Chapter 3. Integrating AHP, clustering and association rule mining

This chapter presents a novel product recommendation methodology that combines group decision-making and data mining. This method has presented in Liu and Shih (2005). How to make recommendations to customers is discussed in following sections.

3.1 WRFM-based method

The proposed recommendation methodology primarily utilizes AHP, clustering, and association rule mining techniques, as shown in Figure 2.

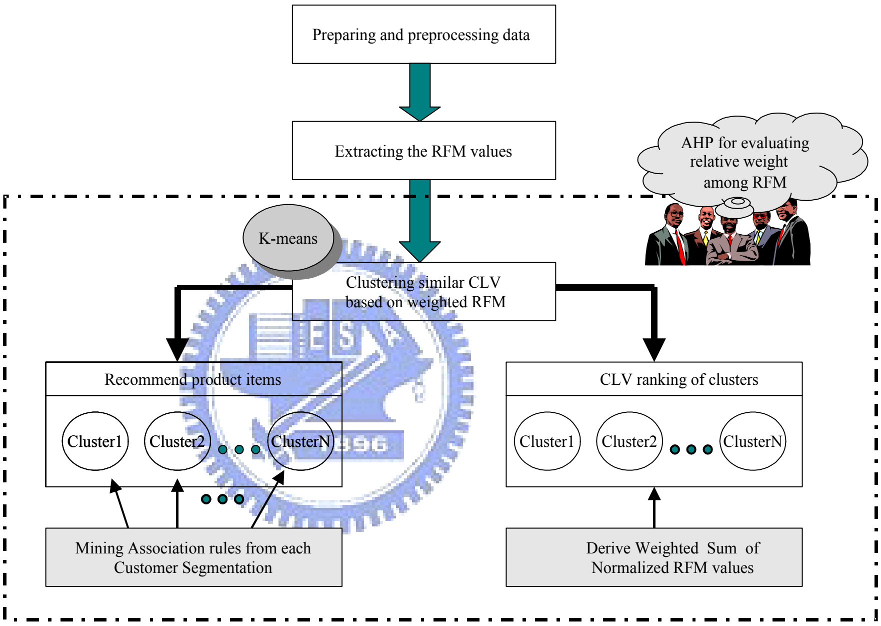


Figure 2. Recommendation methodology of WRFM-based method

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 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 (Saaty, 1980; 1994). K-means clustering then was employed to group customers with similar lifetime value or loyalty, according to weighted RFM. The similarity among customers based on the weighted RFM values is presented in Chapter 3.1.1. 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 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. 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 is 34. 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. Table 3 shows some CLV expresses in terms of RFM.

Table 3. RFM values for each customer

Customer no	Recency (days)	Frequency	Monetary (NT. Dollars)
1260003	159	87	313763
1260006	135	44	146444
1260009	111	379	1426665
…	…	…	…
1300050	256	1	7700

3.1.1 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.

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

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.

3.1.2 Determine the relative weighting

The AHP (refer to Chapter 2.6) 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.

Step1: Perform pairwise comparisons. Evaluators (decision makers) are invited to make pairwise comparisons of the relative importance of RFM variables. In this dissertation, 9-point scale is used to evaluate the pairwise comparisons (see Table 2). There are 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 4), and the answers were expressed in the form of a pairwise comparison matrix (Table 5).

Step 2: Assess the consistency of pairwise judgments. In this work, the consistency ratio is less than 0.1. According to this, evaluators make consistence judgments when making pairwise comparisons.

Step3: Computing the relative weights. This work employs *Eigenvalue* computations to derive the weights of the RFM.

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 4. AHP questionnaire for RFM

	Comparative importance									
Criteria	*9:1*	*7:1*	*5:1*	*3:1*	*1:1*	*3:1*	*5:1*	*7:1*	*9:1*	*Criteria*
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 5. Example of RFM pairwise comparison matrix

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

3.1.3 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.

