Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

Customer Behavior Analysis Using Rough Set Approach

Prabha Dhandayudam[1] and Ilango Krishnamurthi[2]

Sri Krishna College of Engineering and Technology, Department of Computer Science and Engineering, Coimbatore, Tamil Nadu, India, [1]prabhadhandayudam@gmail.com, [2]ilango.krishnamurthi@gmail.com

Received 5 December 2011; received in revised form 19 September 2012; accepted 12 December 2012

Abstract

The customer relationship management (CRM) is a business methodology used to build long term profitable customers by analyzing customer needs and behaviors. The customer behavior is analyzed by choosing important attributes in the customer database. The customers are then segmented into groups according to their attribute values. The rules are generated using rule induction algorithms to describe the customers in each group. These rules can be used by the entrepreneur to predict the behavior of their new customers and to vary the attraction process for existing customers. In this paper a new rule algorithm has been proposed based on the concepts of rough set theory. Its performance has been compared with LEM2 (Learning from Examples Module, version 2) algorithm, an existing rough set based rule induction algorithm. Real data set of the customer transaction is used for analysis. Recency(R), Frequency (F), Monetary (M) and Payment (P) are the attributes chosen for analyzing customer data. The proposed algorithm on average achieves 0.439% increase in sensitivity, 0.007% increase in specificity, 0.151% increase in accuracy, 0.014% increase in positive predictive value, 0.218% increase in negative predictive value and 0.228% increase in F-measure when compared to LEM2 algorithm.

Keywords: Clustering, Customer relationship management, K-means, LEM2, Rough set theory, Rule induction, RFM, RFMP

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

1 Introduction

Customer relationship management (CRM) technology is a mediator between customer management activities in all stages of a relationship (initiation, maintenance and termination) and business performance [41]. It helps industries to gain insight into the behavior of customers and their value so that the enterprise can increase their profit by acting according to the customer characteristics. It is classified into operational and analytical. Operational CRM refers to the automation of business processes whereas analytical CRM refers to the analysis of customer characteristics and behaviors. Analytical CRM helps the entrepreneur to discriminate their customers and decide their marketing activities accordingly [30]. It consists of four ideologies namely customer identification, customer attraction, customer retention and customer development. Customer identification is the process in which the customers are grouped and their characteristics are analyzed. Customer attraction is the process in which the customers buy for the next time by providing customer service, coupon distribution, direct mailing and discounts. Customer retention is the process in which the customer's needs are satisfied by introducing new products and rectifying their complaints. Customer development involves in expansion of transaction intensity, transaction value and individual customer profitability. Customer identification is the most important phase in analytical CRM because once the customer is identified correctly; he can be retained and developed further. The customer identification phase consists of customer segmentation and target customer analysis. Customer segmentation involves in segmenting customers into predefined number of customer groups. Target customer analysis involves in analyzing customer behavior or characteristics in each customer group. It helps the entrepreneur to vary the attraction process for existing customers and to predict new customer's behaviors [30]. Data mining techniques are good at extracting and identifying useful information and knowledge from enormous customer databases, and for making different CRM decisions. The application of data mining techniques in CRM is an emerging trend in the global economy [2].

Data mining is a collection of techniques for efficient automated discovery of previously unknown, valid, novel, useful and understandable patterns in large databases. These patterns are used in an enterprise's decision making process [19]. The tasks that can be performed in data mining are clustering, association rule mining, rule induction and classification [21]. Clustering is an unsupervised classification used to group data with similar characteristics. It produces clusters for the given input data where data in one cluster is more similar when compared to data in other clusters. Association rule mining produces dependency rules which will predict the occurrence of an attribute based on the occurrence of other attributes in the data base. Rule induction belongs to supervised learning where data are already clustered into groups and it generates rules by finding regularities in the data in each cluster. Rules are in the form of If-Then condition. If part is called as antecedent and Then part is called as consequent. Antecedent contains conditional variables and consequent contains single decision variable. The conditional variables are the attributes in the given data and the decision variable is the cluster number assigned to the data using clustering algorithm. Rules generated for a cluster constitute the rule set for that cluster. Each data in the cluster should be described by at least one rule in the rule set of that cluster. This property of rule induction algorithm is called completeness. Each rule in the rule set of a cluster should be satisfied only by the data in that cluster. Rule set for a cluster should cover all the data within that cluster and no rule should be satisfied by any data in other clusters. This property of rule induction algorithm is called consistency. The data's satisfied by rules are not mutually exclusive because a data can be described by any number of rules. Classification on the other hand also generates rule for describing data in each cluster. The main difference between classification and rule induction is that the classification rules are mutually exclusive which means each data in the database is described by exactly only one rule [2]. Rule induction is used to describe the characteristics of the data rather than classification rules because in real data set, each data has to be described by all of its possible combinations of attributes value which means only one rule for each data is not sufficient. Clustering and rule induction of data mining technique is used for customer segmentation and target customer analysis of customer identification phase in CRM.

In this paper, an improved rule induction algorithm based on rough set theory has been developed to generate rules for clustered customer's data. The proposed algorithm has been compared with LEM2, a rough set based approach.
The rest of the paper is organized in the following: In Section 2 we describe the overview of customer relationship management, clustering algorithms, rule induction algorithms and LEM2 algorithm. In Section 3 we propose an improved rule induction algorithm based on rough set approach. In Section 4 we compare the prediction results obtained using rule induction algorithms. Finally in Section 5 we conclude the best rule induction algorithm according to the criteria chosen for comparison.

