Journal of the Korean Data & Information Science Society
2015, **26**(1), 229–241

http://dx.doi.org/10.7465/jkdi.2015.26.1.229
한국데이터정보과학회지

VIP-targeted CRM strategies in an open market

Hanjun Lee[1] · Beomsoo Shim[2] · Yongmoo Suh[3]

[13]Business School, Korea University
[2]National Information Society Agency

Received 3 October 2014, revised 22 November 2014, accepted 9 December 2014

Abstract

Nowadays, an open-market which provides sellers and consumers a cyber place for making a transaction over the Internet has emerged as a prevalent sales channel because of convenience and relatively low price it provides. However, there are few studies about CRM strategies based on VIP consumers for an open-market even though understanding VIP consumers' behaviors in open-markets is important to increase its revenue. Therefore, we propose CRM strategies targeted on VIP customers, obtained by analyzing the transaction data of VIP customers from an open-market using data mining techniques. To that end, we first defined the VIP customers in terms of recency, frequency and monetary (RFM) values. Then, we used data mining techniques to develop a model which best classifies and identifies influential factors customers into VIPs or non-VIPs. We also validate each of promotion types in the aspect of effectiveness and identify association rules among the types. Then, based on the findings from these experiments, we propose strategies from the perspectives of CRM dimensions for the open-market to thrive.

Keywords: Association rule, classification, CRM, data mining, open market, promotion mix, RFM, VIP.

1. Introduction

The Pareto principle, commonly called as the 80-20 rule, refers that, for many events, roughly 80% of the effects or output come from 20% of the causes or input (Bunkley *et al.*, 2008). In business field, the principle might be translated as "80% of the sales come from 20% of the customers". Considering the limitation of resource, it is natural for a firm to put more weight on the 20% of the customers, who bring much more profit, rather than on the 80% of the customers. In this context, we can think that sellers in open-markets also might as well take the 20%-targeted approach to their marketing.

There are diverse types of promotion applied in open-markets such as coupon, free delivery, mileage, etc. In terms of promotion effect, however, there seems not much consideration on its value to VIP customers (the 20% customers). As discussed above, which type of promotion would be more attractive to VIP customers is important. Also diversified promotion types

[1] Ph. D. candidate, Business School, Korea University, Seoul 136-701, Korea.
[2] Corresponding author: Senior researcher, National Information Society Agency, Mugyodong 77, Cheonggyecheonro, Junggu, Seoul 100-775, Korea. Email: bumsuya@korea.ac.kr
[3] Professor, Business School, Korea University, Seoul 136-701, Korea.

give us a question that what specific mix of promotion types, which called as "promotion mix" in this article, would precipitate VIP customers' purchase. We conjecture that better understanding on these would lead us to better strategies for VIP-targeted marketing.

Thus, the objective of our study is to propose VIP-targeted Customer Relationship Management (CRM) strategies for open market owners. So, we raised our research questions as follows: 1) who are VIP customers and what are their important characteristics? Then, what are the factors deciding the VIP customers?; 2) what kind of promotions is significantly important to VIP customers?; 3) what kind of association rules among promotions affects VIP customers most?; 4) what CRM strategies can be derived based on the answers to these questions?

To answer these questions, we first, defined VIP customers based on recency, frequency, and monetary values and developed a classification model which classifies VIP customers and non-VIP customers, using data mining methods. We selected popular classification models such as decision tree (DT) which is a decision support tool that uses a tree-like graph, logistic regression (LR) which is a type of probabilistic statistical classification model that is used to predict a binary response, and artificial neural network (ANN) which is generally presented as systems of interconnected "neurons" that can compute values from inputs. Each of several promotion types was validated regarding whether it is effective to VIP customers' purchase or not, using chi-square test and logistic regression. Then we identified association rules among promotions from the transaction data of VIP customers. Finally we extracted CRM strategies based on the experimental results.

The rest of the article is organized as follows. In section 2, we introduce the existing literature relevant to our study. Section 3 covers explanation about dataset, research framework and the experiments conducted in this study. Research results are presented in section 4, together with CRM strategies suggested based on our findings. Finally, we conclude in section 5.

2. Literature review

In this section, we review first researches on CRM using various data mining techniques and then researches based on RFM model.

