A Data Mining Technique to Find Optimal Customers for Beneficial Customer Relationship Management

G. Babu and T. Bhuvaneswari
1Bharathiar University, Coimbatore, Tamilnadu, India,
2MGR University, Chennai, Tamilnadu, India

Abstract: Problem statement: Modern companies and organizations are efficiently implementing a CRM strategy for managing a company interactions and relationships with customers. CRM systems have been developed and designed to support the areas of marketing, service process and sales. Many literature studies are available to preserve the customer relationship but small drawbacks occur in the existing methods. One method to maintain the customer relationship is frequency based method i.e., the company will give declination to the customer based on the historical data that is the customers how many times come to that company. These methods are not effective. Because the customers give revenue to that company is less. So the company revenue is affected. Approach: In this study, we propose a data mining and artificial technique to maintain the customer relationship between company and customers. Accomplishing this process, we maintain a historical database and then we use data mining ARM technique to mine the customer information from this database. We then present an artificial intelligence PSO technique to provide an offer to the selected customers. This offer does not affect the company revenues as well as satisfies the customers. This process will make a best relationship between the customers and organization and to satisfy the customers forever with company’s rules. Results: The simulation results have shown the performance of apriori algorithm for diverse combination of profit lengths of each customer. The performance of selected customer has been analyzed for five years. Finally the comparison results shows the selected customers are optimal comparing to other customers. Conclusion: The proposed method optimally selects the customers and also avoids the problems in the earliest methods. So, the relationship between customer and company is maintained successfully by using the proposed method.

Key words: Customer Relationship Management (CRM), rule mining, apriori algorithm, Particle Swarm Optimization (PSO), data mining, data objects, General Motors (GM)

INTRODUCTION

A data stream is a huge limitless sequence of data elements continuously generated at a rapid rate. Use of knowledge discovery algorithms that require only one scan over the stream is necessitated by the continuous characteristic of streaming data. An adaptable trade-off between processing time and mining precision should be supported by the data mining that is employed over data streams (Kim et al., 2010). Data mining refers to extracting or mining of knowledge from huge quantity of data (Mary and Iyengar, 2010; Raza, 2010). Classification, regression, clustering and dependence modeling are some examples for data mining tasks (Aghaebrahimi et al., 2009). Of these an important role is played by clustering in data mining. The clustering process forms valuable groups or clusters by classifying data objects, records, documents, such that characteristics of any two object is identical if they belong to the same group and different if they belong to different groups.

In recent times, numerous data mining applications (Syurahbil et al., 2009) and models have been created for diverse fields, like marketing, banking, finance, production, health care and other kinds of scientific data (Suguna and Thanushkodi, 2011; Sarlak and Fard, 2009). Parameters that create defects in manufacturing processes can be identified using data mining methods utilized in other complicated fields like Customer Relationship Management (CRM) (Batmaz et al., 2006). Retaining current customers is five times less expensive than capturing new customers (Krivobokova, 2009).
Also, the gross income generated by repeat customers can be twice as much as that generated new customers. Companies have ascertained that investing in customers who are valuable or potentially valuable by reducing their investments in non-valuable customers is more effective than evenly treating all customers.

The act of maintaining long-term relationship with customers (Sudhahar et al., 2006) is necessary for companies because of these kinds of findings and also due to the fact that customers desire to be served consistent with their independent and unique needs. Consequently, companies are resorting to Customer Relationship Management (CRM) techniques and CRM-supported technologies. Because effectiveness of company depends on CRM, it can be described as a strategy that permits proactive and profitable long-term relationships with customers by utilizing organizational knowledge and technology (Sudhahar et al., 2006; Fatt and Khin, 2010).

Customer relationship management is considered to be one of the aforementioned tools, which is a business strategy, has the propensity to select and manage the most precious relationships with the customers. Customer relationship management contains three components, namely customer, relationship and management. It attempts to expose a single view to its customers, when seeking a customer-centric viewpoint (Sarlap and Fard, 2009). All customer-oriented processes in a company are covered by the wide range of topics present in Customer Relationship Management (Torggler, 2009).

