A Review: Customer Relationship Management Classification using Data Mining Technique

Easha Mangal [1], Mr. Anurag Jain [2]

Student, M.Tech Department, Radharaman Institute of Technology & Science, Bhopal, India[1]

Associate Professor, CSE, Radharaman Institute of Technology & Science, Bhopal, India[2]

Abstract: In business process, the role of Customer Relationship Management (CRM) is very important. CRM provides a customer classification and prediction, which is used for the optimization of business process. In new trends of CRM, one of them is Social customer relationship management (SCRM). In the system we are using a data mining technique to implement customer classification in CRM as we need to analyze the mass volume of data. The research reviews the basic concepts of CRM and Data Mining Techniques. This paper describes many data mining techniques used for the CRM model.

Keywords: Data mining, Customer relationship management, CRM, Data mining techniques.

I. INTRODUCTION

In the present time marketing system changed from a product oriented to a customer oriented model. CRM can be defined as the process of predicting customer behaviour and selecting actions to influence that behaviour to benefit the company. Customer expectations are always progressively and business services must increase along with their expectations. CRM utilized as a business tool to distinguish, select, obtain and build up its beneficial clients. The major issue here is data openness. On the off likelihood that the exploration procedure isn't begun from within the association, the researchers can not have access to customer's data. In today's competitive marketing world customer is a very important quality to an enterprise. The enterprises can get an edge in the drastically increasing competitive environment if they efficiently manage their customer relationship.

Article is fastidiously reviewed to eliminate those articles that aren't associated with application of information mining in customer relationship management. The aim of the paper is to review literature associated with the appliance of data mining techniques for CRM in numerous educational journals. CRM and its techniques. Secondly, the data processing techniques used for CRM are summarized.Classification analysis addresses this by making a model to assign customers to teams supported a group of celebrated characteristics of that individual. The model is made by a knowledge set with each cluster membership and therefore the set of individual characteristics and applied to new unclassified customers.

The training data set of the KDD Cup 2009 orange small data set is used in the research. This data set contains 190 numeric attributes and 40 nominal attributes. Among this 190 numeric attributes, 16 attributes are empty and 132 attributes are sparse with more than 90% missing rate.

The data set is a complex CRM problem having the following challenges:
Large number of instances   : 50000

Large number of features   : 230
Large number of missing values: 60% and above
Imbalanced class (49110 majority instances, 890 minority instances)   Nominal features with high cardinality.[2]

TABLE I: CHARACTERISTICS OF DATA SET

Data Set	KDD Cup 2009
Instances	50000
Features	230
Output Class	2
Total features	231
Positive Class Instances	49110
Negative Class Instances	890
Missing values	yes

II. DATA MINING AND CRM

Data mining refers to obtaining or "mining" knowledge from great deal of data. Data mining as an equivalent word for additional popularly used term, knowledge discovery from data or "KDD" .The goal of this method is to search out pattern that was antecedently unidentified data [11].

Data mining will offer client insight that is significant for establishing an efficient CRM strategy, it can result in customized interactions with customers and therefore augmented satisfaction and profitable customer relationships through data analysis. It can support associate degree and optimized customer management through all the phases of customer life cycle, from the customer acquisition and institution of a powerful relationship to the hindrance of attrition and winning back of lost customers. Marketers attempt to induce a bigger market share of their customers. in plainer words, they're blame for obtaining ,developing, and keeping the customers.

Operational, collaborative and analytical are three types of CRM, [8]:

- Operational CRM which generally refers to products and services that permits an organization to be concerned with their customers. It provides support for various business processes, which can include s, marketing, sales and service. Some examples of Operational CRM are Contact centres, web sites and data aggregation systems.

- Collaborative CRM is communication with customers and covers direct dealings with customers including response and issue reporting. Dealing can take place through web pages, email, Automated Voice Response. This type of CRM greatly improves on services existing.

