Combing Customer Profiles for Members' Repurchase Rate Predictions

Yi-Hsin Wang
Department of Information Management[1]
Chang Gung University Of Science and Technology,
261,Wen-hwa 1st Rd, Kwei-shan, Taoyuan County, Taiwan.
Email: yishing@gw.cgust.edu.tw

Rui-Dong Chiang*, and Huan-Chen Chu
Department of Computer Science and Information Engineering[2]
Tamkang University
No.151, Yingzhuan Rd., Tamsui Dist., New Taipei City, Taiwan
Email: 081863@mail.tku.edu.tw; HuanChen.Chu@gmail.com

***Abstract*—Customer relationship management (CRM) leverages historical users' behaviors in order to dawn effort of enhancing customer satisfaction and loyalty. Thus, constructing a successful customer profile plays a critical role in CRM. 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 repurchase index (RI) and the preference drifts to propose a Behavioral Repurchase Prediction (BRP) model, and to predict the members' repurchase rates in the specific category of the e-shop. The marketers of the e-shop can target the registered members with high repurchase rates and design corresponding marketing strategies. The experimental results with a real dataset show that our model can effectively predict the registered members' repurchase rates.**

***Index Terms*—CRM, Customer profile, Concept drift detection, Repurchase rate prediction**

I. INTRODUCTION

With increasing number of Internet users and the rapid growth of networking technologies, Electronic commerce (E-commerce) is perceived as one of the killer applications of the computer and communication technologies. Internet technology enables companies to track multiple activities of customers. It can lead to customizing communications, products, services, and prices. Thus, it is vitally important for the company to anticipate visitors' concerns and provide relevant products when customers visit the site of a company.

To meet various needs, companies tend to adopt differentiated and customer-oriented marketing strategies to gain competitive advantage[1]. CRM is one of the fastest growing business technology initiatives since the web. It emphasizes that companies find it more profitable to retain existing customers, by developing long-term relationships that meet their need, than attracting new customers[2]. The Pareto Principle (80/20 Rule) is commonly employed in CRM, in which "80% of the profits are produced by top 20% of profitable customers"[3, 4]. Furthermore, marketing management can solidify ephemeral relationships with customers into long-term ones if it can discover and predict changes in customer behavior. Thus, to analyze customers' historical behaviors to recognize users' preferences and to build customer profiles are becoming one of the most issues for CRM[5-10]. In this study, we expect to exploit repurchase rates (the possibility of each registered member purchase at least once one product with respect to a specific category for the e-shop) within a given time frame. Since our data only contains transaction records, we borrowed the method of [11-13] and apply the purchasing power (the frequency of purchase and the monetary value) with respect to a specific category of e-shop to construct a preference model for repurchase rate predictions.

Customer preferences change over time. Often the cause of changes is hidden, which makes the learning task complicated[14]. Hidden changes can induce some radical changes in the target concept, and this is known as concept drift[15]. There are two basic types of concept drift which may occur in the real world: (1)sudden concept drift, and (2)gradual concept drift[16]. 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, when someone graduates from school and is ready for taking up an occupation, his preference gradually focuses on professional knowledge.

In the paper, we would like to predict members' repurchase rates within a given time frame. Since the majority of research in Recommender System has focus on the mechanism of recommendation, there has thus far been relatively little discussion on the repurchase rates of the recommended members into the area. When a member has different interest in different category of the e-shop, we can promote the products of the specific category, which the member has higher interest in. The main contributions in this study are listed as follows:

doi:10.4304/jsw.8.6.1316-1326

1. The first one is that the proposed approach takes the different types of past repurchase patterns as basis to capture the members' concept drift over time.
2. The second one is that the proposed approach combines the repurchase index (RI) and the preference drifts to propose a Behavioral Repurchase Prediction(BRP) model to predict the members' repurchase rates.

The experimental results with a real dataset show that the model can achieve our goal.

II. 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 [17]. In addition, Lee et al. has proposed that customers of online stores go through four main shopping steps: product impression, click-through patterns, basket placement, and purchase[18]. 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[11-13, 19, 20]. Among them, Cho et al. proposed the following approach to build customer profiles that is depicted in Equation (1) [11]:

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

In the 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.
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

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.

In general, the model above is used in the Recommenders system. If we directly apply it to the repurchase 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.

III. PROBLEM STATEMENTS

In our study, since our data only contains transaction records, we borrowed the method of [11-13] and apply the purchasing power (the frequency of purchase and the monetary value) to construct a repurchase index (RI) for the members' repurchase rate predictions. This is shown in Equation (2).

$$RI_{i,j} = \frac{P_{i,j}}{Max(C_j)} + \frac{M_{i,j}}{Max(P_j)} \quad (2)$$

$P_{i,j}$ represents the total number of purchase records for member *i* within a given time period *j*; $M_{i,j}$ represents the monetary value of purchase records for member *i* within a given time period *j*; $Max(P_j)$ and $Max(M_j)$ are the maximum values for the number of purchases and monetary value within the given month *j*; $RI_{i,j}$ is normalized to avoid a single extreme value skewing the RI model for member *i*, and $\frac{P_{i,j}}{Max(P_j)}$ and $\frac{M_{i,j}}{Max(M_j)}$ are limited between 0 and 1 respectively. We have used members' past feedback in the previous months to predict the repurchase rates of the next month. Basically, we expect that a higher RI will imply a higher repurchase rate.

In this investigation, we have applied RI_{j-1} to observe a repurchase rate in month *j*. The repurchase rate for RI_{j-1} is defined in Equation (3).

$$\textit{the repurchase rate for the same } RI_{j-1} = \frac{\textit{the total number of repurchased customers in } Month_j \textit{ for the same } RI_{j-1}}{\textit{the total number of customers in } Month_{j-1} \textit{ for the same } RI_{j-1}} \quad (3)$$

Taking data in the game category for example, we assumed that the dataset covers four months(from April to July), and we took 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 RI for the game category. Then, we observed the repurchase rate of the members.