2.1. Researches on CRM using data mining techniques

Since 1980s, the concept of CRM (customer relationship management) has become more and more popular, especially in marketing domain. Though it is really difficult to make a universally approved definition of CRM, it can be explained as a comprehensive strategy for acquiring, retaining and partnering with selective customers to create value for both a company and its customers. Many previous CRM-related researches have applied data mining techniques to analyse and understand customer behaviours and characteristics, and most of them have worked well (Bortiz *et al.*, 1995; Fletcher *et al.*, 1993; Langley *et al.*, 1995; Lau *et al.*, 2003; Salchenberger *et al.*, 1992; Su *et al.*, 2002; Tam *et al.*, 1992; Zhang *et al.*, 1999). In this section, we review previous researches mainly on classification and association rules for a variety of tasks in CRM domain.

Classification works have been performed for various purposes in CRM domain. For example, Kim *et al.* (2006) used decision tree for classifying customers and developed strategies based on customer lifetime value. Hwang*et al.* (2004) adopted logistic regression for segmenting customers based on their customer loyalty. Yu *et al.* (2005) identified interesting visitors analyzing log data. Dennis *et al.* (2001) developed customer knowledge management framework using K-means. Kim *et al.* (2004) showed a system for customer targeting based on ANN and genetic algorithm. Kim (2006) applied logistic regression and ANN to feature selection for predicting churn. Baesens (2004) identified the slope of the customer lifecycle using Bayesian network classifier.

One of the other data mining techniques known to be useful for customer analysis is association rule. Researches on utilizing association rules have been performed for various purposes in CRM domain. For instance, Adomavicius *et al.* (2001) examined association rules for one-to-one marketing. Aggarval *et al.* (2002) and Kubat *et al.* (2003) identified association rules from market basket analysis. Changchien *et al.* (2004) conducted research using both ANN and association rules for on-line personalized sales promotion.

Tables 2.1 and 2.2 provide a summary of previous researches which performed classification and association rules for various purposes in CRM domain, respectively.

Table 2.1 Previous classification and/or prediction researches in CRM domain

Specific Task	Data Mining Techniques	Reference
Customer segmentation and strategy development based on customer lifetime value	Decision tree	Kim *et al.* (2006)
Customer segmentation based on customer value	Logistic regression	Hwang *et al.* (2004)
Identification of interesting visitors through web log analysis	Decision tree	Yu *et al.* (2005)
Suggestion of customer knowledge management framework	K-means	Dennis *et al.* (2001)
Suggestion of a system for customer targeting	ANN and genetic algorithm	Kim *et al.* (2004)
Feature selection for predicting churn	Logistic regression and ANN	Kim *et al.* (2006)
Identification of the slope of the customer-lifecycle	Bayesian network classifier	Baesens *et al.* (2004)

Table 2.2 Previous association rule researches in CRM domain

Specific Task	Data Mining Techniques	Reference
VIP transaction analysis	Association rule	Shim *et al.* (2012)
One to one marketing	Association rule	Adomavicius *et al.* (2001)
Market basket analysis	Association rule	Aggarval *et al.* (2002); Kubat *et al.* (2003)
On-line personalized sales promotion	ANN and association rule	Changchien *et al.* (2004)

2.2. Researches based on RFM model

RFM model is introduced by Hughes (1994) on his article about database marketing. The model is widely used to analyze customer behavior and to define market segments. RFM stands for three variables, Recency, Frequency, and Monetary value. These variables are defined as follows.

· Recency: The time interval between the last purchase and current

· Frequency: The number of purchases over a specific period of time

· Monetary value: The amount of cost on purchase

One of the biggest advantages of the RFM model is its simplicity: no specialized statistical software is required and the results are easily understood by business people. Newell (1997) also announced the effectiveness of the RFM model for customer segmentation. Thus, RFM model-based data mining researches have often been conducted in CRM domain. For example, Hosseini *et al.* (2010) adopted K-means algorithm to classify the customer loyalty based on RFM values. Cheng *et al.* (2009) used K-means and rough set theory to segment customer value based on RFM values. Shim *et al.* (2012) used ANN, decision tree, and logistic regression to classify customers into VIP and non-VIP. Khajv and Tarokhb (2011) applied K-means to customer segmentation based on customer lifetime value.

Table 2.3 shows a summary of previous data mining researches based on RFM values in CRM domain.

