Expert Systems with Applications 36 (2009) 4176–4184

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Classifying the segmentation of customer value via RFM model and RS theory

Ching-Hsue Cheng, You-Shyang Chen *

Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 640, Taiwan

ARTICLE INFO

Keywords:
CRM (customer relationship management)
Customer value analysis
K-means algorithm
Rough set theory
RFM (recency, frequency and monetary) model

ABSTRACT

Data mining is a powerful new technique to help companies mining the patterns and trends in their customers data, then to drive improved customer relationships, and it is one of well-known tools given to customer relationship management (CRM). However, there are some drawbacks for data mining tool, such as neural networks has long training times and genetic algorithm is brute computing method. This study proposes a new procedure, joining quantitative value of RFM attributes and K-means algorithm into rough set theory (RS theory), to extract meaning rules, and it can effectively improve these drawbacks. Three purposes involved in this study in the following: (1) discretize continuous attributes to enhance the rough sets algorithm; (2) cluster customer value as output (customer loyalty) that is partitioned into 3, 5 and 7 classes based on subjective view, then see which class is the best in accuracy rate; and (3) find out the characteristic of customer in order to strengthen CRM.

A practical collected C-company dataset in Taiwan's electronic industry is employed in empirical case study to illustrate the proposed procedure. Referring to [Hughes, A. M. (1994). *Strategic database marketing*. Chicago: Probus Publishing Company], this study firstly utilizes RFM model to yield quantitative value as input attributes; next, uses K-means algorithm to cluster customer value; finally, employs rough sets (the LEM2 algorithm) to mine classification rules that help enterprises driving an excellent CRM. In analysis of the empirical results, the proposed procedure outperforms the methods listed in terms of accuracy rate regardless of 3, 5 and 7 classes on output, and generates understandable decision rules.

1. Introduction

Due to the complication and diversification of business operation, information of company is essential and vital forces for advantage competition and going-concern. Particularly, the growing of information technology (IT) in rapid changing and competitive environment today motivates the activity of transaction, which increasingly facilities the markets competition. Based on this relationship, information serves as central to face the opportunities and challenges of day-to-day operation for companies. It is very difficult for companies that strengthen business's competitive advantage if information only becomes to support the functions within company when facing to the heavy challenges coming from outsides surroundings. Thus, how to enhance the market competitive power for companies is an interesting issue because of the more the competitive power, the more the probability for going-concern. The key point gaining profit of companies is to integrate the upstream members of supply chain via an effective IT in order to reduce cost, and reinforce the downstream customer relationships via an excellent CRM in order to gain more profit. CRM becomes the focal point of company profits and more and more important for companies because customers are main resources of profits. Therefore, this study insists on that an excellent CRM with customers for companies is a critical for gaining more profit.

The fulfillment of customer requirements is one of key factors for the success of business operation. CRM is to achieve the needs of customers and to enhance the strength with customers for company (Thompson & Sims, 2002). However, the effective and efficient utilization of IT to support the CRM process is short path for successful CRM. Although understanding the situations of customers is somewhat different, the companies that all provide products and services for customers to satisfy their demands are similar to mine valuable information of customers, to realize the customer value maximization, to increase customer loyalty and finally to obtain plenty profits for themselves (Joo & Sohn, 2008). Therefore, a large number of companies apply the different tools such as computer software package, statistical techniques, to enhance a more efficient CRM, in order to let companies understanding more about their customers.

Nowadays, by utilizing data mining tools for assisting CRM, some techniques, which include decision trees (DT), artificial neural networks (ANN), genetic algorithms (GA), association rules (AR), etc., are usually used in some fields such as engineering, science, finance, business, to solve related problems with customers (Witten & Frank, 2005). A decision tree is a flow-chart-like tree structure, where each

* Corresponding author.
E-mail address: g9523804@yuntech.edu.tw (Y.-S. Chen).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.04.003

internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent class or class distributions (Han & Kamber, 2001). An artificial neural network is a large number of highly interconnected processing elements (neurons) that uses a mathematical model, computational model or non-linear statistical data modeling tools for information processing to capture and represent complex input/output relationships. Genetic algorithms, which were formally introduced in the United States in the 1970s by John Holland at University of Michigan, are search algorithms applied to solve problems on a computer based on the mechanics of natural selection and the process of natural evolution (Holland, 1973; Miller, Todd, & Hegde, 1989). Association Rules based on co-occurrence can be used to address relationships that customers which buy *X* tend to buy *Y*, and to support related activity of business operations such as product promotions, CRM programs, and inventory control. The problem of mining association rule has firstly been stated in 1993 and is one of the most used tools to find the relationships among items (products) in large databases today (Dunham, 2003).

In recent years, data mining has not only a great popularity in research area but also in commercialization. Data mining can help organizations discovering meaningful trends, patterns and correlations in their customer, product, or data, to drive improved customer relationships and then decrease the risk of business operations. The basic data mining techniques include classification, clustering, association rules, regression analysis, sequence analysis, etc. Other data mining techniques include rule-based reasoning approach, genetic algorithms, decision trees, fuzzy logic, inductive learning systems, statistical methods, and so forth (Witten & Frank, 2005).

Generally, no tool for data mining in CRM is perfect because there are some uncertain drawbacks in it. For example, in decision trees, too many instances lead to large decision trees which may decrease classification accuracy rate and do not clearly create the relationships which come from the training examples. In artificial neural networks, number of hidden neurons, number of hidden layers and training parameters need to be determined, and ANN has long training times in a large dataset especially. Moreover, ANN served as "black box" which leads to inconsistency of the outputs, is a trial-and-error process. In genetic algorithm, GA also has some drawbacks such as slow convergence, a brute computing method, a large computation time and less stability. In association rules, major drawback is the number of generated rules is huge and may be a redundancy.

