Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
The Journal of Systems and Software 77 (2005) 181–191
The Journal of Systems and Software
www.elsevier.com/locate/jss

Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences

Duen-Ren Liu [a,*], Ya-Yueh Shih [a,b]

[a] Institute of Information Management, National Chiao Tung University, Hsinchu, 300 Taiwan, ROC
[b] Department of Information Management, Minghsin University of Science and Technology Hsinchu, Taiwan, ROC

Received 15 January 2004; received in revised form 28 July 2004; accepted 8 August 2004
Available online 7 October 2004

Abstract

Recommending products to attract customers and meet their needs is important in fiercely competitive environments. Recommender systems have emerged in e-commerce applications to support the recommendation of products. Recently, a weighted RFM-based method (WRFM-based method) has been proposed to provide recommendations based on customer lifetime value, including Recency, Frequency and Monetary. Preference-based collaborative filtering (CF) typically makes recommendations based on the similarities of customer preferences. This study proposes two hybrid methods that exploit the merits of the WRFM-based method and the preference-based CF method to improve the quality of recommendations. Experiments are conducted to evaluate the quality of recommendations provided by the proposed methods, using a data set concerning the hardware retail marketing. The experimental results indicate that the proposed hybrid methods outperform the WRFM-based method and the preference-based CF method.

Keywords: Recommender system; Data mining; Product recommendation; Customer lifetime value (CLV); Collaborative filtering

1. Introduction

Recommender systems have emerged in e-commerce applications to support product recommendation (Schafer et al., 2001), which provide individual marketing decisions for each customer (Peppers and Rogers, 1997). Recommender systems are technologies that assist businesses to implement one-to-one marketing strategies. Recommender systems rely on customer purchase history to determine customer preferences and to identify products that customers may purchase. Supporting product recommendation services can strengthen the relationship between the buyer and seller and thus increase profit. Schafer et al. (2001) presented a detailed taxonomy of recommender systems in E-commerce, and elucidated how they can provide personalization to establish customer loyalty. Generally, recommender systems provide several merits, including increasing the probability of cross-selling; establishing customer loyalty and fulfilling customer needs by discovering products in which they may be interested in purchasing.

Various recommendation methods have been proposed for recommender systems. Collaborative filtering (CF) has been successfully used in various applications. The CF method utilizes preference ratings given by various customers to determine recommendations to a target customer based on the opinions of other similar customers. A typical CF method employs K-nearest neighbors approach to derive top-N recommendations (KNN-based CF method). The GroupLens system (Resnick et al., 1994) applied CF method to recommend Usenet News and movies. Video recommender (Hill et al.,

* Corresponding author. Tel.: +886 35712121x57405; fax: +886 35723792.
E-mail address: dliu@iim.nctu.edu.tw (D.-R. Liu).

0164-1212/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2004.08.031

1995) also used the collaborative approach to generate recommendations on movies. Examples of music recommender systems are Ringo (Shardanand and Maes, 1995) and MRS (Chen and Chen, 2001). Siteseer (Rucker and Polanco, 1997) provided Web page recommendations based on the bookmarks of the user's virtual neighbors. Content-based filtering is another approach different from collaborative filtering. Conventional content-based filtering provides recommendations by matching customer profiles (e.g. interests) with content's features (e.g. product's attributes). NewsWeeder (Lang, 1995) is an example of content-based recommender systems. Moreover, Changchien and Lu (2001) developed a procedure for mining association rules to support on-line product recommendations. Amazon.com employed item-to-item collaborative filtering to provide recommendations of those products that are similar to the customer's purchased and rated products (Linden et al., 2003).

Although various approaches for making recommendations have been presented, few consider customer lifetime value (CLV) and the effect on product recommendations. Firms are increasingly recognizing the importance of the lifetime value of customers (Berger and Nasr, 1998). Generally, RFM (Recency, Frequency, and Monetary) method has been used to measure CLV (Miglautsch, 2000; Kahan, 1998). In fiercely competitive environments, identifying the CLV or loyalty ranking of customer segments is important for helping decision-makers to target markets more clearly. Additionally, the effect of CLV on recommendations should be investigated to make more effective marketing strategies. Recently, Liu and Shih (2004) proposed a weighted RFM-based method (WRFM-based method) that integrates AHP and data mining to recommend products based on customer lifetime value. The WRFM-based method employs association rule mining to identify recommendation rules from customer groups that are clustered according to weighted RFM values. Their experiments demonstrated that the WRFM-based method outperforms the typical KNN-based CF method, and can identify effective rules for making recommendations to customers with high lifetime value or loyalty. However, generating recommendation rules for less loyal customers is difficult. Similar to the WRFM-based method, a preference-based CF method can be derived that employs association rule mining to extract recommendation rules from customer groups which are clustered according to customers' purchase preferences. A pilot experiment of this study revealed that the preference-based CF method may suggest some useful recommendations that the WRFM-based method can not provide, and thus may improve the quality of recommendations to less loyal customers. Accordingly, this study proposed hybrid recommendation methods that incorporate the advantages of the WRFM-based method and the preference-based CF method.

The rest of this study is organized as follows. Section 2 outlines the background and reviews related work on customer lifetime value, market segmentation and recommendation methods. Section 3 illustrates the proposed methods. Section 4 describes the experimental setup and the criteria to evaluate the quality of recommendations. Experimental results are also presented to verify the proposed approach. Finally, Section 5 draws conclusions, summarizing the contributions of this work and outlining areas for future research.

2. Background

2.1. Customer lifetime value analysis and RFM evaluation

Customer lifetime value (CLV) is typically used to identify profitable customers and to develop strategies to target customers (Irvin, 1994). Measuring RFM is an important method to assess customer lifetime value. Bult and Wansbeek (1995) explained the RFM terms as follows: (1) *R* (Recency): period since the last purchase, and a lower value corresponds to a higher probability of the customer's making a repeat purchase; (2) *F* (Frequency): number of purchases made within a certain period; higher frequency indicates higher loyalty and (3) *M* (Monetary): the amount of money spent during a certain period; a higher value indicates that the company should focus more on that customer.