Table 2.3 Previous data mining researches based on RFM values

Specific Task	Data Mining Techniques	Reference
Classification of customer loyalty	K-means	Hosseini *et al.* (2010)
Segmentation of customer value	K-means and rough set theory	Cheng *et al.* (2009)
Classification of VIPs	ANN, Decision tree, Logistic regression	Shim *et al.* (2012)
Customer segmentation based on customer lifetime value	K-means	Kim *et al.* (2006)

3. Experiments

This section describes the experiments we conducted to get answers to our research questions mentioned earlier. First, we introduce the dataset obtained from our target open market. Then, we explain research framework in detail. We use Weka 3.6.6 for VIP classification, R 2.12 and SPSS 12.0 for promotion validation, and SAS Enterprise Miner 9.1 for the identification of association rules among promotions.

3.1. Dataset

3.1.1. Target open market

Open market is a place where sellers or consumers can freely compare various products and make transactions on online channel. It has flourished due to its great merits such as various opportunities for consumers to compare products and purchase them with ease at a reasonable price. The most typical open market is e-Bay in USA. In this paper, we analysed transaction data from a top-ranked open market in Korea, G-market. The open market deals with various items from cheap commodities to luxury electronic products. Its amount of revenue is approximately more than 6,000,000,000 dollars per year. The size and revenue have grown increasingly since its opening in 2000. However, the marketing team of the open market has little knowledge about their loyal VIP customers and the kinds of promotion strategies that would be effective when applied to VIPs. Such knowledge related to CRM is believed to allow the target open market to grow continuously.

3.1.2. Data description

The original dataset consists of 24,925 transactions in which 1,000 customers and 6,230 sellers (open market owners) involved. The dataset was collected from August, 2010 to

August, 2011 and contains data on customers, sellers, products and transactions in separate tables. In customer table, there are customer's sex, age, ID, region, registration date, total number of purchase, and the total amount of purchase information. Seller table contains information such as sellers'ID, registration date, type, level, region, the total amount of sales, and purchase satisfaction by each customer. Product table includes information such as product ID, category, price, option price, market price, coupon, shipping fee, product satisfaction, and number of comments. Lastly, information such as transaction ID, seller ID, customer ID, purchase date, purchase options, quantity, mileage and pay method (credit card or cash) exist in transaction table.

It seems to be necessary to explain some of the attributes of the tables. Option is a kind of promotions which lead customers to make additional payment by providing diverse optional goods related to the original product. For example, when buying bicycle, optional goods such as helmet, lock and other accessories are suggested as options. Even essential goods such as CPU, HDD and memory are provided as optional goods when purchasing a personal computer. In our target market, the latter case is much more common. Option is a means for sellers to exhibit a product as if it is sold lower than in other competitors by showing lower initial price for a basic product. Market price represents the product's price in an off-line market. It promotes customers' purchase by giving information about how much they can save when purchasing it in an open market. A coupon which is a well-known type of promotion can be regarded as a discount ticket. Shipping fee is a fee charged for delivery and some sellers provide free delivery service as a promotion.

3.2. Research framework

Figure 3.1 depicts the framework of our research. After integrating the original four tables, we pre-processed the data, identified VIPs, built models which classify VIPs and non-VIPs, and identified association rules among promotions from VIP customers' transaction data. After that, we derived CRM strategies by examining classification models and the identified association rules. Details of each step are described below. Note that all the experimental results are discussed in section 4.

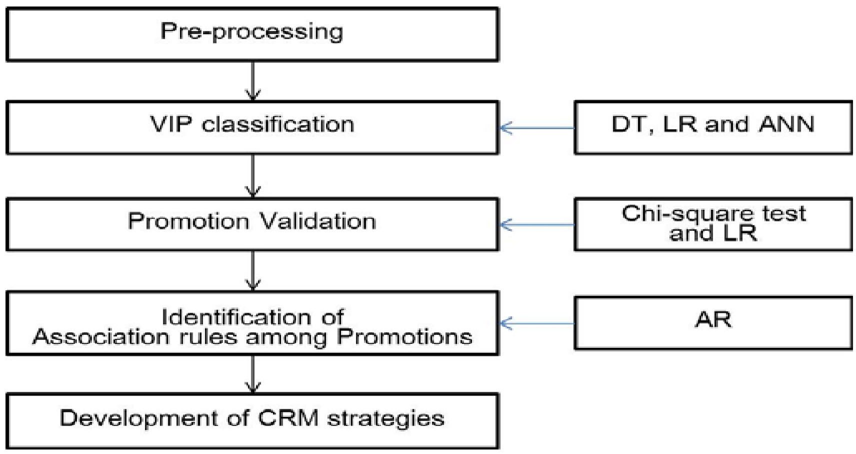

Figure 3.1 Framework for experiments

3.2.1. Pre-processing

We first joined using MS SQL Server 2008, four tables (e.g., customer, seller, product, and transaction tables). The result table consists of 34 attributes including transaction ID as a primary key. We scan the whole records and delete totally 1,262 transaction records which have missing values in many columns. Finally, 23,663 transaction records from 991 customers were selected for our experiments.

