Data Mining for Customer Relationship Management: A Review

Mihin Sumaria
Student, Department of Computer Engineering, Dwarkadas J.Sanghvi College of Engineering
Mumbai, India
mihin.sumaria@gmail.com

Nishita Sheth
Student, Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
ns7604@gmail.com

Mrs. Sindhu Nair
Asst. Professor, Department of Computer Engineeringl Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
sindhu.nair@djsce.ac.in

Abstract— In today's competitive environment, the only way to proliferate one's business, apart from providing with quality product, is to maintain sturdy relations with the customer. In order to maintain a stable profit inflow and retain customers, companies today employ customer relationship management techniques to ensure that existing customers remain faithful to the brand, and prospective customers are motivated to affiliate themselves to the brand. In this paper, we study various data mining techniques employed in different papers, for identifying trends among customers, and hence improving customer relations.

Keywords-*Customer Relationship Management; Data Mining; Supply Chain Management; Clustering; Association Rule Mining; Decision Tree;*

*****

I. INTRODUCTION

A. CRM

The definition for CRM according to Gartner Inc. is, "A business strategy, the outcomes of which optimizes the profitability, revenue and customer satisfaction by organizing around customer segments, fostering customer-satisfying behaviours, and implementing customer centric processes. By definition then CRM technologies enable greater customer insights, increased customer access, more effective interactions and integration throughout all customer channels and back office enterprise functions." [1][2] Customer relationship management, in simple terms, is fostering good relations with the customers so as to ensure that the loyal ones keep returning for business and attract prospective customers for the same. The process involves storing and analysing customer data to formulate new plans and to keep reinventing existing strategies to maximize gains. CRM can be used to determine trends among customers in a given area and employ effective decision making policies to cater this particular area [4].

The components of a CRM system are follows [2][3]:

- Operational - automation of rudimentary business process.
- Analytical - using business intelligence to sift through customer data and determine customer patterns.
- Collaborative - interacting with targeted customers.

B. SCM

Supply Chain Management is designing and implementing optimal plans to manage the delivery of goods from source (manufacturers) to destination (customers). It maintains an organised flow of products and services so as to meet the needs of a group of customers. SCM plays a vital role in gaining an upper-hand over other competitors, i.e. if they make parallel adjustments to customers' patterns [7]. An ideal example of a supply chain are the online retailers who manage stock of products, from suppliers to retailers, based on the demand perceived on the online portal, so as to make sure that there is no customer dissatisfaction in case of dearth of goods.

.

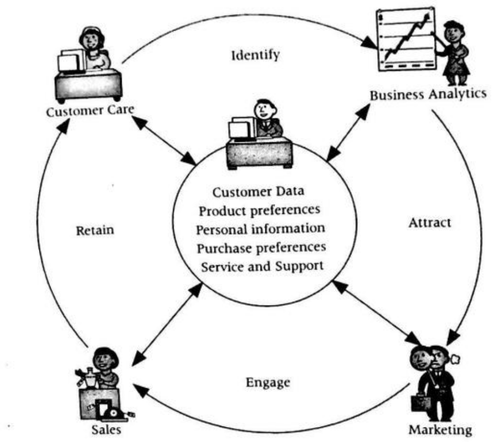

Figure 2: CRM Lifecycle [2]

C. Synergy between CRM and SCM

The online retailers today, like the ones mentioned in the above example, record customer history and can use this information to improve supply chain techniques. This information is recorded as a part of customer relationship management, which shows that the two processes can be used in synergy to make the business more profitable. One of the attributes that can be effectively managed by this synergy is stock availability, which can be altered on the basis of demand and supply ratio [7].

The rest of the paper is structured as follows. Section II discusses data mining and the various techniques that it encompasses. Section III gives a brief overview of the selected papers for this survey. Section IV explains how data visualization is a pivotal part of the CRM process in an organization. Section V provides with the possible future scope, and Section VI is a conclusion to this literature survey.