CRM (Babu and Bhuvaneswari, 2010) systems that build and maintain profitable long-term customer relations by facilitating the automation of business processes in the marketing, sales and service fields incorporates the information and communication technology constituents of a broad CRM strategy. In CRM, effective management of customer information has become increasingly important. CRM concentrates on the creation, simulation, evaluation and optimization of customer-related decision-making and thereby improved the quality of service provided to their customers. Marketing, sales or service (Muthaiyah, 2004) processes collect these data.

Its objectives are to increase profitability, revenue and customer satisfaction (Shen and Pethalakshmi, 2006). Often, information related to customers, products and markets are analyzed by means of simulations, predictions and classification of customers, using analytical tools such as OLAP and Data Mining after consistently storing such data in data warehouses. As useful findings about the market or customer behavior are provided by the results, these analyses can form the background for further CRM activities (Sarlap and Fard, 2009). Nowadays, the business profit of enterprises is improved supported by Customer Relationship Management (CRM). Enterprises normally provide personalized products and services to customers to increase their satisfaction for stable and long term relationship by knowing the customers utilizing CRM (Hsu and Lin, 2008). In the study 2, a few recent studies available in the literature are reviewed.

Related works: Cheng and Chen (2009) their study has proposed a procedure for successfully extracting meaning rules to improve these drawbacks by combining quantitative value of RFM attributes and K-means algorithm into rough set theory (RS theory). The proposed method has first obtained quantitative value as input features by utilizing the RFM model. Then, customer values have been clustered using K-means algorithm; and in the end, classification rules that assist enterprises driving an outstanding CRM have been mined by employing rough set (the LEM2 algorithm). The proposed procedure has been proved to surpass the methods listed in terms of precision rate irrespective of 3, 5 and 7 classes on output and produce comprehensible decision rules by analyzing the experimental results.

Al-Mudimigh et al. (2009a) have presented the most excellent combination of ERP with Customer Relationship Management (CRM). The three main parts present in their model have been outer view-ERP, inner view-ERP and knowledge discovery view. The CRM EPR and knowledge discovery have been used for the purpose of gathering the queries of customer, evaluate and combine the data and provide forecasts and recommendations for the improvement of an organization respectively. They have utilized MADAR data and implemented Apriori Algorithm on it for the practical implementation of the presented model.

Their proposed approach has revealed its capability in forecasting with reasonable precision by constructing churn prediction model using customer demography, billing information, call detail records and service changed log in an analysis of result from telecom provider.

Al-Mudimigh et al. (2009b) have gathered the data from central database in cluster format using characteristics and background of ERP which has been
based on the steps taken against the queries created by the customers. In addition, Apriori Algorithm has used the clustered data to extract new rules and patterns for the improvement of an organization. Data mining applications have been comprehensively implemented on ERP framework for forecasting the solution of future queries.

Mishra and Mishra (2009) have discussed that Customer Relationship Management (CRM) could help maintain competitiveness in the present economy by enabling organizations to handle customer interactions more effectively. Effective CRM implementation has been an intricate, costly and seldom technical project. In a trans-national organization with operations in diverse sections, they have effectively implemented CRM from process point of view. Their analysis has been expected to assist such organizations in comprehending transition, constraints and the implementation process of CRM.

Berndt et al. (2005) have concentrated on the implementation of a one-to-one programme especially in the financial services atmosphere inside a developing economy. They have examined the phases in the implementation of CRM as proposed by Peppers, Rogers and Dorf (1999b) and analyzed the consequence on customer service in a developing market. The results have revealed that constructive relationships exist between these phases and customer service.