- Analytical CRM is the analysis of customer data for a host of various purposes. Basically it is used to design and complete targeted marketing campaigns which optimize marketing effectiveness. This type of CRM takes into account product and service decision making, new product development and pricing.

CRM Dimensions

According [7], there are four dimensions of CRM:

Customer Identification*:* Customer Identification is also called Customer acquisition. The intend of this phase is to get to know which are the most liable to be an important customer and target him. Target customer examination looks the productive customer through the investigation of customers' fundamental characteristics, where the division of the subdivision of the customer database into small groups of customers who are nearly comparative.

Customer Attraction: After identifying the target customer, surely the organization will focus all the effort and resources to attract the profitable customer. This will be done by using the direct marketing, which motivates the customer to order their desired items by various channels. Some examples of direct marketing are E-mails or coupons.

Customer Retention: The top way to keep the customers, and restore customers is to raise their satisfaction. Customer retention is considered as the CRM core. Customer maintenance or retention components are one to one marketing, reliability programs and objection management.

Customer Development: This phase aims to grow the size of the customer transactions with the organization. Basics of customer development include customer lifetime value analysis, up/cross selling and market basket analysis. The customer lifetime value means the total returns which are expected to gain by the customer during his terms with association. Up/Cross selling are promotion's activities aim to market related services according to the customer use. Market analysis maximizes customer transactions by helpful activities in the customers' purchasing behaviour.

III.CLASSIFICATION TECHNIQUE FOR CRM

A. Decision Tree:
Decision tree learning uses a decision tree as a predictive model which maps remarks about an item to conclusions about the item's target value. It is one of the predictive modelling techniques used in statistics machine learning and data mining. New expressive names for such tree models are regression trees or Classification trees. In these tree structures, leaves symbolize class labels and branches symbolize conjunctions of features that lead to those class labels. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making. Decision tree learning is a method commonly used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. A decision tree is a plain depiction for classifying examples. Decision tree learning is a successful technique for supervised classification learning.

Decision trees used in data mining are of two main types:

- Classification tree analysis can be used when the predicted outcome is the class to which the data belongs.
- Regression tree analysis can be used when the predicted outcome is considered an actual number (e.g. the price of a house, or a patient's length of stay in a hospital).

B. *Neural Networks:*
Everyone can see that the human brain is superior to a computer at various tasks. A good example is the processing of visual data: a 1-year-old kid is vastly improved and quicker at perceiving articles thus on than even the most sophisticated AI system running on the supercomputer. Thus brain has a lot of several other features that would be advantageous in artificial systems. This is the actual inspiration for studying neural calculation. This is an substitute perception to the common one (based on a programmed instruction sequence), that was introduced by von Neumann and hare being used as the resource of nearly all machine computation. It is stimulated by the knowledge from neuroscience; however it doesn't assempt to be organically reasonable in detail. Neural systems are way to computing that involves creating numerical structures with the capacity to learn. The techniques are the consequences of scholastic examinations to model nervous system learning. Neural systems have the remarkable ability to get significance from entangled or uncertain information and can be utilized to concentrate patterns and distinguish patterns that are too immense to possibly be seen by either people or other computer methods. An artificial neural network is an interconnected group of nodes, similar to the vast network of neurons in the human brain. A trained neural network can be considered as an "expert" in the class of information which given to analyse. Now this expert can be used to present projections specified new situations of attention and respond "what if" questions. Neural systems use a set of processing nodes similar to neurons in the brain. These processing nodes are interconnected in a network that can recognize patterns in data once it is showing to the data, i.e., the network learns from practice same as people do. This recognizes neural networks from conventional computing programs that just follow commands in a set sequential order.

