Expert Systems with Applications 39 (2012) 7736–7742

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns

Beomsoo Shim *, Keunho Choi, Yongmoo Suh

Business School, Korea University, Anam-Dong 5-Ga, Seongbuk-Gu, Seoul 136-701, Republic of Korea

ARTICLE INFO

Keywords:
Data mining
CRM strategy
Association rule
Sequential pattern
RFM

ABSTRACT

As dot-com bubble burst in 2002, an uncountable number of small-sized online shopping malls have emerged every day due to many good characteristics of online marketplace, including significantly reduced search costs and menu cost for products or services and easily accessing products or services in the world. However, all the online shopping malls have not continuously flourished. Many of them even vanished because of the lack of customer relationship management (CRM) strategies that fit them. The objective of this paper is to propose CRM strategies for small-sized online shopping mall based on association rules and sequential patterns obtained by analyzing the transaction data of the shop. We first defined the VIP customers in terms of recency, frequency and monetary (RFM) value. Then, we developed a model which classifies customers into VIP or non-VIP, using various data mining techniques such as decision tree, artificial neural network, logistic regression and bagging with each of these as a base classifier. Last, we identified association rules and sequential patterns from the transactions of VIPs, and then these rules and patterns were utilized to propose CRM strategies for the online shopping mall.

1. Introduction

Internet technologies have provided many competitive advantages such as agility, selectivity, individuality and interactivity (Weiber & Kollmann, 1998). The Internet enables customers to search products and services meeting their needs with less time than before. Online shopping malls, where buyers place orders over the Internet, have emerged to become a prevalent sales channel. After dot-com bubble burst in 2002, an uncountable number of small-sized online shopping malls have emerged every day due to many good characteristics of online marketplace, including significantly reduced search costs and menu cost for products or services and easily accessing products or services in the world.

However, all of small-sized online shopping malls have not flourished. Many of them even vanished. As such, owners of the online shopping malls are concerned about how to manage their shops successfully. So is the owner of a small-sized online shopping mall in Korea, for which we want to develop CRM strategies. The shop deals with imported items such as those used in living room, kitchen, bathroom, etc. Since it started, the online shopping mall has shown increased revenue for a while. These days, however, the revenue curve changes from steep to gentle slope because the owner of the online shopping mall does not have knowledge how to develop CRM strategies that fit to them (Anderson, Jolly, & Fairhurst, 2007). To make the online shopping mall thrive, good CRM strategies are critical. Therefore, our study aims to deal with following research questions: (1) Who are the VIPs of our target online shopping mall? (2) What are the VIPs' purchase patterns? (3) What kinds of CRM strategies can increase revenue?

To address these research questions, we (1) define VIP in terms of recency, frequency and monetary (RFM) value, and develop a classification model which classifies customers into VIP and non-VIP, using data mining techniques such as decision tree, artificial neural network, logistic regression, and bagging; (2) identify association rules and sequential patterns hidden in the transaction data of solely VIP customers; (3) propose CRM strategies for the small-sized online shopping mall based on the identified rules and patterns, respectively.

The rest of this paper is organized as follows. Section 2 presents literature review about data mining techniques for CRM and definition of RFM. Section 3 describes research methods used in this study. Section 4 explains experiments conducted in this study with respect to dataset and research framework, followed by step by step explanation in detail. Section 5 describes the experimental results of classification analysis and association rule and sequential pattern analysis, and derives CRM strategies for the target shop. Finally, Section 6 ends the paper with conclusions.

* Corresponding author. Tel.: +82 70 8223 6041; fax: +82 2 4183 0102.
E-mail addresses: bumsuya@korea.ac.kr (B. Shim), keunho@korea.ac.kr (K. Choi), ymsuh@korea.ac.kr (Y. Suh).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.080

2. Literature review

This study proposes CRM strategies which can be helpful to a small-sized online shopping mall. In this section, previous CRM-related researches using data mining techniques are reviewed.

