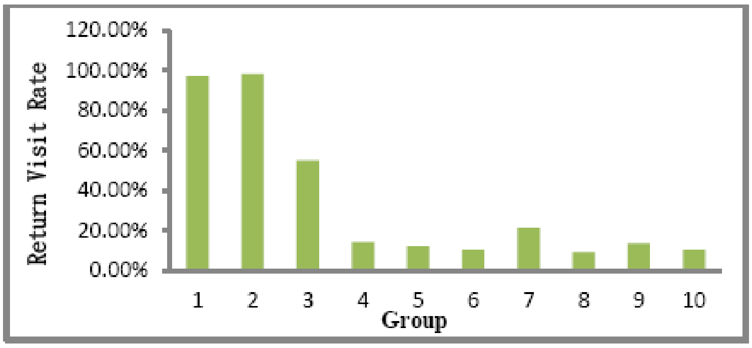


FIGURE 2. Return visit rate in July for CPI_{4-6}

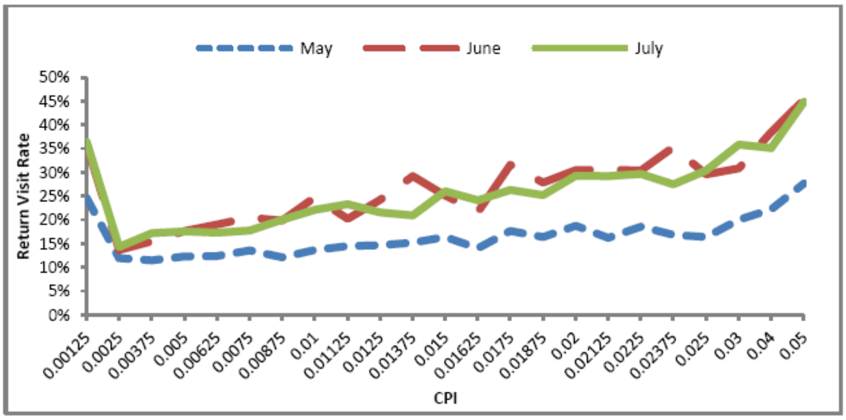


FIGURE 3. The return visit rates of the members in May, June and July

TABLE 4. The return status for different types of members with identical CPI_{4-6} in Group 7

	April	May	June	Return visit rate of July
(1)	X	X	O	13.5%
(2)	X	O	X	8%
(3)	O	X	X	5.3%

rates in July. Figure 3 shows the relationship with the return visit rate in July based upon different CPI_{4-6} in June. In Figure 2, the return visit rate of Group 7 is higher than Group 4 and Group 5. Moreover, as shown in Table 4, we observed that the return visit rates of the three different types of members with the same CPI in Group 7 are different, where type (1), type (2) and type (3) are for the members who revisited the music service category only in June, May and April respectively.

TABLE 5. The past behaviors patterns in the previous 3 months

	Month_{j-3}	Month_{j-2}	Month_{j-1}
Type 1	O	O	O
Type 2	X	O	O
Type 3	O	X	O
Type 4	X	X	O
Type 5	O	O	X
Type 6	X	O	X
Type 7	O	X	X
Type 8	X	X	X

Even though the training data time was extended to three months, for these types of members, we still could not accurately predict their return visit rate by CPI. In other words, extending the training period did not contribute to reflect the members' past behaviors with CPI. The above results confirm that the return visit rates of members are deeply affected by their past behaviors. In fact, the different past behaviors would result in a different return visit rate despite the same CPI. For this reason, we considered the time factor into the model to differentiate members with different behavior patterns.

As shown in Figure 1, we observe that members within the top 50% CPI accompany nearly 80% $\sim$ 100% return visit rates. From the marketing point of view, these kinds of members are usually loyal members, and they are obviously referred to as target members for marketing with a higher priority. The members within the top 50% CPI in Figure 1 are approximately 14% of the total number of members in the music service category. Moreover, for the members who did not return to the music service category in June, their CPI_6 are zero, we will discuss this case in the next section. Relatively, to discover the potential member is another focus in marketing. In our study, we primarily focus on discovering the potential member within the bottom 50% CPI in Figure 1 and the members who did not return to the music service category in the current month, that is, the remaining 86% of the total number of members in the music service category. Furthermore, since there are some CPI in which corresponding members are few, their return visit rate predictions will be distorted. For example, for some CPI, the total number of its corresponding members is one. If this member returns to the portal in the next month, the return visit rate is 100%; others, the return visit rate is 0%. For this situation, we consider that the adjacent CPI would be mapped into a similar return visit rate. Hence, for some CPI in which corresponding members are scarce, we merge CPI with their neighbors (lower CPI) until the total number of members is more than 100.

As mentioned above, the different past behaviors may imply different return visit rates. Figure 3 yields the members and their corresponding return visit rates in May, June and July. The x-axis presents the ranked CPI, and the y-axis signifies the corresponding return visit rates. Since the members' return visit rates are inconsistent and the total number of return members corresponding to every CPI is different, it would be difficult in observation to identify the relationship between CPI and its corresponding return visit rates. Thus, we observe the return visit rates which are corresponding to CPI in the previous month and the past behavior patterns in the previous 3 months, $Month_{j-1} \sim Month_{j-3}$. Table 5 lists the combination of the different types of past behavior patterns in the previous 3 months.

One point is worth making about Table 5, since the members for Type 5, Type 6 and Type 7 did not return to the music service category of the portal in $Month_{j-1}$, their

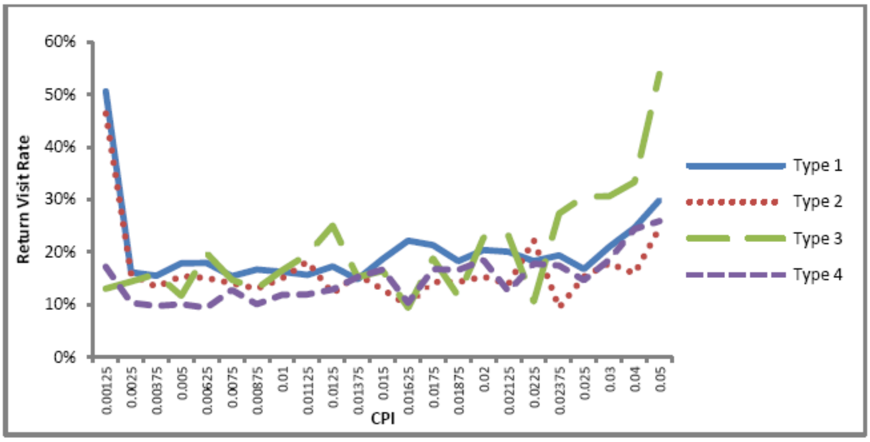


FIGURE 4. The return visit rates of the members (Type 1 $\sim$ Type 4) in May

CPI_{j-1} are zero and we cannot apply their CPI_{j-1} to observe the return visit rates. Thus, we focus on the members, whose CPI_{j-1} are not zero, that is, Type 1 $\sim$ Type 4. Take the return visit rates of the members (Type 1 $\sim$ Type 4) in May for example, as shown in Figure 4, it is clear that the relationship between the past behavior patterns, CPI and the return visit rates are inconsistent. Therefore, we cannot apply CPI_{j-1} to predict members' return visit rates due to the inconsistent of the return visit rates every month and the members, whose CPI_{j-1} are zero.

