DATA MINING TECHNIQUES FOR CUSTOMER RELATIONSHIP MANAGEMENT

Reetika Chaudhary
Department of Computer Engineering, MPSTME, NMIMS University
Mumbai, India
reetikachaudhary.nmims@gmail.com

Shubha Puthran
Assistant Professor, Computer Engineering Department
MPSTME, NMIMS
Mumbai, India

Abstract
Consumer behaviour is the key factor to predict the trends of retail of products and services, and henceforth increasing profits of an organization. This forecasting is done by analysis of big data, which is tonnes of raw data gathered from various sources. The data collected is then segmented and modelled into customer profiles, which then can be updated. Data mining methods like K-Nearest Neighbours (Clustering), Artificial Neural Networks, Decision trees, Association and Sequential Rule discovery are used to manipulate data and extract results. On the basis of performance evaluation, study of all the above methods gives complete information.

Index Term - Customer relationship management, data mining techniques, segmentation, profiles, k-nearest Neighbours, neural networks, decision trees, association rule, sequential rule discovery.

I.   INTRODUCTION

Customer Relationship Management (CRM) is the practice of cleverly discovering, marketing to, offering to and servicing clients. It is a comprehensively utilized term that covers ideas utilized by organizations and open establishments to oversee their associations with clients and stake holders. In consumer world, the amount of data about consumers, suppliers has been exploding through millions of networked sensors and consumer video surveillance systems. The unedited, user generated information of these platform include both structured and unstructured data which fall into a category of Big Data analysis. Moreover, consumers are no longer what they used to be. Today's consumers have evolved beyond being merely 'buyers'. So, more insights information is necessary for analysing a consumer behaviour. In this aspect, Big Data has become a central role for making data driven decision making processes[1]. The nature of data collected in the big data platform is significantly different. To analyse the Big Data, both in terms of quantity and the nature of the data, is a challenging task.

Data Mining is the non-inconsequential extraction of novel, verifiable, significant information from expansive information sets, here it can play a significant role in analysing customer data[2]. Information Mining is a procedure that employments an assortment of information examination and displaying strategies to find examples and connections in information that may be utilized to make precise forecasts. It helps to choose the right prospects on whom to centre, offer the privilege extra items to existing clients and distinguish great clients who may be going to abandon. The generation of insights helps in improving the value supply chain, develop customized and targeted marketing campaigns and products and services[3]. The outcome is enhanced due to an extraordinarily enhanced capacity to every individual contact in the best way.

II. CUSTOMER SEGMENTATION AND PROFILING

Customer relationship management includes the terms: Segmentation and Profiling. Customer Segmentation is process of dividing customers into groups on the basis of shared common attributes and behaviour. Customer profiling is characterizing customers on basis of their attributes like age, sex, lifestyle and more. It is a way of applying external data to a population of existing customers[4]. This helps in identifying 'good' and 'bad' customers (in terms of value assets to company) and to know the degree of their loyalty and profitability.

A. Customer Segmentation

Segmenting means putting customers into different segments according to their characteristics. It classifies each customer into groups predefined as desired. By doing this the whole data set is broken into smaller segments which falls into a range of attributes, and hence makes it easier to channel the resources. Though, there are some difficulties which can be faced while segmenting[5]:

- Quality of data- If the data is too big, small, vague or incomplete (missing fields) then it is difficult to derive meaningful sentences from it. Furthermore, the resulting segment would have some type of errors like if some fields are used to group by and its information is missing, then it is difficult to decide which customer goes into which group.
- Continuous process- Database keeps updating frequently. So, segmentation process has to be applied on data set continuously to maintain sync with the incoming data. Effective segmentation strategies influences the behaviour of customers, thereby necessitating revision of classes.
- Over segmentation- This is a difficulty which arises when, a segment becomes too small or big to justify treatment as separate segments.

B. Customer Profiling

Client profiles are the basis for providing better services to existing clients and retain them. These services are kept in constraint, according to the project scope. The profiles can also be used to attract new customers using external data, like demographic data of a product from different sources. This profiling is done by collecting information about customer such as demographic and behavioural data.

Depending on the result desired, company has to select the types of profiles it want for its use. A customer profile has fields like name, age, sex, address, and city.

Customer features used for profiling are[6]:

- Geographical: Where the clients reside.
- Ethnicity: What culture the customers follow and where they originally are from, like which country or state.
- Income: This decides how much they are willing to spend on the retailing products.

Age, lifestyle, shopping frequency etc. can also be used to make it in more depth.

These customer segments and profiles are helpful for training the model developed for predicting the consumer behaviour. They themselves can be used as a result of a sales target.

III.   DATA MINING TECHNIQUES

Data mining techniques have to be applied on the database to extract meaningful information from it. There are various types of techniques and which one to use depends upon the company extracting the information i.e. for what purpose the information is needed and how to manipulate it to get better results.

A. Classification

It is a technique that put similar entities into the same groups based on similar data characteristics and dissimilar entities are put in other groups. Similarity is measured according to a distance measure function generally Euclidean distance measure is used. Distance measures identify if the clusters are useful or not and to what extent.