Numerous studies have discussed the evaluation of CLV. Hughes (1994) presented a method for RFM scoring that involved using RFM data to sort individuals into five customer quintiles. Different marketing strategies can thus be adopted for different customers. Practically, RFM variables may be differently weighted in different industries (Stone, 1995). However, Stone (1995) determined the RFM weightings subjectively, without employing a systematic approach. The analytic hierarchy process (AHP), a mathematical technique for multi-criteria decision-making (Saaty, 1994), enables decision-makers to make decisions. Liu and Shih (2004) had employed AHP to evaluate RFM weightings (relative importance) according to the subjective perceptions of decision-makers.

2.2. Market segmentation

Data mining techniques have been applied to various application domains. Clustering technique is usually used to segment markets (Punj and Stewart, 1983; Chen et al., 1996), which seeks to maximize variance among groups while minimizing variance within groups. Many clustering algorithms have been developed, including K-means, hierarchical, fuzzy c-means and others. K-means clustering (MacQueen, 1967) is a method commonly used to partition a set of data into *k* groups. The

K-means algorithm assigns data samples to clusters by the minimum distance assignment principle (Hearst, 1998), which assigns a data sample d_i to the cluster c_j such that the distance from d_i to the center of c_j is the minimum over all k clusters.

2.3. Association rules for product recommendation

Association rule mining (Agrawal et al., 1993; Srikant and Agrawal, 1995; Yun et al., 2003) is a widely used data mining technique to generate recommendations in recommender systems. Accordingly, this work employs association rule mining to discover the relationships among product items based on patterns of co-occurrence across customer transactions.

2.3.1. Association rule mining

Association rule mining aims to find an association between two sets of products in the transaction database. Agrawal et al. (1993) formalized the problem of finding association rules that satisfy minimum support and minimum confidence requirements. Let I be a set of product items and D be a set of transactions, each of which includes a set of products that are purchased together. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \Phi$. X is the antecedent (body) and Y is the consequent (head) of the rule herein. Two measures, support and confidence, are used to indicate the quality of an association rule. The support of a rule is the percentage of transactions that contain both X and Y, whereas the confidence of a rule is the fraction of transactions that contain X, that also contain Y.

The support of an association rule indicates how frequently that rule applies to the data. Higher support of a rule corresponds to a stronger correlation between the product items. The confidence is a measure of the reliability of an association rule. The higher the confidence of a rule corresponds to a more significant correlation between product items. The *apriori* algorithm (Agrawal et al., 1993; Agrawal and Srikant, 1994) is typically used to find association rules by discovering frequent *itemsets* (sets of product items). An *itemset* is considered to be frequent if the support of that *itemset* exceeds a user-specified minimum support. Moreover, association rules that meet a user-specified minimum confidence, can be generated from the frequent *itemsets*.

2.3.2. Association rule based recommendation

Sarwar et al. (2000) described the method of association rule-based recommendation as follows. For each customer, a customer transaction is created to record all the products previously purchased by the customer. The association rule mining algorithm is then applied to find all the recommendation rules that satisfy the given minimum support and minimum confidence constraints. The top-N products to be recommended to a customer u, is then determined as follows. Let X_u be the set of products previously purchased by customer u. The method first finds all the recommendation rules $X => Y$, for which $X \subseteq X_u$. Then, for each extracted recommendation rule, all the products in Y that have not yet been purchased by customer u are candidate products for recommendation. Each candidate product is associated with the confidence of the corresponding recommendation rule. The candidate products are sorted by associated confidence value, where the N highest ranked candidate products are selected as the recommendation set.

2.4. Collaborative filtering

Collaborative filtering is a successful recommendation method, which has been widely used in various applications. A typical KNN-based collaborative filtering (CF) method (Resnick et al., 1994; Shardanand and Maes, 1995; Sarwar et al., 2000) employs nearest-neighbor algorithm to recommend products to a target customer u based on the preferences of *neighbors*, that is, those customers having similar preferences as customer u. Notably, preferences generally are defined in terms of customer purchasing behavior, namely, purchased/non-purchased (binary choice) of shopping basket data, or taste, namely, preference rating on product items. This work focused on product recommendation of retail transaction data that contains binary choice of shopping basket data.

The typical KNN-based CF method is detailed as follows. Customer preferences, namely, customer purchase history, are represented as a customer-item matrix $\boldsymbol{R}$ such that, r_{ij} is one if the ith customer purchased the jth product; and is zero otherwise. The similarity of preferences among customers can be measured in various ways. A common method is to compute the Pearson correlation coefficient defined as Eq. (1)

$$corr_P(c_i, c_j) = \frac{\sum_{s \in I} (r_{c_i,s} - \bar{r}_{c_i})(r_{c_j,s} - \bar{r}_{c_j})}{\sqrt{\sum_{s \in I} (r_{c_i,s} - \bar{r}_{c_i})^2 \sum_{s \in I} (r_{c_j,s} - \bar{r}_{c_j})^2}} \quad (1)$$

The notations $\bar{r}_{c_i}$ and $\bar{r}_{c_j}$ denote the average number of products purchased by customers c_i and c_j, respectively. Moreover, the variable I denotes the set of products. Additionally, the $r_{ci,s}$ and $r_{cj,s}$ indicate whether customers c_i and c_j purchased product item s.

Customers are ranked by their similarity measures in relation to the target customer u, as determined using the Pearson correlation coefficient. The k most similar (highest ranked) customers are selected as the k-nearest neighbors of customer u. Finally, the top-N recommended products are determined from the k-nearest neighbors of u, as follows. The frequency count of products is calculated by scanning the purchase data of the k-nearest neighbors. The products then are sorted based on frequency count. The N most frequent products that have not yet been purchased by target customer u are selected as the top-N recommendations.