4. Proposed Method.

4.1. The relationship between the past behavior patterns, CPI and the cumulative average return visit rate. As mentioned in problem statement, the return visit rates every month is inconsistent for every CPI. Thus, we use the cumulative average return visit rate which is depicted in Equation (4) to solve this problem.

$$\begin{aligned}&\text{the cumulative average return visit rate for the same } CPI \text{ in } Month_j \\ &= \frac{\sum_{k=1}^{j-1} \text{the number of return members for the same } CPI_k \text{ in } Month_{k+1}}{\sum_{k=1}^{j-1} \text{the number of members for the same } CPI_k \text{ in } Month_k}\end{aligned} \tag{4}$$

The cumulative average return visit rates for every CPI in May, June and July are shown in Figure 5. The results indicate that higher CPI is uncertain to imply a higher cumulative average return visit rate. However, it is clear that every CPI will imply a similar cumulative average return visit rate in Figure 5. Thus, when CPI_{j-1} is not zero, we can use the cumulative average return visit rate of CPI_{j-1} as the basis to predict the return visit rate. Take for example, when a marketer wants to target the members with high return visit rates in May, we can use the cumulative average return visit rate of CPI_4 as the basis for prediction, then rank the members with the cumulative average return visit rates of CPI_4 in April for targeting.

As mentioned above, we can use the cumulative average return visit rate of CPI_{j-1} as the basis for prediction. We further discuss the relation between the cumulative average return visit rate and the members' past behavior patterns. Table 5 lists the combination of the different types of past behavior patterns in the previous three months. One point worth making about Table 5, since the members for Type 5, Type 6, Type 7 and Type 8 did not return to the music service category of the portal in $Month_{j-1}$, their CPI_{j-1} is zero and we cannot apply their CPI_{j-1} to observe the return visit rates.

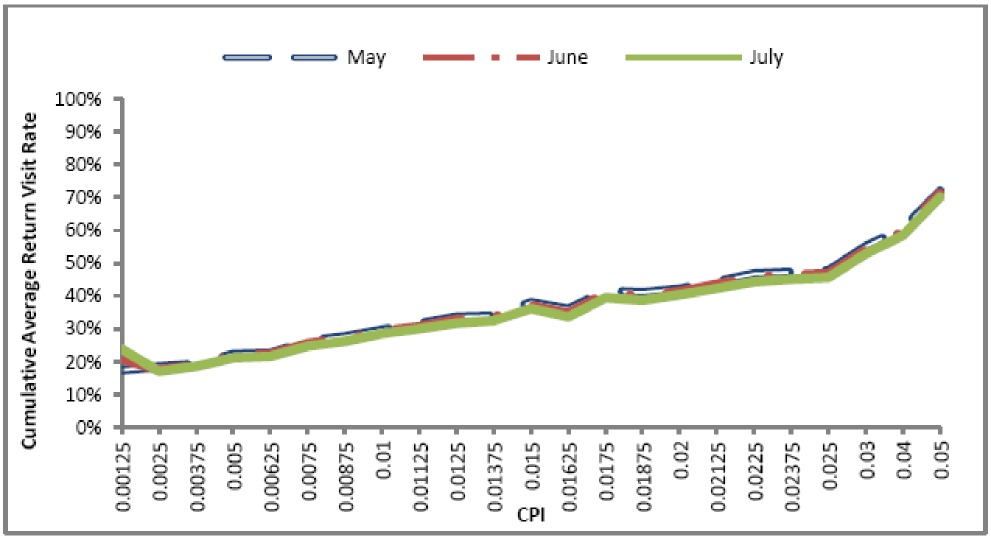


FIGURE 5. The cumulative average return visit rates of the members in May, June and July

Firstly, we focus on the members who did return to the music service category of the portal in $Month_{j-1}$. We differentiate the past behavior patterns of the members into Type 1, Type 2, Type 3 and Type 4, which are listed in Table 5. These types of behavior pattern and their corresponding cumulative average return visit rates for every CPI_{j-1} are presented in Figure 6. Although the same CPI_{j-1} may imply different cumulative average return visit rates and past behavior patterns, the results in Figure 6 show that with the same CPI_{j-1} and past behavior pattern, the cumulative average return visit rates of the members in May, June and July are similar. For example, in Figure 6, it is clear that when CPI_{j-1} is 0.00125, its corresponding cumulative average return visit rates ranges between 12% and 80% for different types of past behavior pattern every month; but within the same past behavior pattern, the same CPI_{j-1} will imply similar cumulative average return visit rates. In other words, when we differentiate the past behavior patterns of the members, we can apply the cumulative average return visit rate, which is corresponding to CPI_{j-1} and their past behavior patterns, as the basis for members' return visit rate predictions.

Next, for the members who did not return to the music service category of the portal in $Month_{j-1}$, we differentiate the past behavior patterns of the members into Type 5, Type 6, Type 7 and Type 8, which are listed in Table 5. For Type 8, which indicates the members who did not return to the music service category at least three recent consecutive months, their return visit rates in May, June and July, are 2.67%, 2.78% and 3.47, respectively. Thus, when we want to target the members with high return visit rates, this type of members should be chosen last. Moreover, for Type 5, Type 6 and Type 7, the corresponding cumulative average return visit rates of these three types of behavior pattern are presented in Figure 7. The results show that the cumulative average return visit rates of the members in May, June and July are similar. As a result, when we combine the past behavior patterns of the members and the cumulative average return visit rate, we can effectively predict the return visit rates for the members, whose CPI_{j-1} is zero.

4.2. **Comparison with the length of the referenced month.** In this section, we further discuss the relation between the different lengths of the referenced month and the usability. Since the marketer often uses the number of target members as the basis