2 Related Works

Customer Relationship Management comprises a set of processes and enabling systems supporting a business strategy to build long term, profitable relationships with specific customers [27]. It is a philosophy of business operation for acquiring and retaining customers, increasing customer value, loyalty and retention, and implementing customer-centric strategies [30]. It is an important technology in every business because the business is customer centric. It consists of identifying, attracting, retaining and developing customers. Customer identification requires a comprehensive understanding of enterprise customers [12]. It includes target customer analysis and customer segmentation. The clustering algorithms are used for customer segmentation and rule induction algorithms are used for target

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

customer analysis. This section describes the overview of customer relationship management, clustering algorithms, rule induction algorithms and LEM2 algorithm.

2.1 Customer Relationship Management

Customer segmentation gives a quantifiable way to analyze the customer data and distinguish the customers based on their purchase behavior [40]. It is the process of dividing customers into homogeneous groups on the basis of common attributes [37]. It is typically done by applying some form of cluster analysis to obtain a set of segments [5]. In this way the customers can be grouped into different categories for which the marketing people can employ targeted marketing and thus retain the customers. Target customer analysis is used to analyze the customers in each cluster or segment so as to predict the new customer to the appropriate cluster. The customers are segmented and then rules are generated to describe them. These rules can be used to classify the new customers to the appropriate cluster who have similar purchase characteristics. The customer identification is followed by customer attraction which motivates each segment of customers in different way. Customer retention and customer development deals with retaining the existing customers and maximizing the customer purchase value respectively [30].

The attributes which describe the purchasing behavior of the customers are first chosen before customer segmentation because it requires a comprehensive understanding of enterprise customers [12]. RFM model is used to identify and represent the customer characteristics by three attributes namely Recency (R), Frequency (F) and Monetary (M). R indicates the interval between the time that the latest consuming behavior happens and present. F indicates the number of transactions that the customer has done in a particular interval of time. M indicates the total value of the customer's transaction amount in a particular interval of time [40].

In [7] RFM method, K-means clustering algorithm and LEM2 are used to obtain the classification rules. According to [23], customers with the same pattern of purchasing are only clustered and RFM is used to calculate the value of each cluster. Tsai and Chiu (2004) in [36] proposed a market segmentation methodology based on product specific variables such as items purchased and the associative monetary transactional history of customers and they used RFM to analyze the relative profitability of each customer's cluster. In [38] customer behavior is identified using RFM model and grey correlation model is used for customer targeting. Yeh et al. (2009) in [43] extended the traditional RFM model by including two parameters, time since the first purchase and churn probability. In [20] RFM analysis along with K-means clustering is used to study customer's fluctuations over different time frames. In [25]-[26] customer lifetime value (CLV) is calculated using RFM. In [9], [34] WRFM (Weighted RFM) is used instead of RFM. In this weights were assigned to R, F, and M depending on characteristics of the industry. Stone (1995), suggested for placing the highest weight on the Frequency, followed by the Recency, with the lowest weight on the Monetary measure [34]. In Chuang and Shen (2008), Monetary had the most value and Recency had the least value [9].

The attributes chosen to describe the customer behavior and the weightage of the attributes will differ from domain to domain. Here, the RFMP model which has four attributes R, F, M and P with equal weights is used. RFMP model is the modified RFM model where the payment details of the customers are considered. P indicates the average time interval between payment and purchase date. Payment detail of the customer is an important attribute because any two customers with same R, F, M value but different P value cannot be treated equally by the company. The customers are segmented using their consuming behavior via RFMP attributes. This ensures that the standards which cluster customer value are not established subjectively, so that the clustering standards are established objectively based on RFMP attributes [7].

The clustering algorithms for customer segmentation and rule induction algorithms for target customer analysis are discussed in section 2.2 and 2.3 respectively.

2.2 Clustering Algorithms

The customers are segmented using clustering based on their important attributes like R, F, M and P. It is an unsupervised classification where there are no predefined classes. The data in the data set is assigned to one of the output class depending upon its distance to other data. The data within each class forms a cluster. The number of clusters is equal to the number of output classes. The clustering technique produces clusters in which the data inside a cluster has high intra class similarity and low inter class similarity. The similarity is measured in terms of the distance between the data. For a numerical dataset, the distance between two data can be calculated using Euclidean, Manhattan and Minkowski distance.

Euclidean distance is given by

$$d(x, y) = squareroot(\sum_{i}^{n} | (x_i - y_i) |\text{^}2) \quad (1)$$

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

Manhattan distance is given by

$$d(x, y) = \sum_{i}^{n} | (x_i - y_i) | \quad (2)$$

Minkowski distance is given by

$$d(x, y) = (\sum_{i}^{n} | (x_i - y_i) |^{\wedge} \mathrm{p})^{\wedge}(1/\mathrm{p}) \quad (3)$$

In the above equations, n indicates the number of attributes in the given data, x and y are the data in the data set, d(x, y) is the distance between data x and y. In Minkowski distance if p=1 it is similar to Manhattan and if p=2 it is similar to Euclidean. In Euclidean distance the variation in one attribute is different from the variation in another attribute but in Manhattan distance the sum of the variation in each attribute is considered. In our real data set all the attributes R, F, M and P are equally weighted, so the variation in all the attributes is to be equally treated. Thus in this case Manhattan distance is used instead of Euclidean distance.