One Month Data:

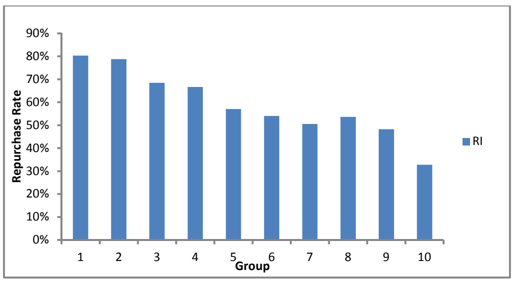


FIGURE 1. Repurchases rate for July for RI_6

Using the data in June as training data, we calculated RI_6 for the members who have purchased in the game category in June, then, we observed the repurchase rates of those in July. However, for the members who did not repurchase in June, their RI_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 RI from high to low and divided the members into ten groups by number of people, with Group 1 having the highest RI and Group 10 having the lowest RI. The y-axis signifies the repurchase rate in July of each group. In FIGURE 1, we observe that the trend of the graph for RI_6 is different from that which we expected because Groups 8 does not conform to the theory that a higher RI corresponds to a higher repurchase rate. As a result, for selecting target members, if the priority option of a marketer is in accordance with the group sequence, 1~10, it may lead to some mistakes and incur some unnecessary promotional costs.

TABLE 2.

THE REPURCHASE STATUS FOR DIFFERENT TYPES OF MEMBER WITH IDENTICAL RI_6 IN GROUP 8

	April	May	June	Repurchase rate for July
(1)	X	X	O	40%
(2)	O	O	O	78.7%

TABLE 2 presents the respective repurchase statuses for two different types of member with the same RI in Group 9. Type (1) is the members who repurchased the products in the game category only in June, and type (2) is the members who continuously repurchased from April to June. The result shows that type (1) and type (2) have different repurchase rates in July. In other words, the problem of RI could not differentiate the members with different past behaviors.

TABLE 3.

THE REPURCHASE STATUS IN JULY FOR DIFFERENT TYPES OF MEMBER WITH $RI_6 = 0$

	April	May	June	Repurchase rate for July
(1)	O	O	X	31.3%
(2)	X	O	X	20.9%
(3)	O	X	X	14.7%

Besides, if the members did not repurchase the products in the game category in June, their RI_6 would be zero. We cannot distinguish the repurchase rates with RI_6 for those members, even if they repurchased in April or May. The result reflected in TABLE 3 has different past behaviors may imply different repurchase rates.

Three Month Data:

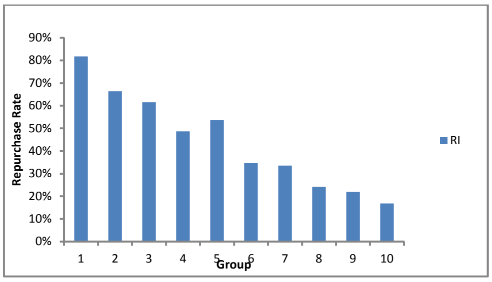


FIGURE 2. Repurchase rates for July for RI_{4-6}

Now, we extended the training data period from one month to three months, i.e. from April to June. Thus, we computed RI_{4-6} and then observed the members' repurchase rates in July. Fig. 3 shows the relationship with the repurchase rate in July based upon different RI_{4-6} in June. In FIGURE 2, the repurchase rate of Group 5 is higher than those of Group 4. Moreover, as shown in we can observe that the repurchase rates of the three different types of members with the same RI in Group 7 are different, where type (1), type (2) and type (3) are for the members who repurchased only in June, May and April respectively.

TABLE 4, we can observe that the repurchase rates of the three different types of members with the same RI in Group 7 are different, where type (1), type (2) and type (3) are for the members who repurchased only in June, May and April respectively.

TABLE 4.

THE REPURCHASE STATUS FOR DIFFERENT TYPES OF MEMBER WITH IDENTICAL RI_{4-6} IN GROUP 7

	April	May	June	Repurchase rate of July
(1)	X	X	O	51.6%
(2)	X	O	X	29.6%
(3)	O	X	X	16.3%

Even though the training data time was extended to three months, we still could not accurately predict their repurchase rate by RI. In other words, extending the training data period did not provide us a true reflection of the members' past behaviors with RI. The above results confirm that the repurchase rates of members are deeply affected by their past behaviors. In fact, the different past behaviors would result in a different repurchase rate despite the same RI. For this reason, we introduced a time factor into the model to differentiate members with different behavior patterns.

As shown in FIGURE 1, we observe that members within the top 45% RI accompany nearly 60% ~ 90% repurchase 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 45% RI in FIGURE 1 are approximately 12% of the total number of members in the game category. Moreover, for the members who did not repurchase in June, their RI_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 55% RI in FIGURE 1 and the members who did not repurchase in the current month, that is, the remaining 88% of the total number of members in the music service category. Furthermore, since there are some RI in which corresponding members are few, their repurchase rate predictions will be distorted. For example, for some RI, the total number of its corresponding members is one. If this member repurchases in the next month, the repurchase rate is 100%; others, the repurchase is 0%. For this situation, we consider that the adjacent RI would be mapped into a similar repurchase rate. Hence, for some RI in which corresponding members are scarce, we merge RI with their neighbors (lower RI) until the total number of members is more than 100.

