ELSEVIER

Integrating AHP and data mining for product recommendation based on customer lifetime value

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

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

Abstract

Product recommendation is a business activity that is critical in attracting customers. Accordingly, improving the quality of a recommendation to fulfill customers' needs is important in fiercely competitive environments. Although various recommender systems have been proposed, few have addressed the lifetime value of a customer to a firm. Generally, customer lifetime value (CLV) is evaluated in terms of recency, frequency, monetary (RFM) variables. However, the relative importance among them varies with the characteristics of the product and industry. We developed a novel product recommendation methodology that combined group decision-making and data mining techniques. The analytic hierarchy process (AHP) was applied to determine the relative weights of RFM variables in evaluating customer lifetime value or loyalty. Clustering techniques were then employed to group customers according to the weighted RFM value. Finally, an association rule mining approach was implemented to provide product recommendations to each customer group. The experimental results demonstrated that the approach outperformed one with equally weighted RFM and a typical collaborative filtering (CF) method.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Recommendation; Marketing; Analytic hierarchy process (AHP); Customer lifetime value; Collaborative filtering; Clustering; Association rule mining

1. Introduction

Intense competition is forcing companies to develop innovative marketing activities to capture customer needs and improve customer satisfaction and retention. The use of the Internet and the explosive growth of e-commerce have expanded marketing activities and made large volumes of customer data available for analysis. Businesses can benefit significantly from analyzing customer data to determine their preferences and thus improve marketing decision support. Providing adequate support to meet customer needs can boost the success of on-line e-stores [18] and web site success depends on enhancing information and service quality to serve customers better [21].

Recently, IT has been utilized to help companies maintain competitive advantage [36]. Data mining techniques [9] are a widely used information technology for extracting marketing knowledge and further supporting marketing decisions [4,5,33]. The applications include market basket analysis, retail sales analysis, and market segmentation analysis. Lin et al. [19] applied data mining techniques to extract inter-organizational retailing knowledge from POS information

* Corresponding author. Tel.: +886-3-5712-121;
fax: +886-3-572-3792.
E-mail address: dliu@iim.nctu.edu.tw (D.-R. Liu).

in retail store chains. Moreover, Hui and Jha [14] employed it to provide customer service support. The knowledge can support marketing decisions and customer relationship management.

The buying patterns of individual customers and groups can be identified via analyzing customer data [38], but also allows a company to develop one-to-one marketing strategies that provide individual marketing decisions for each customer [24]. Recommender systems are technologies that assist businesses to implement such strategies. They have emerged in e-commerce applications to support product recommendation [31]. The systems use customer purchase history to determine preferences and identify products that a customer may wish to purchase. Schafer et al. presented a detailed taxonomy of recommender systems in e-commerce, and determined how they can provide personalization to establish customer loyalty. Generally, recommender systems increase the probability of cross-selling; establish customer loyalty; and fulfill customer needs by discovering products in which they may be interested.

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 customers. The Group-Lens system [26] applied the CF method to recommend Usenet news and movies. Video recommender [12] also used this approach to generate recommendations on movies. Examples of music recommender systems are Ringo [32] and MRS [8]. Siteseer [27] provided recommendations based on the bookmarks of the user's virtual neighbors. Content-based filtering provides recommendations by matching customer profiles (e.g. interests) with content's features (e.g. product attributes). NewsWeeder [17] is an example of content-based recommender systems. Changchien and Lu [7] developed a procedure for mining association rules to support on-line product recommendations. Amazon.com [20] employed item-to-item collaborative filtering to provide recommendations of those products that are similar to the customer's purchased and rated products. However, few have considered customer lifetime value (CLV).

From the perspective of niche marketing, all customers are not equal (they have different lifetime value or purchase behaviors), even if they purchase identical products or services; market segmentation is therefore necessary. Firms are increasingly recognizing the importance of the lifetime value of customers [3]. Several studies have considered the use of CLV. Generally, recency, frequency, and monetary (RFM) methods have been used to measure it [16,23]. The concept has been applied to cluster customers for niche marketing [11].

Our work proposes a novel product recommendation methodology that combines group decision-making and data mining. The analytic hierarchy process (AHP) [28] was applied to evaluate the importance (weight) of each RFM variable, according to a group of decision-makers. Clustering was then employed to group customers based on their weighted RFM value. Finally, association rule mining was used to provide product recommendations for each group of customers.

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 [15]. Measuring RFM is an important method for assessing customer lifetime value. Bult and Wansbeek [6] defined the terms as: (1) R (Recency): period since the last purchase; 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 greater loyalty; (3) M (Monetary): the 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. Goodman [10] suggested that the RFM method avoided focusing on less profitable customers, allowing resources to be diverted to more profitable customers. Hughes [13] proposed a method for RFM scoring that involved using RFM data concerning to sort individuals into five customer groups. Different marketing strategies could then be adopted for different customers. Stone [35] suggested that different weights should be assigned to RFM variables depend-

ing on the characteristics of the industry. In analyzing the value of customers who used credit cards, he suggested placing the highest weighting on the Frequency, followed by the Recency, with the lowest weighting on the Monetary measure. However, he determined the RFM weightings subjectively, without employing a systematic approach.