2.1. Data mining for customer relationship management

Since the early 1980s, the concept of customer relationship management has gained its importance in marketing domain. Although it is difficult to make a totally approved definition of CRM, we can generally describe it as a comprehensive strategy including processes of acquiring, retaining and partnering with selective customers to create value for both the company and customers (Parvatiyar & Sheth, 2001). Many previous CRM-related researches have used data mining techniques to analyze and understand customer behavior and characteristics, and most of these have worked well (Langley & Simon, 1995; Lau, Wong, Hui, & Pun, 2003; Salchenberger, Cinar, & Lash, 1992; Tam & Kiang, 1992; Zhang, 1999). They have shown that data mining techniques can be used to elicit untapped useful knowledge from a large customer data. This section reviews previous researches which utilize classification, and association rules and/or sequential patterns analyses for various tasks in CRM domain.

Classification tasks have been carried out for various purposes in CRM domain. For example, Kim, Jung, Suh, and Hwang (2006) adopted decision tree to classify the customers and developed CRM strategies based on customer lifetime value. Hwang, Jung, and Suh (2004) used logistic regression to segment customers based on their customer loyalty. Yu, Ou, Zhang, and Zhang (2005) identified interesting visitors through web log classification. Dennis, Marsland, and Cockett (2001) proposed customer knowledge management framework using K-means. Kim and Street (2004) proposed a system which makes use of ANN and genetic algorithm for customer targeting. Kim (2006) used logistic regression and ANN for feature selection to predict churn. Baesens et al. (2004) identified the slope of the customer lifecycle based on Bayesian network classifier. Sinha and Zhao (2008) adopted decision tree and logistic regression for churn prediction by incorporating domain knowledge into data mining. Tsai and Lu (2009) used hybrid neural network for churn prediction.

Other data mining techniques that are useful for the analysis of customer data are association rules and/or sequential patterns analyses. Researches on the utilization of association rules and sequential patterns have been conducted for various purposes in CRM domain. For example, sequential patterns are used to predict future complaint (LarivièRe & Van den Poel, 2005) and to predict network banking churn (Chiang, Wang, Lee, & Lin, 2003). Adomavicius and Tuzhilin (2001) examined association rules for one to one marketing. Aggarwal, Procopiuc, and Yu (2002) conducted market basket analysis to identify association rules. Changchien, Lee, and Hsu (2004) used both ANN and association rules obtained from market basket analysis to develop on-line personalized sales promotion. Tsai and Chen (2010) used association rules to select variables for churn prediction.

Table 1 provides a summary of previous researches which performed classification and Tables 2 and 3 a summary of association rules and/or sequential pattern analyses for various purposes in CRM domain.

2.2. RFM model definition

The RFM analytic model is proposed by Hughes (1994) which differentiates important customers based on the values of three variables, i.e., recency (R), frequency (F) and monetary value (M). They are defined as follows:

- R refers to the time interval between the last purchasing behavior and current.
- F refers to the number of transactions over a certain period of time.
- M refers to the amount of money spent on products or services over a certain period of time.

RFM are very effective values for customer segmentation (Newell, 1997). According to the literature (Wu, Kao, Su, & Wu, 2005), researches showed that the bigger the values of R and F are, the more likely the corresponding customers are to make a new trade with companies. Moreover, the bigger M is, the more likely the cor-

Table 1
Previous classification and/or prediction researches in CRM domain.

Task	Data mining techniques	Reference
Customer segmentation and strategy development	Decision tree	Kim et al. (2006)
Customer segmentation	Logistic regression	Hwang et al. (2004)
Identification of interesting visitors through web log classification	Decision tree	Yu et al. (2005)
Suggestion of customer knowledge management framework	K-means	Dennis et al. (2001)
Customer targeting	ANN and genetic algorithm	Kim and Street (2004)
Churn prediction	Logistic regression and ANN	Kim (2006)
Identification of the slope of the customer-lifecycle	Bayesian network classifier	Baesens et al. (2004)
Churn prediction	Decision tree and logistic regression	Sinha and Zhao (2008)
Churn prediction	ANN and SOM	Tsai and Lu (2009)

Table 2
Previous AS rules and/or SE patterns researches in CRM domain.