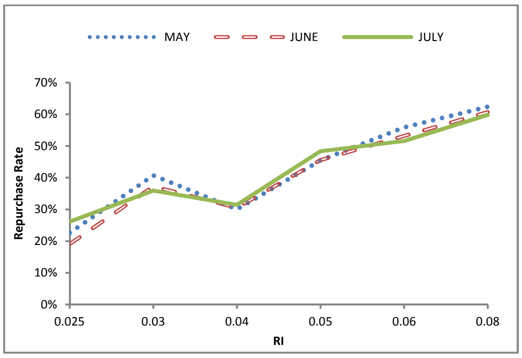


FIGURE 3. The repurchases rate of the members in May 2009, June 2009 and July 2009

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

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

One point is worth making about TABLE 5, since the members for Type 5, Type 6, and Type 7 did not repurchase the products of the game service category of the e-shop in $Month_{j-1}$, their RI_{j-1} are zero and we cannot apply their RI_{j-1} to observe the repurchase rates. Thus, we focus on the members, whose RI_{j-1} are not zero, that is, Type 1 ~ Type 4. Take the repurchase rates of the members (Type 1 ~ Type 4) in May for example, as shown in FIGURE 4, it is clear that the relationship between the past behavior patterns, RI and the repurchase rates are inconsistent. Therefore, we cannot apply RI_{j-1} to predict members' repurchase rates due to the inconsistent of the repurchase rates every month and the members, whose RI_{j-1} are zero.

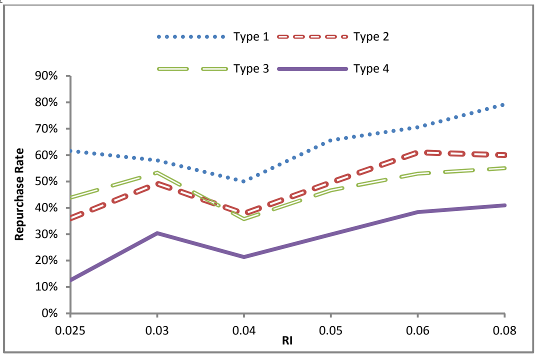


FIGURE 4. The repurchase rates of the members (TYPE 1 ~ TYPE 4) in May 2009

IV. Proposed Method

A. *The Relationship between The Past Behavior Patterns, RI and The Cumulative Average Repurchase Rate*

As mentioned in problem statement, the repurchase rates every month is inconsistent for every RI. Thus, we use the cumulative average repurchase rate which is depicted in Equation (4) to solve this problem.

$$\text{the cumulative average repurchase rate for the same RI in Month}_j = \frac{\sum_{k=1}^{j-1} \text{the number of repurchased members for the same RI}_k \text{ in Month}_{k+1}}{\sum_{k=1}^{j-1} \text{the number of members for the same RI}_k \text{ in Month}_k} \quad (4)$$

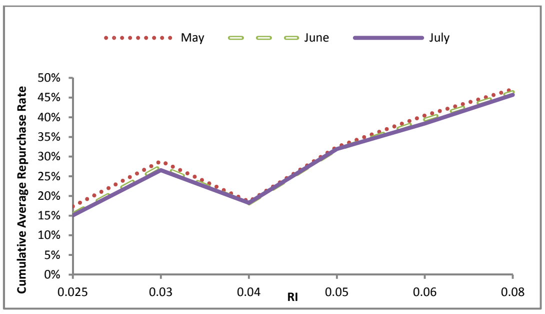


FIGURE 5. The cumulative average repurchase rates of the members in May 2009, June 2009 and July 2009

The cumulative average repurchase rates for every RI in May, June and July are shown in FIGURE 5. The results indicate that higher RI is uncertain to imply a higher cumulative average repurchase rate. However, it is clear that every RI will imply a similar cumulative average repurchase rate in FIGURE 5. Thus, when RI_{j-1} is not zero, we can use the cumulative average repurchase rate of RI_{j-1} as the basis to predict the repurchase rate. Take for example, when a marketer wants to target the members with high repurchase rates in May, we can use the cumulative average repurchase rate of RI_4 as the basis for prediction, then rank the members with the cumulative average repurchase rates of RI_4 in April for targeting.

As mentioned above, we can use the cumulative average repurchase rate of RI_{j-1} as the basis for prediction. We further discuss the relation between the cumulative average repurchase 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 repurchase the products of the game category of the e-shop in $Month_{j-1}$, their RI_{j-1} is zero and we cannot apply their RI_{j-1} to observe the repurchase rates.

Firstly, we focus on the members who did repurchase 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 repurchase rates for every RI_{j-1} are presented in FIGURE 6. Although the same RI_{j-1} may imply different cumulative average repurchase rates and past behavior patterns, the results in FIGURE 6 show that with the same RI_{j-1} and past behavior pattern, the cumulative average repurchase rates of the members in May, June and July are similar. For example, in FIGURE 6, it is clear that when RI_{j-1} is 0.03, its corresponding cumulative average repurchase rates ranges between 30% and 57% for different types of past behavior pattern every month; but within the same past behavior pattern, the same RI_{j-1} will imply similar cumulative average repurchase rates. In other words, when we differentiate the past behavior patterns of the members, we can apply the cumulative average repurchase rate, which is corresponding to RI_{j-1} and their past behavior patterns, as the basis for members' repurchase rate predictions.