2.2. Market segmentation

Clustering [25] seeks to maximize variance among groups while minimizing variance within groups. Many clustering algorithms have been developed, including K-means, hierarchical, fuzzy c-means, etc. We used the K-means method to group customers with similar lifetime value according to weighted RFM. K-means clustering [22] is a method commonly used to partition a set of data into groups. This scheme proceeds by selecting m initial cluster centers and then iteratively refining them. (1) Each instance d_i is assigned to its closest cluster center; (2) each cluster center C_j is updated to the mean of its constituent instances. The algorithm has converged when the assignment of instances to clusters no longer changes.

2.3. Association rule mining

Association rule mining, which identifies associations among a set of product items frequently purchased together, is a widespread approach for market basket analysis [1,34]. It attempts to find association rules that satisfied minimum support and minimum confidence requirements. Appendix A provides the formalization of association rule mining. The support of an association rule indicates how frequently that rule applies to the data. Higher support corresponds to a stronger correlation between the product items. The confidence is a measure of the reliability of an association rule. It corresponds to a more significant correlation between product items. The *apriori* algorithm [2] 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. Association rules that meet a user-specified minimum confidence, can be generated from the frequent *itemsets*.

2.4. Association rule based recommendation

Sarwar et al. [30] described the method of association rule-based recommendation as: for each customer, a customer transaction is created to record all the products previously purchased by a 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, are then determined according to the recommendation rules. A detailed illustration is provided in Appendix B.

2.5. Collaborative filtering

A typical collaborative filtering (CF) method employs nearest-neighbor algorithms to recommend products to a target customer u based on the preferences of *neighbors*, that is, those customers having similar preferences to customer u. Preferences generally are defined in terms of customer purchasing behavior or taste (preference rating on products). Appendix C lists one common approach to compute the similarity of preferences among customers based on the Pearson correlation coefficient.

Customers are ranked by their similarity measures in relation to the target customer u. The k most similar (highest ranked) customers are selected as the k nearest neighbors of customer u. 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. Methodology: integrating AHP, clustering and association rule mining

The proposed recommendation methodology primarily utilizes AHP, clustering, and association rule mining techniques, as shown in Fig. 1. The rationale of the proposed approach is that if customers have had similar purachasing behavior or purchases, then they are very likely also to have similar RFM values. However, RFM values could be similar given very
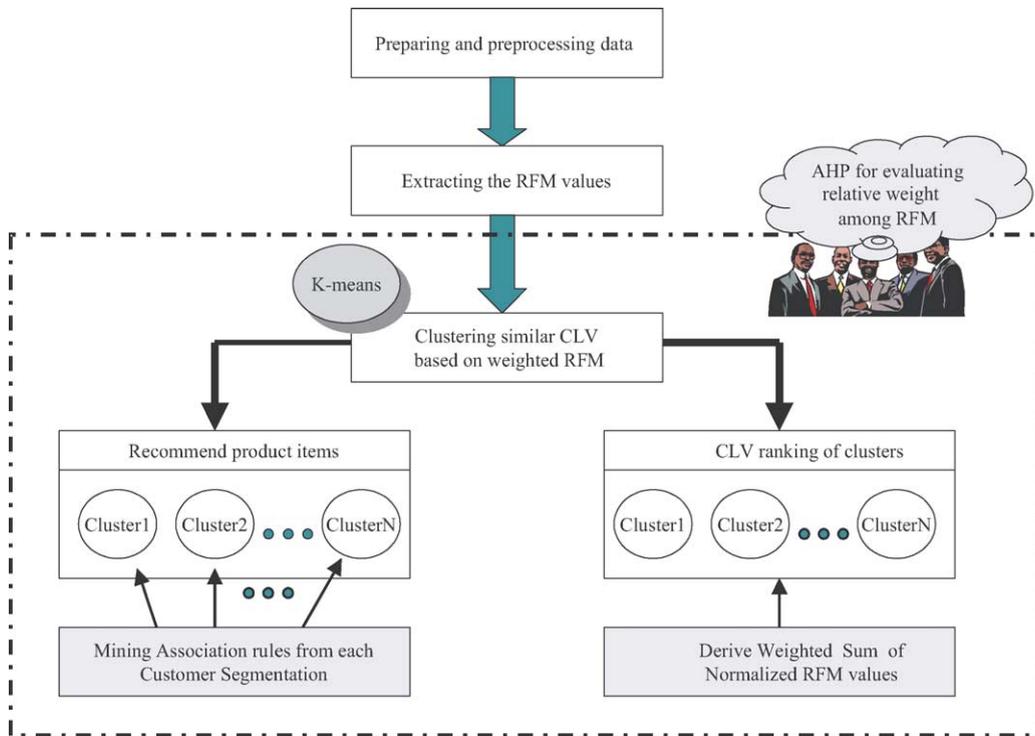
Fig. 1. Recommendation methodology.

different product purchases. Thus, the approach developed here employed two steps to identify similar purchase patterns. First, RFM values were used to cluster customers into groups with similar RFM values: The weighting (relative importance) of each RFM variable was evaluated using AHP. K-means clustering then was employed to group customers with similar lifetime value or loyalty, according to weighted RFM. Second, an association rule mining approach was applied to extract recommendation rules, namely, frequent purchase patterns from each group of customers. The extracted frequent purchase patterns represent the common purchasing behavior of customers with similar product purchases. Therefore, the approach presented in this work recommends products to customers based on frequent purchase patterns of customers with similar product purchases.