Task	Data mining techniques	Reference
Prediction of future complaint	Sequential pattern	LarivièRe and Van den Poel (2005)
Network banking churn analysis	Sequential pattern	Chiang et al. (2003)
One to one marketing	Association rule	Adomavicius and Tuzhilin (2001)
Market basket analysis	Association rule	Aggarwal et al. (2002)
On-line personalized sales promotion	Association rule and ANN	Changchien et al. (2004)
Variable selection for churn prediction	Association rule and ANN	Tsai and Chen (2010)

Table 3
Previous data mining researches based on RFM values in CRM domain.

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 and Chen (2009)
A purchasing pattern segmentation of customers	Sequential pattern	Chen et al. (2009)
Analysis of the characteristics of customers	K-means and clustering	Huang et al. (2009)
Mining association rules of customer values	Association rules	Chiang (2010)

responding customers are to buy products or services from the same companies again.

Data mining researches have been carried out based on RFM values in CRM domain. For example, Hosseini, Maleki, and Gholamian (2010) adopted K-means algorithm to classify the customer loyalty based on RFM values. Cheng and Chen (2009) used K-means and rough set theory to segment customers based on RFM values. Chen, Kuo, Wu, and Tang (2009) identified purchasing patterns in the form of sequential patterns. Huang, Chang, and Wu (2009) adopted K-mean and clustering to analyze the characteristics of customers based on RFM values. Chiang (2010) mined association rules of customer values.

3. Research method

3.1. Research framework

Fig. 1 depicts the framework of our research. Details of each step are described in below.

3.2. VIP classification

First, we developed models each of which classifies customers into VIP or non-VIP, using various data mining techniques such as decision tree, artificial neural network, logistic regression and bagging of each of these data mining techniques as a base classifier. DT, ANN, and LR are data mining approaches that have been heavily used for classification and/or prediction in support of marketing decision marking, and have shown good performance (Chien & Chen, 2008; Kim, 2006; Kim & Street, 2004; Kim et al., 2006; Yu et al., 2005). As an alternative to a single classifier approach, bagging has been considered in the recent years (Frosyniotis, Stafylopatis, & Likas, 2003; Kang & Doermann, 2003; Roli, Kittler, & Windeatt, 2004). The key idea of bagging is to combine a number of classifiers such that the resulting combined system achieves higher classification accuracy than the original single classifiers.

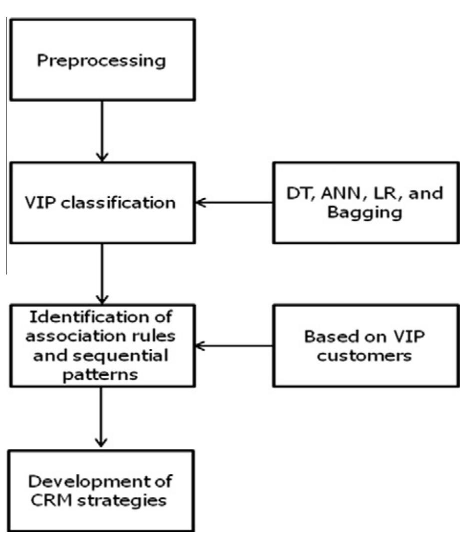

Fig. 1. Framework for experiments.

3.3. Identification of association rules and sequential patterns

Once the VIP classification model is built, we attempted to find association rules and sequential patterns among the item categories as well as among the item subcategories from the transaction data of only customers classified as VIPs by the classification model.

4. Experimental design

This section explains the experiments conducted in this paper. First, we introduce our target small-sized online shopping mall together with its transaction data. Then, we provide our research framework, with step by step explanation. Our experiments are conducted using Weka 3.6 and SAS Enterprise Miner 9.1.

4.1. Data description and pre-processing

4.1.1. Target online shopping mall

A small-sized online shopping mall[1] in Korea provided its transaction data for analyses. It distributes and sells items imported from other countries by online. The online shopping mall deals with cute items that are used in living room, kitchen, bathroom, etc. Its revenue was approximately 10,000 dollars per month and had 3445 registered members as of August 2008. The size and the revenue have been grown rapidly since the online shopping mall opened on August 2008. However, the owner of the online shopping mall has little knowledge about their loyal customers and their purchasing patterns. Without actionable knowledge related to CRM, such as who loyal customers are and what their purchasing patterns are, it might be difficult to expect continuous increase in sales.