3. Recommendation methods

This study proposes two hybrid methods for recommending products. These two methods incorporate the advantages of the WRFM-based method (Liu and Shih, 2004) and the preference-based CF method. The core concept of the WRFM-based method is to group customers based on weighted RFM, and then extract recommendation rules from each customer group. Section 3.1 summarizes this method. The preference-based CF method is similar to the WRFM-based method except that preference-based method groups customers by purchase preferences. The hybrid1 method is proposed to group customers separately based on CLV and purchase preferences. Then, recommendation rules extracted from the WRFM-based method are used to recommend products to loyal customers; recommendation rules extracted from the preference-based CF method are used to recommend products to less loyal customers. Furthermore, previous research grouped customers by lifetime value or purchase preference separately. Accordingly, this work proposes a hybrid2 method that groups customers by considering both CLV and purchase preferences, and then extracts recommendation rules from each group to support recommendations.

3.1. Weighted RFM-based method

This method (Liu and Shih, 2004) primarily integrated AHP, clustering and association rule mining techniques for product recommendation. It employs the AHP to evaluate the weighting (relative importance) of each RFM variable, and specifically asks decision-makers to make intuitive judgments about ranking ordering to make pairwise comparisons. K-means clustering is then employed to group customers with similar lifetime value or loyalty based on weighted RFM. Finally, the association rule mining is applied to extract recommendation rules from each group of customers.

3.1.1. Assessing RFM weightings

The AHP (Saaty, 1994) is used to determine the relative importance (weights) of the RFM variables, w_R, w_F and w_M, respectively. The three main steps are as follows: asking evaluators (decision makers) to make pairwise comparisons of the relative importance of RFM variables; assessing the consistency of pairwise judgments; and employing *Eigenvalue* computation to derive the weights of RFM variables.

3.1.2. Clustering customers with similar lifetime value

The RFM values of each customer are normalized. The normalized RFM values of each customer are then multiplied by the relative importance of RFM variables, w_R, w_F and w_M, which are determined by the AHP. The similarity among customers can be measured by computing the Pearson correlation coefficient based on the weighted RFM values of customers, as defined in Eq. (2)

$$corr_{WRFM}(c_i, c_j) = \frac{\sum_{s\in V}(WRFM_{c_i,s} - \overline{WRFM}_{c_i})(WRFM_{c_j,s} - \overline{WRFM}_{c_j})}{\sqrt{\sum_{s\in V}(WRFM_{c_i,s} - \overline{WRFM}_{c_i})^2 \sum_{s\in V}(WRFM_{c_j,s} - \overline{WRFM}_{c_j})^2}} \tag{2}$$

In Eq. (2), $\overline{WRFM}_{c_i}$ and $\overline{WRFM}_{c_j}$ are the average weighted RFM (WRFM) value of customer c_i and c_j, respectively. The variable V denotes the set of RFM variables. The variables $WRFM_{c_i,s}$ and $WRFM_{c_j,s}$ indicate the weighted value R (F or M) of customer c_i and c_j, respectively, $s \in [R, F, M]$. The K-means method is then applied to cluster customers based on the weighted RFM values.

3.1.3. Recommendation based on association rules

For each customer, a customer-transaction is created to record all the products previously purchased by the customer. The customer transactions are grouped according to the clusters of customers. Association rule mining is then used to extract the recommendation rule set RS_j from transactions associated with each cluster, rather than from all customer transactions. The cluster C_j to which a customer, u, belongs is first identified. Then, RS_j, the recommendation rule set extracted from C_j is used to select the top-N candidate products to be recommended to customer u. Let X_u represents the set of products previously purchased by customer u. For each recommendation rule $X => Y$ in RS_j, if $X \subseteq X_u$, then all products in $Y - X_u$ are the candidate products for recommendation to customer u.

3.2. Hybrid1 approach

The WRFM-based method is capable of identifying effective recommendation rules for customers with high lifetime value or loyalty. However, generating recommendation rules for less loyal customers is difficult because such customers typically have low purchase frequencies and spend less money; they are also unlikely to have made recent purchases. The preference-based CF method uses clustering to group customers with similar purchase preferences. A pilot experiment of this work yields a meaningful result that shows the preference-based CF method would improve the quality of recommendations for less loyal customers. Accordingly, the proposed hybrid1 method incorporates the merits of

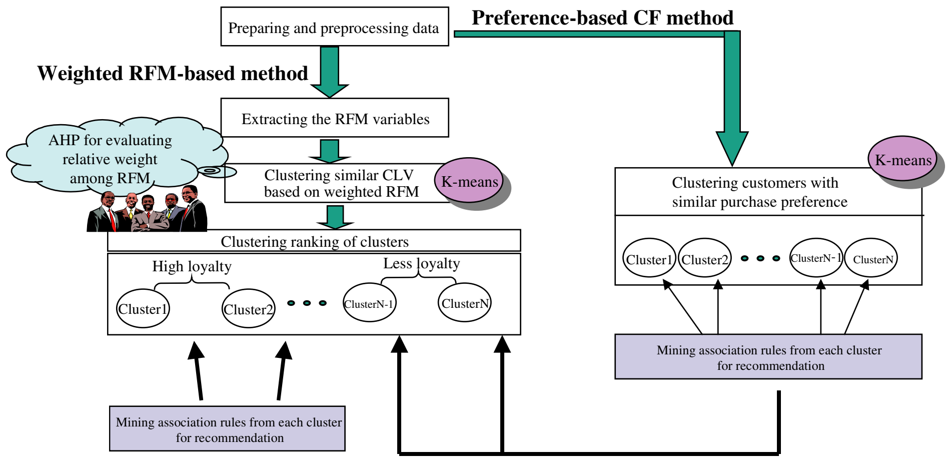

Fig. 1. Hybrid1 method for product recommendation.

the WRFM method and the preference-based CF method to recommend products, as shown in Fig. 1.