A case study was used to illustrate the methodology. It concerns a hardware retailing company that manufactures wheels, casters, platforms, and hand trucks for industrial, medical, hospital and institutional use. This company produces over 3000 products. Its deci-

sion-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 consumer transactions, approximately 70,000 rows, have been collected. The data set was preprocessed to extract customer transactions. Unreasonable records, such as those of customers who have a non-zero purchase but have never made any transactions, were also removed. RFM values of the 895 customers were extracted from the database to measure the customers' CLV.

3.1. AHP approach

The AHP was used to determine the relative importance (weights) of the RFM variables, w_R, w_F, and w_M. The three main steps of the AHP are as follows.

3.1.1. Step1: perform pairwise comparisons
This asks evaluators (decision makers) to make pairwise comparisons of the relative importance of RFM variables using the scale shown in Table 1.

Table 1
Relative degree of importance for pairwise comparisons

Comparative importance	Description	Explanation
1	Equally importance	Two activities contribute equally to the objective
2	Intermediate between equal and weak	Experience and judgment slightly favor one activity over another
3	Weak importance of one over another	Experience and judgment slightly favor one activity over another
4	Intermediate between weak and strong	Experience and judgment strongly favor one activity over another
5	Essential or strong importance	Experience and judgment strongly favor one activity over another
6	Intermediate between strong and demonstrated	An activity is strongly favored and its dominance is demonstrated in practice
7	Demonstrated importance	An activity is strongly favored and its dominance is demonstrated in practice
8	Intermediate between demonstrated and absolute	The evidence favoring one activity over another is of the highest possible order of affirmation
9	Absolute or extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation

3.1.2. Step 2: Assess the consistency of pairwise judgments

Evaluators may make inconsistent judgments when making pairwise comparisons. Before the weights are computed, the degree of inconsistency is measured by an inconsistency index. Perfect consistency implies a zero inconsistency index. However, perfect consistency is seldom achieved, since humans are often biased and inconsistent, when making subjective judgments. Therefore, an inconsistency index of less than 0.1 is acceptable. If the inconsistency index exceeds this, then the pairwise judgments may be revised before the weights of RFM are computed.

3.1.3. Step 3: Computing the relative weights

This determines the weight of each decision element. This work employs Eigenvalue computations to derive the weights of the RFM.

In our study, the three groups of evaluators judge the RFM weightings: three administrative managers, two business managers in sales, and one marketing consultant, and five customers who had previously made at least one purchase. These groups were invited to evaluate the relative importance of the RFM variables. Data were gathered by interviewing the evaluators. Interviews were conducted using a questionnaire (Table 2), and the answers were expressed in the form of a pairwise comparison matrix (Table 3).

According to the assessments, the relative weights of the RFM variables are 0.731, 0.188 and 0.081, respectively. The implication of the RFM weightings is that recency is the most important variable; thus evaluators must mainly concentrate on whether customers purchase regularly. If some perform no transaction for a long period, they may have been lost or transferred to a new vendor.

Table 2
AHP questionnaire for RFM

Criteria	Comparative importance									Criteria
	9:1	7:1	5:1	3:1	1:1	3:1	5:1	7:1	9:1	
Recency	9	7	5	3	1	3	5	7	9	Frequency
Recency	9	7	5	3	1	3	5	7	9	Monetary
Frequency	9	7	5	3	1	3	5	7	9	Monetary

Table 3
Example of RFM pairwise comparison matrix

	Recency	Frequency	Monetary
Recency	1	5	7
Frequency	1/5	1	3
Monetary	1/7	1/3	1

3.2. Clustering customers with similar lifetime value

Customers with similar lifetime values, in terms of weighted RFM, were next clustered using the K-means method. This must specify the number of clusters, m, in advance. The parameter was set to 8, since eight $(2 \times 2 \times 2)$ possible combinations of inputs (RFM) can be obtained by assigning ↓ or ↑, according to the average R (F, M) value of a cluster being less than or greater than the overall average R (F, M). The RFM values of customers were normalized as follows. The profit form, $x' = (x - x^S)/(x^L - x^S)$, was used to normalize the F (frequency) and M (monetary) values, since F and M positively influenced CLV or loyalty. The cost form, $x' = (x^L - x)/(x^L - x^S)$, was used to normalize the R value, since it negatively impacted CLV. x' and x represented the normalized and original R (F, M) values, while x^L and x^S represented the largest and smallest R (F, M) value of all customers. The normalized RFM values of each customer were then multiplied by the relative importance of RFM variable, w_R, w_F and w_M, which were determined by the AHP. The K-means method was then applied to cluster the customers into eight groups, according to the weighted RFM values.