4.1.2. Data description

The entire data was collected from August, 2008 till August, 2010. The dataset consists of five tables such as demographic table, bulletin table, comment table, order management table, and order product table. Demographic table contains information such as member ID, age, member name, address, the first day of registration of a member. Bulletin table has information about names of the writer, the number of click on bulletin, etc. Comment table has similar information to bulletin table. It consists of names of commenter, the number of reading comments, etc. In order management table, there is information about member ID, order ID, delivery address, zip code, telephone number, e-mail address, pay method (card or cash), pay amount, shipping fee, and mileage used. Order product table consists of order ID, category goods, product names, and product price. Having discussed with the domain experts, we selected 9 attributes such as age, reg_channel,

[1] URL here (http://welesfamily.com/).

card_rate, delivery, comment, bulletin, reg_duration, mileage_used, and shipping_fee from those table. The number of customers registered is 3445, and the number of their purchasing items is 14,782, respectively.

4.1.3. Pre-processing

After deleting the duplicated records or those with many missing or inaccurate values, we obtained 2295 customers who have purchased products at the shop, and 11,033 transaction data. Then, we defined VIP customers as follows based on the values of RFM variables each customer has.

We computed and normalized *R*, *F* and *M* values of 2295 customers, sorted them in descending order, and selected the top *X* % of customers in each variable. We tried several times to compute the intersection of the three sets, with different values for *X*. When *X* is 70, the total monetary value of the customers (1096 out of 2295 customers) belonging to the intersection accounts for approximately 70% of the total revenue of the shop. We defined the customers that belong to the intersection as VIPs. The meanings, and data types and ranges of values of *R*, *F*, *M* variables (after normalization) and other variables are explained in Table 4. Examples of records of the resulting dataset are shown in Table 5.

Table 5
Examples of records of our dataset.

ID	dlr4488	ejknock	hot1132	jiuk486	luck777m
R	87.14	4.88	89.49	91.02	4.31
F	82.04	28.07	88.49	89.58	23.67
M	91.84	84.00	71.05	86.96	66.08
Age	41	36	22	33	34
reg_channel	n	Search	rec	n	n
card_rate	0	1	0	1	1
Delivery	c	a	c	c	c
Comment	0	1	2	3	1
Bulletin	10	1	1	5	0
reg_duration	387	62	52	248	323
mileage_used	4000	0	0	13,410	0
shipping_fee	12,500	0	0	0	5000

4.2. Experiment for VIP classification

Prior to the model construction, we conducted feature selection by employing the wrapper approach with backward elimination. 12 input variables were evaluated using Gain Ratio attribute evaluator based on ranker search method, to select more influential variables when classifying the instances into VIP or non-VIP. The descending order of importance of 12 input variables is as follows: *R*, *F*, *M*, reg_duration, shipping_fee, age, comment, bulletin, card_rate, delivery, reg_channel, mileage_used.

With Weka 3.6 data mining tool which has been used widely, C4.5 algorithm was used to build decision tree model with 0.25 confidence factor for pruning in this study. Having tried several times to conduct ANN, we set learning rate, momentum, epoch, and the number of hidden-layer to 0.1, 0.9, 50, and 1, respectively. In bagging method, we used 10 classifiers built from 10-fold cross-validation.

Comparisons of the classification results of models from 10-fold cross-validation are made in Section 5.1.

4.3. Experiment for identification of association rules and sequential patterns

In this study, we used the transaction data of only VIP customers. It contains 7271 transaction data of 1096 VIP customers. As shown in Table 6, each instance of transaction data consists of 5 attributes such as member_id, order_id, category, subcategory, and order_date. As is shown in Table 6, when customers purchase several items at the same time, an instance record is generated for each item. We attempted to find association rules and sequential patterns among the categories as well as among the subcategories.

Table 7 shows categories and subcategories of goods sold in the open market. After several attempts with different parameter values, we finally set the minimum support to 4% and the minimum confidence to 40% for categories analyses and set the minimum support to 3% and the minimum confidence to 20% for subcategories analyses. To derive sequential patterns, we set the time window to the whole period of the dataset because the number of instances was too small.

