Churn Prediction in Telecommunication Using Classification Techniques Based on Data Mining: A Survey

Nisha Saini
Student, Department of Computer
Science and Applications,
Kurukshetra University,
Kurukshetra, Haryana, India

Monika
Asstt. Professor, Department of Computer
Science and Applications,
Kurukshetra University,
Kurukshetra, Haryana, India

Abstract- The rapid growth of the market in telecom industries is leading to a bigger subscriber base for service provider. More competitors along with innovative business models and better services are increasing the cost of customer acquisition. Hence, providers need to put more efforts for prediction and prevention of churn. This review paper aims to present most commonly used data mining techniques used for churn identification.

Keywords - Churn Prediction, Data mining, Classification, Decision Tree, Neural Networks.

I.   INTRODUCTION

In telecommunication industry, customers are able to choose among multiple service providers and can actively switch from one service provider to another. A number of factors contribute to the risk of customers switching to competitors. These factors include stiff competition in telecom, increasing innovation, low switching costs and deregulation by the government. Customer churn is a costly risk if not managed properly. As the size of organization increases, cases of churn also increase making it difficult to manage. Previous studies have shown that cost of acquiring a new customer is much more than the cost of retaining a customer. Costs associated with customer churn include loss of revenue, cost of customer retention and reacquisition, advertisement costs, etc. *[1]*.Data mining methods are used to discover unknown patterns in huge data sets and they can be used in many CRM applications such as credit card fraud detection, credit score, churn prediction etc. Classification is one of the data mining tasks, which aims at classifying unknown cases based on a set of known examples. The objective of churn prediction is to find out which type of customers of a telecommunication will churn and when. Data mining techniques such as Decision Trees, Nearest Neighbor, Support Vector Machine and Artificial Neural Networks are used in churn analysis in order to predict whether a particular customer will churn and also the reasons for a particular customer churn *[2][3]*.

II.   TYPES OF CHURNERS

As shown in figure 1, there are two main types of churners: voluntary and involuntary. Voluntary churners are more difficult to determine and occur when a customer initiates termination of a service customer. Involuntary churners are easier to identify and include those customers which are disconnected by the company for fraud, non-payment or under-utilization of services.

Voluntary churn can further be categorized as deliberate and incidental churn. Incidental churn happens because of any unexpected incident in customer's life such as change in financial condition churn and location churn etc. Deliberate churn happens because of change in technology, economics, service quality factors, social and psychological factors, and convenience. Deliberate voluntary churning is the major concern for most of the telecommunication companies and hence need to be dealt with properly *[4]*.
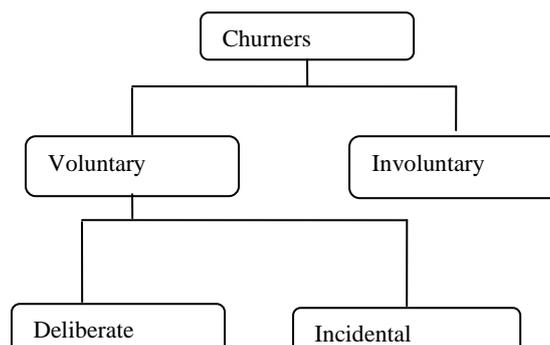

Fig. 1 Churn Taxonomy

III. LITERATURE SURVEY

According to research papers studied from decade of 90's till date, following observations were made:

Clement et al. *[1]* have presented new features categorized as contract-related, call pattern description, and call pattern changes description features derived from traffic figures and customer profile data. The given features were evaluated using naïve bayes and bayesian network and obtained results were compared to results obtained using decision tree. Results have shown that probabilistic classifiers have shown higher true positive rate than decision tree but decision tree performs better in overall accuracy. Kamalraj and Malathi in *[2]* have proposed a new model to be developed for churn prediction. The authors have proposed the use of a combination of techniques in order to make better business decisions.

Essam et al. in *[4]* have introduced a simple model based on data mining to track customers and their behavior against churn. A dataset of 500 instances with 23 attributes has been used to test and train the model using 3 different techniques i.e., Decision trees, SVM and Neural networks for classification with k-means for clustering. Results indicate that SVM has been stated as the best suited method for predicting churn in telecom.

Vladislav and Marius in *[5]* have presented quality measures of six churn prediction models including regression analysis, naïve bayes, decision trees, neural networks etc. They have also pointed out the links between churn prediction and customer lifetime value. According to the authors, new prediction models need to be developed and combination of proposed techniques can also be used. Umman Tuğba Şimşek Gürsoy *[6]* have compared regression techniques with decision tree based techniques. Results have shown that in logistic regression analysis churn prediction accuracy is 66% while in case of decision trees the accuracy measured is 71.76%. Hence decision tree based techniques are better to predict customer churn in telecom.