C. *Genetic Programming:*

Genetic programming (GP) has been immensely used in study in the past 10 years to resolve data mining classification issues. The reason genetic programming is so widely used due to fact that prediction rules are very logically represented in GP. GP has verified to produce better results with worldwide search problems like classification. The exploration space for classification can be described as having numerous peaks", this causes local search algorithms, like simulated annealing, to perform defectively. GP consists of stochastic search algorithms based on abstractions of the processes of Darwinian evolution. Each candidate solution is represented by an individual in GP. The solution is coded into chromosome like structures that can be mutated and/or combined with some other individual's chromosome. Each individual contains a fitness value, which measures the quality of the individual, in other words how close the candidate solution is from being optimal. Based on the fitness value, individuals are selected to mate. This process creates a new individual by combining two or more chromosomes, this process is called crossover. They are combined with each other in the hope that these new individuals will evolve and become better than their parents. Additionally to matting, chromosomes can be mutated at random. The running time of GPs is usually controlled by the user. There are many parameters used to determine when the algorithm should stop, and each data set can have very different settings. In all cases, the best individual is stored across generations and is returned when the algorithm stops. The most commonly used parameter is number of generations. Another stop parameters used is minimum expected hit ratio, in which case the algorithm will run until a candidate solution has a hit ratio greater than expected. This however can cause the algorithm to run forever. Combinations of stop conditions can also be used to ensure stoppage.

D. *Statistical Algorithms:*

Quinlan introduced C4.5 AND ID3 algorithms for inducing Classification Models, which also called Decision Trees, from data. We have a set of records. All records have the same composition, consists of a several values pairs. One of these values indicates the category of record. Now the problem is to create a decision tree on the basis of answer to question about the non-category attributes which predicts accurately the value of the category attribute. The category attributes take only the values {true, false}, or {success, failure}, or something similar. In any other case, one of its values will denote failure.

The basic conceptions of ID3 are these:

- In the decision tree each node represents a non-categorical attribute and each arc a possible value of attribute. A leaf of the tree shows the expected value of the categorical attribute for the records described by the path from the root to that particular leaf.
- In this at each node should be correlated the non-categorical attribute which is most informative amongst the attributes not yet measured in the path from the root.

- To measure how informative is a node Entropy is used. ID3 algorithm selects the best feature based on the conception of entropy and information gain for developing the tree. C4.5 algorithm acts alike to ID3 but improves a few of ID3 behaviours:
- A prospect to use incessant data. Using unknown values which have been marked by "?".
- Prospect to use attributes with dissimilar weights.
- Pruning the tree is created.

IV.LITERATURE SURVEY

"Application of data mining techniques in CRM" was published by E.W.T. Ngai et al. This provides an academic database of literature between the period of–2006 covering 24 journals and Nine hundred articles. Findings of this paper point to that the research area of customer retention received more research attention. Of these, most are related to one-to-one advertising and faithfulness programs respectively, [1] .

According to [3],Leela Rani Komma Reddy & G Loshma presented a framework of an evolving information system based on knowledge from data mining, and has discussed the framework by targeting on knowledge of classification. Their work is mainly focused on the research of the customer categorization and prediction in Customer Relation Management concerned with data mining based on Back propagation technique.

Nishant Kumar, Suraj Singh focused on the research of the customer prediction in Customer Relation Management related with data mining based on C4.5 classification algorithm, which have a try to the optimization of the business process, [4].

According [5],Velu C. M build-up a software based intelligent model by using data mining technique Genetic Algorithm (GA) , Artificial Neural Network (ANN), and Fuzzy Logic to discover the pattern of the customers , this concluded in classification of the customers and found that 98 % above classification is correctly.

P. Isakki alias Devi et al., developed a method to design retail promotions, informed by product associations observed in the same groups of customers. Clustering and association rule is used find to identify customer behaviour. It can easily predict the sales. The customer with similar purchasing behaviour are first grouped by means of clustering techniques such as K-means method and for every cluster an association rule (Apriori algorithm) to identify the products that are brought together by the customers. Association rules are used to discover the relationship and knowledge of the database. It proved that apriori algorithm is the most well known association rule mining algorithm because it is easily found in the frequent item datasets from the database.. This paper focused on getting more customer satisfaction [6].