5. Results

5.1. Results of classification analysis

A total 72 experiments in each dataset were conducted for classification analyses. In Table 8 showing the classification accuracy as the number of input variables decrease, we can find out that the highest classification accuracy was acquired when we included the most influential 3 attributes such as R, F and M in our dataset as was expected. Fig. 2 showing the classification accuracy with only RFM values, we can find out that DT (Decision Tree) and bagging-DT show the best performance among all data mining techniques. Although at first glance the results of classification accuracy might look too high, it is expected because the target class of the classification algorithm is defined in terms of three (i.e., RFM) of the input variables of the algorithm. Also, note that previous researches to classify VIP customers based on RFM values such as Hosseini et al. (2010) and Cheng and Chen (2009) yielded similar classification accuracy 99.72% and 97.98%, respectively.

Table 4
Independent attributes in this study.

Attribute	Meaning	Data type and range of values
R	The time interval between the last purchasing behavior and current	Real (0–100)
F	The number of transactions over a certain period of time	Real (0–100)
M	The amount of money spent on products or services	Real (0–100)
Age	Customer's age	Integer
reg_channel	Channel of registration (search or recommendation)	String
card_rate	Rate of paying in credit cards (ex. 0.25 means one use in four)	Real (0–1)
Delivery	Delivery (a: 1 day delivery, b: 1 to 2 day delivery, c: 2 to 3 day delivery, d: more than 3 day delivery)	Character
Comment	The number of posting comments on products	Integer
Bulletin	The number of activities on bulletin board	Integer
reg_duration	The number of days after registration	Integer
mileage_used	The amount of used mileage in total	Integer
shipping_fee	The amount of paid shipping fee in total	Integer

Table 6
Examples of instances for association rules and sequential patterns.

Member_id	Order_id	Category	Subcategory	Order_date
angel9270	20090320-0000063	Dishes	Bottle	2009-03-20
angel9270	20090331-0000235	Dishes	Cup	2009-03-31
angel9270	20090331-0000235	Dishes	Bowl	2009-03-31
zzzmia	20090822-0000018	Dishes	Tray	2009-08-22
zzzmia	20090822-0000018	Cooking goods	Cooker	2009-08-22
zzzmia	20091101-0000038	Dishes	Vessel	2009-11-01

Table 7
Category and subcategory of goods.

Surrogate of nomination	Category	Subcategories
A	Dishes	Bottle, bowl, coaster, cup, dosirak, knife, plate, spoon, tray, vessel, box
B	Cooking goods	Cooker, cutting board, kettle, pan
C	Kitchen goods	Apron, napkin, oil paper, kitchen towel
D	Bathroom goods	Body washer, toilet cover, basin, bath gloves, soap, tooth brush, bath towel
E	Living room goods	Basket, air freshener, bag, blanket, box, calendar, case, clock, doll, handkerchief, hanger, mat, toy, pouch, purse, shoes
F	Washing goods/cleansing supplies	Broom, scouring pad, trash can
G	Office supplies	Calculator, note, file, mouse pad, paste, pencil case, scotch tape

Table 8
Accuracy comparisons of each mining technique.

The number of features	DT (%)	ANN (%)	LR (%)	Bagging		
				DT (%)	ANN (%)	LR (%)
12	99.83	99.19	86.23	99.83	99.33	86.31
11	99.83	98.87	86.12	99.83	99.07	86.18
10	99.39	98.54	85.58	99.45	98.78	85.76
9	99.39	98.93	85.45	99.45	99.13	85.62
8	99.69	99.08	85.08	99.71	99.25	85.22
7	99.54	99.45	86.21	99.62	99.65	86.28
6	99.54	98.71	85.99	99.62	98.92	86.06
5	99.54	99.24	86.51	99.62	99.37	86.57
4	99.69	99.45	86.59	99.71	99.65	86.65
3	99.83	99.56	86.67	99.83	99.78	86.75
2	82.00	82.26	76.12	82.26	82.27	76.17
1	81.13	81.09	73.73	81.31	81.22	73.73

