Enhanced Boosted Trees Technique for Customer Churn Prediction Model

Kiranjot Kaur, Sheveta Vashisht

[1](M.Tech Scholar, Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab)
[2](Asst. Professor, Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab)

Abstract: - As the competition grows in the market, the organizations are more concern about customers than products. To be in the competition, organizations always want to retain their profitable customers. To predict the customers' churn behavior, data mining techniques are used. Many algorithms have been proposed to predict churning customers. In this paper, enhanced boosted trees technique is used for customer churn prediction. This approach will improve the performance of existing boosted trees technique. The process of customer churn prediction is discussed in this paper. It will be used to identify churn customers and effective marketing strategies could be planned for this group of customers.

Keywords: - *Boosted Trees, Churn Model, Classification Technique, Customer Churn Prediction, Data Mining*

I   INTRODUCTION

Data mining is the process of getting knowledge form the data to make decisions for the growth of an organization. Data mining tools perform data analysis and find interesting and useful data patterns. It is used in areas such as machine learning, database and data warehouse technology, statistics, information retrieval, neural networks, pattern recognition, knowledge-based systems, high-performance computing, artificial intelligence, image and signal processing, spatial or temporal data analysis, and data visualization [1]. The discovered knowledge can be applied to process control, query processing, decision making, and information management.

In recent years, management of organizations is moving from "Product-Centric" to "Customer-Centric" [2]. They are not only provides products to meet the need of customers but also improving their services to increase the loyalty and satisfaction of the customers. Acquiring the new customers is more expensive than retaining the existing customers. To retain the customers, organizations are more concern about the customer behavior analysis. The major factors of success include learning costumers' purchase behavior, developing marketing strategies to discover latent loyal customers [3]. The customer behavior is analyzed to making the marketing strategies and public policy. The stored data contains the information of about the spending behavior of customer, how much they buy, which day at what time he/she does the shopping, and what they buy most often, in that locality etc. The purchasing sequences can determine the changes in customers' preferences over time.

Consumer behavior is the study of individual, or group about their process of selecting and using the product, services, ideas or experiences to satisfy needs. New technologies of data mining can be used for Customer Relationship Management (CRM) and with this different marketing strategies are devised for different set of customers [4]. Researchers proposed hybrid techniques which used decision theory to check whether their proposed model is financially beneficial for an organization.

Data mining techniques are effective tool for analyzing consumer behaviors. There are seven powerful techniques with are useful for this purpose [7] such as Cluster Detection, Memory-Based Reasoning, Market Basket Analysis, Generic Algorithms, Link Analysis, Decision Trees, Neural Nets. Customer churn prediction model can be designed with CRISP-DM based on RFM and Random Forest Technique [8]. For the segmentation of churn customers researchers used random forest technique and boosted tree as a hybrid technique. Few researchers proposed a methodology to predict customers' purchasing behavior using purchase sequences of customers. Then sequential purchase patterns are extracted using association rule. Using purchase transactional records of customers, profile of the customers is build that describes the likes and dislikes of the customers [3]. Then SOM technique is used to detect customer purchase sequences. They predict the customer behavior using customer's purchase sequence base on transaction data.

The rest of the paper is organized as follows: Section II describes the customer churn. Section III describes literature review. Section IV describes the existing boosted trees technique. Section V represents a novel approach to enhance the performance of existing boosted trees technique. Section VI describes the methodology for predicting churn customers using enhanced boosted trees technique. In the last section, conclusion is presented.

II   CUSTOMER CHURN

When customer of one supplier is willing to change his or her supplier then this phenomenon is called customer churn. There are two types of churning customers [5]:

2.1   Non-Voluntary Churners: This category includes those customers who have had their service withdrawn by the organization.

2.2   Voluntary Churners: This category is divided into two subcategories such as:

2.2.1   Incidental Churn: This type of customers unintentionally leaves the organization because of reasons such as change in the location of their occupation.

2.2.2   Deliberate Churn: This type of customers decides to end the relationship with organization to seek the better services and cheaper good quality products.

In the process of customer churn management, company identifies the group of deliberate churners so that any proactive approach is used to retain those customers. According to Van Den Poel, customer churn management is affected by two approaches [6]:

2.3   Passive Approach: In this approach, when customer requests for ending the relationship with organization then required measured are taken to retain the customer.

2.4   Active Approach: In this approach, organization uses data mining techniques to identify those customers who want to end the relationship with organization and implement special incentive programs to retain the customers.