Table 4 presents the result, listing eight clusters, each with the corresponding number of customers and their average R, F and M values. The last row also shows the overall average for all customers. These, for each cluster, were compared with the overall averages. If the average R (F, M) value of a cluster exceeded the overall average R (F, M), then an upward arrow ↑ was included. The last column of Table 4 shows the RFM pattern for each cluster.

Each cluster represents a market-segmentation. Customers in clusters with the pattern R ↓ F ↑ M ↑ are considered to be loyal, purchased recently, purchase frequently, and spend regularly with the firm. They are *gold* customers. Clusters with the pattern R ↓ F ↓ M ↓ 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 ↑ F ↓ M ↓ include those who very rarely visited the site and made very few transactions. They are *valueless* customers, and may only make purchases during sales. Enterprises reduce prices to attract such customers.

Analysis of variance is used to determine whether RFM variables could be used to distinguish the eight clusters (whether statistically significant). The analysis rejected the null hypothesis H_0 because the P-values were significant ($P < 0.05$). The result confirmed that these eight clusters can be significantly distinguished by recency, frequency, and monetary.

Table 4
Eight clusters generated by K-means clustering

Cluster	Number of customers	Recency (days)	Frequency	Monetary (NT dollars)	Type
1	212	79	36	199010	R ↓ F ↓ M ↓
2	150	69	54	306065	R ↓ F ↑ M ↑
3	190	66	95	593861	R ↓ F ↑ M ↑
4	123	92	41	152007	R ↑ F ↓ M ↓
5	47	147	18	100483	R ↑ F ↓ M ↓
6	100	108	23	130096	R ↑ F ↓ M ↓
7	28	162	10	71536	R ↑ F ↓ M ↓
8	45	135	25	67403	R ↑ F ↓ M ↓
Overall average		89	48	270837	

Table 5
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

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

3.3. CLV ranking

The CLV ranking was derived to help develop more effective strategies for retaining customers and thus identify and compare market segments. The ranking of clusters proceeds as follows. The RFM values of each customer were normalized. Table 5 shows 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 were 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 was computed as the weighted sum of C_R^j, C_F^j, and C_M^j, that is, $C_I^j = w_R C_R^j + w_F C_F^j + w_M C_M^j$, where w_R, w_F and w_M are the relative importance of the RFM variables from AHP. Finally, the CLV ranking of the clusters was derived according to their integrated rating. The ranking indicated that cluster three had the highest rank, followed by cluster two. Customers in a cluster with a higher rank are more loyal.

3.4. Recommendation based on association rules

For each customer, a customer-transaction was created to record all the products previously purchased by him or her. The transactions were grouped according to the clusters of customers. Association rule mining was 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, belonged was first identified. Then, RS_j, the recommendation rule set extracted from C_j was used to select the top-N candidate products to be recommended to customer u. Let X_u represent the set of products previously purchased by customer u. For each recommendation rule $X \Rightarrow Y$ in RS_j, if $X \subseteq X_u$ then all products in Y–X_u are the candidate products for recommendation to customer u. All candidate products were sorted and ranked according to the associated confidence of the recommendation rules. The N highest ranked candidate products were selected as the top-N recommended products.

4. Experimental evaluation

4.1. Experimental setup

The proposed method was experimentally compared with three other methods—the non-weighted RFM method, the non-clustering method, and the typical CF method. The non-weighted RFM method does not consider the relative importance of RFM variables. The method initially sets $w_R = w_F = w_M$, and then uses K-means clustering to cluster customers according to the RFM values of customers. Association rule-based recommendation was applied to each cluster to recommend the top-N products. The non-clustering method did not perform clustering before making an association rule-based recommendation. The recommendation rules were extracted by mining association rules from the entire set of customer transactions. The typical CF method uses the preferences on product purchases to compute the similarity between customers, and then employs the k-nearest neighbor (k-NN) approach to derive top-N recommendations.

Various experiments were performed to compare the quality of recommendations made by the proposed

method with those of the other three methods. In comparing the weighted with the non-weighted RFM method, clusters with the same order of CLV ranking were compared.

The hardware retailing data set was divided into a 75% training set and a 25% testing set. The training set included product items purchased by customers in a specified period and was used to extract recommendation rules by association rule mining. The minimum confidence level was set to 0.8 and the minimum support to 0.1. Identifying all frequent *itemsets* was difficult, since the average number of product items purchased by customers exceeded 60. Hence, association rule mining explored only frequent *itemsets* with sizes less than or equal to three. Testing data were used to verify the quality of the recommendations of the various methods.