Clustering is mainly classified into hierarchical and partitioning algorithms. The hierarchical algorithms are further sub divided into agglomerative and divisive. Agglomerative clustering treats each data point as a singleton cluster and then successively merges clusters until all points have been merged into a single cluster. Divisive clustering treats all data points in a single cluster and successively breaks the clusters till one data point remains in each cluster. Partitioning algorithms partition the data set into predefined k number of clusters [14]. K-means algorithm is one of the most commonly used clustering algorithms [7]. It is a partitioning clustering algorithm which partitions the database D of n objects into a set of k clusters. The output differs when the initial centers for clusters are varied. The distance between objects in same cluster is less when compared to the distance between objects in different cluster. Each object is placed in exactly one of the k non-overlapping clusters [1]. The steps in K-means algorithm are as follows:

1. Initialize centers for k clusters randomly
2. Calculate distance between each object to k-cluster centers using the Manhattan distance formula given by Equation 1
3. Assign objects to one of the nearest cluster center
4. Calculate the center for each cluster as the mean value of the objects assigned to it
5. Repeat steps 2 to 5 until the objects assigned to the clusters do not change

2.3 Rule Induction Algorithms

The rule induction algorithms are used to generate rules to describe the characteristics of the customers in each segment. Decision trees (DT), artificial neural networks (ANN), genetic algorithms (GA) and rough set theory (RST) are used to produce rules [39]. DT is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent cluster number [21]. ANN 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 [7]. GA, 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 [22], [29]. In DT, too many instances lead to large decision trees and decrease classification accuracy rate. In ANN, number of hidden neurons, number of hidden layers and training parameters need to be determined, and has long training times. Moreover, ANN served as *black box* which leads to inconsistency of the outputs, is a trial-and-error process. GA also has some drawbacks such as slow convergence, a brute computing method, a large computation time and less stability. With respect to rough set theory, the advantages are they do not require any preliminary or additional parameter about the data, less expensive or time to generate rules, ability to handle large amounts data, yield understandable decision rules and stable [31]-[33]. It can be used to make decisions in any underlying business [42]. In the experimental results of [7], accuracy rate 1is more in LEM2 when compared to DT and ANN.

RST introduced by Pawlak in 1982 is a knowledge discovery tool that can be used to help induce logical patterns hidden in massive data. Some of the applications of RST in the field of knowledge discovery are dimensionality reduction, clustering, rule induction and discretization. The concept LERS (Learning from Examples using Rough Sets) was developed for rule induction. The basic algorithms based on LERS are LEM1, LEM2 and AQ. LEM1 algorithm computes global covering of attributes for producing rules. LEM2 algorithm on the other hand computes local covering and then converts into a rule set, so it gives better results compared to LEM1. AQ algorithm developed by R.S.Mickalski generates cover for each concept by computing stars and selecting from them single complexes to the

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

cover. In the worst case the time complexity of computing conjuncts of partial stars is O(nm) where n is the number of attributes and m is the number of data in the data set. So for large data set, AQ is not efficient when compared to LEM2. LEM2 of LERS is most frequently used since it gives better results [8], [10], [14], [16] and [18]. Extensions of LEM2 are MLEM2 and LEM3. MLEM2 extends LEM2 capability by inducing rules from data with both symbolic and numerical attributes including data with missing attribute values. It produces the rules sets with the smallest number of rules but needs an additional tool to simplify conditions using numerical attributes [17]. LEM3 is based on incremental learning of production rules from examples so the memory space requirement is minimal but it uses the same rule generating procedure of LEM2 [6]. The variable precision rough set model (VPRS), introduced in [44], is a generalization of the original rough set data analysis in the direction of relaxing the strict boundaries of equivalence classes. It assumes that rules are only valid within a certain part of the population, and it is able to cope with measurement errors. In [3], [28] and [35] approaches based on VPRS are dealt. In [13], extension of the rough set theory based on the dominance principle is dealt. This method is mainly based on substituting the indiscernibility relation by a dominance relation in the rough approximation of decision classes. However, the decision rules induced from the lower approximations of the Dominance-based Rough Set Approach (DSRA) are sometimes weak in that only a few objects support them. For this reason, a variant of DSRA, called VC-DRSA, has been proposed in [4]. It allows some inconsistency in the lower approximations of sets by a parameter called consistency level. It is more general than the classic functional or relational model and is more understandable for users because of its natural syntax and because it considers the inconsistency of real-life. The problem domain considered in the paper has complete and consistent data, so the algorithms based on LERS has been concentrated and the LEM2 algorithm has been taken for comparison.

2.4 LEM2 Algorithm

It is a rule induction algorithm based on rough set theory. It is used to find regularities hidden in the data and express in terms of rules. The clustering algorithm output is given as an input so that rules are generated for each cluster. Rules are in the form of

if (attribute-1, value-1) and (attribute-2, value-2) and ... and (attribute-n, value-n)
then (decision, value)

In the database, each row is called as a case and each column is called as an attribute. Attributes are independent variables and decision is a single dependent variable. Here, Recency, Frequency, Monetary, Payment are attributes and cluster number is the decision variable. The set of all cases labeled by same decision value is called a concept. A case x is covered by a rule r if and only if every condition (attribute-value pair) of r is satisfied by the corresponding attribute value for x. A concept C is completely covered by a rue set R if and only if for every case x from C there exists a rule r from R such that r covers x. R contains set of rules for each decision value. R is complete if and only if every concept from the data set is completely covered by R. A rule r is consistent if and only if for every case x covered by r, x is a member of the concept C indicated by r. R is consistent if and only if every rule from R is consistent with the data set. Rule induction produces complete and consistent rule set [18].