III   BOOSTED TREES TECHNIQUE

Boosted tree is used for segmentation of churn customers. Boosted trees method is very effective predictive data mining technique. It is used as boosting application method for decision trees. It assigned a weight for each leaning object. After training the previous classifier, weight of the learning objects is updated so that next classifier pay more attention to the object if it is not accurately classified by previous classifier. The assigned weight is used to vote for each classifier. If there is less error rate of classifier then more weight assigned to its vote. This training process is repeated. The weight of classifiers which voted for an object of a class is added. The class which gains higher total weight is the final class and it will introduced as the predictive class for that object.
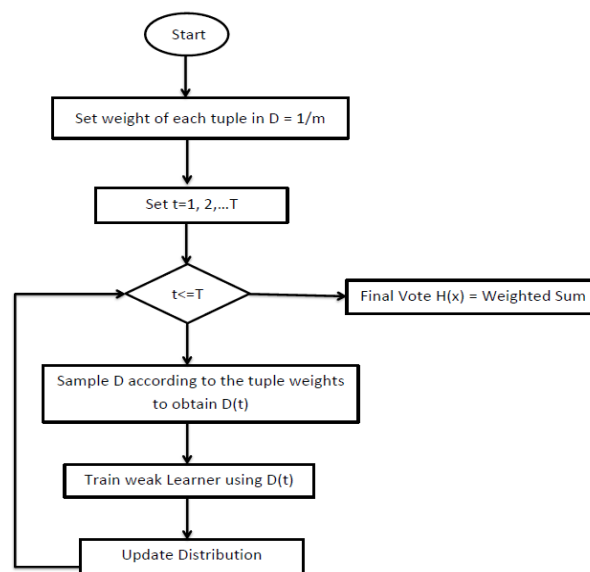
Fig 1. Diagram of Boosted Trees Technique

Here 'D' is number of data sets used in training. 't' is a variable used for iterating the dataset. It is used to keep track of items. 'm' is the a number used to divide data sets so that classifiers can be defined. 'T' is transaction set which contains all the transactions. 'H' is final decision of the tree.

IV   PROPOSED WORK

This paper proposes the enhancement in the performance of existing boosted tree algorithm by removing repeated scanning for scoring of the classifiers. We are proposing an algorithm to generate the boosted tree whose trees are fuzzy decision trees. Each tree in the hierarchy will be fuzzy tree.

In existing Boosted tree algorithm, building classifiers and then weighted vote to the decision of the classifier takes a long time. We are reducing this time by increasing the accuracy of the classification using fuzzy logic and removing noise from the test data before starting the processes. As smaller differences in data items can lead to drastic decision changes. The Fuzzy logic will help in decreasing decision change if data contains small changes.

The iterations are used to find the strong classifier. At each step of the iteration a new weak classifier is added and weighted according the learner's accuracy and/or the step in the iterative process. A characteristic inherent in fuzzy trees is that the classification of an example can derive in two or more leaves due to the overlapping of the fuzzy sets. As a result of increased accuracy and data partitions more data will be consumed in lesser number of classifiers. The lesser number of classifiers will result into lesser decision time and increased performance of the algorithm.

Meanwhile, the data is reweighted by assigning more importance to the still misclassified observations; as a consequence newest classifiers focus more on the population that was not correctly classified in previous steps.
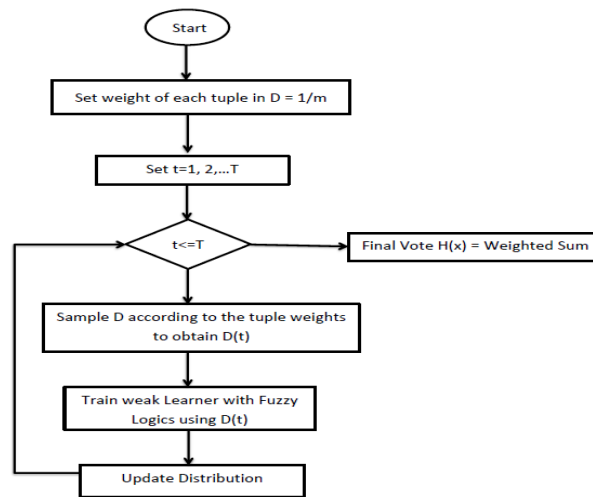

Fig 2. Diagram of proposed enhanced boosted trees approach

The above Fig 2 represents flow of enhanced boosted tree algorithm. The algorithm starts with initializing the weight of each tuple in dataset D. Choose set of transactions from D as a training set. Training set is then passed through the Classifier to train it for decision making. Update the weight of correctly classified transaction or tuple and normalize all the transactions such that sum of weights of all the transactions remains same. So weights 'w' are assigned to the classifier's vote. Then again choose a set of transactions according to their weights and train weak classifiers and generate T classifiers in each round. At final stage, sum all the weights of votes of classifiers for class c. which class has higher weight that class would be predictive class.

V   METHODOLOGY

In the first step, samples are selected for better understanding and predicting customer attrition. In second step, necessary features for predicting customer attrition are collected. In third step, data preprocessing is performed. In data preprocessing, data may collected from different sources, having some missing values in it so there is a need to integrated, cleaned and transformed the data before supplied to the data mining algorithm. As the quality of output is directly affected by the quality of input values. In fourth and fifth step, there is training and evaluation of predictive models to choose the one best model for predicting customer attrition.
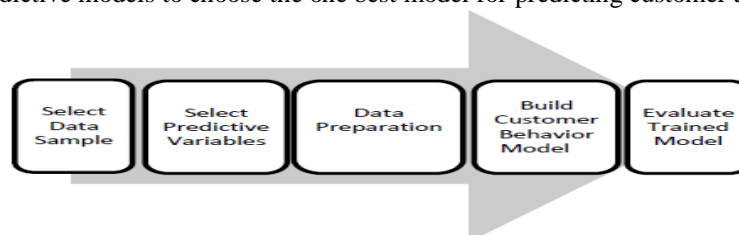

Fig 3. Customer Behavior Analysis Model

2.5 Data Sample

In this collected the data of customers, their purchasing items and amount of money spend by them. We can integrate the data from multiple sources. We can use on or off web data for this purpose.