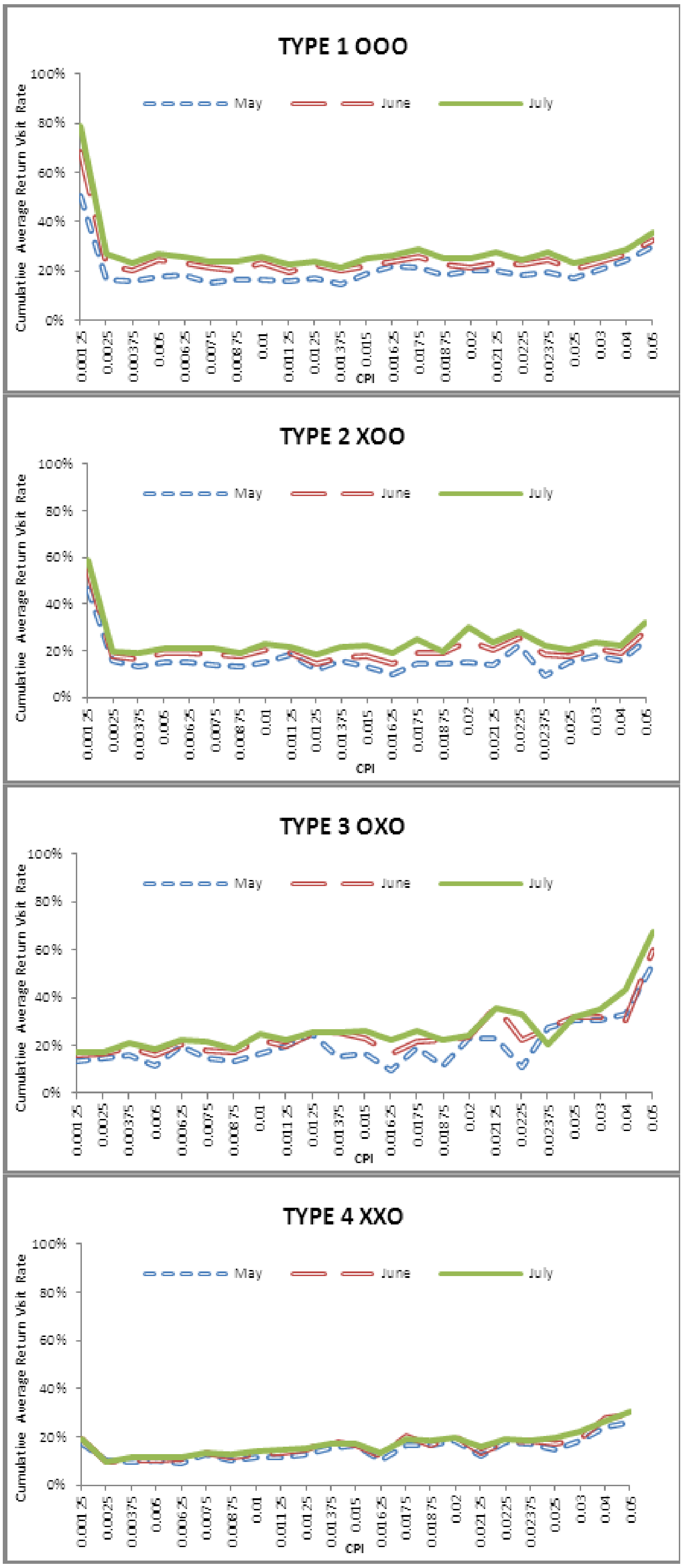


FIGURE 6. The cumulative average return visit rates of the members (Type 1 ~ Type 4) in May, June and July

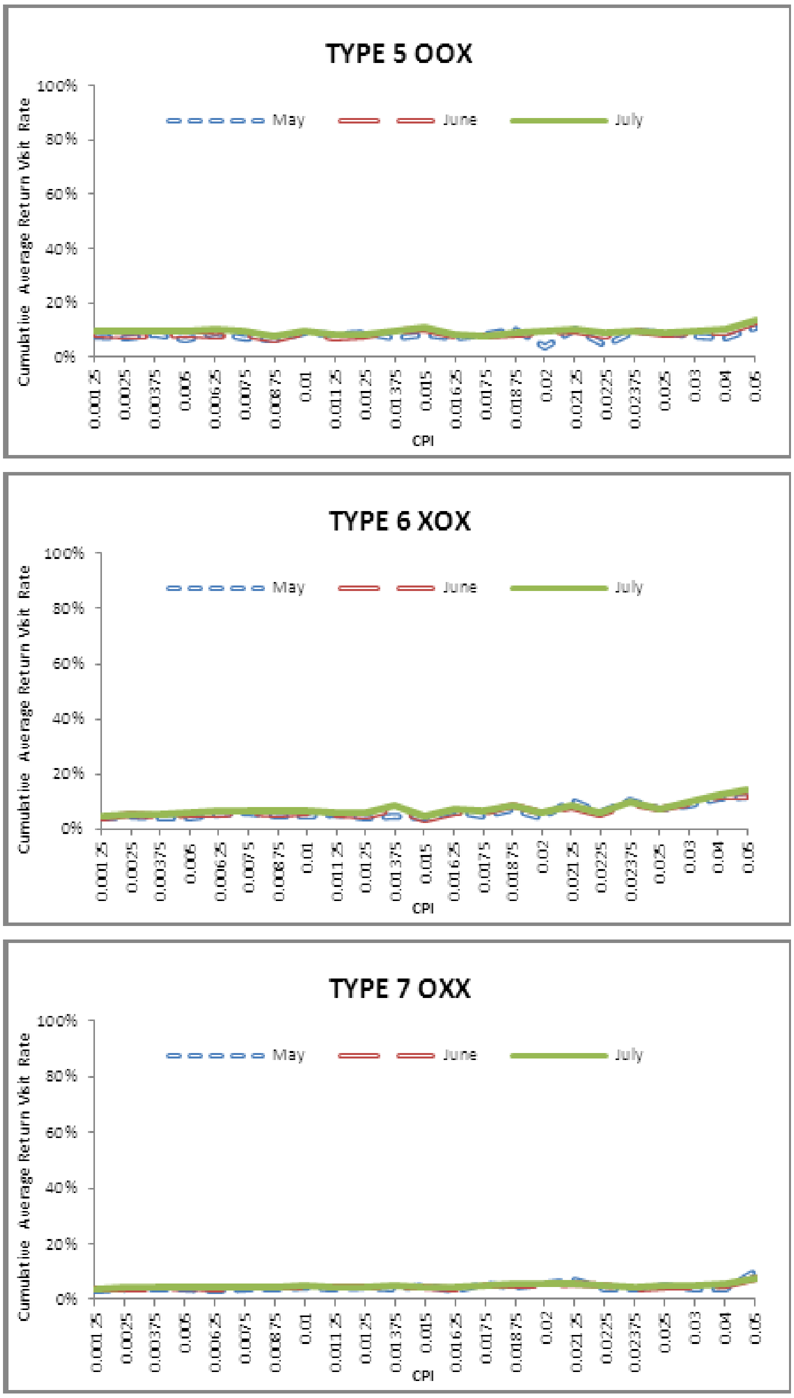


FIGURE 7. The cumulative average return visit rates of the members (Type 5 $\sim$ Type 7) in May, June and July

for marketing, we adjust the different lengths of the referenced month to observe their corresponding return visit rates for the usability estimation. As shown in Figure 8, we compare the different lengths of the referenced month, with the x-axis presenting the cumulative number of members, the y-axis signifying the corresponding return visit rates, and the parameter n indicating the length of the referenced month. In order to observe the members with high return visit rates, we only represent the top members with high return visit rates. The result in June indicates that when the cumulative number of members is approximately 750 and the length of the referenced month is four, we can

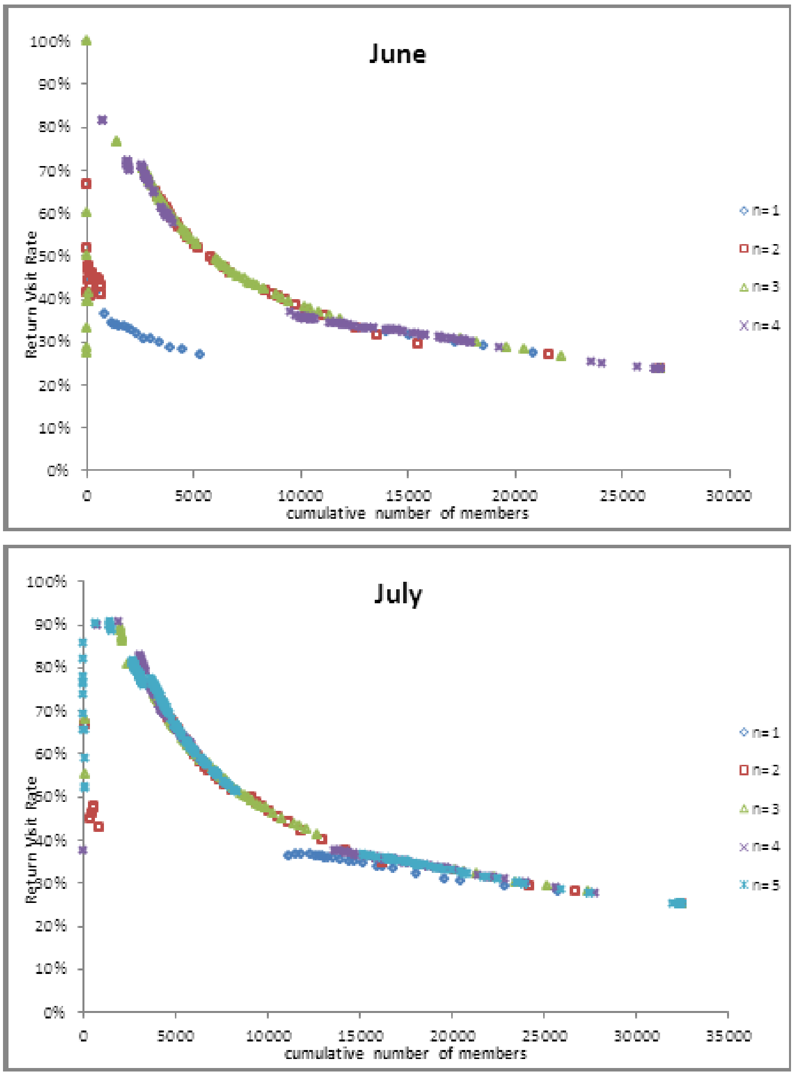


Figure 8. The different lengths of the referenced month and the corresponding return visit rates in June and July