A block of an attribute-value pair t = (a, v), denoted [t], is the set of all examples that for attribute a have value v. A concept, described by the value w of decision d, is denoted [(d, w)], and it is the set of all examples that have value w for decision d. Let B be a concept and let T be a set of attribute-value pairs. Concept B depends on a set T if and only if

$$\phi \neq [T] = \bigcap_{t \in T} [t] \subseteq B \tag{4}$$

Set T is a minimal complex of concept B if and only if B depends on T and T is minimal. Let τ be a nonempty collection of nonempty sets of attribute-value pairs. Set τ is a local covering of B if and only if the following three conditions are satisfied:

1. each member of τ is a minimal complex of B,
2. $\bigcup_{T \in \tau} [T] = B$ and
3. τ is minimal (τ has the smallest possible number of members)

For each concept B, the LEM2 algorithm induces production rules by computing a local covering τ. Any set T, a minimal complex which is a member of τ, is computed from attribute-value pairs selected from T (G) of attribute-value pairs relevant with a current goal G, i.e., pairs whose blocks have nonempty interaction with G. The initial goal G is equal to the concept and then it is iteratively updated by subtracting from G the set of examples described by the set of minimal complexes computed so far. Attribute-value pairs from T which are selected as the most relevant,

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

i.e., on the basis of maximum of the cardinality of [t] ∩ G, if a tie occurs, on the basis of the small cardinality of [t]. The last condition is equivalent to the maximal conditional probability of goal G given attribute-value pair t. For a set X, |X| denotes the cardinality of X [15]. The procedure of LEM2 is as follows:

begin
G:=B;
$\tau := \phi$;
while G ≠ ϕ
begin
T := ϕ;
T (G) := {t | [t] ∩ G ≠ ϕ};
while T = ϕ or [T] ⊃ B
begin
select a pair t∈T(G) such that |[t] ∩ G| is maximum; if a tie occurs, select a pair t∈T(G) with the smallest cardinality of [t];
if another tie occurs, select first pair;
T := T ∪ {t};
G := [t] ∩ G;
T(G) := {t | [t] ∩ G ≠ ϕ};
T(G) := T(G) – T;
end {while};
for each t∈T do
if [T – {t}] ⊆ B then T := T – {t};
$\tau := \tau \cup \{T\}$;
G:= B - $\bigcup_{T \in \tau} [T]$;
end {while};
for each T ∈ τ do
if $\bigcup_{S \in \tau - \{T\}} [S] = B$ then $\tau := \tau - \{T\}$;
end {procedure}

The algorithm is run exactly |d| times, where |d| is the number of decision classes. The number of decision classes indicates the number of clusters produced by K-means algorithm. The while loop (G ≠ ϕ) is performed at most n times because we may have the whole set as the upper approximation to every decision class. Here n is the number of objects in the training set. To select a pair t ∈ T(G) as the best one, we have to iterate n * m times so that all possible pairs of attributes and values are examined. Here m is four which indicates the number of attributes in the training set. T contains m elements at most and τ contains n elements at most. So the computational complexity of for loop (for each t ∈ T) is m*n. Therefore the total computational complexity of LEM2 is equal to O (|d| * n * (n * m) * (m * n)) which is simplified as O (|d| * m^2 * n^3).

3 Proposed Algorithm

In LEM2 algorithm, the rules generated for each cluster is complete and consistent but it doesn't produce all the consistent rules in a cluster because once a consistent rule is discovered, the objects satisfying that rule is eliminated and rules are discovered for the rest of objects. Due to this the number of rules produced for a particular cluster becomes less and consequently the chances of predicting the customer to the correct cluster becomes less. In order to overcome this disadvantage the proposed rule induction algorithm produces all the consistent rules and complete rules for the objects in the cluster. Target cluster is the cluster for which rules are generated. Remaining clusters are the clusters other than target cluster. A block of an attribute-value pair t = (a, v), denoted [t], is the set of all examples that for attribute a have value v. A block of n attribute-value pair $t_1 = (a_1, v_1)$, $t_2 = (a_2, v_2)$, and so on, $t_n = (a_n, v_n)$ denoted $[t_1, t_2, \ldots t_n]$, is the set of all examples that for attribute a_1 have value v_1, for attribute a_2 have value v_2, and so on, a_n have value v_n. A block of size 1 has one attribute – value pair. A block of size n has n attribute – value pairs. For a set X, |X| denotes the cardinality of X. The procedure for improved rule induction algorithm is as follows:

begin
U - Set of all objects in the data set
B - Set of all objects in the target cluster
C := U – B (set of all objects in U but not in B)
G := B;

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

```
        temp := ϕ ;
        while G ≠ ϕ
        begin
                T(G) := {t | [t] ∩ B ≠ ϕ and |[t] ∩ B| ≠ |[t] ∩ G| and [t] ∩ C = ϕ }
                for each pair t ∈ T(G) do
                        temp := temp ∪ { [t] ∩ B};
                G := B – temp;
        end {while};
end{procedure}
```

In the while loop of $G \neq \phi$, find [t] having block size 1 and then block size 2 and so on until block size m. Here m indicates four which denotes the number of attributes in the data set. In each iteration of while loop, G contains set of objects whose T(G) contains set of all attribute – value pairs which satisfies the following three conditions:

1. $[t] \cap B \neq \phi$
2. $|[t] \cap B| \neq |[t] \cap G|$
3. $[t] \cap C = \phi$

The first condition chooses the pairs whose blocks have nonempty interactions with B. The last condition chooses the pairs whose blocks have empty interactions with C. The first and last conditions are required to satisfy the consistency property of rule generating algorithm. The second condition chooses the pairs whose cardinality of blocks satisfied in B is not equal to cardinality of blocks satisfied in C. This condition is required so that the rules generated are not redundant. The covering property of rule generating algorithm is satisfied by choosing $G \neq \phi$ as the while loop condition. In each iteration of while loop, rules are generated as the attribute-value pairs of t.