Ou and Banerjee (2009) have described that the past research has frameworks and fragmented evidence of factors that results in effective Customer Relationship Management (CRM). Environmental factors that spread outside organizational boundaries have also been added in addition to customers by CRM systems. Thus, in addition to appropriate organizational factors, extra-organizational environmental factors must also be addressed by CRM initiatives. The applicability of the Work System (WS) framework as a guide to evaluate the CRM initiatives carried out by Shanghai General Motors (GM) has been validated by their framework. Their observations have suggested that comprehensive management of the individual slices of the WS framework and their interaction and interdependencies is essential for effective CRM operation. The analysis have also suggested that the success of CRM depends remarkably on the cultural values of the not only the organization, but also its customers.

In this study, we propose an effective CRM system for maintain the relationship between customer and company by using ARM and PSO technique. To accomplish this, we need to maintain a historical database. To extract the required information from the database, we will use an ARM algorithm. Determine the best possible customers from the mining result using PSO.

MATERIALS AND METHODS

The proposed customer relationship management methodology: The main aim of our proposed technique is to maintain a good relationship between customers and companies through monetary offers that are based on the revenues provided by those customers. So, this relationship is maintained by using the historical database which contains the information about the customers stored during their visit to our company for buying products or some service purposes. The information stored in the database is processed by using the data mining and PSO (Nacy et al., 2009). Our proposed study is composed into three stages:

- Querying
- Association Rule Mining using Apriori
- PSO- bases customer selection process

The structure of the proposed CRM system: The Customer information in the historical database has four fields: Date, Number of transactions, Customer ID and Profit. The customer information is stored in the field Customer ID and their date of arrival is stored in the field date, number of transactions done by the particular customer is stored in the field Number of transactions. Profit is the company revenue given by the particular customer.
Fig. 1: Structure of proposed CRM system

Table 1: Historical database

<table>
<thead>
<tr>
<th>Date</th>
<th>Number of transactions</th>
<th>Customer id</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/2/2000</td>
<td>1</td>
<td>101</td>
<td>20</td>
</tr>
<tr>
<td>3/2/2000</td>
<td>2</td>
<td>102</td>
<td>70</td>
</tr>
<tr>
<td>4/2/2000</td>
<td>3</td>
<td>104</td>
<td>80</td>
</tr>
<tr>
<td>5/2/2000</td>
<td>4</td>
<td>106</td>
<td>20</td>
</tr>
<tr>
<td>6/2/2000</td>
<td>5</td>
<td>101</td>
<td>70</td>
</tr>
<tr>
<td>7/2/2000</td>
<td>6</td>
<td>106</td>
<td>80</td>
</tr>
<tr>
<td>8/2/2000</td>
<td>7</td>
<td>105</td>
<td>90</td>
</tr>
<tr>
<td>9/2/2000</td>
<td>8</td>
<td>102</td>
<td>80</td>
</tr>
<tr>
<td>10/2/2000</td>
<td>9</td>
<td>104</td>
<td>20</td>
</tr>
<tr>
<td>11/2/2000</td>
<td>10</td>
<td>101</td>
<td>70</td>
</tr>
<tr>
<td>12/2/2000</td>
<td>11</td>
<td>102</td>
<td>20</td>
</tr>
<tr>
<td>13/2/2000</td>
<td>12</td>
<td>101</td>
<td>70</td>
</tr>
<tr>
<td>14/2/2000</td>
<td>13</td>
<td>104</td>
<td>80</td>
</tr>
<tr>
<td>15/2/2000</td>
<td>14</td>
<td>101</td>
<td>90</td>
</tr>
<tr>
<td>16/2/2000</td>
<td>15</td>
<td>101</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 2: A single customer database after query process

<table>
<thead>
<tr>
<th>Number of transactions</th>
<th>Customer id</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>101</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
<td>70</td>
</tr>
<tr>
<td>10</td>
<td>101</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>101</td>
<td>80</td>
</tr>
<tr>
<td>13</td>
<td>101</td>
<td>90</td>
</tr>
</tbody>
</table>