1) K - Nearest Neighbour

It is a predictive technique used for classification models. In this the training data is the model. In this a subset of cases is found out that are most similar to the input and then it predicts the outcome.

There are 2 important points:

- K is the no of nearest neighbour i.e. the number of nearest cases to be used.
- A metric to measure the similarity of the cases.

K-NN is given input which is a test case, it places that cases into a class by examining the measure of similarity and k specified. Then the distance is computed from the new case to each close case[7]. The new case is predicted to have the same outcome as the k close cases in the dataset.

It is easy to understand and easy to perform but there are many disadvantages. It uses whole training set as model which makes the model very huge and time consuming. Moreover the measure function Euclidean distance used cannot work well with multidimensional data.

2) Self Organising Maps

A self organising map is a kind of neural network structure which provides a lattice of lower order generally 2 dimensional grid from multidimensional data. This grid exists in a space different from input space. It tries to find clusters such that any two clusters that are close to each other in the grid space is also close to each other in input space as well. It performs unsupervised training of data. The nodes in the network converge to form clusters to represent groups of entities with similar attributes. The number and composition of clusters then help predict the outcome generated by training process[8].

The SOM network is a type of feed forward neural network with zero hidden layer; input layer is connected to the kohonen or output layer. It uses Euclidean distance for measure.

SOM method is very efficient in visualising multidimensional data into a map of fewer dimensions, and hence easy to interpret. But it is quite difficult to handle and its output cannot be easily understood.

B. Artificial Neural Network
An artificial neural system (ANN), regularly just called a "neural system" (NN), is a scientific model or computational model in view of natural neural systems, is an imitation of organic neural framework. It comprises of an interconnected gathering of counterfeit neurons and forms data utilizing a connectionist way to deal with calculation. By and large an ANN is a versatile frame work that progressions its structure based on external or internal data that streams through the system during the learning stage[9]. In this there are 2 ways possible:

- Multilayer perceptron network- it comes under single layer feed forward networks and also called simple perceptrons. The perceptron network consist of three units namely, input unit, hidden unit and output unit. In this system, the data moves in one and only course, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the system. The information preparing can reach out over various (layers of) units, however no feedback associations are available, that is, and associations reaching out from output of units to inputs of units in the same layer or past layers[9]. It is best for forecasting and specially used in areas of finance and banking.
- Back propagation network- It is a multilayer feedforward network that consist of units like input units, hidden units and output units with continuous differentiable activation functions. This is the network where the error is back propagated back to the hidden unit and input unit. The net input at hidden-layer and input-layer is calculated first then net-input between hiddenlayer and output-layer is calculated. Then input is mapped to the output. These all layers computation are inbuilt in the tool of neural network which is used for the optimization results[9].

Neural networks are fast and can work well with incomplete data as well. But since it is a black box technique it is unable to justify the found relationship. It only works with numerical data normalized between 0 and 1, this means all the tuple values have to be modified according to it which is risky and tedious.

C. Decision Trees
Decision trees tests many cases and then arrive at the best sequence for predicting the target. Each test creates branches and this goes on until testing terminates in a leaf node. The path from root to leaf is the rule that classifies the target. It is a supervised training process and this makes the decision trees which is known as induction[10]. Most decision goes through two phases:

- Splitting: The tree growing phase is an iterative process which keeps splitting the tree into smaller subsets. The split starts with the root node, and the subsequent iterations of splitting is done on the derivative nodes. At each split, all the attributes are analysed and the best decision is chosen. It follows greedy algorithm to split. The rules which stop the splitting are based on factors, like

maximum depth of the tree, or minimum number of elements in a node.
- Pruning: Once the tree is developed, it can be optimised by removing unwanted nodes or subtrees formed due to overfitting, or rules that are evaluated as inappropriate. This removal of nodes is known as pruning. It makes the tree more general and widely usable.

A decision tree is good as it helps the user identify which factors to use and how they are associated with the prediction of outcome. It is easy to understand and is unaffected by missing values in data.

D. Association and Sequencing Rule Discovery
Association rule discover rules as the knowledge that expresses the strength of connection among phenomena. For instance, if the rule is "when a customer bought item R, also he/she buys item S with high probability" obtained from the data, then this rule clarifies that how much item R contributes to sale item S. In addition, it leads to forecasting the sales volume with a certain degree that a customer buys item S when the customer bought item R[11].

The index of support denotes frequency that all elements included in the association rule appear in the database, and the index of confidence denotes accuracy of the association rule in the database[12].

An association rule is generally defined as {A,B,C....} -> {D,F,G...} where A,B,C,D,F,G are items i.e. attribute- value pairs. Then,

Support of A -> B = number of transactions containing both item sets A, B/ number of all transactions

Confidence of A -> B = number of transactions containing both item sets A, B / number of transactions containing item set A

Apriori algorithm generates an association rule at high speed. It is a high-speed algorithm to discover the association rule with support and confidence by setting a threshold to minimum support and minimum confidence[13]. Suppose that a database is denoted as D, a large item set is as L and the candidate item set is as C. Then, the database is scanned and the support in each candidate item set is examined. And assumed that L (k) is fulfilled minimum support in C(k) . Finally, it generates C(k+1) from L(k) . This process is continued until the candidate item set becomes empty. Hence, the final database with the association is obtained and can be used directly[14].