4.2. Evaluation metrics

Two metrics, precision and recall, are commonly used to measure the quality of a recommendation. These are also used measures in information retrieval [29]. Product items can be classified into products that customers are interested in purchasing, and those that they are not interested in purchasing. A recommendation method may recommend interesting or uninteresting products. The recall-metric indicated the effectiveness of a method for locating interesting products. The precision-metric represented the extent to which the product items recommended by a method really are interesting to customers.

Recall is the fraction of interesting product items that can be located.

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

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

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

Items interesting to customer u were those products purchased by u in the test set. Correctly recommended items were those that match interesting items. However, increasing the number of recommended items tended to reduce the precision and increase the recall. An F1-metric [37] could be used to balance the trade-off between precision and recall. F1 metric assigned equal weight to precision and recall and was given by,

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

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

4.3. Experimental results

4.3.1. Comparing weighted RFM with non-clustering method

The quality of the top-*all* recommendation generated by the weighted RFM method was analyzed for

Table 6
Quality of recommendation by weighted RFM and non-clustering (top-*all*)

CLV ranking	Weighted-RFM			Non-clustering		
	Precision	Recall	F1-metric	Precision	Recall	F1-metric
1	0.433	0.893	0.580	0.431	0.783	0.550
2	0.385	0.878	0.532	0.420	0.710	0.515
3	0.368	0.828	0.491	0.330	0.674	0.437
4	0.321	0.804	0.446	0.272	0.751	0.382
5	0.282	0.847	0.413	0.247	0.623	0.351
6	0.219	0.758	0.324	0.180	0.453	0.248
7	0.192	0.741	0.286	0.145	0.721	0.232
8	0.184	0.674	0.285	0.143	0.625	0.227
Overall average	0.346	0.836	0.476	0.326	0.697	0.430

each cluster. The top-*all* recommendation recommended all candidate products to the customer. Table 6 presented the CLV ranking of clusters and the average performance values—Precision, Recall and F1-metric for each cluster. The average performance value of a cluster was computed over the customers in the cluster. The last row in the table gave the overall average for all customers. For the non-clustering method, clusters generated by the weighted RFM method were used to compute the average performance values of each cluster. The weighted RFM method extracted recommendation rules from customer-transactions in a cluster, while the non-clustering method extracted them from the entire training set. As presented in Table 6, the performance values (precision, recall, and F1-metric) for weighted RFM generally exceeded those for the non-clustering method. The weighted RFM method yields better recommendations.

4.3.2. Comparing weighted RFM with non-weighted RFM method

The top-*all* recommendation quality by the proposed methodology, weighted RFM, was compared with that by the non-weighted RFM. The clusters generated by weighted and non-weighted RFM are different. The two methods were compared using clusters of the same CLV ranking order. Table 7 shows the result. For all clusters, the F1-metrics of weighted RFM exceeded those of non-weighted RFM, except for cluster six. The overall average precision, recall and F1 metrics of weighted RFM exceeded those of non-weighted RFM. Thus the weighted RFM method

outperforms the non-weighted RFM method. For weighted and non-weighted RFM, the relationship between CLV rank and F1-metric was positive. The F1 metrics of more highly ranked clusters generally exceeded those of the lower-ranked clusters; the clusters with a higher CLV rank included more loyal customers. This result implies that the proposed methodology is more effective for more loyal customers. However, those with a lower CLV ranking may not receive improved recommendations.

4.3.3. Effect of CLV ranking and top-N recommendations

Earlier experimental results indicated that, the F1-metrics of clusters were generally positively as compared with the CLV rankings. The quality of recommendation for clusters with a high CLV ranking exceeded that for clusters with a lower CLV ranking. This experiment examined the effect of varying N, the number of recommended items. Fig. 2 compares the F1 metrics of the weighted RFM (WRFM) with non-weighted RFM (non-WRFM) for top-4, top-10, top-30 and top-50 recommended product items. The analytical results indicated that the positive relationship between CLV ranking and recommendation quality may not have applied for small N (top-4 and top-10). This implies that appropriately selecting the number of recommended items is critical in product recommender systems.

Fig. 3 presents the effect of top-N on the quality of recommendation, when the weighted RFM method was used. For clusters with a high CLV rank (1, 2 or 3), the F1 metrics stopped rising at a large N (18–30).

Table 7
Quality of recommendations for weighted RFM and non-weighted RFM (top-*all*)