II. DATA MINING

Data mining is the process of discerning unknown and useful information, unusual occurrences, and trends in data stored for decision making or survey. It usually involves dividing existing data into test, training and validation data, such that an appropriate model can be built upon the training data to minimize the error and predict with accuracy the outcome for test data, and the validation dataset is used to check whether the prediction model is within a given error interval, to ensure the precision of the model [4]. Data Mining involves a lot of pre-processing before the actual analysis takes place. It starts with identifying a sources for raw data which can be found in the real world, and integrating this data to create a warehouse. This raw data is then standardized for the ease of processing. For example, male and female customers can be represented as 1 and 0 respectively under a single attribute, gender. Next steps involves discarding erroneous and incomplete data, so as to ensure that these errors are not reflected in the knowledge acquired after data analysis. This data is now normalised and loaded into data warehouses to fetch and process data in an efficient and secured manner. This standardized, normalized, cleansed data is then analyzed and visualized [8].

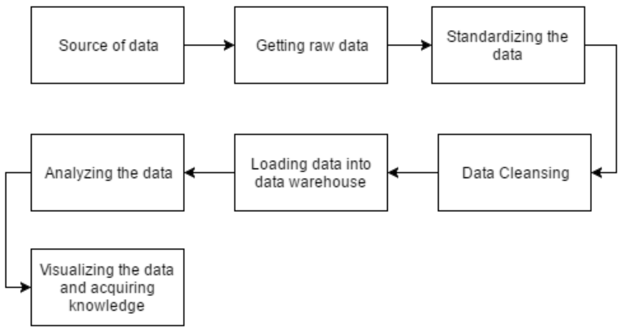

Figure 2: Steps for Data Mining

The various data mining techniques can be broadly classified into unsupervised and supervised models. Supervised models need human input from time to time, to train the data mining model. Unsupervised models try to discern associations without any external user input. While providing labelled data can be time consuming, unsupervised data mining models can be expensive when it comes to computing resources [10].

The approaches that we will be reviewing in this paper are as follows:

A. Decision tree

Decision Trees are supervised structured models of data used for classifying, and is used to break down complex problems into smaller sub-problems to aid decision making. The variables that factor into decision making can be integers, categorical variables, any numeric value or even structured data types [8]. In a decision tree, each internal node acts as a predicate which helps dividing the sample space base on a specified condition. Every edge in the decision tree is the result of the aforementioned predicate, and every leaf node represents a class of objects. The decision tree algorithms that are widely used are ID3 algorithm, C4.5 algorithm, CHAID algorithm, CART, etc.

Advantages of decision trees in data mining are [8][14]:

- The class labels used in the decision tree make it easy to comprehend and facilitate decision making possible in the absence of expert analysts.
- A decision tree can be extended to any possible limit based on the new scenarios encountered by an organization in the consumer market.
- Decision trees can be used in case of redundant, empty data sets and can be used to identify unusual data or outliers.
- They are relatively faster than other decision making techniques and inexpensive to compute.

Disadvantages of decision trees in data mining are [8]:

- The generated output in case of large number of attributes or classes can be difficult to comprehend.
- In case of correlated data, classes that are supposed to be mutually disjoint, may overlap to a great extent.

B. Association Rule Mining

Association is correlation among occurrences in data and the likelihood of them occurring again in the future. An association rule involves an antecedent and a consequent, the latter usually occurring due to the former, so association rule mining, which is an unsupervised model of data mining, is procuring such implications from the given data. To isolate significant associations measures such as support, confidence, lift and conviction are used [15].

- Confidence is the number of times an association rule has been found to be true.
- Support is the number of times both the antecedent and consequent have occurred simultaneously.
- Lift is the ratio of support in case of the association rule being true, to the support in case of the two events being completely disparate.
- Conviction can be stated as the ratio of number of times an event occurs without the other such that the two events are independent, to the number of time the association rule has been found to be false.

Based on these measures, minimum support and confidence constraints are used to observe association rules in the data.

These associations are mined in the temporal or spatial domains depending on the problem definition given. Association rule mining can be used for discerning trends among buying patterns of customers, such as purchasing ink cartridges on the purchase of a printer. This kind of information can be used to facilitate supply chain management and inventory arrangement. Widely used association rule mining algorithms are Apriori, FP-growth, AIS, etc.
Advantages:

- In case of large databases, association rule mining reduces the number of scans on the database for related and frequent attributes.