The algorithm is run exactly $|d|$ times, where $|d|$ is the number of decision classes. The number of decision classes indicates the number of clusters produced by K-means algorithm. The while loop ($G \neq \phi$) is performed at most n times because we may have the whole set as the upper approximation to every decision class. Here n is the number of objects in the training set. To calculate T (G), all the objects are examined so the computation is n. To select a pair $t \in T$ (G) as the best one, we have to iterate n * m times so that all possible pairs of attributes and values are examined. Here m is the number of attributes in the training set. Therefore the total computational complexity of proposed algorithm is equal to O ($|d|$ * n * n * (n * m)) which is simplified as O ($|d|$ * m * n^3). The proposed algorithm is an improved algorithm in terms of time complexity because LEM2 algorithm computation is m times more than the proposed algorithm. RAM usage for both LEM2 and proposed algorithm are same since both requires the entire clustered output values to be in main memory for analysis.

4 Experimental Results

Real data set of the customer transaction is used for the clustering and rule induction algorithms. The data set is collected from a fertilizer manufacturing company. It consists of 12,028 records of customer transaction for a period of three months for 3278 customers. In each transaction, party id, date of purchase, amount of purchase and payment of purchase are used to define R, F, M and P values. For each distinct party id, R is calculated as the interval between the last purchase and present, F is calculated as the number of his/her transaction records, M is calculated as the sum of his/her purchase amount and P is calculated as the average time interval (in terms of days) between his/her payment date and his/her purchase date for each transaction in the data set. The data set now has only four attributes namely R, F, M and P for 3278 customers. The values of R, F, M and P are normalized as given below:

For normalizing R or P

1. The data set is sorted in descending order of the R or P
2. Divide the data set into five equal parts of 20% record in each
3. Assign numbers 1,2,3,4,5 to first, second, third, fourth, fifth part of records respectively

For normalizing F or M

1. The data set is sorted in ascending order of the F or M

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

2. Divide the data set into five equal parts of 20% record in each

3. Assign numbers 1,2,3,4,5 to first, second, third, fourth, fifth part of records respectively

After normalization, the values of R, F, M and P are from 1 to 5. The normalized data set is now used by k-means clustering algorithm to segment the 3,278 customers into three groups or clusters. The number of actual cluster required is given by the business people. This number is determined by them according to the number of different scheme to be introduced as their promotional activity. Here the company requires three clusters so we segment the customers into three clusters. As a result, cluster1 contains 1,114 customers, cluster2 contains 1,064 customers and cluster3 contains 1,100 customers. LEM2 and proposed rule induction algorithms are used to generate rules for training data (two-third in each cluster). The test data (remaining one-third in each cluster) is given as input for LEM2 and proposed rule induction algorithm to predict the cluster value according to their generated rules for training data. The training and testing data are mutually exclusive. In training data, cluster1 contains 743 customers, cluster2 contains 709 customers and cluster3 contains 733 customers. In test data, cluster1 contains 371 customers, cluster2 contains 355 customers and cluster3 contains 367 customers. The performance criteria for prediction using rule induction algorithms are false positive (FP), false negative (FN), true positive (TP), true negative (TN), sensitivity, specificity, accuracy, precision, positive predictive value (PPV), negative predictive value (NPV), F-measure.

False Positive (FP) is the number of objects that don't belong to a cluster but are allocated to it. False Negative (FN) is the numbers of objects that belongs to a cluster but are not allocated to it. True Positive (TP) is number of objects that are correctly predicted to its actual cluster. True Negative (TN) is the number of objects that get predicted to a cluster but actually don't belong to [19]. Sensitivity is also called as true positive rate or recall. Sensitivity relates to the test's ability to identify positive results. It measures the proportion of actual positives which are correctly identified as such. Specificity relates to the ability of the test to identify negative results. It measures the proportion of negatives which are correctly identified. Accuracy is defined as proportion of sum of TP and TN against all positive and negative results. Positive predictive value or precision is defined as proportion of the TP against all the positive results (both TP and FP). Negative predictive value is defined as proportion of the TN against all the negative results (both TN and FN) [11].The F-measure can be used as a single measure of performance of the test. The F-measure is the harmonic mean of precision and recall [24]. The formulas are given below:

$$\text{Sensitivity or recall} = \frac{TP}{(TP + FN)} \tag{5}$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \tag{6}$$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FN + FP + TN)} \tag{7}$$

$$\text{Positive Predictive Value or Precision} = \frac{TP}{(TP + FP)} \tag{8}$$

$$\text{Negative Predictive Value} = \frac{TN}{(TN + FN)} \tag{9}$$

$$\text{F-measure} = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{10}$$

It is observed that the k-means clustering algorithm produces nearly 1000 customers in each cluster. So, LEM2 and proposed rule induction algorithms are repeated numerous times where training data (two-third) and test data (one-third) are randomly chosen from the data set such that training and testing data are mutually exclusive. For many runs, it produces the previously seen run value. So the twenty runs which produces different values are presented in the Table 1.The performance criteria for prediction are calculated for all the twenty cases. The Table 1 shows the false positive, false negative, true positive, true negative produced by the rule induction algorithms for all the twenty cases. The objective of the rule induction algorithm is to minimize false positive, false negative and to maximize true positive and true negative. From the Table 1 it is observed that the proposed rule induction algorithm has minimum FP, minimum FN, maximum TP and maximum TN for all the twenty cases when compared to LEM2.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