target the members with 81% return visit rates; when the cumulative number of members is approximately 1390 and the length of the referenced month is three, we can target the members with 76.6% return visit rates. In the remaining cumulative number of members, whether the length of the referenced month is three or four, the return visit rates of the target members are similar.

Moreover, the results using just one or two months as the referenced months are unable to target precisely the members with high return visit rates. In conclusion, as shown in Figure 8, when we use more months as the referenced months, we can target the members more precisely. However, one thing worth noting in July is that, compared with the results which take four months as the referenced months, although the result which using five months as the referenced months can target the members more precisely, the enhancement of accuracy is limited. Since we use n months as the referenced months, there will be 2^n different kinds of past behavior patterns, that is, when we use five months as the referenced months, there will be 32 kinds of past behavior patterns. The lengths of the referenced months can be adjusted according to the marketing strategy. In order

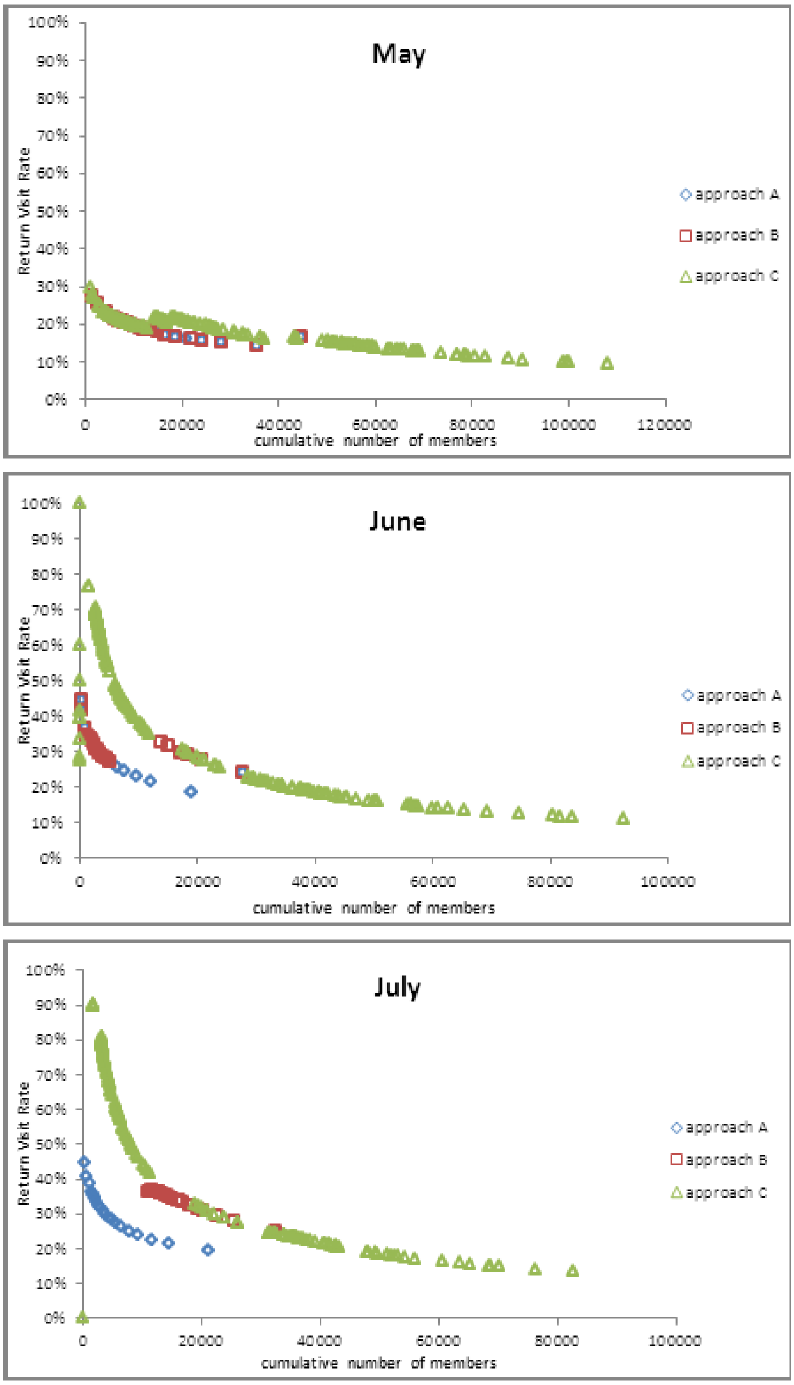


FIGURE 9. Comparison of three different approaches in May, June and July

to decrease the execution complexity, our research takes three months as the referenced months.

4.3. **The BP model.** Based on the discussion in Sections 4.1 and 4.2, for the same CPI_{j-1} and past behavior pattern, the trends of the cumulative average return visit rates are similar. Next, we apply three different approaches to predict the members' return visit rates in a $Month_j$, such as follows: (approach A) we use ranked CPI_{j-1} as the basis; (approach B) we use the ranked cumulative average return visit rate, which corresponds to CPI_{j-1} as the basis; (approach C) we differentiate the past behavior patterns of the members, then apply the cumulative average return visit rate, which corresponds to CPI_{j-1}

and the past behavior patterns, as the basis. We use the first seven months data (from January to July) as the training data to compute CPI_4, CPI_5 and CPI_6, then apply the three above-mentioned approaches to predict the return visit rates in May, June and July, respectively. The results are shown in Figure 9, where the x-axis presents the cumulative number of members, and the y-axis signifies the corresponding cumulative average return visit rates. For the result in May, since there is only one month for calculation of the cumulative average return visit rate, some members with high return visit rates are not chosen at the beginning. However, for the result in June and July, since there are enough months for calculation of the cumulative average return visit rate, the results can be presented fairly. The results reflect that approach C can achieve our goal with clear support that when we apply the ranked cumulative average return visit rate as the basis.

In the studies discussed above, we propose a novel algorithm of the Behavioral Preference (BP) model to predict the members' return visit rates. As shown in Figure 10, firstly, we compute the CPI of members every month based on Equation (2) to construct the member profiles. Then, for each type of past behavior pattern, we compute the cumulative average return visit rate by $Month_{j-1}$ based on Equation (4). Next, since our goal is to target the members with the high return visit rates, we rank the cumulative average return visit rate by $Month_{j-1}$ as the basis to predict the monthly return visit rate. In other words, when we want to predict the members' return visit rates in July, we should apply their ranked cumulative average return visit rates in June, which corresponds to every CPI_5 and each type of past behavior pattern, as the basis for prediction. As shown in Figure 11, the x-axis presents the cumulative number of members, and the y-axis signifies the corresponding actual and predictive return visit rate of the cumulative number of members. It is obvious that the actual and predictive return visit rates in May, June and July are increasingly similar.