| CLV ranking	Weighted RFM			Non-weighted RFM		
	Precision	Recall	F1-metric	Precision	Recall	F1-metric
1	0.433	0.893	0.580	0.397	0.912	0.543
2	0.385	0.878	0.532	0.366	0.903	0.519
3	0.368	0.828	0.491	0.351	0.822	0.482
4	0.321	0.804	0.446	0.320	0.802	0.442
5	0.282	0.847	0.413	0.168	0.838	0.257
6	0.219	0.758	0.324	0.216	0.820	0.334
7	0.192	0.741	0.286	0.177	0.734	0.264
8	0.184	0.674	0.285	0.176	0.659	0.273
Overall average	0.346	0.836	0.476	0.317	0.844	0.445
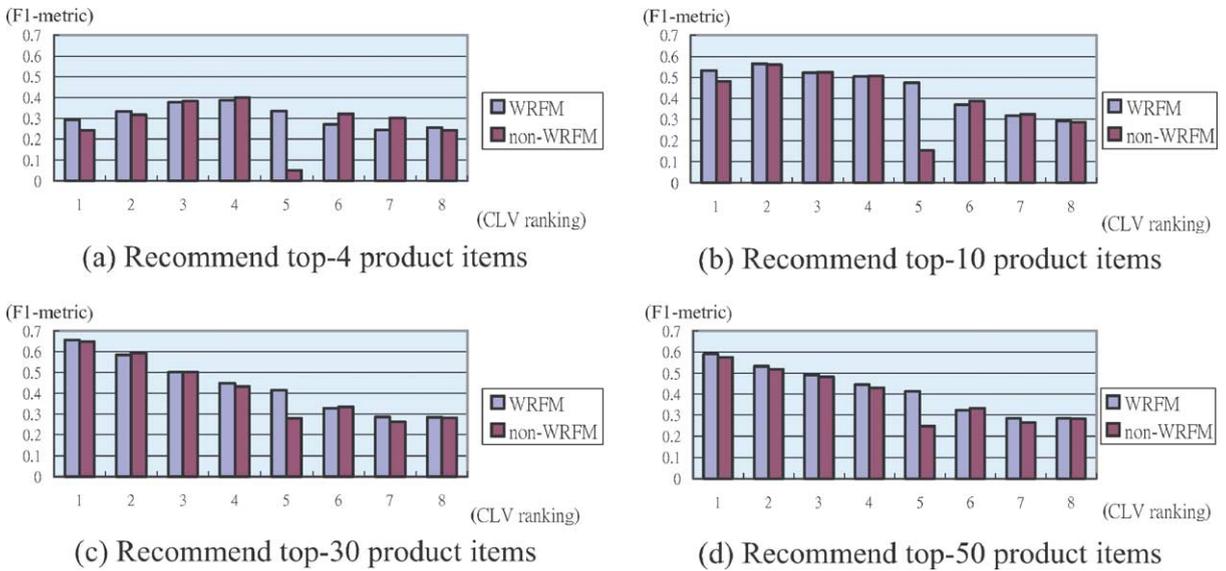
Fig. 2. Comparisons under various top-*N*.

Thus, recommending more items helped to increase the F1 metric and improved the quality of recommendation for clusters with a high CLV rank—for more loyal customers. For clusters with a low CLV rank, such as 6 and 7, the F1 metrics stopped rising at a small *N* (6–14). Thus recommending more product items may not improve the quality of the recommendation for less loyal customers.

4.3.4. Comparing weighted RFM with typical CF method

Experiments were conducted to compare the weighted RFM method with the typical CF method. The typical CF method has been widely used and is a representative recommendation method. The method uses product purchase preferences to compute similarity among customers, and then employs the *k*-nearest neighbor (*k*-NN) approach to derive top-*N* recommendations. Table 8 lists the overall average F1 metrics of weighted RFM and the typical CF method, respectively, for different *k* and *N*. From Table 8, the F1 metrics of weighted RFM exceeded those of the typical CF method. This result indicated that the proposed method provided better recommendations.

An RFM-based *k*-nearest-neighbor method was used to evaluate its effect on recommendation quality. The method resembles the typical CF method that
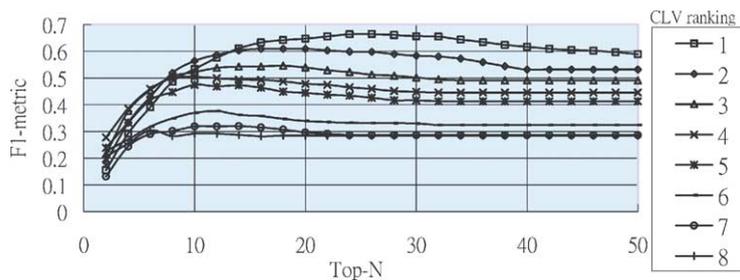

Fig. 3. Effect of top-*N* recommendations vs. CLV rankings (weighted RFM; eight clusters).

Table 8
F1 metrics for weighted RFM and typical CF method

Top-*N*	Weighted RFM	Typical CF method				
		90-NN	100-NN	110-NN	130-NN	150-NN
Top-4	0.333	0.285	0.286	0.291	0.300	0.296
Top-6	0.413	0.376	0.381	0.380	0.386	0.392
Top-10	0.499	0.484	0.487	0.488	0.491	0.491
Top-20	0.524	0.514	0.515	0.517	0.516	0.517
Top-30	0.504	0.497	0.498	0.498	0.501	0.503
Top-40	0.484	0.467	0.467	0.467	0.470	0.470
Top-50	0.477	0.422	0.422	0.422	0.424	0.425

selected k-nearest neighbors to obtain top-N recommendations. However, the RFM-based k-NN method used the weighted RFM values of customers to compute the similarity measures between customers rather than using product purchase preferences. Table 9 lists the experimental result, and shows the F1 metrics of the RFM-based k-NN method and the typical CF method. The RFM-based k-NN method performed better than the typical CF method. The relative importance of RFM variables contributed to improving product recommendation quality.