According [9],authers combined three diverse classifiers and exploit their specialties to carefully design steps that deal with heterogeneous and partially missing data. Then they combined the heterogeneous information using

AdaBoost with separate clear-cut and numerical decision trees that handles the missing values directly. They exposed their probabilistic relations with the selective na¨ıve Bayes by compressing the numerical and the categorical features. Finally, in their work over fitting is carefully prevented during the individual training and the post-processing steps using cross-validation.

According [10], authers showed that tree-based methods are more capable of handling or ignoring missing values. They also showed that most successful method was a Logistic Model Tree with AUC as split criterion using predictions from boosted decision stumps as features.

V. CONCLUSION

In this review paper, we have demonstrated that data mining can be coordinated into the customer relationship management and improved the procedure of CRM with advancement. In marketing, to hold their customer, CRM is one of the best driving methodologies. This study demonstrates that data mining strategies in CRM which enhance the productivity of CRM and give a superior forecast capacity to the associations. Data mining will have significant effect on customer relationship management and will exhibit challenges for future studies. In future, the data analyst should compare or even combine available techniques in order to achieve the best possible outcomes/results.

REFERENCES

[1] E.W.T. Ngai , Li Xiu , D.C.K. Chau, " Application of data mining techniques in customer relationship management: A review and classification" published in the International Journal " Expert Systems with Applications 36 (2009).

[2] S.Ummugulthum Natchiar, Dr.S.Baulkani,"Customer Relationship Management Classification Using Data Mining Techniques", International Conference on Science, Engineering and Management Research, (ICSEMR 2014)978-1-4799-7613-3/14©2014 IEEE.

[3] Leela Rani Komma Reddy & G Loshma, "Classification and Pridiction in Customer Relationship Management using Back Propagation", International Journal of Computer Science and Informatics ISSN, Volume-2, Issue-3, 2012.

[4] Nishant Kumar, Suraj Singh "Decision Making Using Classification" , International Journal Of Research In Computer Engineering And Electronics.VOL : 4 ISSUE :4.

[5] Velu C. M. and Kashwan K. R. "Pareto Classification of Data Mining for Customer Relationship" , International Conference on Information and Network Technology (ICINT 2012) IPCSIT ,Vol.-37.

[6] P. Issakki Alias Devi, S.P. Rajagopalan, "Analysis of Customer Behavior using Clustering and Association Rules", International Journal of Computer Applications, Vol. 43, No.23, April 2012, pp.19-27 .

[7] Raghda M. Nasr , Sahar Nagaty, and Mohammed B. , " Survey and Challenges of Mining Customer Data Sets to Enhance Customer Relationship Management",Senousy International Journal of Computer Applications Volume 107 – No 13, December, , pp.34-35.

[8] Dr. C. Sunil Kumar , "Increasing the Efficiency of CRM Process Using Data Mining Practices"',International Journal of Advance Research in Computer Science and Management Studies Volume 3, Issue 7, July 2015, , pp.107.

[9] Hung-Yi Lo, Kai-Wei Chang, Shang-Tse Chen, Tsung-Hsien Chiang, Chun- Sung Ferng, Cho-Jui Hsieh, Yi-Kuang Ko, Tsung-Ting Kuo, Hung-Che Lai,Ken-Yi Lin, Chia-Hsuan Wang, Hsiang-Fu Yu, Chih-Jen Lin, Hsuan-Tien Lin,Shou-de," Ensemble of Three Classifiers for KDD Cup 2009 Expanded Linear Model, Heterogeneous Boosting, and Selective Na¨ıve Bayes", ,csie.ntu.edu.twAn.

[10] Patrick Christian, Buck Christian, Pavlo Golik, Niklas, Michael ,Johannes Laudenberg ,Christian Oberd¨orfer christian., Pascal Steingrube pascal, Arne Mauser, Logistic Model Trees with AUC Split Criterion for the KDD Cup 2009 Small Challenge, JMLR: Workshop and Conference Proceedings 1.

[11] Jiawei Han and Micheline Kamber, "Data mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.