Two kinds of clustering are conducted in hybrid1 method to group customers into weighted RFM-based clusters and preference-based clusters, respectively. The weighted RFM-based clusters are created by grouping customers with similar lifetime value according to weighted RFM values, while the preference-based clusters are created by grouping customers with similar purchase preferences. Notably, the CLV ranking of the weighted RFM-based clusters represent the loyalty ranking of customer groups. The association rule mining approach is then applied to extract recommendation rules from each group of customers, derived separately from weighted RFM-based clustering and preference-based clustering. Finally, recommendation rules extracted from weighted RFM-based clusters are used to recommend product items to loyal customers; and recommendation rules extracted from preference-based clusters are used to recommend product items to less loyal customers. The following subsections detail the hybrid1 method.

3.2.1. Data preparing

A data set is used to elucidate the proposed methodology. The case concerns a hardware retailing company that manufactures wheels, casters, platforms and hand trucks for industrial, medical, hospital and institutional use. This company presently produces over 3000 products. Its decision-makers must target customer groups and develop market strategies to satisfy customer needs and thereby increase the market share of the company. Two years of data on purchase transactions, approximately 70,000 records, have been collected. For each customer, a customer transaction is created to record all the products previously purchased by the customer. The average number of product items purchased by customers exceeds 60. The data set is preprocessed to extract customer transactions. Unreasonable records such as those of customers who have a non-zero amount of purchases but have never made any transactions are also removed. In this study, 895 customer transactions are extracted from the database. RFM values of customer transactions are extracted to measure the customers' CLV.

3.2.2. Determining the relative weights of the RFM variables

According to the assessments obtained by the AHP, the relative weights of the RFM variables are 0.731, 0.188 and 0.081, respectively. The implication is as follows. Recency is the most important variable, since the unit price of hardware products is relatively low, and thus evaluators mainly concentrate on whether customers purchase continually. If some customers perform no transaction for a long period, the customers may have been lost or have transferred to new vendors.

3.2.3. Grouping customers with similar CLV

K-means method is used to group customers into WRFM-based clusters based on the weighted RFM values. Notably, the RFM values of customers are normalized and then multiplied by the relative importance of RFM variable. The similarity among customers can be measured by computing the Pearson correlation coefficient based on the weighted RFM values, as defined in Eq. (2).

The method must specify the number of clusters, m, in advance. The parameter m is set to 8, since eight $(2 \times 2 \times 2)$ possible combinations of inputs (RFM) can be obtained by assigning $\downarrow$ or $\uparrow$. If the average R (F, M) value of a cluster exceeds the total average R (F, M), then an upward arrow $\uparrow$ is shown; in the opposite case, a downward arrow $\downarrow$ is shown. Table 1 presents the clustering result of the hardware retailing data set, listing eight clusters, each with the corresponding number

Table 1
Eight clusters generated by K-means clustering

Cluster	WRFM-based method				
	No. of customers	Recency	Frequency	Monetary	Type
1	212	79	36	199010	$R\downarrow F\downarrow M\downarrow$
2	150	69	54	306065	$R\downarrow F\uparrow M\uparrow$
3	190	66	95	593861	$R\downarrow F\uparrow M\uparrow$
4	123	92	41	152007	$R\uparrow F\downarrow M\downarrow$
5	47	147	18	100483	$R\uparrow F\downarrow M\downarrow$
6	100	108	23	130096	$R\uparrow F\downarrow M\downarrow$
7	28	162	10	71536	$R\uparrow F\downarrow M\downarrow$
8	45	135	25	67403	$R\uparrow F\downarrow M\downarrow$
Overall average		89	48	270837	

of customers and the average R, F and M values. The last row also shows the overall average WRFM values of overall customers. The average RFM values of each cluster are compared with the overall averages. The last column of Table 1 shows the RFM pattern for each cluster. Each cluster represents a market-segmentation.

Each cluster represents a market-segmentation. Customers in clusters with the pattern $R\downarrow F\uparrow M\uparrow$ are considered to be loyal, purchased recently, purchase frequently, and spend regularly with the firm. They are gold customers. Clusters with the pattern $R\downarrow F\downarrow M\downarrow$ may include new customers who have only recently visited the company. Customers in such clusters may be trying to develop closer relationships with the company. These customers may become gold customers. Finally, clusters with the pattern $R\uparrow F\downarrow M\downarrow$ include those who very rarely visited the site and made very few transactions. They are valueless customers, and may only make purchases during sales.

3.2.4. Grouping customers with similar purchase preferences

K-means method is also employed to group customers into preference-based clusters based on purchase preferences. The similarity among customers is measured by computing the Pearson correlation coefficient based on purchase preferences, as defined in Eq. (1).

3.2.5. Ranking CLV for each cluster

The purpose of CLV ranking is to clarify the high loyal customers and less loyal customers. The analytical result is derived to help market practitioners develop more effective strategies for retaining customers and thus clearly identify and compare market segments. The CLV ranking of clusters proceeds as follows. The average normalized RFM values of each cluster, denoted as C_R^j, C_F^j, and C_M^j, respectively, for $j = 1$ to m (the number of clusters). C_R^j, C_F^j, and C_M^j are computed by averaging the normalized RFM values of customers in cluster j. Let C_I^j be the integrated rating of cluster j. C_I^j is computed as the weighted sum of C_R^j, C_F^j and C_M^j; that is, $C_I^j = w_R \times C_R^j + w_F \times C_F^j + w_M \times C_M^j$ where w_R, w_F and w_M are the relative importance of the RFM variables, as determined by the AHP. Finally, the CLV ranking of the clusters is derived according to their integrated rating.

