Volume 7, Issue 3, March 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Survey of Data Mining Techniques Used for Real Time Churn Prediction

Jean Claude Turiho[1, 2*], Wilson Cheruiyot[1], Anne Kibe[1], Irénée Mungwarakarama[1, 2]
[1]School of Computing and Information Technology/JKUAT, Kenya
[2]Faculty of Computing and Information Sciences/UNILAK, Rwanda

Abstract: ***Customer churn is perhaps the biggest challenge in telecommunication industry. Customer churn is the term which indicates the customer who is in the stage to leave the company or Customer churn means a customer can leave their service or service provider and move to another service or service provider. This rate is increasing day by day in all telecommunication companies. Data mining is one of the techniques which provide different methods and application to find out those customers who are going to churn and how to prevent them. The aim of this paper is to compare different algorithms of Data Mining which have been used for making distinction between customers into non churn and churn, so that appropriate steps can be taken into consideration in order to retain the churn customers to the company as customers are more valuable to the survival and development of the company.***

Keyword: Churn customers, churn prediction, data mining, telecommunication industry, and data.

I. INTRODUCTION

Today, all over the world developed and developing countries, telecommunication marketplace is fronting a severe loss of revenue due to fierce competition and loss of potential customers. In all the case, [1] states the following to keep the competitive advantages and acquire as many customers as possible; most operators invest a huge amount of revenue to expand their business in the very beginning. Therefore it has become vital for the operator to acquire the amount invested and gain at least a minimum profit within a very short period of time, because it is very much challenging and tedious issue to keep the customers intact for a long duration due to the competition involved in this business field. To survive in the market, a telecom operator usually offers a variety of retention policies to attract new customer according to the results supplied by data mining. This is the major cause of subscribers leaving one network and moving to another one which suits their needs. According to telecom market, the process of subscribers (either prepaid or postpaid) switching from one service provider is called 'Customer Churn'. Churn prediction is a method of differentiating churners and non-churners, so that appropriate steps can be taken to retain them. To control the churn customers in telecommunication Company, it is necessary to develop and apply day to day an effective model for churn prediction.

Data mining techniques have been used to discover necessary rules from warehouses or from other information resources. After term definition in Section II, where Churn customers, Churn prediction, data mining, telecommunication operator, data are defined; different data mining techniques used up to date are identified in Section III. Section IV reveals developed (near) real time churn prediction techniques. Section V concludes the paper.

II. TERMS DEFINITION

Churn customers

In [2] churn customers' leads to the loss of company as they are moving from one company to another, where they found some extra profit. It is not easy to identify the customers who are tending to move or leaving the company. Therefore, company prefers to use various Data mining techniques in order to hold the churn customers as customers are the valuable persons for them. Churn means the role of customers who are about to relocate their usage of service to a competitive service agency.

Churn prediction

Prediction of customers who are at risk of leaving a company is called as churn prediction in telecommunication. The company should focus on such customers and make every effort to retain them. This application is very important because it is less expensive to retain a customer than acquire a new [3]. Churn prediction methods gives the prediction regarding customers who planning to churn soon whereas churn operations helps on the other hand which aims to recognize such churners and to execute some beneficial actions to minimize the churn effect.

Data mining

In [4], the following definition is given: Data mining is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules.

Telecommunication industry

The combined telephone, computer, and cable TV industry segments are referred to as the broader industry of telecommunications. Together they provide equipment and wired and/or wireless connections needed for communicating via telephone and connecting to the Internet. Some retail and service companies that provide access to phone and Internet services and sell equipment and accessories, such as phones, cell phones, tablets, and computers, are also considered part of the telecommunications industry.

Data

Ackoff defines data as following: [5] data is raw. It simply exists and has no significance beyond its existence (in and of itself). It can exist in any form, usable or not. It does not have meaning of itself. In computer parlance, a spreadsheet generally starts out by holding data.

III. DATA MINING TECHNIQUES FOR CHURN DETECTION

Authors in [7] 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 Data Mining Evolutionary Learning is impractical on lager dataset with high dimension.

Adem Karahoca et al. [8] 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., [9] 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. [10] 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. [11] 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 [12] used logistic regression, G. Nie et al. [13] used logistic regression and decision tree model and A. Keramati, S.M.S. Ardabili [14] 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. [15] 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 [16] 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. [17] 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 [18] 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. [19] 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. [20] 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. [21] used decision tree, logistic regression and neural network to predict churn and suggested that demographic features have the lowest effect on the churn prediction when compared to the billing and usage details.

Anuj Sharma et al. [22] 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. [23] used neural network, logistic regression to obtain higher prediction accuracy in churn. Gang Cui [24] focused on BP neural network to predict the influence of customer retention for churn management.

Y.-H. Lee et al. [25] 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 [26] 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., [27] 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+.

According [28] 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.

Sadaf Nabavi, Shahram Jafari [29] proposed customer churn prediction model using a standard CRISP-DM (Cross Industry Standard Process for Data Mining) methodology based on RFM (Recency, Frequency, Monetary) and random forest and boosted trees techniques, the database of one of the biggest holdings of the country, Solico food industries group, is explored. Using this model, the customers tending to turn over are identified and effective marketing strategies will be planned for this group. Customer behavior analysis indicated that length of relationship, the relative frequency and the average inter purchase time were among the best predictors.

A Multi-Layer Perceptron Approach for Customer Churn Prediction, Mohammad Ridwan Ismail, et al. [30] proposed a Multilayer Perceptron (MLP) neural network approach to predict customer churn in one of the leading Malaysian's telecommunication companies. The results are compared against the most popular churn prediction techniques such as Multiple Regression Analysis and Logistic Regression Analysis. The result has proven the supremacy of neural network (91.28% of prediction accuracy) over the statistical models in prediction tasks.