This must specify the number of clusters, m, in advance. The parameter was set to 8, since eight (2x2x2) 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 6 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 6 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 6. 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	

3.1.4 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 7 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, $C_I^j = w_R \times C_R^j + w_F \times C_F^j + w_M \times C_M^j$, where w_R, w_F and w_M are the relative importance of the RFM variables 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 ranking are with higher CLV.

Table 7. 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

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

3.1.5 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.

3.2 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 is 34. 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.

3.3 Experimental results

3.3.1 Comparing weighted RFM with non-clustering method

The quality of the top-*all* recommendation generated by the weighted RFM method was analyzed for each cluster. The top-*all* recommendation recommended all candidate products to the customer. Table 8 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 8, the performance values (precision, recall, and F1-metric; referred to Chapter 2.5) for weighted RFM generally exceeded those for the non-clustering method. This implies that the weighted RFM method yields better recommendations than non-clustering method.

Table 8. Quality of recommendation by weighted RFM and non-clustering (top-*all*)

Clusters	Weighted-RFM			Non-clustering		
CLV ranking	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

3.3.2 Comparing weighted RFM with non-weighted RFM method

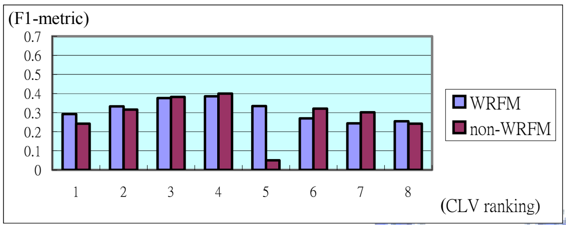
The top-*all* recommendation quality by the proposed methodology, weighted RFM, was compared with that by the non-weighted RFM. The clusters generated by weighted and non-weighted RFM are different. The two methods were compared using clusters of the same CLV ranking order. Table 9 shows the result. 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 ranking 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 ranking 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.

Table 9. Quality of recommendations by weighted RFM and non-weighted RFM (top-*all*)

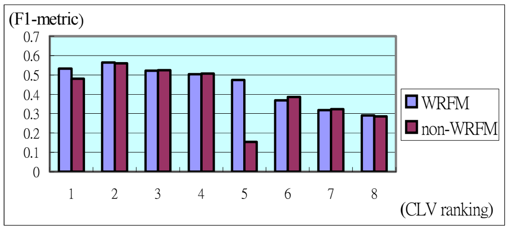
Clusters	Weighted RFM			Non- Weighted-RFM		
CLV ranking	Precision	Recall	F1-metric	Precision	Recall	F1-metric
1	0.433	0.893	0.580	0.397	0.912	0.543
2	0.385	0.878	0.532	0.366	0.903	0.519
3	0.368	0.828	0.491	0.351	0.822	0.482
4	0.321	0.804	0.446	0.320	0.802	0.442
5	0.282	0.847	0.413	0.168	0.838	0.257
6	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

3.3.3 Effect of CLV ranking and top-*N* recommendations

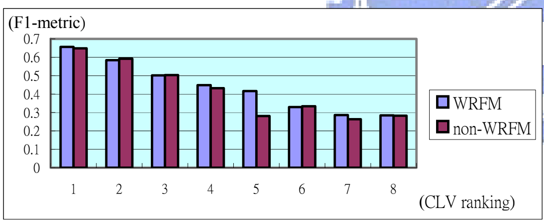
Earlier experimental results indicated that, the F1-metrics of clusters were generally positively as compared with the CLV rankings. The quality of recommendation for clusters with a high CLV ranking exceeded that for clusters with a lower CLV ranking. This experiment examined the effect of varying *N*, the number of recommended items. Figure 3 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.



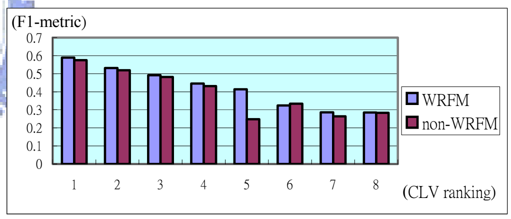
(a) Recommend top-4 product items



(b) Recommend top-10 product items



(c) Recommend top-30 product items



(d) Recommend top-50 product items

Figure 3. Comparisons under various top-*N*

Figure 4 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). Thus, recommending more items helped to increase the F1 metric and improved the quality of recommendation for clusters with a high CLV ranking and for more loyal customers. For clusters with a low CLV ranking, 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.