Disadvantages [9]:

- It is expensive to calculate values of constraints for the measures of confidence and support, and to improve algorithmic efficiency in association rule mining.
- Association rule mining is difficult to comprehend for someone not well versed with reading analyzed data.

C. Clustering

Clustering can be defined as categorizing of entities into groups, called clusters that share a strong resemblance with one another, as compared to entities from other clusters. Clustering is especially important to find trends in a focused part of a sample space, by reducing the dimensions of the data. Data entities can be plotted in euclidian space, and the distance between any two entities will signify the similarity between the two. Clustering, thus, is grouping objects close to one another in the euclidian space. Clusters can be mutually disjoint or overlapping depending on the algorithms employed. Fuzzy algorithms might categorize an entity into more than one cluster with varying degrees, and such algorithms are known as soft clustering algorithms. Hard clustering algorithms such as k-means categorize each entity into one cluster only or none at all. These clusters can be displayed using dendrograms, scatterplots, etc. [8].

The usual cluster algorithms are iterative in nature, and keep repeating unless there is no further change in the clusters. The similarity among objects that is used to determine the contents of the clusters is quantified using similarity metrics. The similarity metrics widely used by clustering algorithms are Euclidean distance metrics, manhattan metric, minkowski metric, mahalanobis' distance metric, chebychev metric, etc.
There are several types of clustering algorithms - hierarchical, partitioning and density-based.

Hierarchical clustering algorithms makes use of dendrograms to cluster data entities, such that all these data entities are represented by the root node and internal nodes are the clusters. These clusters are joined or split using arcs. Advantage of hierarchical clustering is that it can produce a dendrogram even if distinct clusters are non-existent [8].

These dendrograms can be made in a bottom-up manner, i.e. agglomerative, or top-down manner, i.e. divisive manner.

Partitioning clustering algorithms are the most rudimentary form of clustering algorithms, and it categorizes, a set of given objects into several disjoint clusters. The number of clusters needs to be assumed beforehand, and passed on to the algorithm as a parameter. These clusters are then iteratively updated to optimize a given criterion for clustering, so as to minimize the disparity among entities in the same cluster.

The following are the various clustering algorithms used for customer relationship management in the reviewed papers:

- K-means algorithm: This algorithm starts with defining a fixed number of clusters, say k, with each cluster having a centroid selected randomly from the data set D. Each entity is placed in one of the clusters, one with a centroid which is the closest to it. This is reiterated until centroids stop changing their position. The distance metric used in this algorithm is the euclidean distance metric, where the distance of each entity is measured from the selected set of centroids. The quality of a cluster can be measured by the in-cluster variance [5][10][11]. The sum of squared error is calculated using the formula below, (1), for every centroid and entity pair, which is to be optimized. In the given equation, dist(x,y) is the euclidean distance between the points x and y. p and c_i represent a point in cluster C_i and its centroid respectively. The resulting clusters are disjoint and as far away as possible from each other. It has linear complexity and can be used on large datasets [6]. This algorithm can be disadvantageous if the dataset contains outliers, or unusual data, as it will shift the center of the clusters, and skew the arrangement of the clusters. Also, the random selection of data in the beginning, might result in different sets of clusters if the algorithm is run on the same data more than once [5].

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} dist(p, c_i)^2 \qquad (1)$$

- K-medoids algorithm: To overcome the disadvantage of k-means, k-medoids makes use of the median object of a cluster to represent the cluster, instead of using the mean of all objects. Each of the remaining entities are added to the clusters, based on their correlation to its median object. The next step, selecting an object other than the median object, computing the total number of swaps with the median object, and if the number of swaps is less than 0, then the selected object is the new median object, is repeated in an iterative fashion until no further changes are possible [12]. The formula below is used to calculate the sum of error for all the objects in the dataset, where o_i represents the median object [5]. This algorithm is less sensitive to unusual data or outliers, in comparison to k-means.