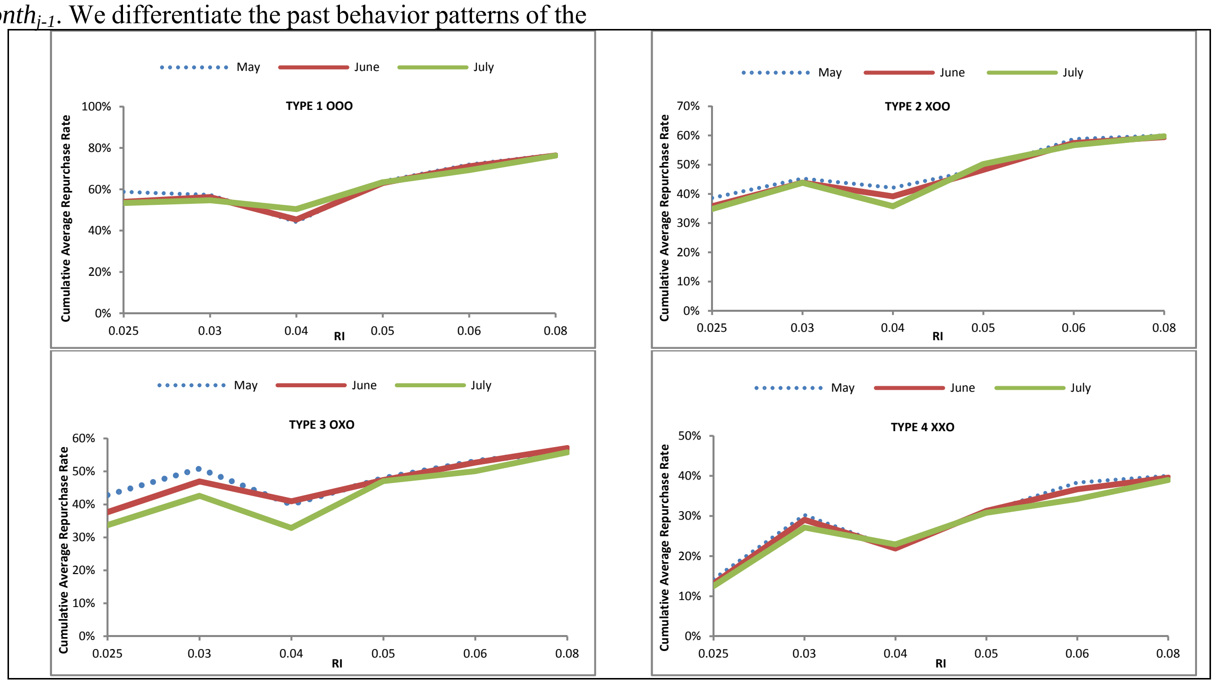
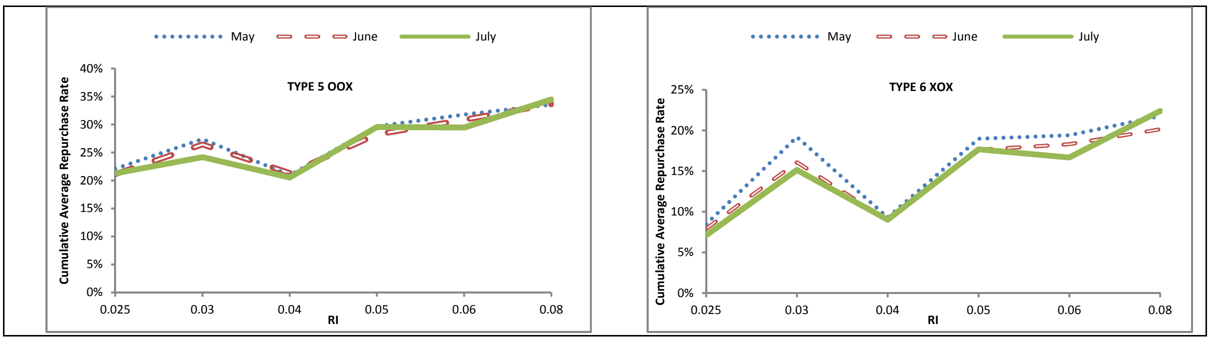


FIGURE 6. The cumulative average repurchase rate of the members (TYPE1 ~ 4) in May 2009, June 2009 and July 2009



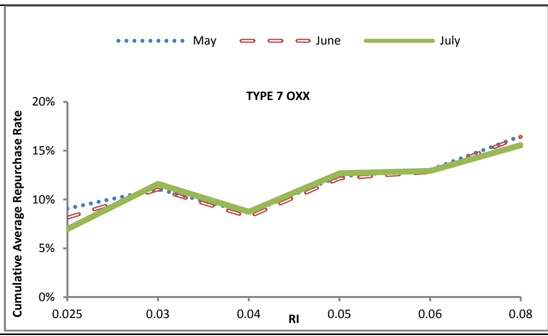


FIGURE 7. The cumulative average repurchase rates of the members ((TYPE5 ~7)) in May 2009, June 2009 and July 2009

Next, for the members who did not repurchase 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 repurchase at least three recent consecutive months, their repurchase rates in May, June and July, are 1.61%, 2.36% and 3.07, respectively. Thus, when we want to target the members with high repurchase rates, this type of members should be chosen last. Moreover, for TYPE 5, TYPE 6 and TYPE 7, the corresponding cumulative average repurchase rates of these three types of behavior pattern are presented in FIGURE 7. The results show that the cumulative average repurchase 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 repurchase rate, we can effectively predict the repurchase rates for the members, whose RI_{j-1} is zero.

B. *The BRP Model*

Based on the discussion in section 4.1, for the same RI_{j-1} and past behavior pattern, the trends of the cumulative average repurchase rates are similar. Next, we apply three different approaches to predict the members' repurchase rates in a $Month_j$, such as follows: (approach A) we use ranked RI_{j-1} as the basis; (approach B) we use the ranked cumulative average repurchase rate, which corresponds to RI_{j-1} as the basis; (approach C) we differentiate the past behavior patterns of the members, then apply the cumulative average repurchase rate, which corresponds to RI_{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 RI_4, RI_5 and RI_6, then apply the three above-mentioned approaches to predict the repurchase rates in May, June and July, respectively. The results are shown in FIGURE 8, where the x-axis presents the cumulative number of members, and the y-axis signifies the corresponding cumulative average repurchase rates. As shown in FIGURE 8, the x-axis presents the cumulative number of members, and the y-axis signifies the corresponding repurchase rate of the cumulative number of members. The results reflect that approach C can achieve our goal with clear support that when we apply the ranked cumulative average repurchase rate as the basis.