Table 1: FP, FN, TP and TN for rule induction algorithms

Case	False Positive		False Negative		True Positive		True Negative	
	LEM2	Proposed	LEM2	Proposed	LEM2	Proposed	LEM2	Proposed
1	1	0	17	10	1076	1083	2185	2186
2	0	0	14	10	1079	1083	2186	2186
3	1	1	13	7	1080	1086	2185	2185
4	0	0	10	6	1083	1087	2186	2186
5	0	0	14	10	1079	1083	2186	2186
6	1	1	5	2	1088	1091	2185	2185
7	1	1	11	7	1082	1086	2185	2185
8	0	0	8	6	1085	1087	2186	2186
9	0	0	16	10	1077	1083	2186	2186
10	0	0	15	10	1078	1083	2186	2186
11	1	0	12	8	1081	1085	2185	2186
12	0	0	11	7	1082	1086	2186	2186
13	0	0	9	3	1084	1090	2186	2186
14	1	1	10	7	1083	1086	2185	2185
15	0	0	17	9	1076	1084	2186	2186
16	0	0	15	8	1078	1085	2186	2186
17	0	0	8	3	1085	1090	2186	2186
18	1	1	10	6	1083	1087	2185	2185
19	1	0	14	7	1079	1086	2185	2186
20	0	0	11	8	1082	1085	2186	2186

Sensitivity, specificity, accuracy, PPV, NPV and F-measure are calculated using formula 5 to 10 respectively for each algorithm in all the twenty cases. The output is tabularized in Table 2 and Table 3. The objective of the rule induction algorithm is to maximize sensitivity, specificity, accuracy, PPV, NPV and F-measure. From the Table 2 and 3, it is observed that the proposed rule induction algorithm has equal or maximum value than LEM2 in all the twenty cases. The proposed algorithm on average achieves 0.439% increase in sensitivity, 0.007% increase in specificity, 0.151% increase in accuracy, 0.014% increase in positive predictive value, 0.218% increase in negative predictive value and 0.228% increase in F-measure when compared to LEM2 algorithm. The percentage increase in each performance criteria might seems to a smaller value but in real data set where customers are in terms of thousands not in hundreds the proposed algorithm has significant improvement than LEM2. For example, the average accuracy obtained using LEM2 is 99.622% and that of proposed algorithm is 99.773%. LEM2 accuracy for 3278 customers is 3265(i.e. 99.622*3278/100) and that of proposed algorithm is 3270(i.e. 99.773*3278/100). Here five more customers are predicted correctly using proposed algorithm when compared to LEM2.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

Table 2: Sensitivity, specificity and accuracy for rule induction algorithms

Case	Sensitivity		Specificity		Accuracy	
	LEM2	Proposed	LEM2	Proposed	LEM2	Proposed
1	0.98445	0.99085	0.99954	1.00000	0.99451	0.99695
2	0.98719	0.99085	1.00000	1.00000	0.99573	0.99695
3	0.98811	0.99360	0.99954	0.99954	0.99573	0.99756
4	0.99085	0.99451	1.00000	1.00000	0.99695	0.99817
5	0.98719	0.99085	1.00000	1.00000	0.99573	0.99695
6	0.99543	0.99817	0.99954	0.99954	0.99817	0.99909
7	0.98994	0.99360	0.99954	0.99954	0.99634	0.99756
8	0.99268	0.99451	1.00000	1.00000	0.99756	0.99817
9	0.98536	0.99085	1.00000	1.00000	0.99512	0.99695
10	0.98628	0.99085	1.00000	1.00000	0.99543	0.99695
11	0.98902	0.99268	0.99954	1.00000	0.99604	0.99756
12	0.98994	0.99360	1.00000	1.00000	0.99665	0.99787
13	0.99177	0.99726	1.00000	1.00000	0.99726	0.99909
14	0.99085	0.99360	0.99954	0.99954	0.99665	0.99756
15	0.98445	0.99177	1.00000	1.00000	0.99482	0.99726
16	0.98628	0.99268	1.00000	1.00000	0.99543	0.99756
17	0.99268	0.99726	1.00000	1.00000	0.99756	0.99909
18	0.99085	0.99451	0.99954	0.99954	0.99665	0.99787
19	0.98719	0.99360	0.99954	1.00000	0.99543	0.99787
20	0.98994	0.99268	1.00000	1.00000	0.99665	0.99756
Average	0.98902	0.99341	0.99982	0.99989	0.99622	0.99773

LEM2 produces only the minimal set of rules whereas proposed rule induction algorithm produces all the possible set of consistent rules to describe the records in the cluster. Thus the proposed algorithm characterizes the customers in each cluster clearly by producing all the consistent rules but eliminates redundant or duplicate rules. Since the number of rules to describe the customer is increased, the prediction accuracy is also improved. This statement is proved experimentally by comparing the performance measure. Hence the chances of judging a customer wrongly is reduced and allotting scheme to the customer is done correctly, which help the business to improve their customer life time value. Though the proposed algorithm produces more rules than LEM2, the computation complexity is m times less than LEM2 algorithm where m indicates the number of attributes considered for analysis. True Positive, True Negative, False Positive and False Negative are the parameters required to calculate the performance criteria measures of prediction. The complexity of calculating these parameters are linear with respect to the number of generated rules. The number of rules generated for each cluster or segment is very less when compared to the number of customers dealt. So this calculation complexity is negligible when compared to the rule induction algorithm complexity. Thus, the proposed algorithm is an improved algorithm in terms of cost benefit analysis.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 8 / ISSUE 2 / AUGUST 2013 / 21-33