5. **Experiments and Discussion.** Our data set came from a well-known portal site in Taiwan. The period of the dataset covers the browsing logs and purchasing records from

Input:

n: the length of reference month

j: the month for prediction

Output:

list: the list of members with ranked return visit rates

Step1. Before $Month_j$, computing CPI of member *i* every month separately based on Equation (2) and constructing the member profile

Step2. Computing the cumulative average return visit rate in $month_{j-1}$ for all past behavior patterns separately based on Equation (4) and applying the cumulative average return visit rate which is corresponding to CPI in $Month_{j-1}$ as the basis to predict the return visit rate in $Month_j$

Step3. Listing the members with ranked predictive return rates

FIGURE 10. The algorithm of the BP model

January 2009 to November 2009. The total useful number of click transactions (including both the number of browsing clicks and purchase records) was 135,872,495, which were contributed by 2,676,068 registered members. Its main revenue comes from providing linkage to other websites and consignment sales such as prepay cards and game cards, and the linkage on the portal site was presented as service item lists such as movies, music, games, travel, and financial management. We applied our method to the several service categories, and the results demonstrated that the method can be practically implemented and provide satisfactory results. However, because of the limitation of space, we used the music service category as a case study to conduct the follow-up experiments and discussions. The total useful number of click transactions in the music service category was 4,669,613, which were contributed by 344,644 members. On average, approximately 17% of the total number of members would return to the music service category every month. Moreover, we took the first seven months data (from January to July) as the training data to build our model, and the last four months data as the testing data.

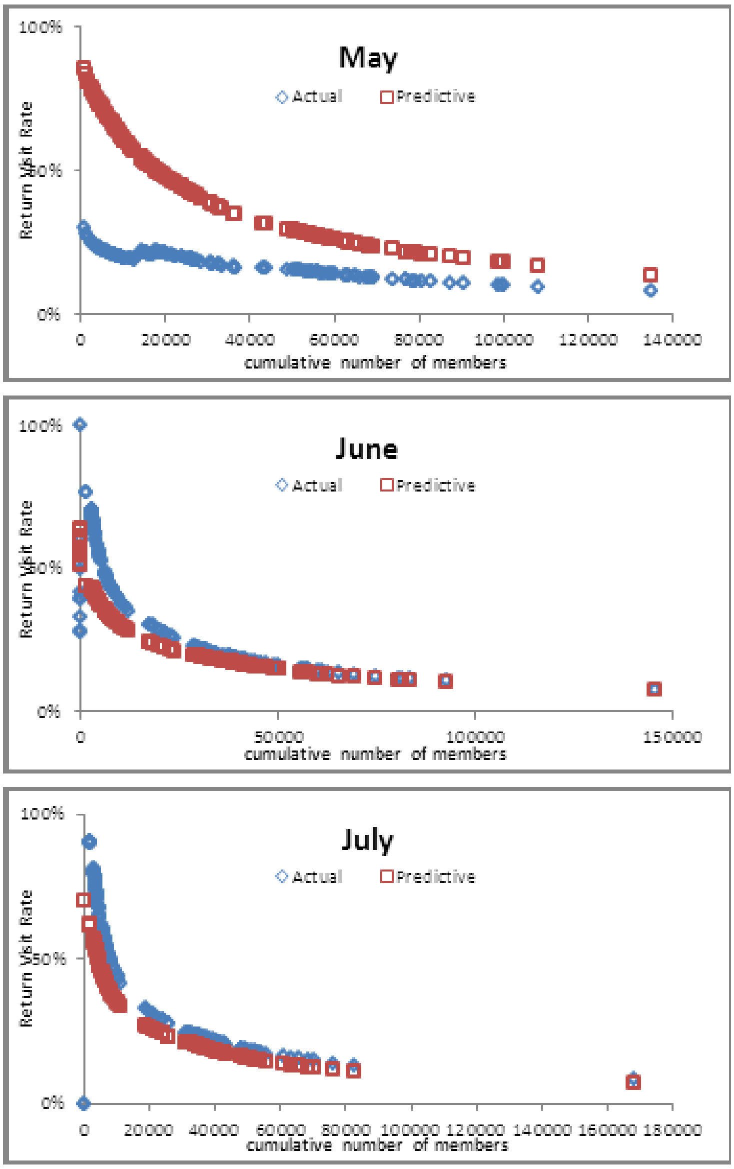


FIGURE 11. The actual and predictive return visit rate in May, June and July

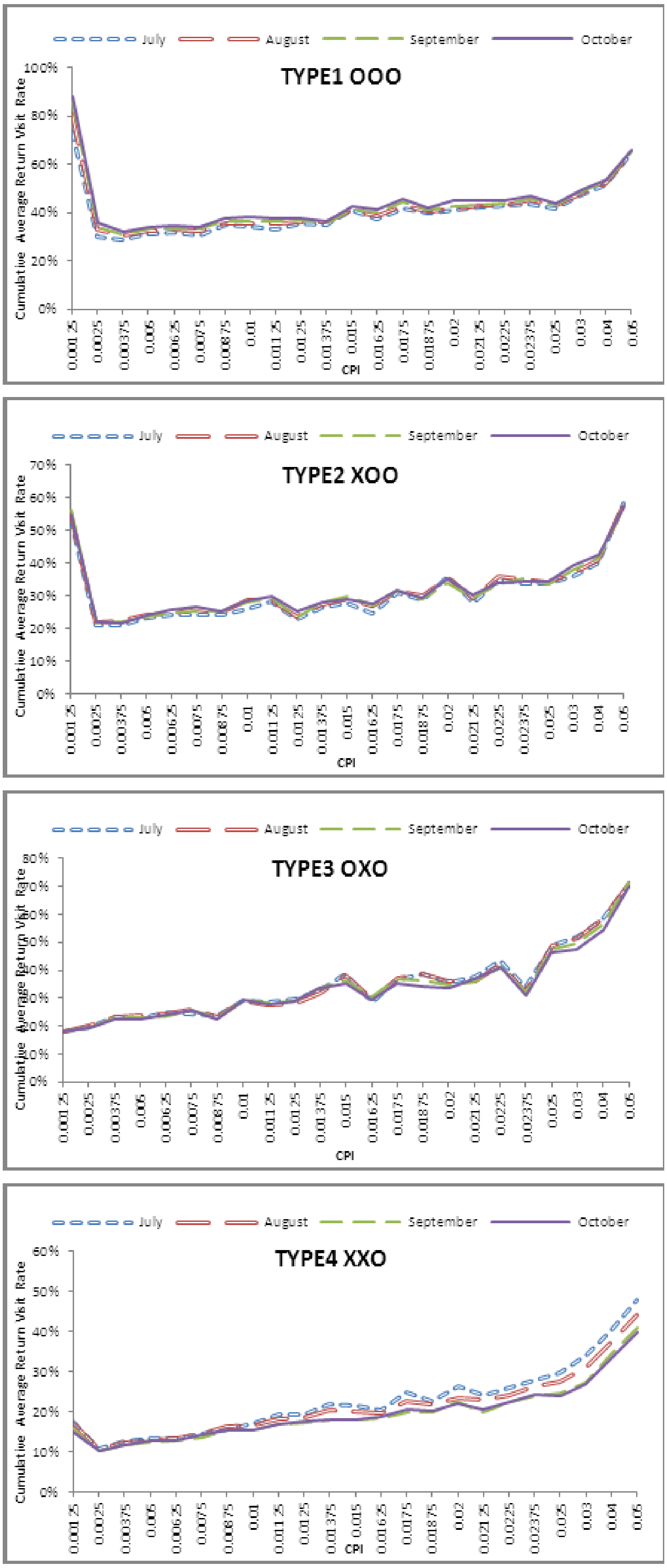


Figure 12. The cumulative average return visit rates of the members (Type 1 $\sim$ Type 4) in July, August, September and October