International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 6, September 2016

Table 1: Comparison table

Technique	Rule based discovery	Decision trees	Neural networks	Cluster detection
Advantages	Easy to understand. Apt for data that is complete with data relationships. Unsupervised learnings.	Easiest to understand. Apt when a particular target value is set on a complex set of attributes. It generates rules which are mutually exclusive. Supervised learnings.	Versatile. Both learnings. Good results with complicated (non- linear, noisy, missing values) data. Can process large amount of data.	Easiest to implement. Prior knowledge of the internal structure of a database is not required. Works well with categorical and numerical data. K-NN is supervised. SOM is unsupervised.
Disadvantages	Relationships might not be defined well. Combinatorial explosion because it tries for every possible relationship between all the fields in a database.	Not recommended for predictive estimation where a definite count is required. Not apt for time series data as well.	Cannot explain the relationships discovered within data. Work with numeric data and in [0, 1] otherwise conversion is necessary. Cannot find a solution for too many input features.	Requires a lot of memory. Difficult to find correct distance, measures and weights. K means is not good for clustering similar records, may be hard to interpret solution clusters (SOM).
Applications	For items that have strong relationship connecting them. For retail industries and time series problem.	For classification of records or predictions of outcomes. To generate rules to be sent to other software or translated in natural languages.	For classification and prediction. When there are many input features, select important features for the training phase. Finance, banking.	For a large complex dataset with many variables and internal structures. Mostly used for finding outliers.
Accuracy of prediction	Highest	Medium	Highest	Lowest
Time consumption	Low	Average. As all the steps like splitting, applying stopping criteria and pruning the existing decision tree.	Average. Takes time with the model configuration steps but less time in prior data transformations.	High. Because clustering at each stage takes time. SOM takes less time than K-NN.
Error rate (mean square error)	Low	Average	Least	Highest

[15]

IV.   CONCLUSION

We have studied five data mining techniques in this paper. They develop into machine learning algorithms and strive to provide accurate results. A table (table 1) of comparison is made to depict the advantages and disadvantages and more factors that affect the model building, hence affecting the quality of results. There are many data analysing algorithms but each of them applies for different cases and hence one should use a model and algorithm which is suitable for their purpose and provides accurate results. For CRM sequencing rule based algorithm works well as it keeps track of the product bought by customer and also helps to find loyal customers so that the value of organisation is increased in all.

REFERENCES

[1] Dr. Abdullah S. Al-Mudimigh, Zahid Ullah, Farrukh Saleem, "Data mining strategies and techniques for CRM Systems", IEEE.

[2] Ngai, E. W. T., Xiu, L., Chau and D. C. K, "Application of data mining techniques in customer relationship management: A literature review and classification", Expert Systems with Applications, vol. 2, pp. 2592-2602, 2009.

[3] Ankur Balar, Nikita Malviya, Swadesh Prasad, Ajinkya Gangrude, "Forecasting consumer behaviour with innovative value proposition for organizations using Big data analytics", IEEE, 2013.

[4] KK Tsiptsis, A Chorianopolous, "Data mining techniques in CRM: Inside Customer Segmentation", August 2011.

[5] Catherine Bounsaythip, Esa Rinta-Runsala, "Overview of data mining in consumer behaviour", 2001.

[6] Stephen Jansen, "Customer segmentation and customer profiling for a mobile telecommunications company based on usage behaviour", A Vodafone Case Study, 2007.

[7] Daniel T. Larose, "Discovering knowledge in data: An introduction to data mining", June 2014.

[8] G. Chicco, R.Napoli, F.Piglione, "Application of clustering algorithms and self organising maps to classify electricity customers", PowerTech Conference Proceedings, IEEE, 2003.

[9] Dr. Yashpal Singh, Alok Singh Chauhan, "Neural networks in data mining", JATIT, 2009.

[10] J.R.Quinlan, "Genereating production rules from Decision Trees", IJCAI, 1987.

[11] Junzo Watada, Kozo Yamashiro, "A data mining approach to consumer behavior", IEEE, 2006.

[12] Abhijit Raorane. R.V.Kulkarni, "Data mining techniques: A source for consumer behavior analysis."

[13] Sergey Brin , Rajeev Motwani and Craig Silvertrin, " Beyond market baskets :Generalizing association rules for correlation", SIGMOD Record, ACM, 1997.

[14] Chang Wang, Yanqin Wang, "Discovering consumer's behavior changes based on purchase sequences", IEEE, 2012.

[15] Syeda Farha Shazmeen, Mirza Mustafa Ali Baig, M.Reena Pawar, "Performance evaluation of different data mining classification algorithm and predictive analysis", IOSR-JCE, 2013.