2.6 Predictive Variables

Amount of gathered data may be huge and we want to use this data for specific application. But for data mining tasks whole data is not required to process. So we have to use effective data reducing strategies. For this many variables are selected for customer churn prediction. These variables are also known as efficient predictive factors. Following factors are [8]:

5.2.1 Purchase Frequency: It is number of times a customer makes a purchase in a given period of time.

5.2.2 Length of Relationship (LOR): It indicates the period of time between the customers' first and last purchase in an organization.

5.2.3 Relative Purchase Frequency: It is the ratio of the number of customer's visits to the active relationship length of the customer with the organization.

5.2.4 Average Inter-purchase Time: It is the average time elapsed between two successive purchases of the customer.

5.2.5 Variance of Inter-purchase Time: It checks the regularity of customer's behavior in visits to the organization.

5.2.6 Last Inter-purchase Time Difference: It is the variations of the customers' recency in the last purchase in the organization.

5.2.7 Monetary value: It is the amount of money spends by the customer in the organization. This variable is used for predicting customers purchase patterns.

5.2.8 Relative Monetary value: It is the amount of customers' purchase compared to his interaction time with the organization.

5.2.9 Weight of Sold Items: It indicates the total weight of customers' purchased goods.

5.2.10 Customer Behavior in Product Category: It indicates the customers' purchasing behavior in different product categories.

5.2.11 Number of Categories: It indicates the total number of product groups purchased by the customer.

5.2.12 Customer Type: Customers are divided into two parts such as real and legal. The legal customers are divided into subgroups such as hotels, supermarkets, chain stores, restaurants, schools, villages and etc. The loyal customer increases the possibility of marketing and providing better services in order to encourage them.

2.7   Data Preparation

In this we have to clean and transform the data to initiate data processing. Missing values are removed from the data. This data is provided to classification technique as an input.

2.8   Customer Behavior Model

Classification technique is used to build customer churn behavior model. Enhanced Boosted Trees technique is used as the classification technique to separate the churn customers from the loyal customers. This proposed approach is discussed above in Section IV.

2.9   Evaluation

This is to check the accuracy of our proposed predictive model using lift criterion. The lift criterion is a performance measure which is the result of the ratio between the obtained outcomes using and without using the prediction model. The higher the lift means the model is more accurate and targeted proactive churn management program will be more profitable [10].

VI   CONCLUSION

Organizations need to be more concern about customers and always work for Customer Relationship Management. Losing present customers causes additional costs due to the loss in sale and it generated the need of attracting more new customers. Retention is the most valuable asset for organization. For this purpose data mining techniques are used. So this paper proposed a new approach to enhance the performance of existing boosted trees technique. Then this enhanced approach is used for predicting customer churn. So that proactive measures could be taken by company for churn prevention.

REFERENCES

[1]   J. Han and M. Kamber, *Data Mining: Concepts and Techniques* (Morgan Kaufmann, India, 2006)

[2]   S.Y. Hung, D. Yen and H.Y. Wang, Applying Data Mining to Telecom Churn Management, *Expert System with Applications, Elsevier, 31(3)*, 2006, 515-524.

[3]   C. Wang and Y. Wang, Discovering Consumer's Behavior Changes Based on Purchase Sequences, *Proc. 9th IEEE-International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012*), Sichuan, China, 2012, 642-645

[4]   M. M. Rosario and V. Pedro, Tracking of consumer behavior in e-commerce, *Proc. 16th International Conference on Information Fusion,* Istanbul, Turkey, 2013, 1214-1221

[5]   J. Hadden, A. Tiwari, R. Roy, and D. Ruta, Computer assisted customer churn management: State-of-the-art and future trends, *Computers and Operations Research, 34(10),* 2007, 2902-2917.

[6]   E.W.T. Ngai, L. Xiu and D.C.K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification*, Expert System with Applications, Elsevier, 36(4),* 2009, 2592-2602.

[7]   M. J. Berry and G. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support* (John Wiley & Sons, New York, NY, USA, 1997)

[8]   S. Nabavi and S. Jafari, Providing a Customer Churn Prediction Model using Random Forest Technique, *Proc. 5th IEEE-Conference on Information and Knowledge Technology (IKT)*, shiraz, Iran, 2013, 202-207

[9]   Javad Basiri, Fattaneh Taghiyareh and Behzad Moshiri., A Hybrid Approach to Predict Churn, *Proc. IEEE Asia-Pacific Services Computing Conference,* Hangzhou, China, 2010, 485-491.

[10]   S. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. Mason, Defection detection: Measuring and understanding the predictive accuracy of customer churn models, *Journal of Marketing Research, 43(2),* 2006, 204-211.