

A Survey on Churn Prediction Techniques in Communication Sector

N.Kamalraj

Asst. Prof., Department of Computer Technology
Dr. SNS Rajalakshmi College of Arts and Science
Coimbatore-49

A.Malathi, PhD.

Asst. Prof., PG and Research Department of
Computer Science, Govt. Arts College
Coimbatore-18

ABSTRACT

The speedy augmentation of the market in every sector is leading to superior subscriber base for service providers. Added competitors, novel and innovative business models and enhanced services are increasing the cost of customer acquisition. In such a tedious set up service providers have realized the importance of retaining the on hand customers. It is therefore mandatory for the service providers to inhibit churn- a phenomenon which states that customer wishes to quit the service of the company. To prevent the churn many approaches are used by the researchers. This paper reviews the different approaches used by researchers not only in communication sector but also other sectors which highly depends on customer participation.

Keywords

Churn Prediction, Customer retention

1. INTRODUCTION

In the world of ever increasing competition on the market, companies have become observant that they should put much effort not only to convince customers but also to keep hold of existing customers. Churners are persons who quit a company's service for some reasons. To reduce customer churn, company should be able to predict the behavior of customer accurately and establish links between customer attrition and keep factors under their control. Churn prediction is a binary classification task which differentiates churners and non-churners. In the literature, many prediction algorithms have been applied to predict the customer churn. Finding the churners can help companies retain their customers. One of the major reasons for predicting churn is that it costs less to retain existing customers than to acquire new customers [15]. Customer acquisition is an expensive process and it costs up to five times as much to make a sale to a new customer as it does to make an additional sale to an existing customer [6,16,17,18,19]. Therefore a small improvement in customer retention can lead to a significant increase in profit [10].

In the last few years, there are revolutionary things have happened in the telecommunications industry, such as, new services, technologies and the liberalizations of the market opening up to competition in the market. Since the customer is the major source of profit, a method to promptly manage customer churn gains vital significance for the survival and development of any telecommunication company. For many telecoms companies, figuring out how to deal with Churn is turning out to be the key for continued existence of their organizations [20, 22].

1.1 Customer Churn

Customer churn is a term used in the telecommunication service industry to denote the customer movement from one service provider to another. Reasons for this may be the dissatisfaction

with the quality of service, high costs, unattractive plans, no understanding of the service plan, bad support, etc. Furthermore the customer quits contract without the aim of switching to a competitor. Reasons for this are changes in the situation that makes impossible for the customer from further requiring the service, e.g. financial problems, leading to impossibility of payment or change of the geographical location of the customer to a place where the company is not available or the service is unavailable. Sometimes the company also stops or withdraws the contract due to company policy reason.

In general, Churn means to role of the customers who are about to move their usage of service to a competing service provider. Churn prediction methods gives the prediction about customers who likely to churn in the near future whereas churn management helps on the other side which aims to identify such churners and to carry out some positive actions to minimize the churn effect.

Van den Poel et al., [10] focused four sets of data variables for customer attrition, that is, customer behavior, customer perceptions, and customer demographics and macroenvironment.

Customer behavior recognizes the exact components of the service are utilizing and how often are they employing them. In telecommunications, for example, the number and length of calls, period between calls, the usage of the network for data exchange, etc could be taken by the provider.

Customer perceptions are identified as the way a customer pick up or stop the service and can be calculated with customer surveys and include data link overall contentment, quality of service, problem experience, satisfaction with problem handling, interest given, location convenience, image or reputation of the company, customer perception of dependency to the vendor, etc.

Customer demographic includes age, sex, education, social status, geographical data are also used for churn calculation.

Macro environment variable identifies the changes in the world and the different experience of customer which affect the way they use the service. For instance, in the telecommunication trade people who have survived a natural disaster and could rely on their mobile phones during it are more likely to continue using the service.

2. LITERATURE SURVEY

B.Huang et al., [1] proposed use of new set of features for churn prediction and henley segment was used to divide customer into several groups. The experiments were conducted using several traditional modeling techniques and they found that that DMEL is impractical on lager dataset with high dimension. Adem Karahoca et al., [5] proposed a data mining solution with a neural

network model to predict churners. The x-means and fuzzy c-means algorithm were used for feature selection and used ANFIS learning algorithm for churn prediction. The results proved that ANFIS combine with fuzzy c-means have shown better results than decision tree and ridor. C.-F.Tsai et al., [13] used association rules to extract feature from original one to improve the prediction performance. The results proved that decision tree perform better than NN model when association rules are used. B.Huang et al., [14] proposed genetic algorithm (NSGA) to find number of features subset in different size and dimension. The experiments were carried out using decision tree C4.5 and results proved that the NSGA algorithm is efficient and successful for churn prediction. Y.Huang et al., [20] presented a new approach which based on chi-square method to select features for customer churn prediction and demonstrated the results with five different methods like DT, NB, LR, SVM and DMEL.