For the VIP classification, two experts those who have expertise in CRM were involved in the attribute selection process. They recommended 12 out of 34 attributes which are expected to be related to VIP customer classification. They are customer's purchase related information such as quantity, satisfaction over seller and product, number of comments, shipping fee, option price, coupon used and/or given, mileage given, and pay method as well as demographic information such as age and sex.

Then, we grouped the selected transaction data by customer ID to get the representative value for each of customer's characteristics. Most of attributes are calculated as explained in Table 3.1 Sample records are shown in Table 3.2.

Table 3.1 Attributes for VIP classification

Attribute	Meaning	Data type
Age	Customer's age	Integer
Sex	Customer's sex	Boolean (M/F)
Quantity	Average quantity of products on each purchase	Real
Seller_satisfaction	Average of customer's evaluations on sellers	Real (0 to 5)
Product_satisfaction	Average of customer's evaluation on products	Real (0 to 5)
Number_of_comments	Average of the number of posting comments on products	Real
Shipping_fee	Rate of paying shipping fee (ex.0.25meansonepaymentoutoffourpurchases)	Real (0 to 1)
Option_price	Average of the total amount of option price in each purchase	Real
Coupon_used	Average of the number of given coupons in each purchase	Real
Coupon_given	Average of the number of used coupons in each purchase	Real
Mileage_given	The amount of given mileage in each purchase	Real
Credit_card_usage_ratio	Rate of paying in credit cards (ex. 0.25 means one use in four)	Real (0 to 1)

Table 3.2 Examples of records of the dataset integrated for VIP classification

C_ID	Age / Seller_satisfaction	Sex / product_satisfaction	Quantity / Number_of comments	Shipping_fee / Coupon_given	Option_price / Mileage_given	Coupon_used / Credit_card_usage_ratio
john	33	F	1.05	2150	1550	0.65
	4.31	3.39	104.5	1.32	37.5	0.75
pharos	52	F	1.0	1630	0	0
	4.1	3.93	33.2	1.1	0	1
lov115	16	M	1.33	1100	2473	0.87
	3.55	4.15	1105.3	1.95	850	0

There is no information about who are the VIP customers in the joined table. Thus, in order to recognize VIP customers, we first normalized R, F and M values of each of 991 customers, and then top X% customer groups are identified from each of R, F and M values. Finally, we decided the customers who belong to the intersection of the three groups as VIP customers, after several attempts with different X value. When X value is 50, VIP customers are approximately 30% of the whole customers and they cover about 70% of total revenue.

3.2.2. VIP classification model

Prior to the VIP classification model construction, we carried out feature selection by employing the wrapper approach with backward elimination, using Weka 3.6.6. Gain Ratio attribute evaluator is used to evaluate12 input variables in ranker search method, to select the most influential K variables when classifying customers into VIP or non-VIP. The descending order of importance of the 12 input variables is: Age, Sex, Quantity, Seller_satisfaction, Product_satisfaction, Number_of_comments, Shipping_fee, Option_price, Coupon_used, Coupon_given, Mileage_given, and Credit_card_usage_ratio.

Then, with Weka 3.6.6 data mining tool, we tried to build a decision tree model and an artificial neural network model to classify the customers into VIPs and non-VIPs. Among many different decision tree algorithms, C4.5 algorithm was used to build decision tree model with 0.25 confidence factor for pruning in this study. We also build an ANN model with learning rate, momentum, epoch, and the number of hidden-layer being set to 0.3, 0.2, 500, and 1, respectively. Besides, we developed a linear regression model for the same purpose.

We built classification models by taking wrapper approach to feature selection with backward elimination techniques, where we started with all 12 input variables and then eliminated the least important variable one by one. Having tried 10-fold cross-validation, we compared classification results among the three models in Section 4.1.

3.2.3. Promotion validation

We identified 6 types of promotion which were used in the target open-market. They are Option, Market_price, Coupon, Free_delivery, Mileage and Additional discount. We used five of them for our experiment because the promotion, Additional discount is applied in only a few products. From the integrated transaction records, we selected five attributes, each representing a type of promotion, and modified them into Boolean type, as is explained in Table 3.3.