Table 2
CLV ranking by weighted sum of normalized RFM values

Cluster	Recency C_R^j	Frequency C_F^j	Monetary C_M^j	Integrated rating C_I^j	CLV ranking
1	0.777	0.0151	0.0228	0.573	3
2	0.856	0.0232	0.0352	0.633	2
3	0.883	0.0413	0.0684	0.658	1
4	0.667	0.0174	0.0174	0.492	4
5	0.204	0.0073	0.0115	0.151	7
6	0.527	0.0093	0.0149	0.388	5
7	0.077	0.0033	0.0081	0.058	8
8	0.301	0.0103	0.0075	0.222	6

$C_I^j = w_R \times C_R^j + w_F \times C_F^j + w_M \times C_M^j$ ($w_R = 0.731$, $w_F = 0.188$, $w_M = 0.081$).

Accordingly, the CLV ranking of WRFM-based clusters is derived according to their integrated rating (listed in Table 2). Customers in a cluster with higher ranking are more loyal. The ranking indicates that cluster three has the highest ranking, followed by cluster two; cluster seven has lowest ranking.

3.2.6. Recommendation phase

The association rule mining approach is employed to extract recommendation rules from each group of customers, derived separately from WRFM-based clustering and preference-based clustering. Let C_{WRFM}^i be the WRFM-based customer group i, generated by clustering customers based on WRFM values. Let C_p^j be the preference-based customer group j, generated by clustering customers based on customers' purchasing preferences. Association rule mining is used to extract the recommendation rule set RS_{WRFM}^i from customer transactions associated with each customers group C_{WRFM}^i. Similarly the recommendation rule set RS_p^j is extracted from customer transactions associated with each customer group C_p^j. Then, the top-N candidate products to be recommended to customer u are selected as follows. The customer group C_{WRFM}^i to which customer u belongs to must be identified before the top-N product items can be recommended to that customer. If the customer belongs to a high-loyalty group, then the recommendation rules (RS_{WRFM}^i), extracted from C_{WRFM}^i, are used to recommend the top-N product items; otherwise, the customer group C_p^j to which customer u belongs must be identified, and then the recommendation rules (RS_p^j), extracted from the C_p^j are used to recommend top-N product items.

Two feasible approaches are available to determine the high/low loyalty ranking of a customer group C_{WRFM}^i. One is to set the threshold *integrated_rating* to α. The *integrated_rating* is the weighted sum of the averaged normalized WRFM values. If the *integrated_rating* of C^i_{WRFM} exceeds α, then the loyalty ranking of C^i_{WRFM} returns high; otherwise, it returns low. The other approach is to observe analytical results on an analytical data set by computing the recommendation quality of the WRFM-based method and the preference-based CF method. For each customer group C^i_{WRFM}, the WRFM-based recommendation quality is derived by using the WRFM-based recommendation rule set (RS^i_{WRFM}), while the preference-based recommendation quality is derived by using the preference-based recommendation rule set (RS^j_p). Notably, customers in the same WRFM-based customer group C^i_{WRFM} may belong to different preference-based customer groups. The recommendation rules (RS^j_p), extracted from customer transactions associated with the customer group C^j_p to which customer u belongs, are used to derive the preference-based recommendation quality. The F1-metric described in Section 4.2 can be used to measure the recommendation quality. If the value of WRFM-based F1-metric exceeds that of preference-based F1-metric, the loyalty of C^i_{WRFM} returns *high*; otherwise, it returns *low*.

3.3. Hybrid2 approach

The hybrid1 method cluster customers based on either the customer lifetime value or the purchase preferences separately. The method then uses the preference-based recommendation rule set to improve the quality of recommendations for less loyal customers. This work proposes another hybrid method that clusters customers by integrating dimensions of customer lifetime value and customer preference, as shown in Fig. 2. The relative weighting is adopted to adjust the importance of customer lifetime value and purchase preferences in clustering.

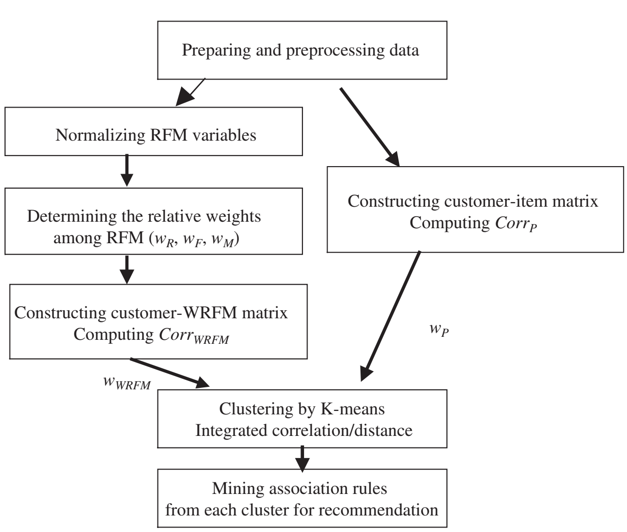

Fig. 2. Hybrid2 method for product recommendation.

The hybrid2 approach initially establishes a customer-WRFM matrix and a customer-item matrix. Then, the WRFM-based and preference-based correlation coefficients are computed using Eqs. (1) and (2), respectively. Subsequently, K-means clustering is used to group customers with similar CLV and preferences based on an weighted correlation coefficients, which is obtained from integrated dimensions of CLV and preferences. Finally, the association rule mining approach is applied to extract recommendation rules from each group derived from K-means clustering.