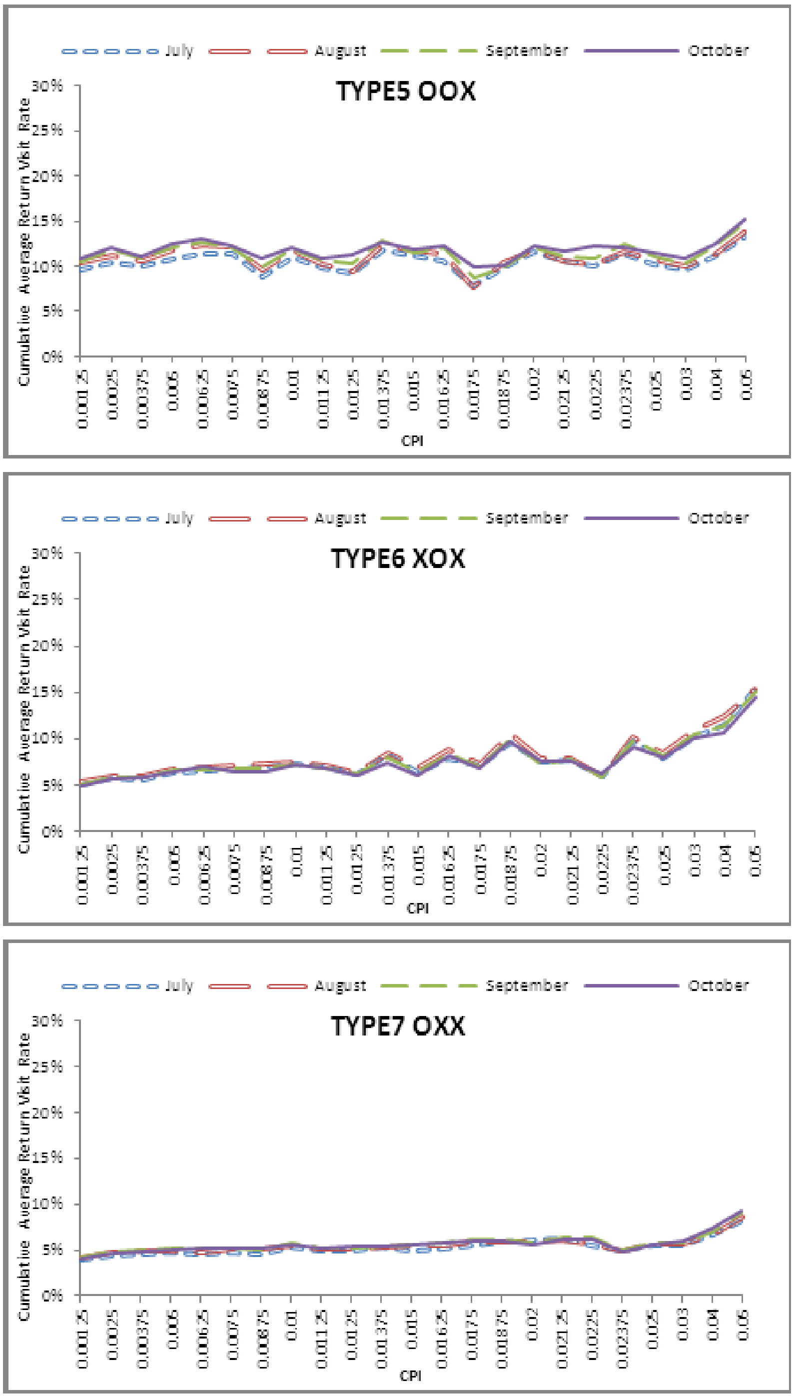
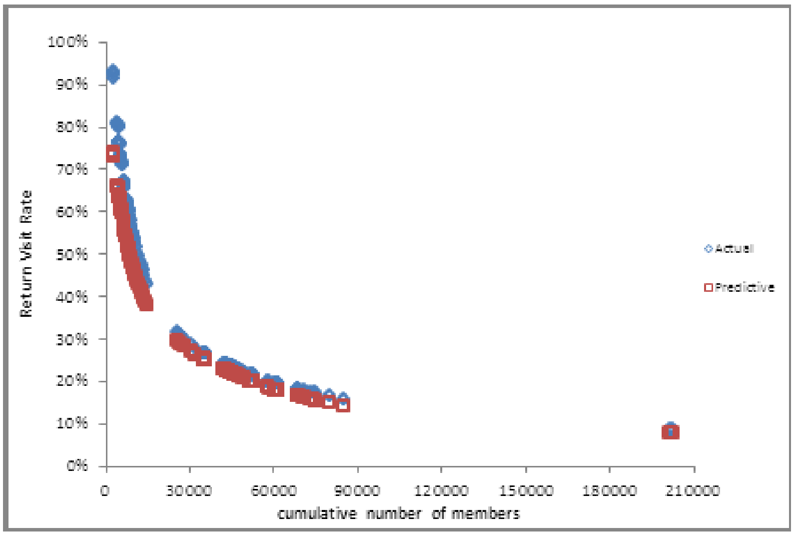
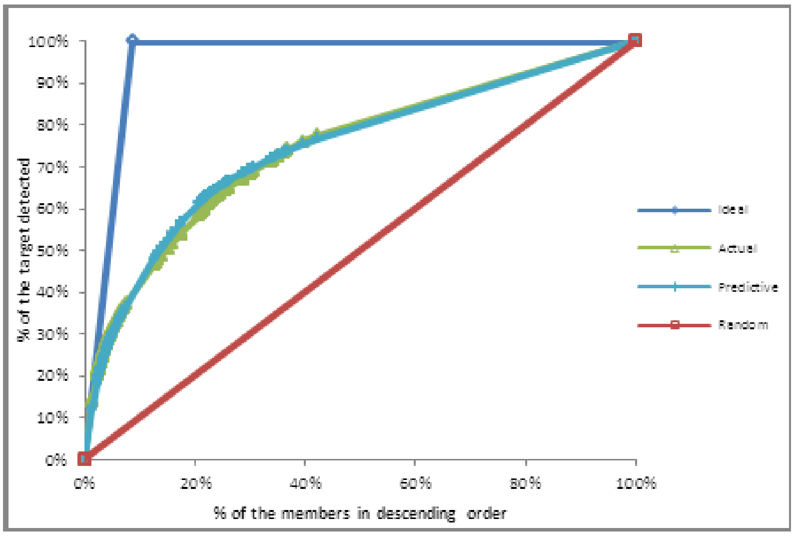


FIGURE 13. The cumulative average return visit rates of the members (Type 5 $\sim$ Type 7) in July, August, September and October

5.1. **Accuracy verification of the BP model.** In Figure 12 and Figure 13, which depict the cumulative average return visit rates of the members in July, August, September and October according to each type of past behavior pattern, the results indicate that the cumulative average return visit rates are similar every month. Furthermore, as shown in part (a) of Figure 14, the actual and predictive return visit rates in August are similar. In the same way, the results in September, October and November can prove the validity. This supports that we can apply the ranked cumulative average return visit rate by $Month_{j-1}$ as the basis to predict the return visit rate by $Month_j$.