For solving the problems of previous paragraph, two methods, K-means algorithm and RS theory, are worth to be explored in this study. K-means is one of the well-known algorithms for cluster analysis and it has been used extensively in various fields including data mining, statistical data analysis and other business applications. Cluster analysis is a statistical technique that are used to identify a set of groups that both minimize within-group variation and maximize between-group variation based on a distance or dissimilarity function, and its aim is to find an optimal set of clusters (Witten & Frank, 2005). With respect to rough set theory (RS theory), five advantages are expressed in the following: (1) the RS theory do not require any preliminary or additional parameter about the data; (2) they can work with missing values, switch among different reducts, and use less expensive or time to generate rules; (3) they offer the ability to handle large amounts of both quantitative and qualitative data; (4) they yield understandable decision rules and own stability; and (5) they can model highly non-linear or discontinuous functional relationships provides a powerful method for characterizing complex, multidimensional patterns (Hashemi, LeBlanc, & Rucks, 1998; Pawlak, 1982). Thus, this study is on the use of some techniques (i.e. RS theory) to cope with these shortcomings and then to improve CRM for enterprises, based on RFM (recency–frequency–monetary) attributes and K-means method for clustering customer value. Generally, the purpose of this study is to generate classification rules for achieving an excellent CRM which are believed to maximize profits with win–win situation for company–customer.

The rest of the paper is organized in the following: In Section 2 we describe an overview of the related works, while Section 3 presents the proposed procedure and briefly discusses its architecture. Section 4 describes analytically the experimental results. Finally, Section 5 concludes the paper.

2. Related works

This study proposes an enhanced rough sets method to verify that whether it can be helpful on operation management in enterprises. This section mainly explores the issue or theoretical parts of operation model and management, and some techniques for clustering customer value. Thus, this study reviews related studies of the customer relationship management, customer value analysis, K-means algorithm, rough set theory and the LEM2 rule extraction method.

2.1. Customer relationship management (CRM)

CRM is a philosophy of business operation for acquiring and retaining customers, increasing customer value, loyalty and retention, and implementing customer-centric strategies. CRM, devoted to improve relationships with customer, focuses on a comprehensive picture on how to integrate customer value, requirements, expectations and behaviors via analyzing data from transaction of customer (Peppard, 2000). Enterprises can shorten sales cycle and increase customer loyalty to build better close relationships with customers and further add revenues by good CRM. Thus, an excellent CRM can help enterprises keeping existing customers and attracting new ones.

Enterprises apply some methods to effectively enhance customer relationships, which include customer relationship management, customer value analysis, enterprise strategy, and positive service mechanisms. Moreover, enterprises also strengthen marketing and sales effectiveness in order to build good CRM. Kalakota and Robinson (1999) explained that the CRM is to integrate the function of the related fields with customer in the enterprise such as marketing, sales, services and technical support for customer needs, and it usually utilizes IT to help an enterprise managing relationships with customer in a systematic way, improving customer loyalty and increasing overall business profits (Kalakota & Robinson, 1999).

It has been estimated that it costs five times as much to attract a new customer as it does to retain a existing one, according to research by the American management Association (Kotler, 1994; Peppers & Rogers, 1996) and this relationship is particularly obvious in the services sector (Ennew & Binks, 1996). Therefore, enterprises understand the importance of developing a good close relationship with existing and new customers. Instead of attracting new customers, they would like to perform as well as possible more business operations for customers in order to keep existing customers and build up long-term customer relationship. Based on this reason, this study ensures that enterprises should be implementing customer value analysis to understand about their customers, to retain valuable customers and finally to bring plenty profits for themselves.

2.2. Customer value analysis

Customer value analysis is a kind of analytic method for discovering customers' characteristics and makes a further analysis of specific customers to abstract useful knowledge from large data. Thus, it is clear that enterprises apply value analysis method to customers for knowing about who are the target customers which contribution is outstanding. Kaymak (2001) further pointed out that the RFM model is one of the well-known customer value analysis methods. Its advantage is to extract characteristics of customers by using fewer criterions (a three-dimension) as cluster attributes so that reduce the complexity of model of customer value analysis (Kaymak, 2001). Moreover, from view of the consuming behavior, Schijns and Schroder (1996) also deemed that the RFM model is a long-familiar method to measure the strength of customer relationship (Schijns & Schroder, 1996). Retention cost is far less costly than acquisition cost (Kotler, 1994; Peppers & Rogers, 1996); therefore, enterprises are intent via using RFM analysis to mine databases for knowing about customers who spend the most money and create the biggest value for enterprises.

In this study, different groups of customers are segmented using their consuming behavior via RFM attributes. By this way, we ensure that the standards which cluster customer value are not established subjectively, so that the clustering standards are established objectively based on RFM attributes. That is, customer segmentation is established only by three attributes (recency–frequency–monetary of purchase). Therefore, this study uses RFM model (fewer attributes) of customer value analysis as cluster attributes, and then improve the close relationships of enterprises with customers.

2.2.1. RFM model definition

The RFM analytic model is proposed by Hughes (1994), and it is a model that differentiates important customers from large data by three variables (attributes), i.e., interval of customer consumption, frequency and money amount. The detail definitions of RFM model are described as follows:

(1) Recency of the last purchase (R).
R represents recency, which refers to the interval between the time that the latest consuming behavior happens and present. The shorter the interval is, the bigger R is.
(2) Frequency of the purchases (F).
F represents frequency, which refers to the number of transactions in a particular period, for example, two times of one year, two times of one quarter or two times of one month. The many the frequency is, the bigger F is.
(3) Monetary value of the purchases (M).
M represents monetary, which refers to consumption money amount in a particular period. The much the monetary is, the bigger M is.