Table 3.3 Promotion attributes

Attribute	Meaning	Data type
Option	Additional payment for option exists or not	Boolean
Market_price	market price exists or not	Boolean
Coupon	Coupon is given or not	Boolean
Free_delivery	Delivery is free or not	Boolean
Mileage	Mileage is given or not	Boolean

To examine whether there are differences between VIP customers and non-VIP customers in terms of each promotion's effect on their purchase, we conducted chi-square test for each promotion. Then, we also conducted a logistic regression analysis to compare how significantly each promotion has influence over VIP customer's purchase, using R2.12 and SPSS 12.0.

3.2.4. Identification of association rules among promotions

Using SAS Enterprise Miner 9.1, we conducted an experiment to identify association rules among the five promotion types. We used only VIP customer's transaction records because we aim to find the effective promotion mix over VIP customers. Each record of promotion data consists of 6 attributes such as Option, Market_price, Coupon, Free_delivery, and Mileage including Transaction_ID. Table 3.4 shows a few examples of promotion records.

After several trials with different parameter values, we finally set minimum support to 10% and minimum confidence to 20% for promotion mix analysis.

Table 3.4 Examples of VIP customers' promotion instances for association rules

Transaction_ID	Option	Market_price	Coupon	Free_delivery	Mileage
04906	T	T	F	T	T
04909	T	F	F	T	T
04910	F	F	F	F	T

4. Results

4.1. Results from building classification models

In this section we will compare three classification models such as DT, LR, and ANN, and identify influential factors. 12 experiments for each of the 3 data mining techniques (DT, LR, ANN) or a total 36 experiments were conducted while decreasing the number of features to find the best classification model and the best set of attributes used to build the model. That is, we used the wrapper method with backward elimination both to select features and to find best classifier.

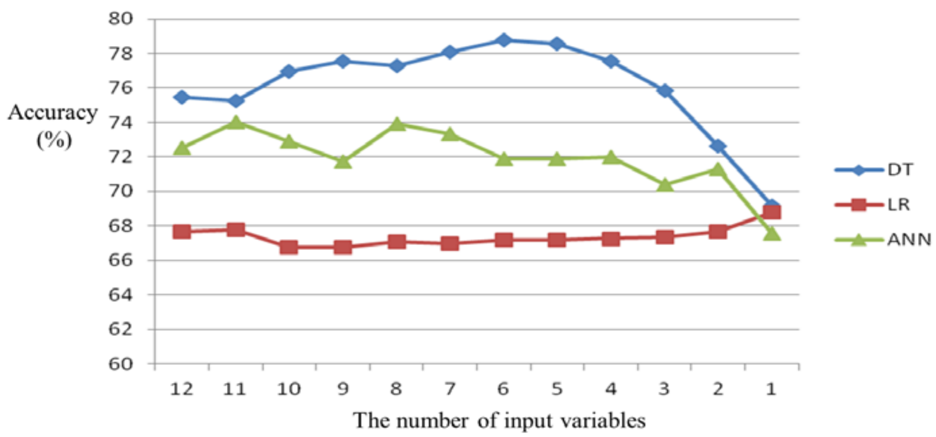

Figure 4.1 Classification accuracy of each mining technique

In Figure 4.1 which shows the classification accuracy as the number of input variables decreases, we can figure out that the decision tree model results in the best classification accuracy among the three models. More specifically, as can be seen in Table 4.1, the highest classification accuracy was acquired with the decision tree model especially when 6 attributes such as Quantity, Seller_satisfaction, Number_of_comments, Coupon_given, Mileage_given, Credit_card_usage_ratio were used. Therefore, it is suggested to use the DT model with those 6 attributes when classifying customers into VIP or non-VIP.