4.3.5. Experiments on three clusters of customers

Experiments were also performed on placing customers into three clusters. Table 10 and Fig. 4 show the experimental results which exhibited trends similar to those of the experiments using eight clusters. The weighted RFM method outperformed the non-clustering, non-weighted RFM and typical CF methods. The F1 metrics of the more highly ranked clusters exceeded those of the lower-ranked clusters. Furthermore, recommending more items helped to increase the F1 metrics and improve the quality of recommen-

Table 9
F1 metrics for RFM-based k-NN and typical CF method

Top-*N*	Neighbors-90		Neighbors-100		Neighbors-110		Neighbors-130		Neighbors-150	
	RFM-based k-NN	Typical CF	RFM-based k-NN	Typical CF	RFM-based k-NN	Typical CF	RFM-based k-NN	Typical CF	RFM-based k-NN	Typical CF
Top-4	0.303	0.285	0.307	0.286	0.311	0.291	0.305	0.300	0.313	0.296
Top-6	0.393	0.376	0.404	0.381	0.409	0.380	0.410	0.386	0.410	0.392
Top-10	0.491	0.484	0.492	0.487	0.500	0.488	0.495	0.491	0.498	0.491
Top-20	0.520	0.514	0.520	0.515	0.516	0.517	0.520	0.516	0.519	0.517
Top-30	0.500	0.497	0.500	0.498	0.499	0.498	0.503	0.501	0.503	0.503
Top-40	0.470	0.467	0.470	0.467	0.470	0.467	0.470	0.470	0.472	0.470
Top-50	0.423	0.422	0.422	0.422	0.424	0.422	0.425	0.424	0.426	0.425

Table 10
F1 metrics of various methods for three clusters under top-30 and 110 nearest neighbors

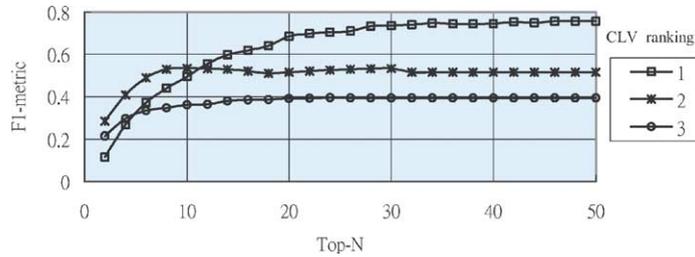| CLV ranking	Weighted RFM	Non-clustering	Non-weighted RFM	Typical CF method
1	0.736	0.617	0.663	0.698
2	0.533	0.469	0.492	0.520
3	0.393	0.363	0.355	0.386
Overall average	0.510	0.451	0.469	0.498

Fig. 4. Effect of top-*N* recommendation vs. CLV ranking (weighted RFM; three clusters).

dation for clusters with a high CLV ranking. However, recommending more product items may not improve the quality of recommendation for customers of lower loyalty.

5. Conclusions

Our work involved the introduction of a novel recommendation methodology that combines AHP, clustering, and association rule-based methods. It clusters customers into segments according to their lifetime value expressed in terms of weighted RFM. Applying AHP to determine the relative importance of RFM variables proved important, since the RFM weights vary with the characteristics of product and industry. Moreover, clustering customers into different groups not only improves the quality of recommendation but also helps decision-makers identify market segments more clearly and thus develop more effective strategies. The experimental results show that the proposed methodology indeed can yield recommendations of higher quality. However, the methodology is not effective for all customer groups. It is more effective for more loyal customers. Recommending more items helps to improve the quality of recommendation for more loyal customers, but may not do so for less loyal customers.

Acknowledgements

Appendix A. Formalization of association rule mining

Agrawal et al. formalized the problem of finding association rules. 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 the consequent (head) of the rule. 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.

Appendix B. Association rule based recommendation

Let X_u be the set of products previously purchased by customer u. First find all the recommendation rules $X \Rightarrow Y$, for which $X \subseteq X_u$; i.e. customer u purchased all the products in X. 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. If the candidate product is associated with multiple rules, then the highest confidence is used. Let P_u be the set of such candidate products. The candidate products in P_u are sorted by associated confidence value. Candidate products with higher confidence are ranked higher, and the N highest ranked candidate products are selected as the recommendation set.

Appendix C. Computing pearson correlation coefficient

Customer purchase history is 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 then be measured by computing the Pearson correlation coefficient defined as:

$$\mathrm{corr}(c_i, c_j) = \frac{\sum_{s \in I}(r_{c_i,s} - \overline{r}_{c_i})(r_{c_j,s} - \overline{r}_{c_j})}{\sqrt{\sum_{s \in I}(r_{c_i,s} - \overline{r}_{c_i})^2 \sum_{s \in I}(r_{c_j,s} - \overline{r}_{c_j})^2}}$$

The notations $\overline{r}_{c_i}$ and $\overline{r}_{c_j}$ 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.