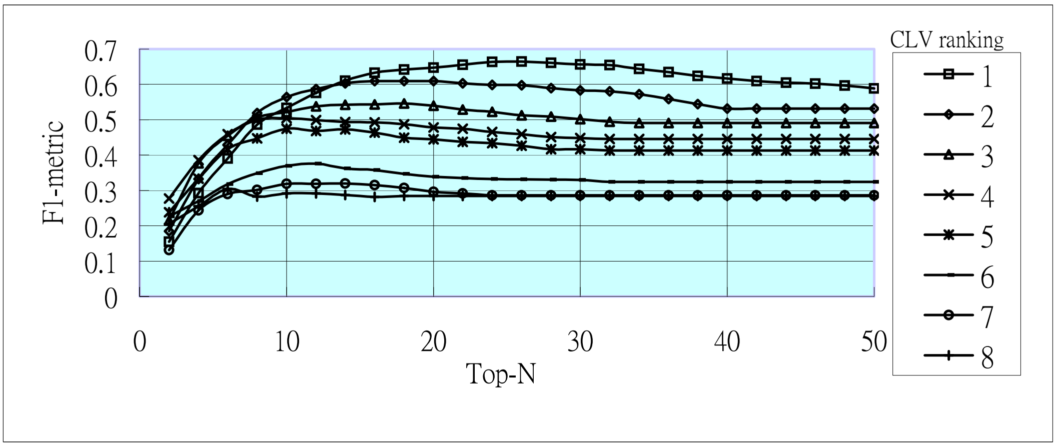


Figure 4. Effect of top-N recommendations vs. CLV ranking (weighted RFM; eight clusters)

3.3.4 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 10 lists the overall average F1 metrics of weighted RFM and the typical CF method, respectively, for different k and N. From Table 10, the F1 metrics of weighted RFM exceeded those of the typical CF method. This result indicated that the proposed method provided better recommendations.

Table 10. F1 metrics of 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

An RFM-based *k*-nearest-neighbor method was used to evaluate its effect on recommendation quality. The method resembles the typical CF method that 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 11 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.

Table 11. F1 metrics of 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

3.3.5 Experiments on three clusters of customers

Experiments were also performed on placing customers into three clusters. Table 12 and Figure 5 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 recommendation for clusters with a high CLV ranking. However, recommending more product items may not improve the quality of recommendation for customers of lower loyalty.

Table 12. F1 metrics of WRFM-based, non-clustering, non-weighted RFM and typical CF methods for three clusters under top-30 and 110 nearest neighbors

CLV ranking	WRFM-based	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

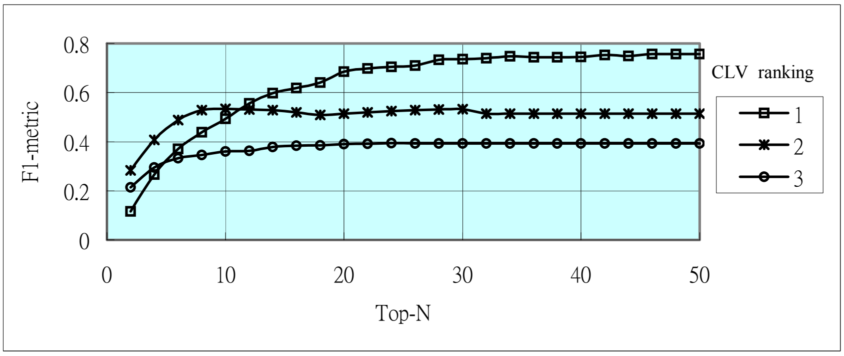


Figure 5. Effect of top-*N* recommendations vs. CLV ranking (weighted RFM; three clusters)

3.4 Discussions

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.