THE ALGORITHM OF THE BRP MODEL

<u>Input:</u>

m: the members in the dataset

t: the target month for prediction

j: the previous month the target month, $j = t-1$

<u>Output:</u>

mList: the list of members with ranked return visit rates

Steps:

1. **For** each member $i \in m$ **do**
2. **For** (j=1; j<t; j++) do
3. Computing $RI_{i,j}$ of member i based on Equation (2)
4. Computing the cumulative average repurchase rate in $Month_j$ for all past behavior patterns separately based on Equation (4) and applying the cumulative average repurchase rate which is corresponding to RI_j for prediction
5. **End For**
6. **End For**
7. Output the target members in *mList* by the ranked cumulative average repurchase rates

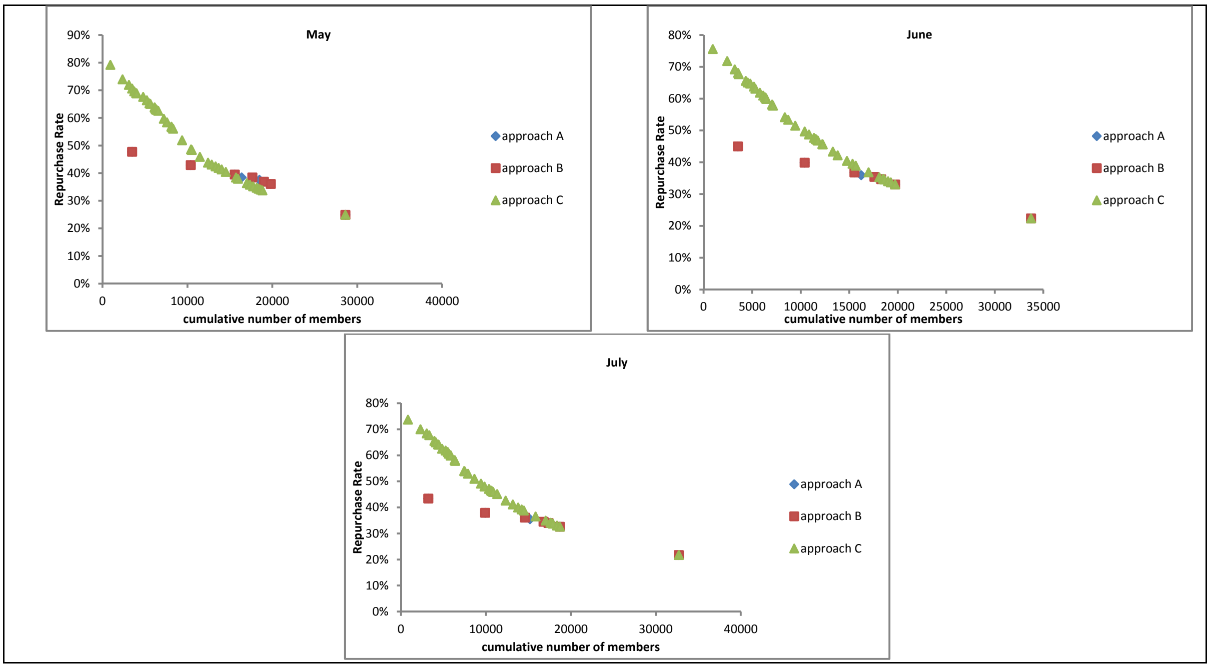


FIGURE 8. Comparison of three different approaches in May 2009, June 2009, July 2009

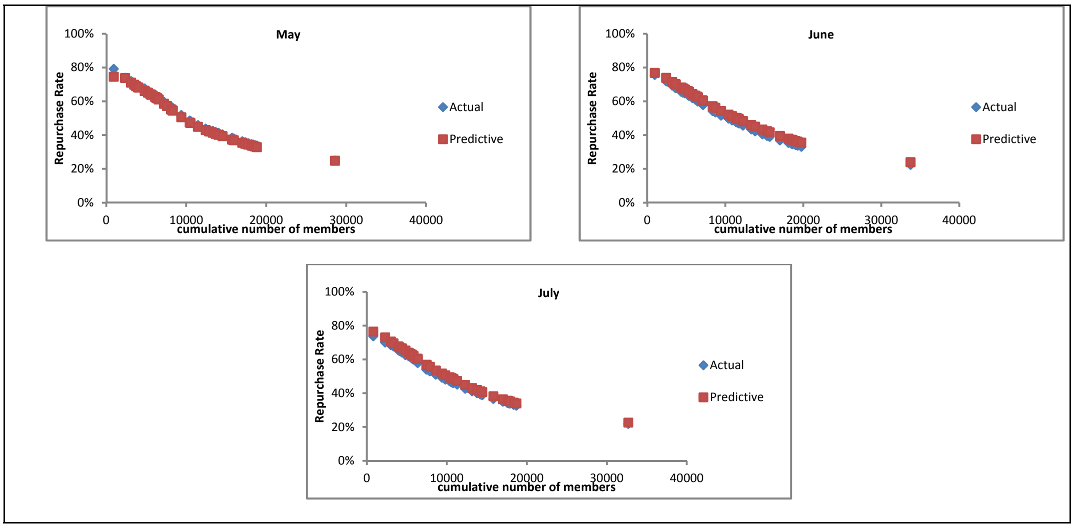


FIGURE 9. The actual and predictive repurchase rate in May 2009,July 2009 and July 2009

In the studies discussed above, we propose a novel algorithm of the Behavioral Repurchase Prediction (BRP) model to predict the members' repurchase rates. As shown above, firstly, we compute the RI 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 repurchase rate by $Month_j$ based on Equation (4). Next, since our goal is to target the members with the high repurchase rates, we rank the cumulative average repurchase rate by $Month_j$ as the basis to predict the monthly repurchase rate. In other words, when we want to predict the members' repurchase rates in July, we should apply their ranked cumulative average repurchase rates in June, which corresponds to every RI_5 and each type of past behavior pattern, as the basis for prediction. As shown in FIGURE 9, the x-axis presents the cumulative number of members, and the y-axis signifies the corresponding actual and predictive repurchase rate of the cumulative number of members. It is obvious that the actual and predictive repurchase rates in May, June and July are increasingly similar.