References

[1] R. Agrawal, T. Imielinski, A. Swami, Mining association between sets of items in large database, in: Proceedings of the ACM-SIGMOD International Conference on Management of Data, Washington, DC, USA, 1993, pp. 207–216.

[2] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on very large Data Bases, Santiago, Chile, 1994, pp. 407–419.

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

[4] I. Bose, R.K. Mahapatra, Business data mining-a machine learning perspective, Information and Management 39 (3), 2001, pp. 211–225.

[5] R.J. Brachman, T. Khabaza, W. Kloesgen, G. Piatetsky-Shapiro, E. Simoudis, Mining business databases, Communications of the ACM 39 (11), 1996, pp. 42–48.

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

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

[8] H.C. Chen, A.L.P. Chen, A music recommendation system based on music data grouping and user interests, in: Proceedings of the ACM International Conference on Information and Knowledge Management, Atlanta, GA, November 2001, pp. 231–238.

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

[10] J. Goodman, Leveraging the customer database to your competitive advantage, Direct Marketing 55 (8), 1992, pp. 26–27.

[11] S.H. Ha, S.C. Park, Application of data mining tools to hotel data mart on the Intranet for database marketing, Expert Systems with Applications 15 (1), 1998, pp. 1–31.

[12] W. Hill, L. Stead, M. Rosenstein, G. Furnas, Recommending and evaluating choices in a virtual community of use, in: Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, Denver, CO, 1995, pp. 194–201.

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

[14] S.C. Hui, G. Jha, Data mining for customer service support, Information and Management 38 (1), 2000, pp. 1–13.

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

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

[17] K. Lang, NewsWeeder: Learning to Filter Netnews, in: Proceedings of the 12th International Conference on Machine Learning, Tahoe City, CA, 1995, pp. 331–339.

[18] T.P. Liang, H.J. Lai, Effect of store design on consumer purchases: an empirical study of on-line bookstores, Information and Management 39 (6), 2002, pp. 431–444.

[19] Q.Y. Lin, Y.L. Chen, J.S. Chen, Y.C. Chen, Mining inter-organizational retailing knowledge for an alliance formed by competitive firms, Information and Management 40 (5), 2003, pp. 431–442.

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

[21] C. Liu, K.P. Arnett, Exploring the factors associated with Web site success in the context of electronic commerce, Information and Management 38 (1), 2000, pp. 23–33.

[22] J.B. MacQueen, Some methods for classification and analysis of multivariate observations, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, 1967, pp. 281–296.

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

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

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

[26] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, GroupLens: an open architecture for collaborative filtering of Netnews, in: Proceedings of the CSCW'94, Chapel Hill, NC, 1994, pp. 175–186.

[27] J. Rucker, M.J. Polanco, Siteseer: personalized navigation for the Web, Communications of the ACM 40 (3), 1997, pp. 73–75.

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

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

[30] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Analysis of recommendation algorithms for e-commerce, in: Proceedings of the Second ACM Conference on Electronic Commerce, Minneapolis, October 2000, pp. 158–167.

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

[32] U. Shardanand, P. Maes, Social information filtering: algorithms for automating 'world of mouth', in: Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, Denver, CO, 1995, pp. 210–217.

[33] M.J. Shaw, C. Subramaniam, G.W. Tan, M.E. Welge, Knowledge management and data mining for marketing, Decision Support Systems 31 (1), 2001, pp. 127–137.

[34] R. Srikant, R. Agrawal, Mining generalized association rules, in: Proceedings of the 21st International Conference on Very Large Data Bases, Zurich, Switzerland, 1995, pp. 407–419.

[35] B. Stone, Successful Direct Marketing Methods, Lincolnwood, NTC Business Books, IL, 1995.

[36] R.W. Stone, D.J. Good, The assimilation of computer-aided marketing activities, Information and Management 38 (7), 2001, pp. 437–447.

[37] C.J. Van Rijsbergen, Information Retrieval, second ed., Butterworths, London, 1979.

[38] J.D. Wells, W.L. Fuerst, J. Choobineh, Managing information technology (IT) for one-to-one customer interaction, Information and Management 35 (1), 1999, pp. 53–62.

Duen-Ren Liu is a professor of the Institute of Information Management, National Chiao Tung University, Taiwan. He received the BS and MS degrees in Computer Science and Information Engineering from the National Taiwan University, Taiwan, in 1985 and 1987, respectively. He received the PhD degree in Computer Science from the University of Minnesota in 1995. His research interests include information systems, electronic commerce, workflow systems, and knowledge management. Dr. Liu is an associate member of the IEEE, and a member of the ACM.

Ya-Yueh Shih is a PhD student of the Institute of Information Management, National Chiao Tung University. She is currently an instructor of the Department of Information Management, Ming Hsin University of Science and Technology. She received her BS and MS degrees in Department of Information Management form the National Yunlin University of Science and Technology, Taiwan, in 1996 and 1998, respectively. Her research interests include data mining, consumer behavior, and electronic commerce.