V. Umayaparvathi and K. Lyakutti *[7]* have used neural networks and Decision trees to build the churn prediction model. Results have shown that Decision trees have the predictive accuracy of 98.88% and an error rate of 1.11167%. Similarly neural networks have shown the predictive accuracy of 98.43% with 1.5616% of error rate. Clearly decision trees have outperformed neural networks for churn prediction. According to the authors, selection of right combination of attributes and fixing the proper threshold values may produce more accurate results. Khalida et al. in *[8]* have used a specific training sample set was used to conduct an experiment on customer churn factor using decision tree. According to the authors, rule information can be easily understood by decision tree. An attempt has been made to identify various factors responsible for customer churn such as area.

Amal et al. *[13]* have reviewed that generally Decision tree based techniques, neural network trees and Regression techniques are applied in churn prediction. Decision tree based techniques outperform all other in terms of accuracy. On the other hand, neural networks outdo other techniques due to size of data sets.

It can be concluded from the existing research in the field of customer churn prediction that decision tree performs best among all data mining techniques.

IV. STEPS INVOLVED IN CHURN PREDICTION

A. Data Collection

Telecom data consist of Customer data and Call Detail Data (CDR). Customer data has various attributes like name, ID, etc. On the other hand CDR is a data record as given by telephone exchange and it contains fields such as the calling number, call type, duration of call, cell ID, switch number etc.

B. Data Preparation

This step can be skipped if a data warehouse is being used as a single source of data needed for analysis purpose. This is because data needs to be cleaned, integrated and prepared in a particular format. Since the data is obtained from multiple sources, it needs to be combined under a single key. Some variables may need to be transformed in accordance with the requirements of data. All these functions can be performed using any database language such as SQL.

C. Selecting a data mining function

Churn prediction is a type of classification problem where the main task is to identify that whether a customer falls into category of churners or not. Classification algorithms such as naïve bayes classifier, k-nearest neighbor, decision tree, support vector machine may be employed for this purpose.

D. Model Building

In this phase data is presented to a data mining software program. Some changes may occur in variables if they are not useful and hence analyst may need to roll back to make changes in the dataset.

E. Interpreting the Results

The results should be presented in a coherent manner and the model should be evaluated with respect to problem solving objectives.

F. Applying the Results

The ultimate goal is to apply what was discovered in order to solve the business problem defined, i.e., customer churn *[3]* *[9]*.

IV. POTENTIAL REASONS FOR CHURN

Various reasons are responsible for churn in telecom. Most important of them are the call charges and network connectivity. In the era of technology, data plans also play a major role in selecting a particular service provider. For the customers who tend to move from one place to another, roaming charges are equally important along with network coverage in a particular area. For any kind of problem, subscribers normally seek help from the customer care. Hence the customer service of companies should be easily available and helpful. The bill payment facilities should also be available in case of postpaid service.

V. DATA MINING TECHNIQUES FOR CHURN PREDICTION

Since churn prediction is a classification problem, therefore various classification algorithms can be used to predict customer churn. Classification algorithms require a data set to be worked upon and after the analysis of data set it can be predicted whether a particular customer will churn or not *[10][11] [12]*.

A. Neural Networks

Neural network is a data mining techniques that has the capability to learn from errors. According to *[13]*, neural networks outperform decision trees for prediction of churn but they need large volumes of data and a lot of time for calculation.

B. Statistical Methods

Statistical methods such as naïve bayes classifier, k-nearest neighbor etc., in data mining is used to process large volumes of data. As suggested by *[13]*, techniques based on regression have given good results in prediction and estimation of churn.

C. Decision Trees

This algorithm uses a divide and conquers approach to construct a tree structure. Nodes in the tree represent features while branches represent values connecting features *[1]*. As stated by *[13]*, decision tree are not suitable for complex relationship between attributes but pruning a decision tree can help to improve its performance.

Comparison of various data mining techniques used for churn prediction can be summarized as shown in Table 1.