3.3.1. Grouping customers with similar CLV and purchase preference

Different from the hybrid1 method, the hybrid2 method clusters customers by integrating dimensions of CLV and purchase preferences. The customer-WRFM matrix and customer-item matrix are first established. For customer-WRFM matrix, the RFM values of customers are normalized and then multiplied by the relative importance of RFM variables. For customer-item matrix, an element r_{ij} represents whether the ith customer had purchased the jth product. Eq. (1) is used to compute the preference-based Pearson correlation coefficient, $corr_P$, while Eq. (2) is used to compute the WRFM-based Pearson correlation coefficient, $corr_{WRFM}$. The integrated correlation coefficient is then derived according to Eq. (3).

$$Corr_{integrated(c_i,c_j)} = w_{WRFM} \times corr_{WRFM}(c_i,c_j) + w_P \times corr_P(c_i,c_j) \tag{3}$$

w_{WRFM} and w_P represent the relative importance (weights) of the dimensions of CLV and purchase preferences, respectively. If $w_{WRFM} = 0$, then the method becomes preference-based CF method, that uses purchase preference to group customers. If $w_{WRFM} = 1$, then the method becomes WRFM-based method, that uses weighted RFM values to group customers. K-means technique is employed to cluster customers based on the integrated correlation coefficients. In general, the integrated correlation coefficient between a centroid c_j of a cluster and a customer c_i is measured using Eq. (3). The centroid of a cluster is represented by both the average WRFM values and the average purchase preferences of customers within the cluster. Customers are assigned to a cluster with maximum integrated correlation coefficient.

The weights of parameters w_{WRFM} and w_P are used to yield an integrated correlation coefficient. The proper weighting values of w_{WRFM} and w_P can be determined by performing some preliminary analytical experiments to evaluate the quality of recommendations under different weight combination (for example, w_{WRFM} equals 0.8 and w_P equals 0.2).

3.3.2. Recommendation phase

Association rule is used to extract recommendation rule set RS_j from transactions associated with each cluster. Each cluster is generated by grouping customers based on weighted correlation coefficients of CLV and purchase preferences. The cluster C_j to which a customer u belongs is first identified. Then, RS_j, the recommendation rule set extracted from C_j is used to select the top-N candidate products to be recommended to customer u.

4. Experimental evaluation

4.1. Experimental setup

Various experiments are conducted to evaluate the proposed hybrid1 and hybrid2 methods for product recommendation, using the hardware retailing data set described in Section 3.2.1. The proposed methods are compared with the WRFM-based method, the preference-based CF method, and the KNN-based CF method. The KNN-based CF method uses preferences for product purchases to compute similarities between customers, and then employs the k-nearest neighbor (KNN) approach to derive the top-N recommendations, as illustrated in Section 2.4.

The hardware retailing data were divided into a 75% data set for training and a 25% data set for testing to verify the quality of the recommendations. The training set includes product items purchased by customers in a specified period, and is used to extract recommendation rules from customer transactions. Moreover, a preliminary analytical experiment was conducted to determine the high/low loyalty ranking of customer groups in hybrid1 method and the proper weighting of w_{RFM} and w_P in hybrid2 method. The training set was also used as the analytical data set in the preliminary analytical experiment, where 65% data set was used for deriving recommendation rules and 10% for analyzing recommendation quality. In the experiments, the minimum confidence is set to 0.8, and the minimum support is set to 0.1. Identifying all frequent *itemsets* is difficult, since the average number of product items purchased by customers exceeds 60. Hence, association rule mining explores only to frequent *itemsets* with sizes of less than or equal to three.

4.2. Experimental metrics

Two metrics, precision and recall, are commonly used to measure the quality of recommendation. These two metrics are also extensively used measures in information retrieval (Salton and McGill, 1983; van Rijsbergen, 1979). Recall is the fraction of interesting product items that can be located.

$$Recall = \frac{number\ of\ correctly\ recommended\ items}{number\ of\ interesting\ items} \tag{4}$$

Precision is the fraction of recommended products (predicted to be interesting) that are really interesting to customers.

$$Precision = \frac{number\ of\ correctly\ recommended\ items}{number\ of\ recommended\ items} \tag{5}$$

Items interesting to customer u are those products purchased by u in the test set. Correctly recommended items are recommended items that match interesting items. However, increasing the number of recommended items tends to reduce the precision and increase the recall. An F1-metric can be used to balance the trade-off between precision and recall (van Rijsbergen, 1979). F1-metric assigns equal weight to precision and recall, and is given by

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \tag{6}$$

Notably, each metric is computed for each customer, and the average value computed for each cluster, as well as the overall average (over all customers) is taken as measures of the quality of the recommendation.

4.3. Experimental results

4.3.1. Evaluation of hybrid1 method

This experiment verifies that the proposed hybrid1 method is a feasible way to enhance the quality of recommendations for less loyal customers. Customers with similar CLV and similar preferences were separately grouped into WRFM-based customer groups (CLV groups) and preference-based groups. The hybrid1 method needs to determine the high/low loyalty ranking of customer groups, which is derived by conducting a preliminary analytical experiment, as described in Section 3.2.6. The analytical experiment used the training set as the analytical data set, where 65% data set was used for deriving recommendation rules and 10% for analyzing recommendation quality. The analytical result shows that the preference-based CF method improves the recommendation quality of customer groups with the seventh and eighth loyalty rankings.