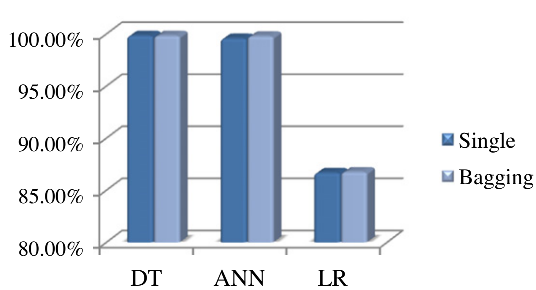

Fig. 2. Classification accuracy of each mining technique with RFM values.

5.2. Results of association rules and sequential patterns analysis

Sixteen association rules were found (see Tables 9 and 10) and 11 sequential patterns were found (see Tables 11 and 12) from VIP transactions. Tables 9 and 10 show the association rules among categories and among subcategories, respectively. Table 9 shows that 'E' (Living room goods) is frequently associated with other categories such as 'A' (Dishes), 'B' (Cooking goods), or 'D' (Bathroom goods). For example, people who purchase 'A', 'B', or 'D' generally tend to purchase 'E' together with each of them. Table 10 indicates association rules among subcategories.

Table 9
Association rules among categories.

Lift	Support (%)	Confidence (%)	Rule
1.08	8.10	61.79	D => E
1.05	5.81	44.31	D => E and A
1.09	5.81	61.93	D and A =>E
1.06	5.81	60.56	B => E
1.09	4.32	62.31	B and A => E
1.06	4.32	45.00	B => E and A

Tables 11 and 12 show the sequential patterns among categories and among subcategories, respectively. From Table 11, we can find that customers who purchase category 'A' and/or 'E' are likely to buy goods of the same category again later. As is shown in Table 12, customers who firstly purchase cup are likely to buy cup again, but those who first purchase plate are likely to buy either cup or plate later.

6. Discussions

According to Swift (2001), Parvatiyar and Sheth (2001), and Kracklauer, Mills, and Seifert (2004), CRM consists of four dimensions such as customer identification, customer attraction, customer retention, and customer development. Customer identification is meant to identify segments of potential customers, each of whom includes customers who are relatively similar. Customer attraction attempts to attract the target customer segments by motivating customers to place orders through various channels (Cheung, Kwok, Law, & Tsui, 2003; He, Xu, Huang, & Deng, 2004; Liao & Chen, 2004). 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. (Chen, Chiu, &

Table 10
Association rules among subcategories.

Lift	Support (%)	Confidence (%)	Rule
1.32	5.17	38.65	Bowl => Cup
1.10	5.06	32.20	Bag => Cup
1.20	4.48	35.00	Box => Cup
1.14	4.32	33.47	Bottle => Cup
1.20	3.94	21.89	Tray => plate
1.20	3.94	21.64	Plate => Tray
1.35	3.78	24.65	Dosirak => Plate
1.35	3.78	20.76	Plate => Dosirak
1.15	3.30	21.02	Bag => Plate
1.11	3.14	20.00	Bag => Tray

Table 11
Sequential patterns among categories.

Support (%)	Confidence (%)	Rule
8.58	44.34	A => E => A
8.31	40.81	E => A => A
7.31	47.34	A => A and E => A
7.31	40.82	A and E => A => A
7.21	42.47	E => E => A
6.21	43.59	A and E => E => A
6.12	47.18	E => A and E => A

Table 12
Sequential patterns among subcategories.

Support (%)	Confidence (%)	Rule
7.85	21.39	Cup => Cup
5.48	24.49	Plate => Plate
4.29	20.18	Plate => Cup
3.65	20.62	Bottle => Cup

Chang, 2005; Jiang & Tuzhilin, 2006). 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. (Etzion, Fisher, & Wasserkrug, 2005; Rosset, Neumann, Eick, & Vatnik, 2003).