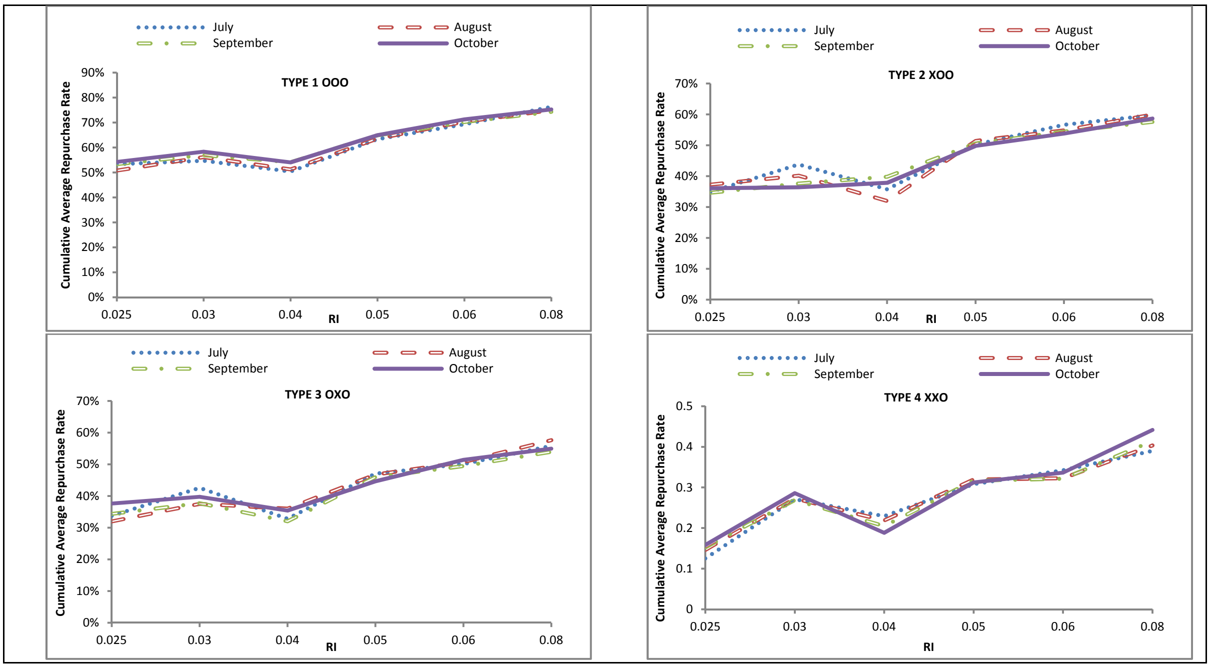


FIGURE 10. The cumulative average repurchase rate of the members (TYPE 1 ~ 4) in July 2009, August 2009, September 2009 and October 2009

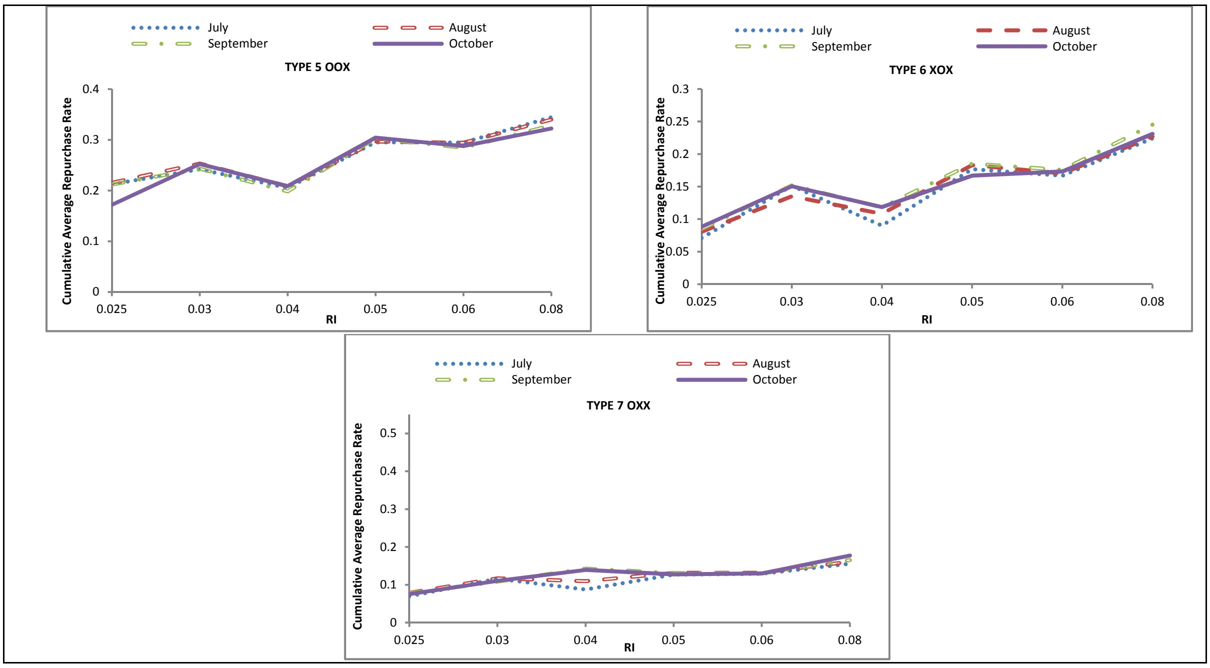


FIGURE 11. The cumulative average repurchase rate of the members (TYPE 5 ~ 7) in July 2009, August 2009, September 2009 and October 2009

V. Experiments and Discussion

Our data set came from a well-known e-shop in Taiwan. The period of the dataset covers the purchasing records from January 2009 to November 2009. It provides abundant services on nearly all conceivable major topics (categories): games, music, fortune-telling service and much more. We also applied our method to the fortune-telling service categories of e-shop, which were contributed by only 5193 registered members. And the results demonstrated that the method can be practically implemented and provide satisfactory results in this small category. However, because of the limitation of space, we used the game category as a case study to conduct the follow-up experiments and discussions. For the game category, the total useful number of transactions records was 7,203,035, which were contributed by 773,513 registered members. Furthermore, 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.