Let us consider that the database D contains a huge set of records, each transaction \( D = \{ \{d_i\}, \{T_j\}, \{id_k\}, \{P_{ijk}\} \} \), where \( d_i \) represents the date on which the customer visited to the company, \( i \) varies from \( 1 \leq i \leq N \times 365 \) and \( N \) is the number of years, \( T_j \) be the number of transactions done in a particular date, \( j \) varies from \( 0 \leq j \leq N_T^{(i)} - 1 \), \( id_k \) is a customer ID and fourth field \( P_{ijk} \) represents the company profit given by the customer. The execution procedure of the proposed CRM system is described as follows: At first, the customer information is mined from the historical database for calculating the frequency and profit of each customer. Here, the mining is done by a personalized querying method. Then, the mined information is used to calculate the company profit and frequency (the number of times the particular customer visit the company). By using association rule mining, the profit and frequency value of each customer is computed. Based on the mining result, the companies provide offers to customer using swarm intelligence technique known as particle swarm optimization. The three stages of our proposed CRM system are explained in detail as below. Figure 1 shows the structure of proposed CRM system.

**Querying:** In this first stage, the customer information is extracted from the database by performing a query process. The extracted customer information is a set of tables which contain set of fields are number of transactions, customer id and customer profit of the respective customer. For each customer in the database, the extraction process is performed to obtain the customer information for the mining process. The pseudo code of the querying process is described as follows:

\[
V \leftarrow C_{id}
\]

**Traversing through \( C_{id} \):**

If \( V = (C_{id} \text{ elements}) \) then

\( C_{id} \text{ elements} \rightarrow X \)

Delete the record from \( V \)

Repeat until all the same records are deleted

As a result of querying process, we obtain the information of the individual customers. For example, consider the historical database given in Table 1 and extracted customer information after the querying process is given in Table 2.

**Association rule mining using apriori algorithm:**

In this stage, we discuss the association rule mining Apriori algorithm that is used in analyzing the customer data to find the frequent items or customers from the transaction database. Consider an association rule is given in the form \( X \rightarrow Y \), where \( X \) and \( Y \) are items. These item sets have no common elements i.e., \( X \cap Y \neq 0 \). \( X \) is the antecedent of the rule and \( Y \) is its consequent. It means that when \( X \) occurs, \( Y \) also occurs with certain probability. Apriori algorithm is the best-known association rule mining algorithm. This algorithm generating frequent items (Tohidi and Ibrahim, 2011) i.e., generating frequently bought products by particular customers and finding the association rules from the frequent item sets (Patel et al., 2005).
Fig. 2: Customer selection based on PSO algorithm

For example we considered the association rule in length 2, the rule antecedent X represents the frequent item profit value and Y represents the associated profit part of the antecedent part.

After mining the frequent customers (items), the total profit values of these frequent customers are calculated by using mathematical operations. Here, the selected frequent items are added by their profit and the profit value is multiplied by number of times the frequent items procure by the particular customer. As a result, the total profit value of each individual customer is obtained. Use the original historical database to calculate the entire profit value for all customers present in the database. We set the threshold based on this historical database profit value for select the customer. The each individual customer profit value is equal to or greater than the threshold value; these customers are select from this process. The profit value is less than the threshold value those customers are not considered. If one customer providing profit value is high but those customers frequent value is less so that the customer getting offer value is less than other customers. Then the selected customers from the above process are optimized by using Particle Swarm Optimization technique.

**PSO-based customer selection process:** In the proposed CRM system, we use the PSO methodology to obtain optimum number of customers among a large customers obtained from ARM. Here, the qualified customer’s are selected based on profit and frequency levels. For optimization problems, PSO (Kanthavel and Prasad, 2011) algorithm is a best approach often used by most researches. Use of this algorithm we will efficiently maintain the customer relationship. The Fig. 2 illustrates the procedure of this proposed algorithm.

Using this procedure we select the optimal customer to maintain the customer relationship. The flow of procedure is discussed below.

PSO (Ramli *et al.*, 2009) defines that each particle has a potential solution to a problem in D-dimensional space. We randomly generate initial particles for customers and velocities for each particle.