Table 1: Comparison Of Various Data Mining Techniques

Classifiers	Method	Parameters	Pros	Cons
Bayesian Method	This algorithm attempts to estimate the conditional probabilities of classes given an observation using the joint probabilities of sample observations and classes.	All model parameters can be approximated with relative frequencies from the training set.	1. Computations are simplified in Naïve Bayes Classifier. 2. Exhibit high accuracy and speed when applied to large databases.	1. The assumptions made in class conditional independence. 2. Lack of available probability data.
Decision Trees	This technique builds a binary classification tree. Each node corresponds to a binary predicate on one attribute; one branch corresponds to the positive instances of the predicate and the other to the negative instances.	Decision Tree Induction uses parameters such as a set of candidate attributes and an attribute selection method.	1. No domain knowledge is required for tree construction. 2. High dimensional data can be handled easily. 3. Representation can be easily understood. 4. Both numerical and categorical data can be processed.	1. Output attribute must be categorical. 2. Limited to one output attribute. 3. Decision tree algorithms are unstable. 4. Tress created from numeric datasets can be complex.
Artificial Neural Network	It is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.	ANN uses a cost function C as a measure of how far away a particular solution is from an optimal solution to the problem to be solved.	1. Requires less formal statistical training. 2. Able to implicitly detect complex nonlinear relationships between dependent and independent variables. 3. High tolerance to noisy data. 4. Availability of multiple training algorithms.	1. More computational burden. 2. Proneness to over fitting. 3. Requires long training time.

VI.   CONCLUSION AND FUTURE WORK

Customer churn is an increasing risk in mobile telecom services. As discussed in the paper, certain factors are responsible for customer churn. These issues need to be addressed in order to predict customer churn and prevent the company for loss of revenues and other costs associated with customer churn. The future work aims at predicting customer churn using decision tree in SPSS environment and comparing the results obtained by three decision tree algorithms namely-CHAID, CART and QUEST.

REFERENCES

[1]   Clement Kirui, Li Hong, Wilson Cheruiyot, Hillary Kirui, "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining", International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, ISSN: 1694-0814, March 2013.

[2]   N.Kamalraj, Dr. A.Malathi, "Applying Data Mining Techniques in Telecom Churn Prediction", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 10, ISSN: 2277 128X, October 2013.

[3]   Pushpa, Dr. G Shobha, "An Efficient Method of Building the Telecom Social Network for Churn Prediction", International Journal of Data Mining & Knowledge Management Process, Vol. 2, No. 3, May 2012.

[4]   Essam Shaaaban, Yehia Helmy, Ayman Khedr, Mona Nasr, "A Proposed Churn Prediction Model", International Journal of Engineering Research and Applications, Vol. 2, Issue 4, ISSN: 2248-9622, June-July 2012.

[5]   Vladislav Lazarov, Marius Capota, "Churn Prediction", Technische Universität München, Eighth ACM SIGKDD International Conference, 2007.

[6]   Umman Tuğba Şimşek Gürsoy, "Customer Churn Analysis in Telecommunication Sector", Istanbul University Journal of the School of Business Administration, Vol. 39, No. 1, ISSN: 1303-1732, 2010.

[7]   V. Umayaparvathi, K. Lyakutti, "Applications of Data Mining in Telecom churn Prediction", International Journal of Computer Applications, Vol. 42, No. 20, ISSN: 0975-8887, March 2012.

[8]   Khalida, Sunarti, Norazrina, Faizin, "Data Mining in Churn Analysis Model for Telecommunication Industry" Journal of Statistical Modeling and Analytics, Vol. 1, No.19-27, ISSN: 2180-3102, 2010,

[9]   D. Ćamilović, "Data Mining and CRM in Telecommunications", Serbian Journal of Management, February 2008.

[10]   V. Vaithiyanathan, K. Rajeswari, Kapil Tajane, Rahul Pitane, "Comparison of Different Classification Techniques using Different Datasets", International Journal of Advances in Engineering and Technology, Vol. 6, Issue 2, ISSN: 2231-1963, May 2013.

[11]   Thair Nu Phyu, "Survey of Classification Techniques in Data Mining", Proceeding of International MultiConference of Engineers and Computer Scientists, Vol. 1, ISBN: 978-988-17012-2-0, March 2009.

[12]   Rahul J. Jadhav, Usharani T. Pawar, "Churn Prediction in Telecommunication Using Data Mining Technology", International Journal of Advanced Computer Science and Applications, Vol. 2, No. 2, Feb 2011.

[13]   Amal M. Almana, Mehmet Sabih Aksoy, Rasheed Alzaharni, "A Survey on Data Mining Techniques in Customer Churn Analysis for Telecom Industry", International Journal of Engineering Research and Applications, Vol. 4, Issue 5, ISSN: 2248-9622, May 2014.