To determine whether the hybrid1 method is effective, the training set was used to extract recommendation rules, and the testing set was used to verify the recommendation quality. Based on the analytical result, recommendation rules extracted from the WRFM-based customer groups (CLV groups) were used to recommend products to the top six loyalty ranking CLV groups, while the recommendations rules extracted from the preference-based customer groups wre used to rec-

Table 3
F1-metrics for various methods under top-20

CLV Ranking	WRFM-based method	Preference-based CF	Hybrid1	KNN-based CF ($k = 100$)
	F1-metric	F1-metric	F1-metric	F1-metric
1	0.645	0.636	0.645	0.634
2	0.608	0.586	0.608	0.596
3	0.553	0.545	0.552	0.545
4	0.473	0.471	0.473	0.456
5	0.458	0.453	0.458	0.422
6	0.412	0.419	0.412	0.408
7	0.387	0.415	0.415	0.377
8	0.310	0.335	0.335	0.308
Overall average	0.524	0.518	0.528	0.515

ommend products to the last two loyalty ranking CLV groups. The hybrid1 method is compared with the WRFM-based method, preference-based CF method and the KNN-based CF method. Table 3 lists the average F1-metrics for each cluster, obtained using various methods under $N = 20$ (top-20 recommendations) and $k = 100$ (100 nearest neighbors). The last row shows the overall average F1-metrics. The F1-metrics obtained by the hybrid1 method exceeds those obtained by the WRFM-based method, preference-based CF method, and the KNN-based CF method, implying that the proposed method, hybrid1 provides better recommendations than the WRFM-based method, preference-based CF method and the KNN-based CF method.

4.3.2. Determine proper weightings of hybrid2 method

The hybrid2 method considered different weightings on the dimensions of CLV and purchase preferences. A preliminary analytical experiment was conducted to determine the proper weightings, w_{WRFM} and w_P ($w_P = 1 - w_{WRFM}$). The analytical experiment used the training set as the analytical data set, where 65% data set was used for deriving recommendation rules and 10% for analyzing recommendation quality. Fig. 3 summarizes the recommendation quality (F1-metric) obtained using the hybrid2 method. If w_{WRFM} equals zero, then the hybrid2 method is the preference-based CF method; if w_{WRFM} equals one, then the hybrid2 method is the WRFM-based method. The hybrid2 method achieved the best recommendation quality when w_{WRFM} equals 0.8 and w_P equals 0.2. Overall, when w_{WRFM} exceeds w_P, the F1-metric of hybrid2 method exceeds that obtained using the WRFM-based method or the preference-based CF method alone. Based on the analytical result, further experiments (described in Section 4.3.3) were conducted to evaluate the hybrid2 method by setting $w_{WRFM} = 0.8$ and $w_P = 0.2$.

4.3.3. Comparing various methods on top-N recommendations

Experiments were conducted to compare various methods by using the training set and the testing set. The methods were compared by varying the N, the number of recommendation items. Table 4 shows the F1-metrics of various methods under different top-N recommendations. In general, both the F1-metrics of the hybrid1 and hybrid2 methods exceed those of the WRFM-based method, preference-based CF method, and the KNN-based CF method. The result implies that the proposed hybrid methods provide better recommendations than other methods. Moreover, the hybrid2 method performs better than the hybrid1 method.

4.3.4. Experiments on three clusters of customers

Experiments were also performed on clustering customers into three clusters. Similarly, 75% data set was

Table 4
F1-metrics of various methods under different N (top-N)

top-N	WRFM-based method	Preference-based CF	Hybrid1	Hybrid2	KNN-based CF ($k = 100$)
top-4	0.333	0.335	0.338	0.342	0.286
top-10	0.499	0.476	0.480	0.486	0.487
top-20	0.524	0.518	0.528	0.533	0.515
top-30	0.504	0.502	0.519	0.525	0.498
top-40	0.484	0.496	0.486	0.496	0.467
top-50	0.477	0.473	0.480	0.489	0.422

W_{WRFM}	Precision	Recall	F1-metric
0.00	0.4271	0.6308	0.5001
0.10	0.4365	0.6329	0.5067
0.20	0.4369	0.6361	0.5074
0.30	0.4452	0.6285	0.5117
0.40	0.4485	0.6294	0.5122
0.50	0.4436	0.6372	0.5121
0.60	0.4428	0.6401	0.5122
0.70	0.4413	0.6475	0.5129
0.80	0.4412	0.6564	0.5133
0.90	0.4388	0.6529	0.5131
1.00	0.4357	0.6473	0.5111

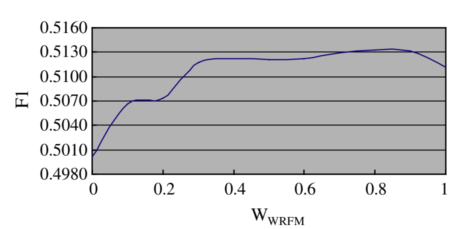

Fig. 3. Analytical result of hybrid2 method under different weightings of w_{WRFM}. (top-20).

Table 5
F1-metrics for three clusters under top-30

CLV ranking	Weighted RFM	Preference-based CF	Hybrid1	KNN-based CF ($k = 100$)
1	0.736	0.713	0.736	0.698
2	0.533	0.529	0.533	0.520
3	0.393	0.416	0.416	0.386
Overall average	0.510	0.508	0.514	0.498

Table 6
Comparisons under various N (top-N); clustering into three clusters

top-N	WRFM-based method	Preference-based CF	Hybrid1 method	Hybrid2	KNN-based CF ($k = 100$)
top-4	0.314	0.311	0.316	0.319	0.286
top-10	0.496	0.484	0.497	0.500	0.487
top-20	0.511	0.504	0.512	0.515	0.515
top-30	0.510	0.508	0.514	0.518	0.498
top-40	0.498	0.476	0.501	0.509	0.467
top-50	0.468	0.470	0.472	0.472	0.422

used as the training set, while 25% data set was used as the testing set. Moreover, the training set was also used as the analytical data set to determine the low/high loyalty ranking of hybrid1 method and the weightings of hybrid2 method. For hybrid1 method, the preference-based CF method can improve the recommendation quality of WRFM-based customer groups with the third loyalty ranking. For hybrid2 method, the best combination of w_{WRFM} and w_P is 0.7 and 0.3, respectively. Tables 5 and 6 present the experimental results on the testing set (25% data set), which exhibit trends similar to those results obtained in the experiments that used eight clusters.

The result also indicates that the proposed hybrid1 and hybrid2 methods provide better recommendations than the WRFM-based method, preference-based CF method and the KNN-based CF method. Additionally, the F1-metric of hybrid2 exceeds that of hybrid1.