M.Owczarczuk [3] used logistic regression, G.Nie et al., [4] used logistic regression and decision tree model and A.Keramati, S.M.S.Ardabili [2] focused on Binomial logistic regression model for churn prediction and identified customer dissatisfaction, service usage, switching cost and demographic variable affects customer churn. B.Shim et al., [7] used decision tree, neural network and logistic regression for customer classification and identified decision tree shows highest hit ratio among them and P.Kisioglu, Y.I.Topcu [8] applied bayesian belief network to find out most important factors that have effects on customer churn in telecommunication industry and CAID algorithm is used to discretize continuous variable in churn.

Y.Xie et al., [11] proposed improved balance random forest model in order to address the limitation in existing algorithm and the results proved that IBRF is better than artificial neural network, decision tree and support vector machines.

Y.Xie et al., [11] proposed improved balance random forest model in order to address the limitation in existing algorithm and the results proved that IBRF is better than artificial neural network, decision tree and support vector machines. P.C.Pendharkar [12] proposed genetic-algorithm based neural network model to predict churn and compared the result with statistical z-score based prediction model.

Javad Basiri et al., [21] proposed a hybrid approach (OWA) based on LOLIMOT and Bagging & Boosting algorithms to improve the prediction accuracy of churn and used chi-square algorithm for feature selection. The Order weighted averaging (OWA) method uses the strength of both LOLIMOT and bagging and boosting classification tree. This method was compared with C5.0, neural network, logistic regression, bayesian network, LOLIMOT and bagging and boosting classification tree. The results proved that the OWA technique outperformed than the other classifier.

Yongbin Zhang et al., [22] attempted to develop behavior-based telecommunication churn prediction system with artificial neural network approach (SOM) and used only customer service usage information for prediction. A.A.Khan et al., [23] used decision tree, logistic regression and neural network to predict churn and suggested that demographic features have the lowest affect on the churn prediction when compared to the billing and usage details.

Anuj Sharma et al., [24] focused artificial neural network approach to predict churn and suggested that neural network can be combined with other techniques like support vector machines, genetic algorithm to develop a new hybrid model to improve the

performance and accuracy for churn prediction. Michael C.Mozer et al., [25] used neural network, logistic regression to obtain higher prediction accuracy in churn. Wai-Ho Au et al., [26] proposed DMEL to handle classification problem in churn prediction and results proved that DMEL outperformed neural networks. Gang Cui [27] focused on BP neural network to predict the influence of customer retention for churn management.

Y.-H.Lee et al., [28] projected kNN-based time-series classification techniques to achieve better performance for churn prediction and results proved that kNN-TSC achieves better performance than the traditional statistical-transformation-based approach. T.S.Zabkowski, W.Szczesny [29] demonstrated neural network and decision tree for customer insolvency in cellular telecommunications and the results proved that neural network models are more stable than decision trees.

W.Verbeke et al., [30] proposed two novel data mining techniques for customer churn prediction. The first one called AntMiner+ uses Ant Colony Optimization gather rules from data and second one named Active Learning Based Approach for support vector machine rule extraction and experiment were conducted with C4.5, RIPPER, SVM, logistic regression. Experiment proved that ALBA combined with C4.5/RIPPER results in higher accuracy than the AntMiner+.

Many prediction algorithms have been used for predicting churn. Some of the well-known algorithms are genetic algorithm, neural networks, decision tree, logistic regression and cluster analysis. Fig. 1 illustrates the share of each algorithm in the churn management systems [9, 21].