Table 4.1 Accuracy comparisons of each mining technique

The number of Features	DT	LR	ANN
12	75.45%	67.67%	72.52%
11	75.25%	67.77%	74.04%
10	76.96%	66.76%	72.92%
9	77.57%	66.76%	71.71%
8	77.27%	67.07%	73.93%
7	78.08%	66.96%	73.33%
6	**78.78%**	67.17%	71.91%
5	78.58%	67.17%	71.91%
4	77.57%	67.27%	72.02%
3	75.85%	67.37%	70.40%
2	72.62%	67.67%	71.31%
1	69.19%	68.78%	67.57%

4.2. Results from promotion validation

Promotion validation is conducted in two steps. First, chi-square test is used to examine whether each promotion has significantly different effect on purchase between VIP and non-VIP customers (Agresti, 2012). Five promotion factors, 2 by 2 contingency tables of VIP and non-VIP customers, and the results of chi-square test are shown in Table 4.2. In the table, 'T' represents 'True', 'F' is for 'False' and, 'Prob' indicates P-value. Each Prob value is the probability of obtaining chi-square test result, assuming that the null hypothesis is true. The null hypothesis is "There is no difference in each factor's influence between VIPs and non-VIPs" According to the results, all of the P-values are less than 0.05, thus the corresponding null hypothesis is rejected at significance level of 0.05. That is, there exists meaningful difference in the aspect of each promotion's effect on whether customers are VIPs or non-VIPs.

Table 4.2 The result of chi square test with promotions

VIP	Option				Market Price			
	T	F	Sum	Prob	T	F	Sum	Prob
T	5,009	10,933	15,942		6,824	9,118	15,942	
F	1,666	6,055	7,721	.000	3,438	4,283	7,721	.012
Sum	6,675	16,988	23,663		10,262	13,401	23,663	
VIP	Free Delivery				Mileage			
	T	F	Sum	Prob	T	F	Sum	Prob
T	7,284	8,658	15,942		8,327	7,615	15,942	
F	3,294	4,427	7,721	.000	2,863	4,858	7,721	.000
Sum	10,578	13,058	23,663		11,190	12,473	23,663	
VIP	Coupon							
	T	F	Sum	Prob				
T	3,170	12,772	15,942					
F	1,876	5,845	7,721	.000				
Sum	5,046	18,617	23,663					

Second, we conducted logistic regression with all promotion types as independent variables, in order to compare how differently each promotion affects VIP customers' purchase. In logistic regression, sign of β value and Exp(β) value represent how influential each of the variable is. If the sign of β value is plus, and the value of Exp(β) exceeds 1, then the variable is regarded to have positive influence over the VIP customers' purchase. As shown in Table 4.3, all of the variables are statistically significant in 95% confidence level and *Option*, *Free_Delivery* and *Mileage* are found to be influential variables. Based on the result, mileage's effect over VIP customers' purchase is bigger as much as 77.1%, compared with when mileage does not exist. Meanwhile, *Market_Price* and *Coupon* are not effective promotion type to VIP customers, rather they are more favorable to non-VIP customers. Logistic regression equation is given as (4.1), where PVIP represents the possibility of being a VIP customer.

$$\log_e \frac{P_{VIP}}{1-P_{VIP}} = 0.417 + 0.4010OPT - 0.59MAR - 0.328COU + 0.104FRE + 0.572MIL \quad (4.1)$$

Table 4.3 The result of logistic regression with promotions

	β	S.E.	Sig.	Exp(β)
OPTION (OPT)	.401	.033	.000	**1.493**
MARKET_PRICE (MAR)	-.059	.028	.039	.943
COUPON (COU)	-.328	.034	.000	.720
FREE_DELIVERY (FRE)	.104	.029	.000	**1.110**
MILEAGE (MIL)	.572	.029	.000	**1.771**
Constant	.417	.027	.000	1.517

4.3. Results of association rules

As shown in Table 4.4, association rules were found from VIPs' transaction data. The table shows that *'OPTION'* and *'MILEAGE'* are frequently associated with other promotion attributes. For example, people who are likely to accept promotion *'MILEAGE'*, *'MARKET_PRICE & MILEAGE'* or *'FREE_DELIVERY & MILEAGE'* generally tend to take each of them together with *'OPTION'* And people taking promotion *'OPTION'*, *'OPTION & MARKET_PRICE'* or *'OPTION & FREE_DELIVERY'* generally tend to accept each of them together with *'MILEAGE'*.