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

Table 3: PPV, NPV and F-measure for rule induction algorithms

Case	PPV		NPV		F-measure	
	LEM2	Proposed	LEM2	Proposed	LEM2	Proposed
1	0.99907	1.00000	0.99228	0.99545	0.99171	0.99540
2	1.00000	1.00000	0.99364	0.99545	0.99355	0.99540
3	0.99907	0.99908	0.99409	0.99681	0.99356	0.99633
4	1.00000	1.00000	0.99545	0.99726	0.99540	0.99725
5	1.00000	1.00000	0.99364	0.99545	0.99355	0.99540
6	0.99908	0.99908	0.99772	0.99909	0.99725	0.99863
7	0.99908	0.99908	0.99499	0.99681	0.99449	0.99633
8	1.00000	1.00000	0.99635	0.99726	0.99633	0.99725
9	1.00000	1.00000	0.99273	0.99545	0.99263	0.99540
10	1.00000	1.00000	0.99318	0.99545	0.99309	0.99540
11	0.99908	1.00000	0.99454	0.99635	0.99402	0.99633
12	1.00000	1.00000	0.99499	0.99681	0.99494	0.99679
13	1.00000	1.00000	0.99590	0.99863	0.99587	0.99863
14	0.99908	0.99908	0.99544	0.99681	0.99495	0.99633
15	1.00000	1.00000	0.99228	0.99590	0.99216	0.99587
16	1.00000	1.00000	0.99318	0.99635	0.99309	0.99633
17	1.00000	1.00000	0.99635	0.99863	0.99633	0.99863
18	0.99908	0.99908	0.99544	0.99726	0.99495	0.99679
19	0.99907	1.00000	0.99363	0.99681	0.99310	0.99679
20	1.00000	1.00000	0.99499	0.99635	0.99494	0.99633
Average	0.99963	0.99977	0.99454	0.99672	0.99430	0.99658

5 Conclusion

Customer relationship management is a technology which helps the entrepreneur to improve their business volume by improving customer relationship. The customer identification is the important phase in CRM. It involves in segmenting the customers and analyzing their behavior for further customer attraction, retention and development. In this paper clustering technique in data mining has been used for customer segmentation and rule induction is used for describing customer behavior in each segment. The entrepreneur can employ different benefit schemes for customer in different clusters or segments. So, classifying a customer to the cluster plays an important role in CRM. For a good rule induction algorithm, the customer's behavior in each cluster should be correctly characterized so that the new customers are predicted to the appropriate cluster. The performance evaluation criteria are chosen based on the prediction accuracy of rule induction algorithm. The proposed algorithm on average achieves 0.439% increase in sensitivity, 0.007% increase in specificity, 0.151% increase in accuracy, 0.014% increase in positive predictive value, 0.218% increase in negative predictive value and 0.228% increase in F-measure when compared to LEM2 algorithm. It has been proved that the time complexity of LEM2 is m times more than the proposed algorithm where m indicates the number of attributes chosen for analysis. Thus, it has been evident from the results that the proposed algorithm is an improved rule induction algorithm which produces better performance in prediction and has less computation when compared to LEM2 algorithm.

References

[1] M. J. Berry and G. S. Linoff, Data Mining Techniques for Marketing, Sales and Customer Relationship Management. New Jersey: Wiley Publishers, 2008.
[2] A. Berson, S. Smith, and K. Thearling, Building Data Mining Applications for CRM. New York: McGraw-Hill Edition, 2000.
[3] M. J. Beynon and M. J. Peel, Variable precision rough set theory and data discretisation: An application to corporate failure prediction, Omega, vol. 29, no. 6, pp. 561-576, 2001.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