5. Conclusions

This work proposed two hybrid recommendation approaches. The hybrid1 method overcomes the drawback of WRFM-based method by using preference-based CF method to improve the quality of recommendation for less loyal customers. Most previous research grouped customers only either by customer lifetime value or by purchasing preferences. The proposed hybrid2 method integrated these two dimensions to group customers and then extracted recommendation rules from each group to improve the quality of recommendation. The experimental results demonstrate that the proposed hybrid1 and hybrid2 methods outperformed the WRFM-based, preference-based CF and the KNN-based CF methods. The hybrid2 method outperformed the hybrid1 method, especially when the CLV was weighted more heavily than purchase preferences.

There are some limitations in our study. First, collaborative filtering may pose the sparsity problem, which refers to a situation in which transactional data is sparse and insufficient to identify similarities in user interests (Sarwar et al., 2000). Our proposed methods adopt collaborative filtering approach, and thus may suffer the sparsity problem. Further work is required to reduce the sparsity problem by considering the customer/product profiles for recommendation. Second, our present work focused on product recommendation of retail transaction data that contains binary choice of shopping basket data; customer preference is represented as one, if the customer purchased the product; and is zero, otherwise. Further investigation is needed to evaluate the effectiveness of our methods on data set with non-binary preference rating. Finally, our proposed methods are verified using one data set from a hardware retailer. More empirical studies should be performed to obtain more tangible conclusions regarding product recommendation. In the future, we plan to evaluate the effectiveness of the proposed approach for other application domains.

Acknowledgment

This research was supported in part by the National Science Council of the Republic of China under the grant NSC 93-2416-H-009-011.

References

Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association between sets of items in large database. In: Proceedings of the ACM-SIGMOD Conference, pp. 207–216.

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. In: Proceedings of the VLDB conference, pp. 407–419.

Berger, P., Nasr, N., 1998. Customer lifetime value: marketing models and applications. Journal of Interactive Marketing 12 (1), 17–30.

Bult, J.R., Wansbeek, T.J., 1995. Optimal selection for direct mail. Marketing Science 14 (4), 378–394.

Changchien, S.W., Lu, Z.C., 2001. Mining association rules procedure to support on-line recommendation by customers and products fragmentation. Expert Systems with Applications 20 (4), 325–335.

Chen, H.C., Chen, A.L.P., 2001. A music recommendation system based on music data grouping and user interests. In: Proceedings of the ACM Conference (Information and Knowledge Management), pp. 231–238.

Chen, M.S., Han, J., Yu, P.S., 1996. Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering 8 (6), 866–883.

Hearst, M., 1998. K-means clustering, UCB SIMS, Fall. Available from <http://www.sims.berkely.edu/courses/is296a-3/f98/lectures/ui-bakground/sld025.htm>.

Hill, W., Stead, L., Rosenstein, M., Furnas, G., 1995. Recommending and evaluating choices in a virtual community of use. In: Proceedings of the ACM (CHI'95), pp. 194–201.

Hughes, A.M., 1994. Strategic Database Marketing. Probus Publishing, Chicago.

Irvin, S., 1994. Using lifetime value analysis for selecting new customers. Credit World 82 (3), 37–40.

Kahan, H., 1998. Using database marketing techniques to enhance your one-to-one marketing initiatives. Journal of Consumer Marketing 15 (5), 491–493.

Lang, K., 1995. Newsweeder: learning to filter Netnews. In: Proceedings of the Machine Learning Conference, pp. 331–339.

Linden, G., Smith, B., York, J., 2003. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing 7 (1), 76–80.

Liu, D.R., Shih, Y.Y., accepted. Integrating AHP and data mining for product recommendation based on customer lifetime value. Information and Management (accepted).

MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability conference, pp. 281–296.

Miglautsch, J., 2000. Thoughts on RFM scoring. Journal of Database Marketing 8 (1), 67–72.

Peppers, D., Rogers, M., 1997. The One to One Future: Building Relationships One Customer at a Time. Bantam Doubleday Dell Publishing.

Punj, G.N., Stewart, D.W., 1983. Cluster analysis in marketing research: review and suggestions for application. Journal of Marketing Research 20, 134–148.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., 1994. GroupLens: an open architecture for collaborative filtering of Netnews. In: Proceedings of the CSCW Conference, pp. 175–186.

Rucker, J., Polanco, M.J., 1997. Personalized navigation for the Web. Communications of the ACM 40 (3), 73–75.

Saaty, T.L., 1994. Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process. RWS Publications, Pittsburgh, PA.

Salton, G., McGill, M.J., 1983. Introduction to Modern Information Retrieval. McGraw-Hill, New York.

Sarwar, B., Karypis, G., Konstan, J., Riedl, J., 2000. Analysis of recommendation algorithms for e-commerce. In: Proceedings of the ACM Conference (Electronic Commerce), pp. 158–167.

Schafer, J.B., Konstan, J.A., Riedl, J., 2001. E-commerce recommendation applications. Journal of Data Mining and Knowledge Discovery 5 (1/2), 115–152.

Shardanand, U., Maes, P., 1995. Social information filtering: algorithms for automating 'world of mouth'. In: Proceedings of the ACM (CHI'95), pp. 210–217.

Srikant, R., Agrawal, R., 1995. Mining generalized association rules. In: Proceedings of the VLDB Conference, pp. 407–419.

Stone, B., 1995. Successful direct marketing methods. NTC Business Books, Lincolnwood, IL.

van Rijsbergen, C.J., 1979. Information Retrieval, Second ed. Butterworths, London.

Yun, H., Ha, D., Hwang, B., Ho Ryu, K., 2003. Mining association rules on significant rare data using relative support. The Journal of Systems and Software 67 (3), 181–191.