According to the literature (Wu & Lin, 2005), researches showed that the bigger the value of R and F is, the more likely the corresponding customers are to produce a new trade with enterprises. Moreover, the bigger M is, the more likely the corresponding customers are to buy products or services with enterprises again. RFM method is very effective attributes for customer segmentation (Newell, 1997).

2.2.2. RFM model weight setup

Although RFM model is a good method that differentiates important customers from large data by three variables, there are two studies, (Hughes, 1994; Stone, 1995), having some different opinions with respect to the three variables of RFM model. Hughes (1994) considered that the three variables are equal in the importance (Hughes, 1994); therefore, the weights of the three variables are identical. On the other hand, Stone (1995) indicated that the three variables are different in the importance due to the characteristic of industry (Stone, 1995). Thus, the weights of the three variables are not equal.

2.3. K-means algorithm

Clustering is the process of grouping a set of physical or abstract objects into groups of similar objects. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters (Han & Kamber, 2001). K-means is one of the well-known algorithms for clustering, originally known as Forgy's method (Forgy, 1965), and it has been used extensively in various fields including data mining, statistical data analysis and other business applications. Thus, this study proposes the K-means algorithm to build clusters by attributes (i.e. R–F–M attributes). The K-means algorithm for partitioning is base on the mean value of the objects in the cluster. MacQueen (1967) suggested the term K-means for describing an algorithm that assigns each item to the cluster with the nearest centroid (mean) (MacQueen, 1967). Based on the concept above, the computing process for K-means is presented as follows:

Step 1: Partition the items into K initial clusters.
Firstly, partition the items (*m* objects) into K initial clusters.
Step 2: Proceed through the list of items.
Assign an item to the cluster whose centroid is nearest (distance is computed by using Euclidean distance with either standardized or un-standardized observations) and re-calculate the centroid for the cluster receiving the new item or for the cluster losing the item.
Step 3: Repeat Step 2 until no more reassigning.
Rather than starting with a partition of all items into K preliminary groups in Step 1, we could specify K initial centroids (seed points) and then proceed to Step 2. The final assignment of items to clusters will be, to some extent, dependent upon the initial partition or the initial selection of seed points. Experience suggests that most major changes in assignment occur with the first reallocation step.

2.4. Rough set theory (RS theory)

Rough set theory, first proposed by Pawlak (1982) in 1982, employed mathematical modeling to deal with class data classification problems, and then turned out to be a very useful tool for decision support systems, especially when hybrid data, vague concepts and uncertain data were involved in the decision process. To use the rough set process, one begins with a relational database, a table of objects with attributes, and attributes values for each object. One attribute is chosen as the decision attribute, then the rest of the attributes are the condition attributes (Pawlak, 1982). Rough sets address the continuing problem of vagueness, uncertainty and incompletion by applying the concept of equivalence classes to partition training instances according to specified criteria. Two partitions are formed in the mining process. The members of the partition can be formally described by unary set-theoretic operators or by successor functions for lower approximation and upper approximation spaces from which both possible rules and certain rules can be easily derived. Vague and imprecise data sets have no clear-cut boundaries. Thus, the rough set theory approach is based on refusing certain set boundaries, implying that every set will be roughly defined using a lower and an upper approximation.

Let $B \subseteq A$ and $X \subseteq U$ be an information system. The set X is approximated using information contained in B by constructing lower and upper approximation sets:

$\underline{B}X = \{x | [x]_B \subseteq X\}$ (Lower approximation),

and $\overline{B}X = \{x | [x]_B \cap X \neq \emptyset\}$ (Upper approximation).

The elements in $\underline{B}X$ can be classified as members of X by the knowledge in B. However, the elements in $\overline{B}X$ can be classified as possible members of X by the knowledge in B. The set $BN_B(x) = \overline{B}X - \underline{B}X$ is called the B-boundary region of X and it consists of those objects that cannot be classified with certainty as members of X with the knowledge in B. The set X is called "rough" (or "roughly definable") with respect to the knowledge in B if the boundary region is non-empty. Rough sets theoretic classifiers usually apply the concept of rough sets to reduce the number of attributes in a decision table (Pawlak, 1991) and to extract valid data from inconsistent decision tables. Rough sets also accept discretized (symbolic) input.

2.5. The LEM2 rule extraction method

A popular method is to induce a decision table transformed into rules (Stefanowski, 1998) that are focused on a minimal set of rules. In RS theory, decision rules are often induced from a given decision table.

Rough set rule induction algorithms were implemented for the first time in a LERS (Learning from Examples) (Grzymala-Busse, 1992) system. A local covering is induced by exploring the search space of blocks of attribute-value pairs which are then converted into the rule set. The rough set LEM2 (Learning from Examples Module, version 2) method is used in generation of decision rules in this study. The algorithm LEM2 (Grzymala-Busse, 1997) for rule induction is based on computing a single local covering for each concept from a decision table.

3. Methodology

This section briefly introduces the research model of this study and the proposed procedure for classifying customer value.

CRM is to achieve the needs of customers and enhance the strength with customers for company (Thompson & Sims, 2002). In recent years, data mining has not only a great popularity in research area but also in commercialization. Nowadays, by utilizing data mining tools for assisting CRM, some techniques, which include DT, ANN, GA, AR, etc., are usually used in some fields such as engineering, science, finance, business, to solve related problems with customers (Witten & Frank, 2005). Generally, no tool for data mining in CRM is perfect because there are some uncertain drawbacks in it. For example, in DT, too many instances lead to large decision trees and decrease classification accuracy rate. In ANN, it has long training times in a large dataset especially, and it is a trial-and-error process. In GA, it has slow convergence, a large computation time and less stability. In AR, it may generate huge rules that may be a redundancy.