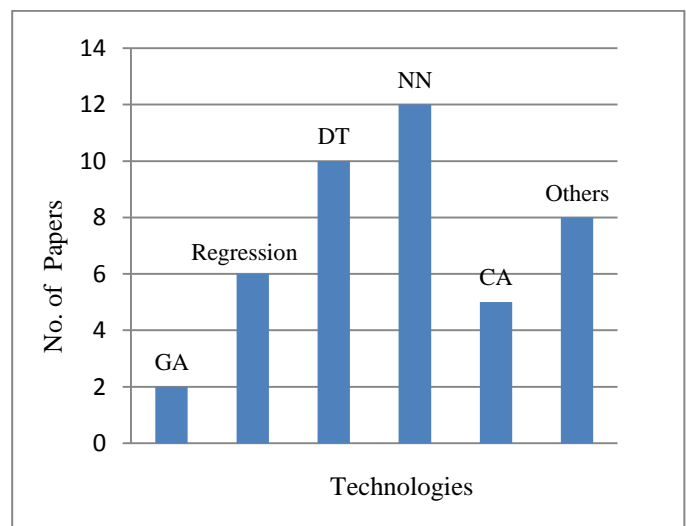


Figure 1: Share of each algorithm in the Churn Prediction

Based on reference materials we can come to the conclusion that the decision tree, neural network, regression and cluster analysis are preferred by the most of researchers where as the other algorithms are not used to that significant level. The amount of work that has been done using genetic algorithm is very small [9], which propose that research into these technologies has not yet been exhausted and the other new technology can also be tried with the churn prediction and management.

3. PROBLEMS AND DIRECTIONS

The aim of this study is to identify the different algorithm used for churn prediction. Once churns are identified company has to take further action to prevent the churn. The managerial department of that company has to use such identified churns and put efforts as how much revenue a service provider is going to get over are to be made to retain period of customer remain. In general, the customer lifetime value is highly connected with the customer decision to stay back. The customer lifetime value can be combined with the churn prediction to reduce the cost for making a excessive retention effort (false positives) and the cost of losing a customer because the model not accurately predict churn (false negatives).

Table 1: Churn Prediction Categories

	Actual Churners	Actual Non Churners
Prediction Churners	True Positive	False Positive
Prediction Non Churners	False Negative	True Negative

Usually the data size used in communication industry is very large, which leads in high dimensionality, making analyze a complex and challenging task. Therefore, before beginning to use a churn prediction method the feature selection technique is used, which selects the attributes that can be of exploited and ignored for prediction.

The final aim of the churn study is to predict what will happen to the customer in the future. To solve this problem we have to identify churn before they become actual churns. Furthermore we have to design a model to classify the churn accurately. Therefore, we must split the time window into two phases, i.e. the observation period and the performance period. In the observation period, we have to design variable to observe the transaction behaviors of a customer, and then we check whether the customer became a churning or not during the performance period. The independent variables are calculated from the data recording the information obtained during the observation period and the dependent variable is calculated from the data recording the information during the performance period.

4. CONCLUSION

In this paper, a detailed discussion on the process of churn prediction of customer in communications is done. From the literature review conducted it is understood that churn prediction in communication is very important for customer retention and become mandatory for all kind of service industry. The purpose of this paper is not to propose a new algorithm, but focuses on the implementation and the understanding of the existing mode. There are many different ways of churn prediction and new techniques continue to emerge not only in communication sector but also in other fields. Future research direction is planned in the direction of developing a new model for churn prediction. Good prediction models have to be constantly developed and a combination of the proposed techniques has to be used. They have to be properly tailor made by the companies for their existence.

5. REFERENCES

[1] Bingquan Huang, Mohand Tahar Kechadi, Brian Buckley "Customer churn prediction in telecommunications", *Expert Systems with Applications* 39 (2012) 1414–1425.

[2] Abbas Keramati, Seyed M.S. Ardabili, "Churn analysis for an Iranian mobile operator", *Telecommunications Policy* 35 (2011) 344–356.

[3] Marcin Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts", *Expert Systems with Applications* 37 (2010) 4710–4712.

[4] Guangli Nie, Wei Rowe, Lingling Zhang, Yingjie Tian, Yong Shi, "Credit card churn forecasting by logistic regression and decision tree", *Expert Systems with Applications* 38 (2011) 15273–15285.

[5] Adem Karahoca, Dilek Karahoca, "GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system", *Expert Systems with Applications* 38 (2011) 1814–1822.

[6] Wouter Verbeke, Karel Dejaeger, David Martens, Joon Hur, art Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach", *European Journal of Operation Research*, 218 (2012) 211-229.

[7] Beomsoo Shim, Keunho Choi, Yongmoo Suh, "CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns", *Expert Systems with Applications* 39 (2012) 7736–7742.