Based on these dimensions of CRM and the results obtained from this study, we decided to suggest CRM strategies against the VIP customers for the online shopping mall as follows:

- The classification model can be made use of to *identify VIP customers*, so that the online shopping mall can exercise marketing activities against them. It will be more cost-effective than doing the same thing against all the customers. Of course, it will be necessary to develop a new classification model based on a new definition of VIP customer as more customer transaction data is accumulated. This cycle of collecting data, building models and using them for marketing activities should be continuous.
- Since category 'E' is highly associated with categories 'A', 'B', and 'D' as shown in Table 9, it is recommended that the online shopping mall redesign their web site so that the page associated with category 'E' can be one-click away from categories 'A', 'B' or 'D'. This will enhance the level of satisfaction of customers, which is good for *customer retention*.
- Although cup and plate are strongly included in most association rules and sequential patterns, the current keywords of the online shopping mall in major search sites do not include or imply cup and/or plate. *To attract prospective customers much more*, it is highly suggested that the online shopping mall properly set current keywords to those which imply cup and/or plate with specific characteristics that are unique to the cups and plates dealt with by the mall.
- Through the purchasing patterns in the form of association rules and sequential patterns found in this study, we can develop personalized promotion strategies, which may be helpful for *customer development*. From Table 10, we suggest the following recommendation or bundling strategy:
  - We can recommend cups to the customers who purchased bowls. The same strategy can be applied to the other association rules among subcategories.
  - Especially, since (tray and plate) and (dosirak and plate) have been sold at the same time, it is highly recommended that they are sold as a bundle of items.

    From Tables 11 and 12, we suggest the following strategies.
  - Having recognized the time gap between the first purchase and the second in a sequential pattern, we can send mails at the right time to the customers who purchased the item in the left hand side of the sequential pattern in order to remind them to buy the item in the right hand side of the sequential pattern.

7. Conclusions

In this study, we defined the VIP based on RFM values. Then, we developed classification models which classify customers into VIP or non-VIP and compared their hit ratios. Decision Tree and the bagging with decision tree show the highest hit ratios among them. Then, we extracted 16 association rules and 11 sequential patterns from the transaction data of VIP customers, and suggested strategies for the small-sized online shopping mall based on these identified rules and patterns. It seems that the strategies suggested by this study can be utilized to enhance the revenue of the online shopping mall more efficiently than before.

This study has a few things to be desired. First, the target online shopping mall is of small size and of short history and thus the dataset we have used for the analysis is not big enough to derive more meaningful results. Second, for the same reason, the VIP customers we have selected account for almost half of the whole customers. The Pareto rule does not hold in this mall. Last, this study should have continued to see if the strategies derived from the actionable knowledge obtained by data mining techniques are effective or not. We plan to do it in the near future. Nonetheless, we believe that such experiments as conducted in this study deserve to be paid attention of small-sized online shopping malls where a huge amount of useful data still remains unused.

References

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

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

Anderson, J. L., Jolly, L. D., & Fairhurst, A. E. (2007). Customer relationship management in retailing: A content analysis of retail trade journals. *Journal of Retailing and Consumer Services, 14*, 394–399.

Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P., & Vanthienen, J. (2004). Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. *European Journal of Operational Research, 156*, 508–523.

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

Chen, M.-C., Chiu, A.-L., & Chang, H.-H. (2005). Mining changes in customer behavior in retail marketing. *Expert Systems with Applications, 28*, 773–781.

Chen, Y. L., Kuo, M. H., Wu, S. Y., & Tang, K. (2009). Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data. *Electronic Commerce Research and Applications, 8*, 241–251.