Table 4.4 Association results of promotion mix only from VIPs' transaction data

Lift	Support (%)	Confidence (%)	Rule
1.16	23.93	41.02	MILEAGE → OPTION
1.16	23.93	67.93	OPTION → MILEAGE
1.2	11.22	42.41	MILEAGE & FREE_DELIVERY → OPTION
1.14	11.22	66.53	FREE_DELIVERY & OPTION → MILEAGE
1.28	11.01	45.25	MILEAGE & MARKET_PRICE → OPTION
1.15	11.01	67.15	MARKET_PRICE & OPTION → MILEAGE
1.02	10.87	49.12	COUPON → MARKET_PRICE
1.02	10.87	22.58	MARKET_PRICE → COUPON

4.4. Development of CRM strategies

According to earlier researches, CRM consists of four dimensions such as customer identification, customer attraction, customer retention, and customer development (Kracklauer *et al.*, 2004; Parvatiyar *et al.*, 2001; Swift, 2001). Customer identification is meant to identify potential customers. Customer attraction attempts to attract the target customer segments by motivating customers to place orders through various channels. Customer retention refers to the activity of preventing the existing customers from switching to competitors by enhancing the level of customer satisfaction through one-to-one marketing, loyalty program, complaints management, etc. Customer development, the ultimate goal of CRM, aims to maximize the revenue by expanding transaction intensity, transaction value and individual customer profitability through customer lifetime value analysis, up/cross selling, market basket analysis, etc.

For each dimension of CRM, we developed CRM strategies for the open market as follows based on the results from this study:

- Strategies for identifying / attracting VIP customers
  - The decision tree classification model could be used to identify VIP customers, in order to practice directly targeted marketing activities focusing on them. It would be even more cost-effective than doing the same things targeting all the customers. Of course, it is necessary to develop a new classification model as more customer transaction data is accumulated. The cyclic process of collecting customers' data, building models and using them for marketing activities needs to be continued.
- Strategies for retaining VIP customers
  - As can be seen in table 4.1, attributes such as *Mileage_given, Seller_satisfaction, Credit_card_usage_ratio, Quantity, Coupon_given, Number_of_comments* are the most influential factors in VIP classification. Therefore, sellers in our target open-market should develop CRM strategies using these factors. It is recommended for sellers to reward VIP customers with additional mileages in light of the fact that they want tangible compensation for their loyalty. And sellers are recommended to give benefits to a customer who purchases many products at once because *Quantity* is influential factor to VIP customers. Also, they should give more proper reward to their VIP customers when they frequently act on online comments to increase motivation of customer participation in making comments. As it was proven from the VIP transaction data that VIP customer's satisfaction over seller is significant, sellers need to provide customers with superb service. VIP customers have a propensity to use credit card in the open market. Thus, it would be suggested that sellers should establish a strategic alliance with credit card companies to give them more benefits when they purchase products with a credit card.
  - Based on the promotion validation results, understanding VIP customers' favor in the aspect of promotion can give the sellers a hint of effective promotion strategies. For example, the seller in the open-market can provide mileage or free delivery, which is found to be attractive to VIP customers, if a VIP customer decides to purchase the goods.
- Strategies for developing VIP customers
  - Since (*OPTION* and *MILEAGE*), (*OPTION* and *MARKET_PRICE* & *MILEAGE*), and (*OPTION* and *FREE_DELIVERY* & *MILEAGE*) are mutually associated with each other in Table 4.4, it is highly recommended that sellers in the open-market should mix promotions to maximize the effectiveness of promotion. In addition, the results shows that VIP customers willingly pay for the additional cost by choosing options, even though they know that they are lured by the initial low price. Mileage or free delivery can be a good supplementary promotion to be provided to them.

5. Conclusions

In this paper, we suggested CRM strategies targeted on VIP customers. For this purpose, we firstly defined VIP customers based on RFM values, and then developed classification models to classify the customers into VIPs and non-VIPs, and identify influential factors using various data mining techniques such as decision tree, artificial neural network, and logistic regression. We also identified five promotion types offered in the open market, and validated each by chi-squared test and logistic regression analysis in the aspect of promotion effect over VIPs' purchase.

To the best of our knowledge, little or no research has been performed so far to examine the effect of individual promotion and promotion mix over VIP customers in an open market. It seems that the strategies we developed can be utilized to enhance the revenue of the open market more efficiently. As an additional research area, we plan to extend our methodologies to explore how to turn non-VIP customers into VIP customers.

The fact that our dataset is relatively small and contains transaction data during a short term is a limitation of our study. It would be better if this study have proved whether the suggested strategies are effective or not in the actual field. Nevertheless, we believe that our research results have meaningful implication to the sellers of open markets.

References

Adomavicius, G. and Tuzhilin, A. (2001). Expert-driven validation of rule-based user models in personalization applications. *Data Mining and Knowledge Discovery*, **5**, 33-58.

Aggarval, C. C. and Yu, P. S. (2002). Finding localized associations in market basket data. *IEEE Transactions on Knowledge and Data Engineering*, **14**, 51-62.