[4] J. Blaszczynski, S. Greco, and R. Slowinski, Multi-criteria classification – A new scheme for application of dominance-based decision rules, European Journal of Operational Research, vol. 181, no. 3, pp. 1030-1044, 2007.
[5] M. Bottcher, M. Spott, D. Nauck, and R. Kruse, Mining changing customer segments in dynamic markets, Expert Systems with Applications, vol. 36, no. 3, pp. 155-164, 2009.
[6] C.-C. Chan, Incremental learning of production rules from examples under uncertainty: A rough set approach, International Journal of Software Engineering and Knowledge Engineering, vol. 1, no. 4, pp. 439-461, 1991.
[7] C.-H. Cheng and Y.-S. Chen, Classifying the segmentation of customer value via RFM model and RS theory, Expert Systems with Applications, vol. 36, no. 3, pp. 4176-4184, 2009.
[8] M. Chmielewski, J. Grzymala-Busse, N. Peterson, and S. Than, The rule induction system LERS-a version for personal computers, Foundations of Computing and Decision Sciences, vol. 18, no. 3-4, pp. 181-212, 1993.
[9] H. M. Chuang and C. C. Shen, A study on the application of data mining techniques to enhance customer lifetime value based on the department store industry, in Proceedings of 7th International Conference on Machine Learning and Cybernetics, California, USA, 2008, pp. 168-173.
[10] J. Dai, Q. Xu, and W. Wang, A comparative study on strategies of rule induction for incomplete data based on rough set approach, International Journal of Advancements in Computing Technology, vol. 3, no. 3, pp. 176-183, 2011.
[11] A. H. Fielding and J. F. Bell, A review of methods for the assessment of prediction errors in conservation presence/absence models, Environmental Conservation, vol. 24, no. 1, pp. 38-49, 1997.
[12] H. Gong and Q. Xia, Study on application of customer segmentation based on data mining technology, presented at the International Conference on Future Computer and Communication, Wuhan, China, June 6-7, 2009, pp. 167-170.
[13] S. Greco, B. Matarazzo, and R. Slowinski, Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems, in New Directions in Rough Sets, Data mining, and Granular-Soft Computing (A. Zhong, S. Skowron, and S. Ohsuda, Eds.). Berlin: Springer-Verlag, 1999, pp. 146-157.
[14] J. W. Grzymala-Busse, LERS-A system for learning from examples based on rough sets, in Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory (R. Slowinski, Ed.). Netherlands: Kluwer Academic Publishers, 1992, pp. 3-18.
[15] J. W. Grzymala-Busse, Selected algorithms of machine learning from examples, Fundamenta Informaticae, vol. 18, no. 1, pp. 193-207, 1993.
[16] J. W. Grzymala-Busse, A new version of the rule induction system LERS, Fundamenta Informaticae, vol. 31, no. 1, pp. 27-39, 1997.
[17] J. W. Grzymala-Busse, MLEM2 – Discretization during rule induction, in Proceedings of the International Conference on Intelligent Information Processing and Web Mining, Zakopane, Poland, 2003, pp. 499-508.
[18] J. W. Grzymala-Busse, Rule induction, in The Data Mining and Knowledge Discovery Handbook (O. Maimon and L. Rokach, Eds.). New York: Springer-Verlag, 2005, pp. 277-294.
[19] G. K. Gupta, Introduction to Data Mining with Case Studies. Delhi: PHI Learning Private Limited, 2009.
[20] A. Hamzehei, M. Fathiana, H. Farvareshb, and M. R. Gholamian, A new methodology to study customer electrocardiogram using RFM analysis and clustering, Management Science Letters, vol. 1, no. 2, pp. 139-148, 2011.
[21] J. Han and M. Kamber, Data Mining: Concepts and Techniques. Boston: Morgan Kaufmann Publisher, 2001.
[22] J. H. Holland, Genetic algorithms and the optimal allocation of trials, SIAM Journal on Computing, vol. 2, no. 2, pp. 88-105, 1973.
[23] S. Huang, E. Chang, and H. Wu, A case study of applying data mining techniques in an outfitter's customer value analysis, Expert Systems with Applications, vol. 36, no. 3, pp. 5905-5915, 2009.
[24] N. Japkowicz and M. Shah, Evaluating Learning Algorithms: A Classification Perspective. Cambridge: Cambridge University Press, 2011.
[25] M. Khajvand and M. J. Tarokh, Estimating customer future value of different customer segments based on adapted RFM model in retail banking context, Procedia Computer Science, vol. 3, no. 1, pp. 1327-1332, 2011.
[26] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study, Procedia Computer Science, vol. 3, no. 1, pp. 57-63, 2011.
[27] R. Ling and D. C. Yen, Customer relationship management: An analysis framework and implementation strategies, Journal of Computer Information Systems, vol. 41, no. 3, pp. 82-97, 2001.
[28] J. S. Mi, W. Z. Wu, and W. X. Zhang, Approaches to knowledge reductions based on variable precision rough sets model, Information Sciences, vol. 159, no. 3, pp. 255-272, 2004.
[29] G. F. Miller, P. M. Todd, and S. U. Hegde, Designing neural networks using genetic algorithms, in Proceedings of the 3rd International Conference on Genetic Algorithms, California, USA, 1989, pp. 379-384.
[30] E. W. T. Ngai, Li Xiu, and D. C. K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, Expert Systems with Applications, vol. 36, no. 2, pp. 2592-2602, 2009.
[31] Z. Pawlak, Rough set, International Journal of Computer and Information Sciences, vol. 11, no. 5, pp. 341-356, 1982.
[32] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data. Netherlands: Kluwer Academic Publishers, 1991.
[33] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences, vol. 177, no. 1, pp. 3-27, 2007.
[34] B. Stone, Successful Direct Marketing Methods. Lincolnwood: NTC Business Books, 1995.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762013000200003

[35] C. T. Su and J. H. Hsu, Precision parameter in the variable precision rough sets model: An application, Omega, vol. 34, no. 2, pp. 149-157, 2006.
[36] C. Tsai and C. Chiu, A purchase-based market segmentation methodology, Expert Systems with Applications, vol. 27, no. 2, pp. 265-276, 2004.
[37] Z. Wang and X. Lei, Study on customer retention under dynamic markets, in Proceedings of 2nd International Conference on Networks Security, Wireless Communications and Trusted Computing, Wuhan, China, 2010, pp. 514-517.
[38] X. Weiwen, C. Liang, Z. Zhiyong, and Q. Zhuqiang, RFM value and grey relation based customer segmentation model in the logistics market segmentation, in Proceedings of 5th International Conference on Computer Science and Software Engineering, Wuhan, China, 2008, pp. 1298-1301.
[39] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques. Boston: Morgan Kaufmann Publishers, 2005.
[40] J. Wu and Z. Lin, Research on customer segmentation model by clustering, in Proceedings of the 7th International Conference on Electronic Commerce, Xi'an, China, 2005, pp. 316-318.
[41] M. Wubben, Fundamentals of customer relationship management, in Analytical CRM, Developing and Maintaining Profitable Customer Relationships in Non-Contractual Settings (F. Schindler and S. Scholler, Eds.). Wiesbaden: Wissenschaft Gabler Edition, 2008, pp. 11-48.
[42] Y. P. O. Yang, H. M. Shieh, G. H. Tzeng, L. Yen, and C.-C. Chan, Business aviation decision-making using rough sets, Lecture Notes in Computer Science, vol. 5306, no. 1, pp. 329-338, 2008.
[43] C. Yeh, K. Yang, and T. Ting, Knowledge discovery on RFM model using Bernoulli sequence, Expert Systems with Applications, vol. 36, no. 3, pp. 5866-5871, 2009.
[44] W. Ziarko, Variable precision rough set model, Journal of Computer and System Sciences, vol. 46, no. 1, pp. 39-59, 1993.