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} dist(p, o_i) \qquad (2)$$

- Self-organizing map: Self organizing map is an unsupervised system based on competitive learning in which the output neurons compete among themselves to be activated, with only one being activated at any instance of time. This competition is generated by having negative feedback paths among the neurons. This results in the neurons being forced to organize themselves. The organizational mapping is such that continuous high dimensional input space is transformed into discrete low dimensional (typically two dimensional) output space on the basis of a feature map. The topological relations between the data patterns are not affected during this transformation [6]. A typical self-organizing map has two layers- an input layer and a computational layer, such that, each neuron is connected to all the source nodes in the input layer. In a self-organizing map, the connection weights in the computational layer are initialized with random values. Next, a discriminant function is made use of for competition among neurons, and the neuron with the smallest value of the discriminant function is declared the winner. This neuron then determines the excited neurons' neighbourhood, and cooperates with these neurons. The cooperating neurons reduce their discriminant function values to improve the response for the encounter of a similar input pattern in the future [13]. Self-organizing maps can be susceptible to data that varies a lot, and to encounter this hierarchical self-organizing maps are made of, where the variables to be clustered are grouped into topics and clustered separately [6].

Advantages of clustering [16]:

- It can easily incorporate changes, and can help marginalize essential features for categorizing entities.

Disadvantages of clustering [16]:

- Some algorithms are susceptible to erroneous or missing data values and can create distortions within the clusters.
- It cannot give a set of discrete clusters, unless the user specifies the number of clusters.
- It is difficult to handle clusters of varying sizes, and to decompose large clusters.

III. Discussion on Selected Papers

The paper, *Application of Data Mining Technology in the Tourism Product's Marketing CRM,* uses basic attributes of a customer, like gender, age, income, with respect to their consumption capacity of an enterprise product. It addresses the problem in two steps. Firstly, processing the available data, i.e.

calculating the gain of each attribute, and building upon it a decision tree for the same. Using this decision tree, the customers are segregated based on their consumption quota. The organization is then able to identify all the profitable customers, and use targeted marketing to retain them [4].

The paper, *The Analysis of CRM Customer Information Based on Data Mining*, makes use of Google maps API to discern geographic coordinates of the customers. These customers are now clustered on the basis of their geographic locations, and the buying patterns of these clusters are analyzed. The paper then illustrates the variability in buying patterns in districts north and south of the Suzhou River. It compares the pace of life, customers' decision speed when a large payment is involved in the aforementioned districts. It suggests to adapt marketing campaign strategies based on these references. It makes use of an improved algorithm based on k-means and k-medoids to cluster the customers [5].

International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 10 01 - 06

The paper, *Building clusters for CRM strategies by mining airlines customer data,* explores clustering methods that can be used to support decision making in an airlines company. The data set used in this problem, makes use of age, gender, country of residence, etc. for over 20,000 customers. The paper then compares the average distance between each cluster obtained by k-means, self-organizing map and hierarchical self-organizing map. The paper concludes by saying that k-means provides more segregated clusters as compared to other algorithms [6].

The paper, *Improving the retailer's profit for CRM using data mining techniques*, uses fuzzy clustering algorithm to categorize a retailer's customers with similar requirements and capabilities into three groups, i.e. valuable, regular, and occasional customers. This categorization takes place on the basis of purchasing capability, and the frequency of customer's visits which is quantified based on a specific threshold value, and is mainly done to achieve customer satisfaction and improve customer relation. The fuzzy clustering module of the Tanagra tool is made use of to achieve this categorization. The same experimental setup was practiced on garment related companies which boosted the sales of those companies, proving that data mining techniques can be used to improve customer relationship and profits of an organization [7].

IV. DATA VISUALIZATION FOR CRM

Traditional CRM data is difficult to grasp, and refer to while making decisions, at an organizational level, because of complex data structures and relations. The advent of technology has made it possible to view this complex data as graphical figures like bar graphs, pie charts, histograms, box plots, etc. which bolsters efficient and swift decision making. Executive level employees in an organization do not have the time to review traditional CRM data, and these visualization techniques reduces the overhead required to make an appropriate decision. While effective visualization can help an organization, substandard CRM visualization can cause problems like lack of adoption by salesperson and compromised decision support systems. However, organizations are now moving towards improved visualization techniques that can summarize the data without withholding any important data, and ones with a logical and intuitive interface [17].