Agresti, A. (2012). *Categorical data analysis*, 3rd ed, John Wiley & Sons, New York.

Baesens, B., Verstraeten, G., Dirk, V. D. P., Michael, E. P., Kenhove, P. V. and Vanthienen, J. (2004). Bayesian network classifiers for identifying the slope of the customer-life cycle of long-life customers. *European Journal of Operational Research*, **156**, 508-523.

Bortiz, J. E. and Kennedy, D. B. (1995). Effectiveness of neural network types for prediction of business failure. *Expert Systems with Applications*, **9**, 503-512.

Bunkley, N. and Joseph, J. (2008) *Pioneer in quality control*, New York Times, http://www.nytimes.com/2008/03/03/business/03juran.html.

Changchien, S. W., Lee, C. F. and Hsu, Y. J. (2004). On-line personalized sales promotion in electronic commerce. *Expert Systems with Applications*, **27**, 35-52.

Cheng, C. H. and Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, **36**, 4176-4184.

Dennis, C., Marsland D. and Cockett, T. (2001). Data mining for shopping centres customer knowledge management framework. *Journal of Knowledge Management*, **5**, 368-374.

Fletcher, D. and Goss, E. (1993). Forecasting with neural networks: An application using bankruptcy data. *Information and Management*, **3**, 159-167

Hosseini, M., Anahita, M. and Mohammad, RG. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, **37**, 5259-5264.

Hughes, A. M. (1994). *Strategic database marketing*, Probus Publishing, Chicago.

Hwang, H., Jung T. and Suh, E. (2004). An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. *Expert Systems with Applications*, **26**, 181-188.

Khajvanda, M. and Tarokhb, M. J. (2011) Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. *Procedia Computer Science*, **3**, 1327-1332.

Kim, S. Y., Jung, T. S., Suh, E. H. and Hwang, H. S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, **31**, 101-107.

Kim, Y. S. (2006). Toward a successful CRM: Variable selection, sampling, and ensemble. *Decision Support Systems*, **41**, 542-553.

Kim, Y. S. and Street, W. N. (2004). An intelligent system for customer targeting: A data mining approach. *Decision Support Systems*, **37**, 215-28.

Kracklauer, A. H., Mills, D. Q. and Seifert, D. (2004). Customer management as the origin of collaborative customer relationship management. *Collaborative Customer Relationship Management - Taking CRM to the Next Level*, 3-6.

Kubat, M., Hafez A., Raghavan, VV., Lekkala, J. R. and Chen, W. K. (2003). Item set trees for targeted association querying. *IEEE Transaction on Knowledge and Data Engineering*, **15**, 1522-1534.

Langley, P. and Simon, H. A. (1995). Applications of machine learning and rule induction. *Communication of the ACM*, **38**, 55-64.

Lau, H. C. W., Wong, C. W. Y., Hui, I. K. and Pun, K. F. (2003). Design and implementation of an integrated knowledge system. *Knowledge-Based Systems*, **16**, 69-76.

Newell, F. (1997). *The new rules of marketing: How to use one-to-one relationship marketing to be the leader in your industry*, McGraw-Hills, New York.

Parvatiyar, A. and Sheth, J. N. (2001). Customer relationship management: Emerging practice, process, and discipline. *Journal of Economic & Social Research*, **3**, 1-34.

Salchenberger, L. M., Cinar, E. M. and Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *Decision Sciences*, **23**, 899-916.

Shim, B. S., Choi, K. H. and Suh, Y. M. (2012). CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns. *Expert Systems with Applications*, **39**, 7736-7742.

Su, C. T., Hsu, H. H. and Tsai, C. H. (2002). Knowledge mining from trained neural networks. *Journal of Computer Information Systems*, **42**, 61-70.

Swift, R. S. (2001). *Accelerating customer relationships: Using CRM and relationship technologies*, Prentice Hall PTR, Upper Saddle River, NJ.

Tam, K. Y. and Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, **38**, 926-947.

Yu, J. X., Ou, Y., Zhang, C. and Zhang, S. (2005). Identifying interesting visitors through web log classification. *IEEE Intelligent Systems*, **20**, 55-59.

Zhang, G., Hu MY. and Patuwo, B. E. (1999). Indro DC, artificial neural networks in bankruptcy prediction: General framework and cross validation analysis. *European Journal of Operational Research*, **116**, 16-32.