A. Accuracy Verification Of The BRP Model

In FIGURE 10 and FIGURE 11, which depict the cumulative average repurchase 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 repurchase rates are similar every month. Furthermore, as shown in part (a) of FIGURE 12, the actual and predictive repurchase 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 repurchase rate by $Month_{j-1}$ as the basis to predict the repurchase rate by $Month_j$.

Next, we use lift curves to show the accuracy of our model. A lift curve presents the proportion of repurchased members detected against the proportion of repurchased 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 repurchased members are selected first and the random curve presents that ß % repurchased are selected from ß% members[21]. As observed in part (b) of FIGURE 12, there are four lift curves in the cumulative gains chart for the repurchase rate in August, which are an ideal curve, a lift curve for a predictive repurchase rate, a lift curve for an actual repurchase rate and a random curve. The results are reflected by using the BRP model, where the actual curve captures approximately the top 20% of the members, which comprises the members with nearly a 56% repurchase rate, compared to a random curve, which only comprises nearly 20% repurchased 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.

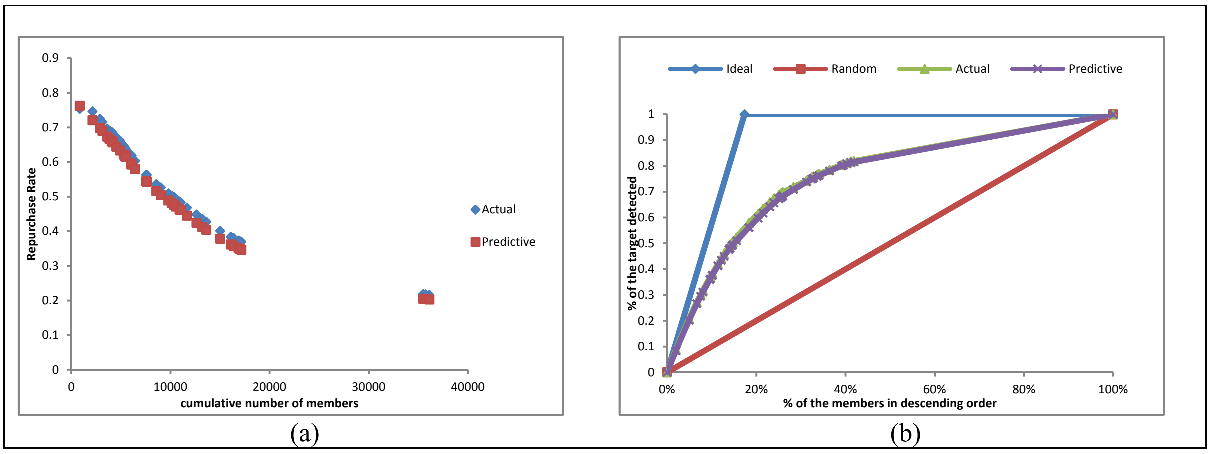


FIGURE 12. The result comparison in August 2009

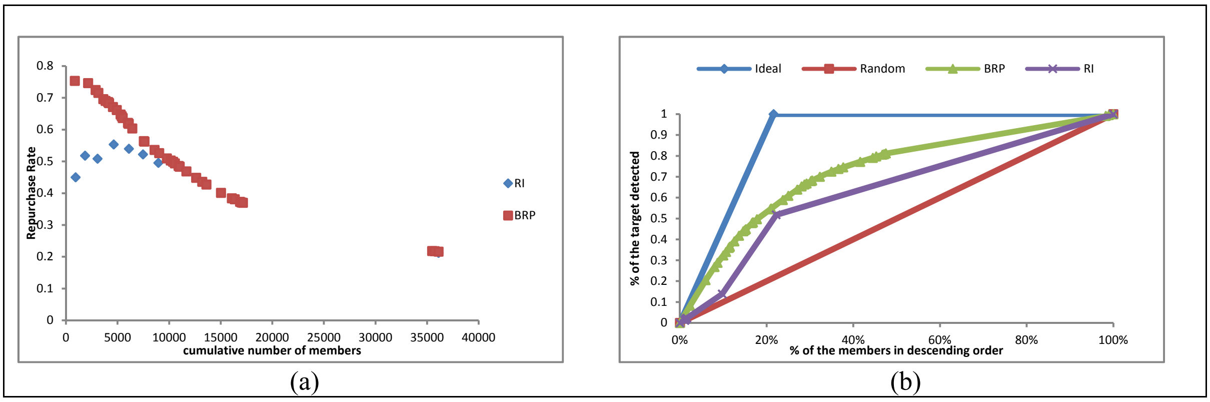


FIGURE 13. Comparison of the RI model and the BRP model in August 2009

B. *Comparison Of The BRP Model And The RI Model*

In this experiment, we compare the accuracy of the BRP model with the RI model. As shown in part (a) of FIGURE 13, when we have to choose 3000 members for targeting, the repurchase rates for the 3000 members in the BRP model are on the verge of 72%, but the repurchase rates for the 3000 members in the RI model are approximately 50%. The result shows that the BRP model is significantly superior to the RI model in usability for repurchase rate prediction. Likewise, the results in September, October, and November can prove validity.

Similarly, we use lift curves to compare the usability of the two models. As observed in part (b) of FIGURE 13, there are four lift curves in the cumulative gains chart for the repurchase rate in August: an ideal lift curve, a lift curve for an actual repurchase rate for the BRP model, a lift curve for an actual repurchase rate for the RI model and a random curve. The results are reflect by using the BRP model, where the actual curve captures approximately the top 10% of the members, which comprises nearly 28% repurchased members, compared to a lift curve for the RI model, which only comprises nearly 13% repurchased members. Moreover, for the top 10%~22% of members in the lift curve for an actual repurchase rate for the RI model, the RI model has to incorporate randomly chosen target members. The BRP model can effectively achieve our goal rather than the RI model. Likewise, the bottom 78% of members cannot be captured by the RI model because those members did not repurchase the products of the game category in July and their RI is zero. In contrast, the BRP model can capture those members for accurate repurchase rate predictions. Similarly, the results in September, October, and November can prove the usability as well.