V. FUTURE SCOPE

In the recent years, the buzzwords 'Social Media' and 'Internet of Things' have been on the rise, and rightly so. CRM software is going to become increasingly social and will track current and potential customers' activity on social media to discern optimal marketing strategies. Even further, more data can be mined with the use of IoT products; products like smart watches, smart home automation system, etc. can help organizations track their customers on a daily basis and also study their habits. This treasure trove of data will provide businesses with higher dimensional data, thus narrowing their targets for better marketing. Further ahead, CRM systems will be integrated and full automated, and will provide without any supervision.

VI. CONCLUSION

This survey informs us that in today's markets maintaining stable relations with customers is of utmost importance, and also how various data mining techniques. Clustering can easily adopt to changes in a changing market, and is pretty flexible in comparison too. Association rule mining is inconvenient when used for large customer databases, as it is expensive to calculate the required constraints and parameters for numerous relations. Also, association rule mining can be difficult to comprehend from a managerial point of view as opposed to clustering. Similarly, decision trees are disadvantageous for voluminous customers' databases, and correlated customer data.

REFERENCES

[1] Brown, S. A., Customer Relationship Management (CRM): Overview, Technical Overview: DPRO-90679, Gartner Inc., October, 2001.

[2] A. Leon, ERP demystified. New Delhi: Tata McGraw-Hill, 2008.

[3] R. Senkamalavalli and T. Bhuvaneshwari, "Data mining techniques for CRM", International Conference on Information Communication and Embedded Systems (ICICES2014), 2014.

[4] S. Pei, "Application of data mining technology in the tourism product's marketing CRM", 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), 2013.

[5] H. Jiang, Q. Yu, C. Liu, Q. Zhu and L. Guo, "The analysis of CRM customer information based on data mining", 2013 Ninth International Conference on Natural Computation (ICNC), 2013.

[6] H. Sofia and R. Henriques, "Building clusters for CRM startegies by mining airlines customer data," *2013 8th Iberian Conference on Information Systems and Technologies (CISTI)*, Lisboa, 2013, pp. 1-5.

[7] K. Deepa, S. Dhanabal and V. Kaliappan, "Improving the Retailers Profit for CRM Using Data Mining Techniques", 2014 World Congress on Computing and Communication Technologies, 2014.

[8] R. Chattamvelli, Data mining methods. Oxford, U.K.: Alpha Science International, 2009.

[9] E. Garcia, C. Romero and T. Calders, "Drawbacks and solutions of applying association rule mining in learning management systems", pp. 15-25, 2007.

[10] J. Han and M. Kamber, Data mining. Amsterdam: Elsevier, 2006.

[11] S. Äyrämö and T. Kärkkäinen, "Introduction to partitioning-based clustering methods with a robust example", 2006.

[12] "Self Organizing Maps: Fundamentals", School of Computer Science, The University of Birmingham, 2004.

[13] K. Naveen Kumar, G. Naveen Kumar and C. Veera Reddy, "Partition Algorithms– A Study and Emergence of Mining Projected Clusters in High-Dimensional Dataset", International Journal of Computer Science and Telecommunications, vol. 2, no. 4, pp. 34-37, 2011.

[14] "Decision tree", Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Decision_tree. [Accessed: 01- Oct- 2016].

[15] "Association rule learning", Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Association_rule_learning. [Accessed: 01- Oct- 2016].

[16] "Data Mining Cluster Analysis", www.tutorialspoint.com. [Online]. Available: https://www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm. [Accessed: 01- Oct- 2016].

[17] B. Boyers, N. Kimla, A. Cluytens, J. Montana, R. Ciglansky, M. Powell, N. Kimla, R. Young and T. Martin, "Why Visualization Matters in CRM - Pipeliner CRM Blog", Pipeliner CRM Blog, 2015. [Online]. Available: http://blog.pipelinersales.com/sales-effectiveness/why-visualization-matters-in-crm/. [Accessed: 01- Oct- 2016].