For solving the problems above, two methods, K-means algorithm and RS theory, are worth to be presented in this study. This study proposes a new procedure, joining quantitative value of RFM attributes and K-means algorithm into RS theory (the LEM2 algorithm), to extract meaning rules, and it can effectively improve these drawbacks above.

3.1. Research model

This study constructs a model for clustering customer value based on RFM attributes and K-means algorithm. The RFM model is regarded as input attributes then to yield quantitative value for K-means clustering. Fig. 1 illustrates research model in this study.

3.2. The proposed procedure

In this subsection, we further explain the proposed procedure for classifying customer value. The proposed procedure can be divided into four processes: (1) select the dataset and preprocess data; (2) use recency, frequency and monetary to yield quantitative value as input attributes for cluster analysis, and cluster customer value as output (named customer loyalty) that is partitioned into 3, 5 and 7 classes based on subjective view by using K-means algorithm; (3) split dataset into training data and testing data, and then extract rules by rough sets (the LEM2 algorithm); and (4) finally, evaluate the results of experiment, compare with different methods and list the comparisons of experimental results in different classes level.

The computing process is introduced step by step as follows:

Step 1: Data preprocessing.
At first, select the dataset for empirical case study. To preprocess the dataset to make knowledge discovery easier is needed. Thus, we firstly delete the records which include missing values or inaccurate values, eliminate the redundant attributes and transform the datum into a format that will be more easily and effectively processed for clustering customer value.

Step 2: Cluster customer value by K-means algorithm.
The following step is to define the scaling of R–F–M attributes based on Hughes (1994) and yield quantitative value of RFM attributes as input attributes, then cluster customer value by using K-means algorithm. The detail process of this step is expressed into two sub-steps.

Step 2-1: Define the scaling of R–F–M attributes.
This sub-step process is divided into five parts introduced in the following:

(1) The R–F–M attributes are equal weight (i.e. 1:1:1).
(2) Define the scaling of three R–F–M attributes, which are 5, 4, 3, 2 and 1 that refer to the customer contributions to revenue for enterprises. The '5' refers to the most customer contribution to revenue and '1' refers to the least contribution to revenue.
(3) Sort the data of three R–F–M attributes by descendant order.
(4) Partition the three R–F–M attributes respectively into 5 equal parts and each part is equal to 20% of all. The five parts are assigned 5, 4, 3, 2 and 1 score by descendant order (see Table 1).

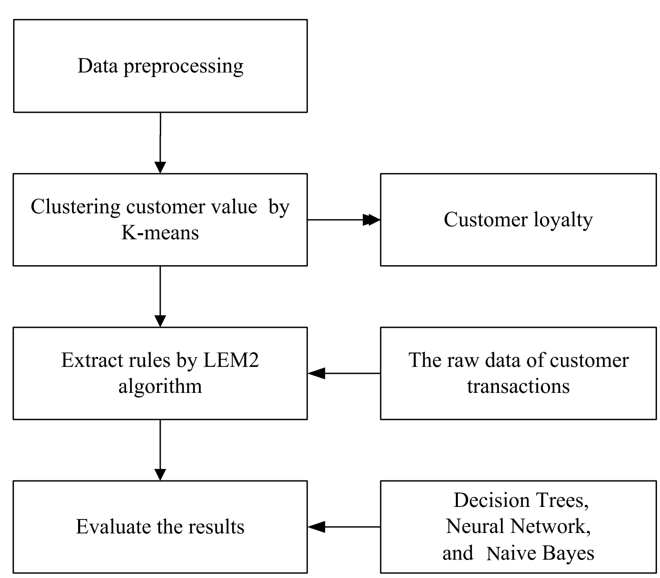

Fig. 1. Research model.

(5) Yield quantitative value of R–F–M attributes according to previous process (4) as input attributes for each customer. There are total 125 ($5 \times 5 \times 5$) combinations since each attribute in R–F–M attributes has 5 scaling (5, 4, 3, 2 and 1) (see Table 2).

Step 2-2: Cluster customer value by K-means algorithm.
According to quantitative value of R–F–M attributes for each customer, partition data (m objects) into K clusters using the K-means algorithm for clustering customer value.
Firstly, let $k = 5$ clusters (i.e. C_1, C_2, ..., and C_5) by clustering methods. Furthermore, we obtain that the center of the cluster C_1, C_2, C_3, C_4 and C_5 is c_1, c_2, c_3, c_4 and c_5, respectively as follows:

$$C_1 = \{S_{11}, S_{12}, \ldots, S_{1P_1}\}, \quad C_2 = \{S_{21}, S_{22}, \ldots, S_{2P_2}\}, \ldots,$$
$$C_5 = \{S5_{51}, S_{52}, \ldots, S_{5P_k}\},$$

where $1 \leqslant k \leqslant m$, S_{ij} denotes the jth element of C_i, P_i denotes the number of elements in C_i and$1 \leqslant i \leqslant k$.Due to the three attributes (R–F–M) (we set R, F, M = 1, 2, 3, respectively), we obtain c_1, c_2,..., and c_5 is:

$$c_1 = (v_{11}, v_{12}, v_{13}),$$
$$c_2 = (v_{21}, v_{22}, v_{23}),$$
$$\vdots$$
$$c_5 = (v_{51}, v_{52}, v_{53}).$$

Secondly, compute the distance D_i between c_i and zero point:

$$D_1 = \sqrt{(v_{11} - 0)^2 + (v_{12} - 0)^2 + (v_{13} - 0)^2},$$
$$D_2 = \sqrt{(v_{21} - 0)^2 + (v_{22} - 0)^2 + (v_{23} - 0)^2},$$
$$\vdots$$
$$D_5 = \sqrt{(v_{51} - 0)^2 + (v_{52} - 0)^2 + (v_{53} - 0)^2}.$$

Thirdly, sort the distance D_1 to D_5 by descendant order and then based on this order, give the cluster C_1 to C_5 a class of customer loyalty. We name the 5 classes as Very High, High, Medium, Low and Very Low. 'Very High' refers to most customer loyalty and 'Very Low' refers to least customer loyalty.
Finally, compute the data of customers to decide to which class is belonging.