(a)



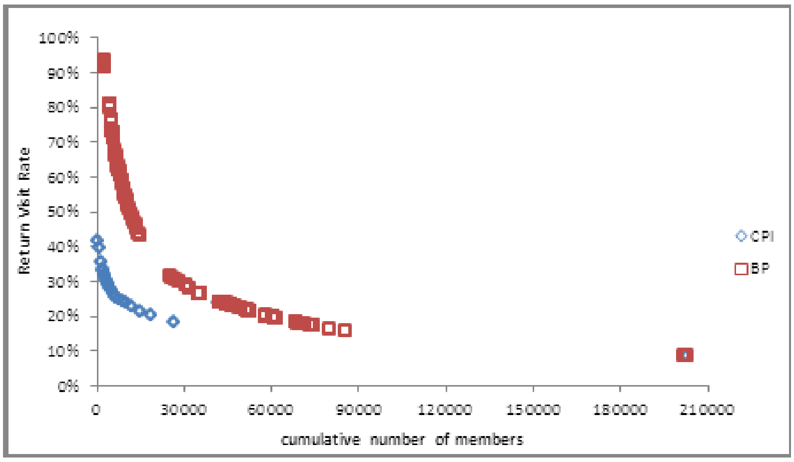
(b)

FIGURE 14. The result comparison in August

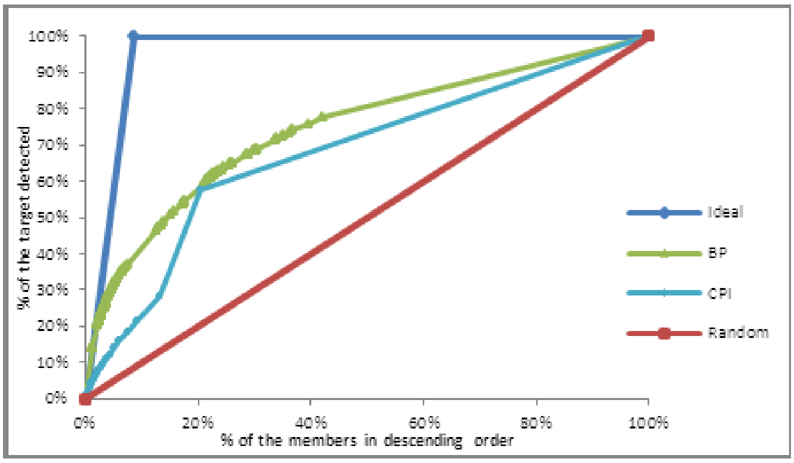
Next, we use lift curves to show the accuracy of our model. A lift curve presents the proportion of revisited members detected against the proportion of revisited members selected. The accuracy of a model can be measured by comparing the lift curve to random and ideal curves, where the ideal curve presents that all revisited members are selected first and the random curve presents that ß% revisited are selected from ß% members [22,23]. As observed in part (b) of Figure 14, there are four lift curves in the cumulative gains chart for the return visit rate in August, which are an ideal curve, a lift curve for a predictive return visit rate, a lift curve for an actual return visit rate and a random curve. The results are reflected by using the BP model, where the actual curve captures approximately the top 25% of the members, which comprises nearly 67% return visited members, compared with a random curve, which only comprises nearly 25% return visited members. Moreover, the trends for the predictive curve and the actual curve are similar. In the same way, the results in September, October and November can prove validity.

5.2. **Comparison of the BP model and the CPI model.** In this experiment, we compare the accuracy of the BP model with the CPI model. As shown in part (a) of

Figure 15, when we have to choose 11000 members for targeting, the return visit rates for the 11000 members in the BP model are on the verge of 50%, but the return visit rates for the 11000 members in the CPI model are approximately 10%. The result shows that the BP model is significantly superior to the CPI model in usability for return visit rate prediction. Likewise, the results in September, October, and November can prove validity.



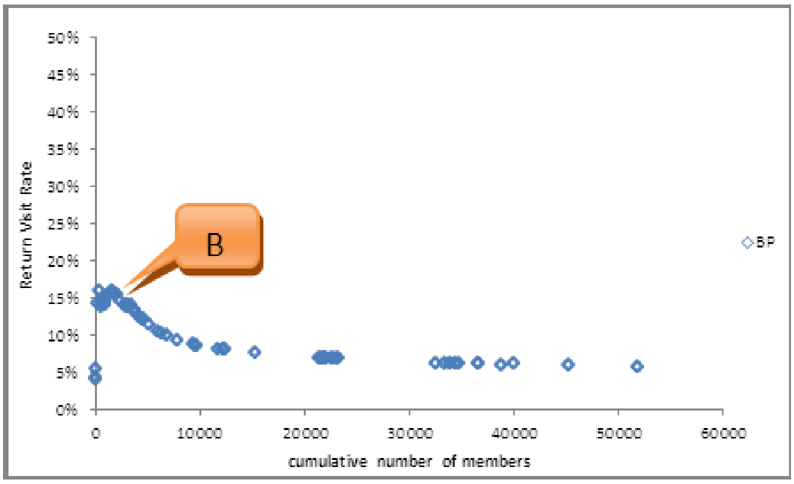
(a)



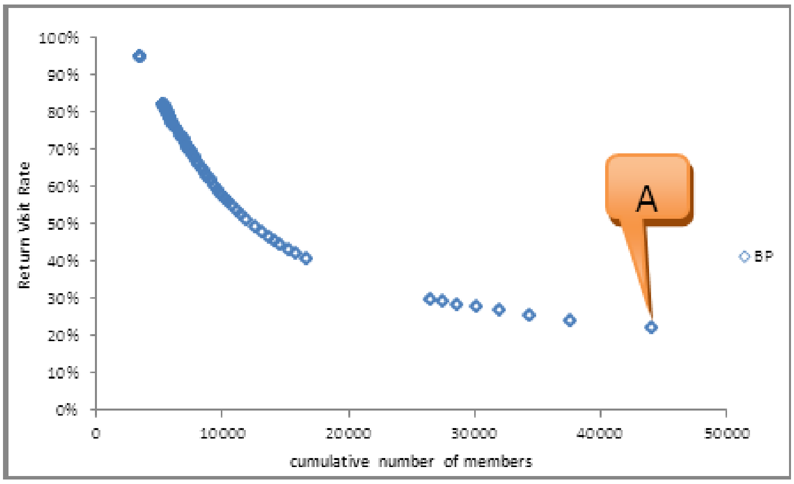
(b)

FIGURE 15. Comparison of the CPI model and the BP model in August

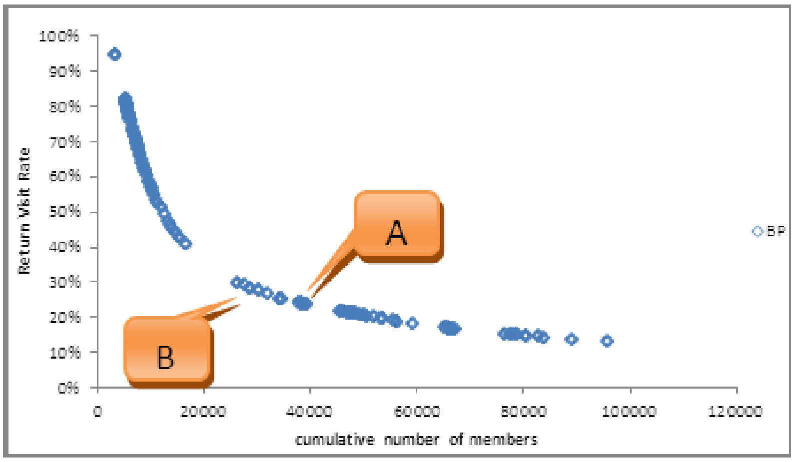
Similarly, we use lift curves to compare the usability of the two models. As observed in part (b) of Figure 15, there are four lift curves in the cumulative gains chart for the return visit rate in August: an ideal lift curve, a lift curve for an actual return visit rate for the BP model, a lift curve for an actual return visit rate for the CPI model and a random curve. The results are reflect by using the BP model, where the actual curve captures approximately the top 10% of the members, which comprises nearly 45% return visited members, compared to a lift curve for the CPI model, which only comprises nearly 21% return visited members. Moreover, for the top 13%~20% of members in the lift curve for an actual return visit rate for the CPI model, the CPI model has to incorporate randomly chosen target members. The BP model can effectively achieve our goal rather than the CPI model. Likewise, the bottom 80% of members cannot be captured by the CPI model because those members did not visit the music service category in July and their CPI is zero. In contrast, the BP model can capture those members for accurate return visit rate