[8] Pýnar Kisioglu, Y. Ilker Topcu, "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey", *Expert Systems with Applications* 38 (2011) 7151–7157.

[9] J. Hadden, A. Tiwari, R. Roy, D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends", *Computers and Operations Research*, vol. 34, no. 10, 2007, pp. 2902-2917.

[10] Dirk Van den Poel, Bart Larivière, "Customer Attrition Analysis For Financial Services Using Proportional Hazard Models".

[11] Yaya Xie, Xiu Li, E.W.T. Ngai, Weiyun Ying, "Customer churn prediction using improved balanced random forests", *Expert Systems with Applications* 36 (2009) 5445–5449.

[12] Parag C. Pendharkar , "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services", *Expert Systems with Applications* 36 (2009) 6714–6720.

[13] Chih-Fong Tsai, Mao-Yuan Chen, "Variable selection by association rules for customer churn prediction of multimedia on demand", *Expert Systems with Applications* 37 (2010) 2006–2015.

[14] Bingquan Huang, B.Buckley, T.-M.Kechadi, "Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications", *Expert Systems with Applications* 37 (2010) 3638–3646.

[15] Roberts,J.H, "Developing new rules for new markets", *Journal of the Academy of Marketing Science*, 28(1), 31–44 (2000).

- [16] Antreas D. Athanassopoulos, “Customer Satisfaction Cues To Support Market Segmentation and Explain Switching Behavior”, *Journal of Business Research* Volume 47, Issue 3, March 2000, Pages 191–207.
- [17] Donath, Bob, “Show Me the Money”, *ISBM Insights* 8 (2) (1998).
- [18] Floyd, T. “Creating a new customer experience”, *Bank Systems and Technology*, 37(1), R8–R13 (2000).
- [19] Slater, S. F., & Narver, J. C. (2000). Intelligence generation and superior customer value. *Journal of the Academy of Marketing Science*, 28(1), 120–127.
- [20] Y. Huang, B. Q. Huang, M. T. Kechadi, “A New Filter Feature Selection Approach for Customer Churn Prediction in Telecommunications”, *Proceedings of the IEEM, IEEE* (2010) 338-342.
- [21] Javad Basiri, Fattaneh Taghiyareh, Behzad Moshiri, “A Hybrid Approach to Predict Churn”, *Proceedings of Asia-Pacific Services Computing Conference IEEE* (2010) pp.485-491.
- [22] Yongbin Zhang, Ronghua Liang, Yeli Li, Yanying Zheng, Michael Berry “Behavior-Based Telecommunication Churn Prediction with Neural Network Approach”, *Proceedings of International Symposium on Computer Science and Society IEEE* (2011), 307 – 310.
- [23] Afaq Alam Khan, Sanjay Jamwal, M.M. Sepehri, “Applying Data Mining to Customer Churn Prediction in an Internet Service Provider”, *International Journal of Computer Applications (0975 – 8887)* Volume 9– No.7, November 2010.
- [24] Anuj Sharma, Dr. Prabin Kumar Panigrahi, “A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services”, *International Journal of Computer Applications (0975 – 8887)* Volume 27– No.11, August 2011.
- [25] Michael C. Mozer, Richard Wolniewicz, David B. Grimes, Eric Johnson, Howard Kaushansky “Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry”, *IEEE Transactions on Neural Networks*, Vol 11, No. 3, May 2000.
- [26] Wai-Ho Au, Keith C. C. Chan, Xin Yao, “A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction”, *IEEE Transactions on Evolutionary Computation*, Vol. 7, No. 6, December 2003.
- [27] Gang Cui, “A Methodologic Application of Customer Retention Based on Back Propagation Neural Network Prediction”, *Proceedings of Computer Engineering and Technology (ICCET) IEEE* (2010), Volume: 3, Page(s): V3-418 - V3-422.
- [28] Yen-Hsien Lee, Chih-Ping Wei, Tsang-Hsiang Cheng, Ching-Ting Yang, “Nearest-neighbor-based approach to time-series classification”, *Decision Support Systems* 53 (2012) 207–217.
- [29] Tomasz S Zabkowski, Wiesław Szczesny, “Insolvency modeling in the cellular telecommunication industry”, *Expert Systems with Applications* 39 (2012) 6879–6886.
- [30] Wouter Verbeke, David Martens, Christophe Mues, Bart Baesens, “Building comprehensible customer churn prediction models with advanced rule induction techniques”, *Expert Systems with Applications* 38 (2011) 2354–2364.