Table 1
The scaling of R–F–M attributes

	R – Recency (%)	F – Frequency (%)	M – Monetary (%)
5 Score	0–20	0–20	0–20
4 Score	20–40	20–40	20–40
3 Score	40–60	40–60	40–60
2 Score	60–80	60–80	60–80
1 Score	80–100	80–100	80–100

Table 2
Quantitative value of R–F–M attributes (Hughes, 1994)

R	F	M	R	F	M	R	F	M	R	F	M	R	F	M
5	5	5	4	5	5	3	5	5	2	5	5	1	5	5
5	5	4	4	5	4	3	5	4	2	5	4	1	5	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5	1	1	4	1	1	3	1	1	2	1	1	1	1	1

Table 3
The description of attributes (input data)

Attributes	Scaling name
Credit amount	High, Medium, Low
Recency	Very High, High, Medium, Low, Very Low
Frequency	Very High, High, Medium, Low, Very Low
Monetary	Very High, High, Medium, Low, Very Low

Step 3: Generate rules by LEM2 algorithm.
After Step 1 to Step 2, we can obtain four attributes (input data), credit amount, recency, frequency and monetary (see Table 3), which are discretized continuous attributes; however, recency refers to the latest trading date of each customer in 2006, frequency refers to the total order number of transactions of each customer in 2006, and monetary refers to average money amount of each customer in 2006 in this study. The following, in addition to a class (output data, see Step 2-2) and an area (an attribute), is to build the decision table. Based on these attributes and a class, generate rules by rough sets (the LEM2 algorithm) to extract classification rules for achieving an excellent CRM for enterprises.

Step 4: Repeat Step 2 to Step 3 with 3 and 7 classes on output.Similarly, we repeat Step 2 to Step 3, and 3 classes on output that are High, Medium and Low replace 5 classes on output.Accordingly, we repeat Step 2 to Step 3 once more, and 3 classes on output are replaced by 7 classes on output, which are Extremely High, Very High, High, Medium, Low, Very Low and Extremely Low.

Step 5: Evaluate the results.Furthermore, evaluate the accuracy rate of the generated rules, compare with different methods, Decision Tree (Quinlan, 1993), Artificial Neural Networks (Rumelhart & McClelland, 1986) and Naive Bayes (Cestnik, 1990), and list the comparisons of experimental results in different class level. Finally, analyze the difference of these different methods, explore the reasons and propose the conclusions.

4. Empirical case study

In this section, we introduce the empirical case (C-company) and the computing process using C-company dataset.

4.1. The introduce of C-company case

C-company, founded in 1991 on Changhua, Taiwan, is an international company and belonging to electronic industry focusing on design manufactory service (DMS) of power supply. Nowadays, its employees are about 400 persons. Based on the partnership with large companies as well as well-known enterprises in global, the C-company has been success to implement enterprise resource planning systems (ERPs) for integrating all data and processes of itself into a unified system since 2004. It builds complete architecture of IT using Oracle 9i database setting up on platform of IBM AIX5.3, supports the global logistic management, integrates total supply chain linkage and becomes the important partner for its suppliers and customers in global. Its datum of customers has over 1000 records and its records of order with customers have over 35,000 today. Due to the rapid growing of IT, this motivates C-company increasing the revenue of sales. Based on this relationship, it is clear that C-company will be a great expansion in scale and more gaining a good fame in the future.

4.2. The computing process using C-company dataset

A practical collected dataset, the C-company for electronic industry in Taiwan, is used in this empirical case study to demonstrate the proposed procedure from 2006/01/01 to 2006/12/31.

The computing process using C-company dataset can be expressed in detail as follows:

Step 1: Data preprocessing.
At first, select the dataset of C-company, which is extracted only from the raw data in 2006, then delete the records which include missing values and inaccurate values, and eliminate the redundant attributes. Next, change the data into appropriate formats. Finally, the dataset remains 401 instances which are characterized by the following seven fields: (i) ID, (ii) Area, (iii) Country, (iv) Recency (Rcen), (v) Frequency (Freq), (vi) Monetary (Money), (vii) Credit amount (Credit). The partial data of C-company dataset is shown in Table 4. However, only the five attributes, area, recency, frequency, monetary and credit, and one class (output), loyalty, are used to build the decision table.

Step 2: Cluster customer value by K-means algorithm.
The following step is based on Hughes (1994) to define the scaling of R–F–M attributes and yield quantitative value of RFM attributes as input attributes, then cluster customer value by using K-means algorithm. The detail process of this step is expressed into two sub-steps as follows:

Step 2-1: Define the scaling of R–F–M attributes.
This sub-step process is mainly divided into five parts introduced in the following:

(1) The R–F–M attributes are equal weight (i.e. 1:1:1).
(2) Define the scaling of three R–F–M attributes, which are 5, 4, 3, 2 and 1.
(3) Sort the data of three R–F–M attributes by descendant order.
(4) Partition the real data of R–F–M attributes respectively into 5 scaling in C-company dataset with 401 instances (see Table 5).
(5) Yield quantitative value of R–F–M attributes based on Table 5 as input attributes for each customer (see Table 6).

Step 2-2: Cluster customer value by K-means algorithm.
According to quantitative value of R–F–M attributes (Table 6) for each customer, partition data (401 objects) into five clusters using K-means algorithm for clustering customer value. The cluster results of K-means see Table 7.