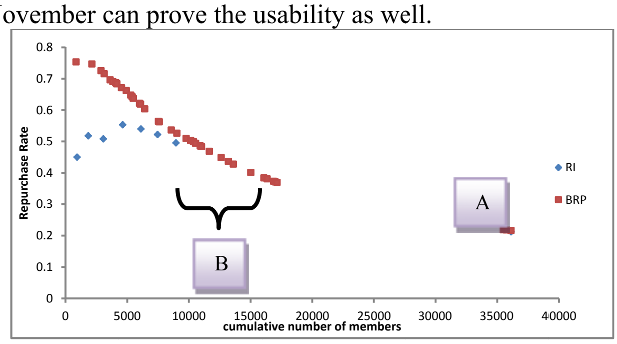


FIGURE 14 Applying the brp model, A choice priority comparison with two types of past behavior patterns

Moreover, taking RI_8 as an example in problem statements, if the members did not repurchase in August, their RI_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 repurchase in August. This kind of potential members should be targeted. As shown in FIGURE 14, Type (A) stands for the members who only purchased the products of the game category in August and type (B) stands for the members who were heavy users but did not repurchase in August. It is clear that the repurchase rate of type (A) is lower than that of type (B). However, in the RI model, their RI_8 for type (B) is zero. If the marketers target potential members for promotion using the RI model, they will neglect the members for type (B). In other words, the BRP model not only solves the problem that RI = 0 but also targets members with the highest repurchase rate more accurately.

CONCLUSIONS

The main purpose of this work is to leverage historical members' past behavior patterns to construct a BRP model to predict the members' repurchase rates. We have demonstrated that the BRP model can be practically implemented and provide adequate results. Besides the repurchase 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.

ACKNOWLEDGEMENT

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

REFERENCES

[1] K. Yada, "CODIRO: A New System for Obtaining Data Concerning Consumer Behavior Based on Data Factors of High Interest Determined by the Analyst," *Soft Computing - A Fusion of Foundations, Methodologies and Applications,* vol. 11, pp. 811-817, 2007.
[2] A. Garrido-Moreno and A. Padilla-Meléndez, "Analyzing the impact of knowledge management on CRM success: The mediating effects of organizational factors," *International Journal of Information Management,* vol. 31, pp. 437-444, 2011.
[3] R. S. Duboff, "Marketing to maximize profitability," *The Journal of Business Strategy,* vol. 13, p. 10, 1992.
[4] B. A. Gloy, J. T. Akridge, and P. V. Preckel, "Customer Lifetime Value: An application in the rural petroleum market," *Agribusiness,* vol. 13, pp. 335-347, 1997.
[5] M. J. Martín-Bautista, D. H. Kraft, M. A. Vila, J. Chen, and J. Cruz, "User profiles and fuzzy logic for web retrieval issues," *Soft Computing - A Fusion of Foundations, Methodologies and Applications,* vol. 6, pp. 365-372, 2002.
[6] H. Sun, Y. Peng, J. Chen, C. Liu, and Y. Sun, "A New Similarity Measure Based on Adjusted Euclidean Distance for Memory-based Collaborative Filtering," *Journal of Software,* vol. 6, 2011.
[7] H. Ye, "A Personalized Collaborative Filtering Recommendation Using Association Rules Mining and Self-Organizing Map," *Journal of Software,* vol. 6, 2011.
[8] S. Gong, "A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering," *Journal of Software,* vol. 5, 2010.
[9] S. Gong, "Employing User Attribute and Item Attribute to Enhance the Collaborative Filtering Recommendation," *Journal of Software,* vol. 4, 2009.
[10] S. Chen, T. Jian, and H. Yang, "A Fuzzy AHP Approach for Evaluating Customer Value of B2C Companies," *Journal of Computers,* vol. 6, pp. 224-231, 2011.

[11] 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.
[12] 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.
[13] 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.
[14] T. C.-K. Huang, "Mining the change of customer behavior in fuzzy time-interval sequential patterns," *Applied Soft Computing,* vol. 12, pp. 1068-1086, 2012.
[15] G. Widmer and M. Kubat, "Learning in the presence of concept drift and hidden contexts," *Machine Learning,* vol. 23, pp. 69-101, 1996.
[16] A. Tsymbal, "The Problem of Concept Drift: Definitions and Related Work," 2004.
[17] 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.
[18] 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.
[19] 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.
[20] X. Wan, Q. Jamaliding, and T. Okamoto, "Analyzing Learners' Relationship to Improve the Quality of Recommender System for Group Learning Support," *Journal of Computers,* vol. 6, pp. 254-262, 2011.
[21] L. Qi, C. Enhong, X. Hui, C. H. Q. Ding, and C. Jian, "Enhancing Collaborative Filtering by User Interest Expansion via Personalized Ranking," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 42, pp. 218-233, 2012.

Yi-Hsin Wang is currently an assistant professor of the Department of Information Management at Chang Gung University of Science and Technology, Taiwan. He received his Ph.D. degrees in the Department of Computer Science and Information Engineering from Tamkang University in Taipei, Taiwan, in 2004. His research interests include fuzzy, relational databases and data mining.

Rui-Dong Chiang is currently a professor of the Department of Computer Science and Information Engineering. He received B.S. degree in hydraulic engineering from Chung Yuan Christian University, Taiwan, in 1981, and the M.S. and Ph.D. degrees in computer science from the University of Southwestern Louisiana in 1986 and 1990, respectively. His Current research interests include fuzzy, relational databases and data mining.

Huan-Chen Chu is working toward a Ph.D degree in Computer Science and Information Engineering from Tamkang University in Taipei, Taiwan. His research interests include temporal data mining, e-commerce, recommender and cyber culture.