(a)



(b)



(c)

FIGURE 16. Applying the BP model, a choice priority comparison with two types of past behavior patterns

TABLE 6. A sample of the two past behavior patterns

	June	July	August	Return visit rate of September
(1)	X	X	O	8.4%
(2)	O	O	X	17.15%

predictions. Similarly, the results in September, October, and November can prove the usability as well.

Moreover, taking CPI_8 as an example in problem statements, if the members did not return the music service category in August, their CPI_8 was zero. For the marketers, these members cannot be targeted for promotion. However, there are some members who were heavy users but just did not visit in August. This kind of potential members should be targeted. Table 6 highlights differences between the two types of members with different past behavior patterns. Type (A) stands for the members who only visit the music service category in August and type (B) stands for the members who were heavy users but did not visit in August. It is clear that the return visit rate of type (A) is lower than that of type (B). However, in the CPI model, their CPI_8 for type (B) is zero. If the marketers target potential members for promotion using the CPI model, they will neglect the members for type (B).

For the sake of preventing the problem mentioned above, consider the graphic representations in Figure 16 (In order to present the results fairly, we do not show the members who did not visit in June, July and August). Part (a) of Figure 16 indicates with the BP model, the return visit rates in September of the members, who did not return in August; part (b) of Figure 16 indicates with the BP model, the return visit rates in September of the members, who did return in August; a choice priority comparison with two types of the past behavior patterns with the BP model is depicted in part (c) of Figure 16. Three things are worth noting in Figure 16. Firstly, the dot that corresponds to type (A) is located at the rear with a lower return visit rate in part (b) of Figure 16. Secondly, the dot that corresponds to type (B) is located in the front with a higher return visit rate in part (a) of Figure 16. Thirdly, in part (c) of Figure 16, the members for type (B) are indeed to be targeted before the members for type (A). In other words, the BP model not only solves the problem that $CPI = 0$ but also targets members with the highest return visit rate more accurately.

6. **Conclusions.** The main purpose of this work is to leverage historical members' past behavior patterns to construct a novel BP model to predict the members' return visit rates. We have demonstrated that the BP model can be practically implemented and provide adequate results. Besides the return visit rate prediction, future work will hopefully apply this model to purchasing behavior prediction so as to provide marketers with useful suggestions for promotions. In addition, it may be beneficial to apply this study as the basis of a hybrid Recommenders System so as to predict members' likely actions and provide useful suggestions for marketing practice.

Acknowledgment. This research was supported by National Science Council, Taiwan, under grant Nos. NSC 101-2221-E-032-050.

REFERENCES

[1] A. C. R. van Riel, V. Liljander and P. Jurriens, Exploring consumer evaluations of e-services: A portal site, *International Journal of Service Industry Management*, vol.12, pp.359-377, 2001.
[2] *Yahoo! Taiwan*, http://tw.yahoo.com/.

[3] *yam*, http://www.yam.com/.

[4] *Hinet*, http://www.hinet.net/.

[5] Y. Chen, D. Pavlov and J. F. Canny, Behavioral targeting: The art of scaling up simple algorithms, *ACM Transactions on Knowledge Discovery from Data*, vol.4, p.17, 2010.

[6] A. Du and B. Fang, Novel approach for web filtering based on user interest focusing degree, *International Journal of Innovative Computing, Information and Control*, vol.4, no.6, pp.1325-1334, 2008.

[7] G. T. Raju and P. S. Satyanarayana, Knowledge discovery from web usage data: A novel approach for prefetching of web pages based on ART neural network clustering algorithm, *International Journal of Innovative Computing, Information and Control*, vol.4, no.4, pp.897-904, 2008.

[8] S. E. Middleton, N. R. Shadbolt and D. C. D. Roure, Ontological user profiling in recommender systems, *ACM Transactions on Information Systems*, vol.22, p.88, 2004.

[9] D. Choi and B. Ahn, Eliciting customer preferences for products from navigation behavior on the web: A multicriteria decision approach with implicit feedback, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol.39, pp.880-889, 2009.

[10] C. Porcel and E. Herrera-Viedma, Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries, *Knowledge-Based Systems*, vol.23, pp.32-39, 2010.

[11] Y. B. Fernández, J. J. P. Arias, A. G. Solla, M. R. Cabrer, M. ı. L. Nores, J. G. Duque, A. F. Vilas, R. P. D. ı. Redondo and J. B. M. noz, A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems, *Knowledge-Based Systems*, vol.21, pp.305-320, 2008.

[12] M. Balabanovic and Y. Shoham, Fab: Content-based, collaborative recommendation, *Communications of the ACM*, vol.40, p.72, 1997.

[13] M. Scheutz and V. Andronache, Architectural mechanisms for dynamic changes of behavior selection strategies in behavior-based systems, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol.34, pp.2377-2395, 2004.

[14] Y. Hu, Y. Koren and C. Volinsky, Collaborative filtering for implicit feedback datasets, *The 8th IEEE International Conference on Data Mining*, 2008.

[15] A. Tsymbal, *The Problem of Concept Drift: Definitions and Related Work*, 2004.

[16] T. Q. Lee, Y. Park and Y. T. Park, A time-based approach to effective recommender systems using implicit feedback, *Expert Systems with Applications: An International Journal*, vol.34, pp.3055-3062, 2008.

[17] J. Lee, M. Podlaseck, E. Schonberg and R. Hoch, Visualization and analysis of clickstream data of online stores for understanding web merchandising, *Data Mining and Knowledge Discovery*, vol.5, pp.59-84, 2001.

[18] A. Albadvi and M. Shahbazi, A hybrid recommendation technique based on product category attributes, *Expert Systems with Applications: An International Journal*, vol.36, pp.11480-11488, 2009.

[19] Y. H. Cho and J. K. Kim, Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce, *Expert Systems with Applications: An International Journal*, vol.26, pp.233-246, 2004.

[20] Y. J. Park and K. N. Chang, Individual and group behavior-based customer profile model for personalized product recommendation, *Expert Systems with Applications: An International Journal*, vol.36, pp.1932-1939, 2009.

[21] K. Palanivel and R. Sivakumar, A study on implicit feedback in multicriterl e-commerce recommender system, *Journal of Electronic Commerce Research*, vol.11, p.17, 2010.

[22] M. Vuk and T. Curk, ROC curve, lift chart and calibration plot, *Metodoloski Zvezki*, vol.3, pp.89-108, 2006.

[23] L. Qi, C. Enhong, X. Hui, C. H. Q. Ding and C. Jian, Enhancing collaborative filtering by user interest expansion via personalized ranking, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol.42, pp.218-233, 2012.