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

Cheung, K. W., Kwok, J. T., Law, M. H., & Tsui, K. C. (2003). Mining customer product ratings for personalized marketing. *Decision Support Systems, 35*, 231–243.
Chiang, D. A., Wang, Y. F., Lee, S. L., & Lin, C. J. (2003). Goal-oriented sequential pattern for network banking churn analysis. *Expert Systems with Applications, 25*, 293–302.
Chiang, W. Y. (2010). *To mine association rules of customer values via a data mining procedure with improved model: An empirical case study*. Expert Systems with Applications.
Chien, C. F., & Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with Applications, 34*, 280–290.
Dennis, C., Marsland, D., & Cockett, T. (2001). Data mining for shopping centres-customer knowledge-management framework. *Journal of Knowledge Management, 5*, 368–374.
Etzion, O., Fisher, A., & Wasserkrug, S. (2005). E-CLV: A modeling approach for customer lifetime evaluation in e-commerce domains, with an application and case study for online auction. *Information Systems Frontiers, 7*, 421–434.
Frosyniotis, D., Stafylopatis, A., & Likas, A. (2003). A divide-and-conquer method for multi-net classifiers. *Pattern Analysis & Applications, 6*, 32–40.
He, Z., Xu, X., Huang, J. Z., & Deng, S. (2004). Mining class outliers: concepts, algorithms and applications in CRM. *Expert Systems with Applications, 27*, 681–697.
Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications, 37*, 5259–5264.
Huang, S. C., Chang, E. C., & Wu, H. H. (2009). A case study of applying data mining techniques in an outfitter's customer value analysis. *Expert Systems with Applications, 36*, 5909–5915.
Hughes, A. M. (1994). *Strategic database marketing*. Chicago: Probus Publishing Company.
Hwang, H., Jung, T., & 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.
Jiang, T., & Tuzhilin, A. (2006). Segmenting customers from population to individuals: Does 1-to-1 keep your customers forever. *IEEE Transactions on Knowledge and Data Engineering*, 1297–1311.
Kang, H. J., & Doermann, D. (2003). Evaluation of the information-theoretic construction of multiple classifier systems. *Document Analysis and Recognition, 2*, 789.
Kim, S. Y., Jung, T. S., Suh, E. H., & 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., & Street, W. N. (2004). An intelligent system for customer targeting: A data mining approach. *Decision Support Systems, 37*, 215–228.
Kracklauer, A. H., Mills, D. Q., & Seifert, D. (2004). Customer management as the origin of collaborative customer relationship management. *Collaborative Customer Relationship Management*, 3–6 [taking CRM to the next level].
Langley, P., & Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM, 38*, 54–64.
LariviÈRe, B., & Van den Poel, D. (2005). Investigating the post-complaint period by means of survival analysis. *Expert Systems with Applications, 29*, 667–677.
Lau, H., Wong, C., Hui, I., & Pun, K. (2003). Design and implementation of an integrated knowledge system. *Knowledge-Based Systems, 16*, 69–76.
Liao, S. H., & Chen, Y. J. (2004). Mining customer knowledge for electronic catalog marketing. *Expert Systems with Applications, 27*, 521–532.
Newell, F. (1997). *The new rules of marketing: How to use one-to-one relationship marketing to be the leader in your industry*. New York: McGraw-Hill.
Parvatiyar, A., & Sheth, J. N. (2001). Customer relationship management: Emerging practice, process, and discipline. *Journal of Economic and Social Research, 3*, 1–34.
Roli, F., Kittler, J., & Windeatt, T. (2004). *Multiple classifier systems*. Springer.
Rosset, S., Neumann, E., Eick, U., & Vatnik, N. (2003). Customer lifetime value models for decision support. *Data Mining and Knowledge Discovery, 7*, 321–339.
Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures∗. *Decision Sciences, 23*, 899–916.
Sinha, A. P., & Zhao, H. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Decision Support Systems, 46*, 287–299.
Swift, R. S. (2001). *Accelerating customer relationships: Using CRM and relationship technologies*. Prentice Hall.
Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science, 38*, 926–947.
Tsai, C. F., & Chen, M. Y. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems with Applications, 37*, 2006–2015.
Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications, 36*, 12547–12553.
Weiber, R., & Kollmann, T. (1998). Competitive advantages in virtual markets-perspectives of ¡°information-based marketing¡± in cyberspace. *European Journal of Marketing, 32*, 603–615.
Wu, C. H., Kao, S. C., Su, Y. Y., & Wu, C. C. (2005). Targeting customers via discovery knowledge for the insurance industry. *Expert Systems with Applications, 29*, 291–299.
Yu, J. X., Ou, Y., Zhang, C., & Zhang, S. (2005). Identifying interesting visitors through Web log classification. *IEEE Intelligent Systems, 20*, 55–59.
Zhang, G. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research, 116*, 16–32.