Randomly generated initial particles are:

\[
P = (p_1, p_2, p_3, ..., p_i) \quad i = 1, 2, 3, ..., N
\]

where, \(p_{ib} < p_i < p_{ub}\)

where, \(P\) is a particle, \(p_b\) and \(p_u\) are upper and lower bound values of the number of particles. The randomly generation of initial particles size \(N\) i.e., \(N\) number of customers are generated randomly.

Hence each particle has a velocity which can be represented as:

\[
V = (v_1, v_2, v_3, ..., v_i) \quad i = 1, 2, 3, ..., N
\]

Velocity is represented as \(V\), \(v_{min}\) and \(v_{max}\) are the minimum and maximum values of velocity respectively.

The all particles are done in above specified particular intervals, no one particles does not exceed the specified interval. Before each and every iteration if we check the particles intervals. The checking process represented as follows:

If:

\[
v_i^{(n+i)} > v_{min}^{(n+i)}, \text{then} \quad v_i^{(n+i)} = v_{min}^{(n+i)}
\]

And:

If:

\[
v_i^{(n+i)} < v_{min}^{(n+i)}, \text{then} \quad v_i^{(n+i)} = v_{min}^{(n+i)}
\]

The duplication does not exist in the generation of particles i.e., the same element are not present in a particle, if we must check the duplication in the PSO process. The duplication checking method is described using the following example.
For example the each particle has five elements as represented as:

\[ p_1 = (g_1, g_2, g_3, g_4, g_5) \]

And the duplication checking process as:

If:

\[ g_1 \neq g_2 \neq g_3 \neq g_4 \neq g_5 \]

Here particle elements are \( g_1, g_2, g_3, g_4, g_5 \) these elements does not exist two times in a particle. Example, particle \( p_1 = (g_1, g_2, g_3, g_1, g_5) \) is an invalid particle because it doesn’t satisfy the above mentioned duplication checking condition.

**Determination of evaluation function:** The evaluation values are calculated for each individual particle to determine the optimal solution. The result of fitness values for all particles, the maximum fitness value is selected as an optimum value, the optimum value initially in pbest (flocal) value and so far gbest (fglobal) value.

Evaluation value can be calculated by:

\[
F = \log \left( \sum_{i=1}^{N} \frac{1}{z} \left( a_i + \left( \sum_{j=1}^{l} f_j \right) + \left( \sum_{k=1}^{M} z_k \right) \right) \right)
\]

Where:

\( N \) = Randomly generated particles elements
\( a_i \) = Consequent element profit value
\( f_j \) = Element frequency value
\( l \) = Frequency length
\( M \) = is the number of years taken in the historical database.
\( z \) = Number of times particular element is presented
\( w_1, w_2, w_3 \) = Weights

Initial iteration the values of velocity are assigned as zero. Use the randomly generated and initial velocity of particle to find the fitness values of these particles. We define pbest and gbest values from this fitness result, pbest value is called local best and gbest value is called global best. All particles have fitness values evaluated by the fitness function to be optimized and have velocities. The particles fly through the problem space by following the current optimum particles.

After finding the best values every particle tries to modify its position and velocity. To modify its position uses two data. First one is the distance between the current particle position and pbest and second one is the distance between the current position and gbest. This modification can be represented by the velocity.

Velocity of each particle can be modified by the following equation:

\[
V_i^{(n+1)} = V_i^{(n)} + C_1 \cdot r_1 \cdot (f_{local} - x_i^{(n)}) + C_2 \cdot r_2 \cdot (f_{global} - x_i^{(n)})
\]

\[
x_i^{(n+1)} = x_i^{(n)} + V_i^{(n+1)}
\]

In the above equation

\( V_i^{(n)} \) = Velocity of ith particle at iteration \( n \)
\( C_1, C_2 \) = Commonly referred as the learning factors
\( r_1, r_2 \) = Random numbers generated in the range of \([0, 1]\)
\( f_{local} \) = Position of the best fitness value of the particle at the current iteration
\( f_{global} \) = Position of the particle with the best fitness value in the swarm
\( x_i^{(n)} \) = Current position of the particle at iteration \( n \)

Each particle knows its best value (pbest) and its position. Moreover, each particle knows the best value in the group (gbest) among the pbests. Particles update their position and velocity for each iteration until it reaches the termination criteria.