Step 3: Generate rules by LEM2 algorithm.
After Step 1 to Step 2, we can obtain four attributes (input data), credit amount, recency, frequency and monetary, which are discretized continuous attributes; however, recency refers to the latest trading date of each customer in 2006, frequency refers to the total order number of transactions of each customer in 2006, and monetary refers to average money amount of each customer for a order in 2006. The following, in addition to a class (output data, see Step 2-2) and an area (an attribute), is to build the decision table (see Table 8).
For verification, comparison and enhancing the accuracy rate, we set pre-conditions in the following: (1) Let Coverage = 0.6 and 0.9 then run the experiment (coverage refers to the parameter value from LEM2 algorithm and the default value of coverage is 0.9); (2) The dataset (401 instances) is split up into two sub-datasets: the 67% dataset (268 instances) is used as a training set, and the other 33% (133 instances) is used as a testing set; and (3) The experiment is repeated ten times with the 67%/33% random split then compute the average accuracy rate. Using these 5 attributes in addition to a class from Table 8 generate rules by rough

Table 4
The partial data of C-company dataset

ID	Area	Country	Recency	Frequency	Monetary	Credit amount
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0455	UK	UK	21-Dec-06	224	81,043,477	
0476	DE	DE	30-Nov-06	222	75,922,177	10000
0563	CH	CH	18-Oct-06	1349	62,275,041	
0656	AMER	US	07-Dec-06	146	42,791,553	5000
0718	AMER	US	25-Sep-06	69	37,565,431	
4001	LOCA	TW	29-Dec-06	189	34,742,150	5000
5240	UK	UK	03-Aug-06	107	19,096,458	
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 6
Quantitative value of R–F–M attributes for C-company dataset

ID	R	F	M
⋮	⋮	⋮	⋮
0004	4	4	2
0005	5	5	4
0006	5	5	4
0007	5	5	3
0008	5	5	4
0009	4	4	2
0010	5	4	2
0011	3	1	1
0012	5	5	3
0017	4	3	1
⋮	⋮	⋮	⋮

Table 7
The cluster results by K-means with 5 classes on output

Cluster center	C_1	C_2	C_3	C_4	C_5
c_1_R	4.52	3.68	1.58	3.58	1.67
c_2_F	4.54	1.96	1.96	2.89	1.13
c_3_M	3.62	4.22	4.30	1.64	1.66
The distance to zero point	7.36	5.93	4.98	4.89	2.61
Loyalty (output)	Very High	High	Medium	Low	Very Low
Number of instances	105	50	77	76	93

Table 5
The real scaling of R–F–M attributes in dataset of C-company

Scaling	Scaling name	R – Recency	F – Frequency	M – Monetary
5 Score	Very High	2006/12/19 ~ 2006/12/31	Over 30 Records	Over 250,001
4 Score	High	2006/11/10 ~ 2006/12/18	6–29 Records	103,001–250,000
3 Score	Medium	2006/09/15 ~ 2006/11/09	3–5 Records	52,151–103,000
2 Score	Low	2006/06/15 ~ 2006/09/14	2 Records	20,001–52,150
1 Score	Very Low	2006/01/01 ~ 2006/06/14	One record	Under 20,000

Table 8
Decision table

Case				Attributes			Decision
Sequence	ID	Area	Credit amount	Recency	Frequency	Monetary	Loyalty
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
249	0743	AMER	High	High	Very High	High	High
250	0744	AMER	High	Low	Very High	Medium	Medium
251	0745	AMER	Low	Very Low	Very High	Low	High
252	0747	ASIA	Low	Very Low	Low	Low	Very Low
253	0751	AMER	High	Medium	Medium	Very Low	Low
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

sets (the LEM2 algorithm) to extract classification rules for achieving an excellent CRM for enterprises. Meanwhile, the results of experiment are displayed in Table 9. Tables 10 and 11 show the first 10 rules for classifying customer value on coverage = 0.9 and coverage = 0.6, respectively.

Step 4: Repeat Step 2 to Step 3 with 3 and 7 classes on output. Similarly, we repeat Step 2 to Step 3, and 3 classes on output that are High, Medium and Low replace 5 classes on output. Table 12 shows the cluster results by K-means with 3 classes on output.Accordingly, we repeat Step 2 to Step 3 once more, and 3 classes on output are replaced by 7 classes on output, which are Extremely High, Very High, High, Medium, Low, Very Low and Extremely Low. Table 13 shows the cluster results by K-means with 7 classes on output.

Step 5: Evaluate the results.
From Table 9, the average accuracy rate of experiment results is 0.9624 and 0.9554 based on coverage = 0.6 and coverage = 0.9, respectively, in this study. Table 14 shows comparisons of experiment results with different methods in term of the accuracy rate in different class level. With the empirical results, the proposed procedure outperforms the methods listed in terms of accuracy rate regardless of 3, 5 and 7 classes on output, and the output of proposed procedure is understandable decision rules.

Table 9
The results of experiment for C-company dataset

Rounds	Coverage = 0.6		Coverage = 0.9	
	Accuracy	Rules	Accuracy	Rules
1 round	0.933	33	0.967	72
2 rounds	0.95	37	0.947	66
3 rounds	0.945	36	0.944	67
4 rounds	0.924	30	0.966	65
5 rounds	1	35	0.989	74
6 rounds	1	33	0.908	73
7 rounds	0.983	34	0.978	69
8 rounds	0.984	41	0.966	58
9 rounds	0.905	36	0.945	69
10 rounds	1	32	0.944	65
Average	0.9624	34.7	0.9554	67.8

Notes: Coverage refers to the parameter value from LEM2 algorithm and the default value of coverage is 0.9.

Table 10
The first 10 rules of C-company dataset result rule set on coverage = 0.9

No.	Rules
1	(Rcen = VeryHigh)&(Freq = VeryHigh)=>(Class = VeryHigh)
2	(Freq = VeryLow)&(Rcen = VeryLow)&(Money = VeryLow)=>(Class = VeryLow)
3	(Freq = VeryLow)&(Money = Low)=>(Class = VeryLow)
4	(Credit = High)&(Rcen = VeryHigh)&(Freq = VeryHigh)=>(Class = VeryHigh)
5	(Freq = High)&(Credit = High)&(Money = Medium)=>(Class = VeryHigh)
6	(Freq = High)&(Money = Medium)=>(Class = VeryHigh)
7	(Rcen = Low)&(Money = High)=>(Class = Medium)
8	(Area = AMER)&(Rcen = Low)&(Money = VeryHigh)=>(Class = Medium)
9	(Money = VeryLow)&(Rcen = High)=>(Class = Low)
10	(Credit = High)&(Rcen = Low)&(Money = VeryHigh)=>(Class = Medium)

Table 11
The first 10 rules of C-company dataset result rule set on coverage = 0.6

No.	Rules
1	(Rcen = VeryHigh)& (Freq = VeryHigh)=>(Class = VeryHigh)
2	(Credit = High)&(Rcen = VeryHigh)&(Freq = VeryHigh)=>(Class = VeryHigh)
3	(Freq = VeryLow)&(Rcen = VeryLow)&(Money = VeryLow)=>(Class = VeryLow)
4	(Credit = High)&(Rcen = Low)&(Money = VeryHigh)=>(Class = Medium)
5	(Freq = VeryLow)&(Credit = High)&(Area = AMER)&(Money = VeryLow) =>(Class = VeryLow)
6	(Rcen = Low)&(Area = AMER)&(Money = VeryHigh)=>(Class = Medium)
7	(Rcen = Low)&(Money = High)=>(Class = Medium)
8	(Credit = High)&(Money = VeryLow)& (Rcen = High)=>(Class = Low)
9	(Rcen = VeryHigh)&(Credit = Low)& (Freq = VeryHigh)=>(Class = VeryHigh)
10	(Freq = High)&(Credit = High)& (Money = Medium)=>(Class = VeryHigh)

Table 12
The cluster results by K-means with 3 classes on output

Cluster center	C_1	C_2	C_3
c_1_R	4.36	2.03	2.26
c_2_F	4.10	1.75	1.58
c_3_M	3.29	4.14	1.38
The distance to zero point	6.83	4.94	3.09
Loyalty (Output)	High	Medium	Low
Numbers of instance	157	126	118

4.3. Discussion and findings

In this study, a discussion about the issue of different classes on output is conducted. From Table 14, the ranking accuracy rate is 0.9798, 0.9624 and 0.9418 in 3, 5 and 7 classes on output, respectively. By this order in accuracy rate, we obtain that the less the classes on output are, the more the accuracy rate is. However, it is an argument that the less the classes on output increase the accuracy rate and similarly increase the number of target customers (about 40% of all in 3 classes on output) that results in the difficulty of management. On the other hand, the more the classes on output decrease the accuracy rate and similarly decrease the number of target customers (about 11% of all in 7 classes on output) that results in easy to management. It is hard to say which situation is best because it is a trade-off selection to decide which classes that are based on subjective view of top management for C-company. Therefore, this trade-off selection is a valuable issue, and it will be discussed further for top management in enterprises and research.

From the experimental and statistical results of empirical case study (C-company), this study elucidates three findings as follows:

Table 13
The cluster results by K-means with 7 classes on output

Cluster center	C_1	C_2	C_3	C_4	C_5	C_6	C_7
c_1_R	4.76	4.46	2.48	3.77	1.47	3.21	1.41
c_2_F	4.57	4.25	3.48	1.82	1.41	1.93	1.20
c_3_M	4.39	2.29	4.16	4.16	4.00	1.41	1.38
The distance to zero point	7.92	6.57	5.96	5.90	4.49	4.01	2.31
Loyalty (Output)	Extremely High	Very High	High	Medium	Low	Very Low	Extremely Low
Number of instances	46	76	44	44	74	56	61

Table 14
Comparisons of experiment results with different methods and classes

Methods	The proposed procedure	Decision tree (Quinlan, 1993)	Neural network (Rumelhart & McClelland, 1986)	Naïve Bayes (Cestnik, 1990)
Accuracy rate in 3 classes	0.9798	0.9245	0.7523	0.8640
Accuracy rate in 5 classes	0.9624	0.9221	0.6796	0.8118
Accuracy rate in 7 classes	0.9418	0.9061	0.5336	0.7425

(1) High superiority on rough set theory: Based on Tables 9 and 14, it is proved that the proposed procedure outperforms the methods listed in terms of accuracy rate regardless of coverage on 0.6 and 0.9. Hence, Results of experiments are presented showing high superiority on rough set theory in this study.
(2) Robustness on rough set theory: According to Table 14, it is clear that the proposed procedure is robust in the different level of classes (3, 5 and 7 classes) on output resulting from minor changes in the accuracy rate (0.9798, 0.9624 and 0.9418, respectively). Similarly, the proposed procedure is robust in important rules that are generated from rough sets LEM2 algorithm because they are mostly consistent in the different level of classes.
(3) Matching the principle of 80/20 (or Pareto principle) (Schmittlein & Schultz, 1998): From the statistical results, the target customers, the customer loyalty (output) is over 'High' and F (frequency) attribute as well as M (monetary) attribute are over 'High' too, are about 26.18% of all customers, which are still on trading with C-company. The trading amount on these 26.18% target customers is 89.59% of all sales in 2006; that is, C-company takes effort on the 26.18% target customers and then obtains 89.59% of all revenues at least. This information above proves that this study can help C-company to focus the target customers via the experimental results of empirical case study. Moreover, the 26.18% target customers are worth of keeping and operating in excellent CRM for C-company. This is evidence for matching the principle of 80/20.