This process will be repeated until the maximum number of iterations is reached. Once the maximum number of iterations is produced the process will be terminated. The last solution pointing the particle (customers) is considered as the best possible customers. The company decides to provide offer to these customers, these customers are the regular customers of this company. Following this procedure we effectively maintain the customer relationship between customer and company.

**RESULTS AND DISCUSSION**

The proposed CRM system is implemented in the working platform of MATLAB (version 7.10). The results have showed that our approach has an efficient performance and we have used the large set of customer information database to estimate the eligible customers. The customer database fields are created by the user.
Database description: The customer database contains the customer information’s such as Date, Number of Transactions, Customer ID and Profit. The database is created for 5 years and the customer ID is created randomly. We randomly generate a number Between 1 to 10, if the generated number is 4, the customer will be provided with new customer ID or the customer will be considered as a regular customer.

The date field contains 1825 days per year and profit field contain the profit amount of company. From this created database, we extract the relevant customer information using above mentioned querying methods.

The querying method result gives a large dataset; we have minimized the number of customers using the proposed apriori algorithm. Using this apriori algorithm, we have produced frequent items and association rules are generated for these frequent items. The generated association rules combination lengths are 2, 3 and 4. We have calculated the total profit value for selected customers and original database customers in the querying process. Based on the threshold value we have selected the eligible customers and again using PSO algorithm the customers are selected from the eligible customers to produce the optimal result. The given Tables 3-5 show the result of apriori algorithm for different combination of profit lengths for each customer. Some of the customer’s length combination results are given below.

The PSO algorithm is applied to the above table, to select the optimal customers. The working procedure of this algorithm is discussed in the proposed part. We have selected the optimal customers based on the customer’s profit and frequency value. Table 6 shows the result of selected optimal customers.

The above table illustrates the customers finally selected for getting offer from the company. Following this procedure we will maintain the customer relationship efficiently. We have designed the comparison result for finally selected customer against the original database customers for five years and the selected customer’s profit, frequency are individually compared to their five year data. Following Fig. 3 shows the comparison result of customer ID 1444 with profit and frequency to their five years data.

Similarly the selected customer ID 1756 comparison result is shown in Fig. 4.

The following comparison result graphs are the selected customer IDs 1688, 861,774 respectively. The comparison results Fig. 5-7 shows the selected customer performance over the five years.

We optimize the selected customer’s result, so the result of selected customers is compared with the remaining customers in the database. First we compare the frequency value of selected customers in opposition to frequency value of remaining customers in an original customer database and another one comparison have made on profit value of the selected customers against the profit value of remaining customers in a database.
The comparison result shows the higher performance of selected customers than remaining customers in both frequency and profit values.

The Fig. 8 shows the selected and remaining customers frequency performance.

From this graph, it is illustrated that selected customers frequency value is higher than the remaining customer’s frequency value in each year. The second graph Fig. 9 shows the results of selected and remaining customers profit value performance. The profit value performance gives very high result to the selected customer’s compared to the remaining customers.
CONCLUSION

In this study, we have developed the efficient CRM system using the data mining and artificial intelligence techniques for maintaining the customer relationship. Based on the customer’s information in the historical database, the CRM system have provided attractive offer to the customers in which they are frequently visited and also have provided high revenue to the company. This study introduces a data mining ARM technique to mine the customer information and PSO algorithms which are used to select the optimal customers. Based on the results, we have designed a graph to evaluate the performance of these proposed methods. These graphs results have explained that the selected customers are optimal when compared to other customers in the database. From this graph, it is demonstrated that the proposed methods effectively maintains the relationship between customer and company. Using this method, companies or organizations maintains and improves the relationship with customers.

REFERENCES