5. Conclusions

This study has proposed a procedure which joins RFM attributes and K-means algorithm into rough sets theory (the LEM2 algorithm) not only to enhance classification accuracy but also to extract classification rules for achieving an excellent CRM for enterprises. Additionally, it can effectively improve some drawbacks of data mining tools. To demonstrate the proposed procedure, this study employs a practical collected C-company dataset in Taiwan's electronic industry, which include 401 instances, as experimental dataset. From Table 9, the proposed procedure outperforms the methods listed in terms of accuracy rate regardless of 3, 5 and 7 classes on output, and the output of proposed procedure is understandable decision rules.

The output of the proposed procedure is a set of easily understandable decision rules which make C-company easier to interpret and know that which customer is more important and which customer is more contribution to revenue for enterprises. Furthermore, this proposed procedure based on RFM attributes and K-means algorithm can help C-company to classify objectively the segmentation of customers. Based on these excellent results of experiment, this study believes to aid C-company focusing the target customers and then gaining maximize profits with win–win situation for company–customer.

With the findings in this empirical case study, we positively conclude that the proposed procedure is more efficient than the listed methods on classifying the segmentation of customer value via RFM attributes, K-means algorithm and RS theory. For future research, other types of datasets can be considered to assess this procedure such as financial industry or healthcare industry even services industry, or other models of customer value analysis (except in RFM model) can be used as attributes for classifying the segmentation of customer. In general, we hope that the proposed procedure can become generalization not specialization for all datasets.

References

Cestnik, B. (1990). Estimating probabilities: A crucial task in machine learning. In *Proceedings of the 9th European conference on artificial intelligence* (pp.147–149). Stockholm.
Dunham, M. H. (2003). *Data mining: Introductory and advanced topics*. New Jersey: Prentice Hall, Upper Saddle River.
Ennew, C. T., & Binks, M. R. (1996). The impact of service quality and service characteristics on customer retention: Small businesses and their banks in the UK. *British Journal of Management, 7*(3), 219–230.
Forgy, E. (1965). Cluster analysis of multivariate data: Efficiency versus interpreability of classifications. *Biomertrics, 21*, 768.
Grzymala-Busse, J. W. (1992). LERS – A system for learning from examples based on rough sets intelligent decision support. *Handbook of applications and advances of the rough sets theory* (pp. 3–18).
Grzymala-Busse, J. W. (1997). A new version of the rule induction system LERS. *Fundamenta Informaticae, 31*(1), 27–39.
Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufmann Publishers.
Hashemi, R. R., LeBlanc, L. A., & Rucks, C. T. (1998). A hybrid intelligent system for predicting bank holding structure. *European Journal of Operational Research, 109*(2), 390–402.
Holland, J. H. (1973). Genetic algorithms and the optimal allocation of trials. *SIAM Journal on Computing, 2*(2), 88–105.
Hughes, A. M. (1994). *Strategic database marketing*. Chicago: Probus Publishing Company.
Joo, Y. G., & Sohn, S. Y. (2008). Structural equation model for effective CRM of digital content industry. *Expert Systems with Applications, 34*(1), 63–71.
Kalakota, R., & Robinson, M. (1999). *e-Business roadmap for success* (1st ed.). New York, USA: Addison Wesley Longman Inc.. pp. 109–134.
Kaymak, U. (2001). Fuzzy target selection using RFM variables. In *IFSA World congress and 20th NAFIPS international conference, Vol. 2* (pp. 1038–1043).
Kotler, P. (1994). *Marketing management: Analysis, planning, implementation, and control*. New Jersey: Prentice-Hall.
MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th berkeley symposium on mathematical statistics and probability* (pp. 281–297). Berkeley: University of California Press.

Miller, G. F., Todd, P. M., & Hegde, S. U. (1989). Designing neural networks using genetic algorithms. In J. D. Schaffer (Ed.), *Proceedings of the 3rd international conference on genetic algorithms* (pp. 379–384). San Mateo, CA: Morgan Kaufmann.
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-Hills Companies Inc.
Pawlak, Z. (1982). Rough sets. *Informational Journal of Computer and Information Sciences, 11*(5), 341–356.
Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Boston: Kluwer Academic Publishers.
Peppard, J. (2000). Customer relationship management (CRM) in financial services. *European Management Journal, 18*(3), 312–327.
Peppers, D., & Rogers, M. (1996). *The one to one future: Building relationships one customer at a time*. NY: Doubleday.
Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.
Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
Schijns, J. M. C., & Schroder, G. J. (1996). Segment selection by relationship strength. *Journal of Direct Marketing, 10*, 69–79.
Schmittlein, D. C., & Schultz, H. F. (1998). Transitioning marketing communications into the 21st century. *Journal of Marketing Communications, 4*, 9–26.
Stefanowski, J. (1998). On rough set based approaches to induction of decision rules. In A. Skowron & L. Polkowski (Eds.). *Rough sets in knowledge discovery* (Vol. 1, pp. 500–529). Heidelberg: Physica Verlag.
Stone, B. (1995). *Successful direct marketing methods*. Lincolnwood, IL: NTC Business Books. pp. 37–57.
Thompson, B., & Sims, D. (2002). CRM improving demand chain intelligence for competitive advantage. *Business Week, 3804*, 75–82.
Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). USA: Morgan Kaufmann Publishers.
Wu, J., & Lin, Z. (2005). Research on customer segmentation model by clustering. *ACM International Conference Proceeding Series*, 113.