IV. REAL TIME CHURN PREDICTION TECHNIQUES

Following the previous paragraphs line, [31] many classifiers have been adopted for churn prediction, including logistic regression, decision trees, boosting algorithms (e.g., variants of adaboost), boosted trees (gradient boosted decision trees) or random forest, neural networks, evolutionary computation (e.g., genetic algorithm and ant colony optimization), ensemble of support vector machines, and ensemble of hybrid methods. Most customer behavior features are extracted from BSS, including call detailed records (call number, start time, data usage, etc.), billing records (account balance, payment records, average revenue per user called ARPU, etc.), demographic information (gender, birthday, career, home address, etc.), life cycle (new entry or in-net duration), package/handset (type and upgrade records, close to contract expiration, etc.), social networks (call graphs), purchase history (subscribed services), complaint records, and customer levels (VIP or non-VIP). The churn management system becomes one of the key components in business intelligence (BI) systems. However, all previous works-based on customer behavior-focused on the use of the historical data stored in Data Warehouse. But recently researchers adopted the way to use the near or real time data for Customer churn prediction.

How can we measure and predict the quality of a user's experience on a telecommunication network in real-time? That is the problem that Ernesto Diaz-Aviles et al. address in [6]. Authors present the real-world solution to tackle the problem and carried out an empirical evaluation that shows that their approach achieves promising results that overall show the potential for measuring user experience at any point in time and without intrusively eliciting explicit feedback from customers.

V. CONCLUSION

Based on reference materials we can come to the conclusion that almost all researchers focused on data mining techniques -logistic regression, decision trees, boosting, boosted trees or random forest, neural networks, evolutionary, ensemble of support vector machines, and ensemble of hybrid methods- using historical data for predicting customer churn. Few among thousands of researchers integrated near or real time analysis to reveal the customer churn and non-churn at any point in time.

REFERENCES

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

[2] Ackoff, R. L. "Ackoff, R. L.From Data to Wisdom." Journal of Applies Systems Analysis, 1989: 3-9.

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

[4] 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* 9, no. 7 (November 2010).

[5] Aishiarya Churi, Mayuri Divekar, Sonal Dashpute, Prajakta Kamble, Reena Mahe. "Prediction of Customer Churn in Mobile Industry using Probabilistic Classifiers." *International Jjournal of Advance Foundation and Research in Science & Engineering (IJAFRSE)* 1, no. 10 (March 2015).

[6] Anuj Sharma, Dr.Prabin Kumar Panigrahi. "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services." *International Journal of Computer Applications* 27, no. 11 (August 2011).

[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*, 2012: 7736–7742.

[8] 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*, 2010: 3638–3646.

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

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

[11] Cui, Gang. "A Methodologic Application of Customer Retention Based on Back Propagation Neural Network Prediction." *Proceedings of Computer Engineering and Technology (ICCET).* IEEE, 2010. 418 - 422.

[12] Ernesto Diaz-Aviles, Fabio Pinelli, Karol Lynch, Zubair Nabi, Yiannis Gkoufas, Eric Bouillet, and Francesco Calabrese, Eoin Coughlan, Peter Holland, Jason Salzwedel. "Towards Real-time Customer Experience Prediction for Telecommunication Operators." *arXiv:1508.02884v2 [cs.CY]*, 2015.

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

[14] Javad Basiri, Fattaneh Taghiyareh, Behzad Moshiri. "A Hybrid Approach to Predict Churn." *Proceedings of Asia-Pacific Services Computing Conference*. IEEE, 2010. 485-491.

[15] Kulkarni, T. H. Hajare and R. V. "APPLICATION OF DATA MINING IN CUSTOMER CHURN ANALYSIS- A Review." (Review of Research) 4, no. 5 (February 2015).

[16] 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* 11, no. 3 (May 2000).

[17] Mohammad Ridwan Ismail, Mohd Khalid Awang, M Nordin A Rahman and Mokhairi Makhtar. "A Multi-Layer Perceptron Approach for Customer Churn Prediction." *International Journal of Multimedia and Ubiquitous Engineering* 10, no. 7 (2015).

[18] N.Kamalraj, A.Malathi. "A Survey on Churn Prediction Techniques in Telecommunication Sector." *International Journal of Computer Applications* 64, no. 5 (February 2013): 39-42.

[19] Nikita Jain, Vishal Srivastava. "Data Mining Techniques: A Survey Paper." *IJRET: International Journal of Research in Engineering and Technology* 2, no. 11 (November 2013).

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

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

[22] 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*, 2011: 7151–7157.

[23] Rahul J. Jadhav, Usharani T. Pawar. "Churn Prediction in Telecommunication Using Data Mining Technology." *(IJACSA) International Journal of Advanced Computer Science and Applications* 2, no. 2 (February 2011).

[24] Riddhima Rikhi Sharma, Rajan Sachdeva. "Review on Prediction of Churn Customer Behavior." *International Journal of Engineering Research & Technology (IJERT)* 6 , no. 01 (January 2017).

[25] Sadaf Nabavi, Shahram Jafari. "Providing a Customer Churn Prediction Model Using Random Forest and Boosted TreesTechniques." *Journal of Basic and Applied Scientific Research*, 2013: 1018-1026.

[26] Tomasz S Zabkowski, Wiesław Szczesny. "Insolvency modeling in the cellular telecommunication industry ." *Expert Systems with Applications*, 2012: 6879–6886.

[27] Wouter Verbeke, David Martens, Christophe Mues, Bart Baesens. "Building comprehensible customer churn prediction models with advanced rule induction techniques." *Expert Systems with Applications*, 2011: 2354–2364.

[28] 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.

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

[30] Yen-Hsien Lee, Chih-Ping Wei, Tsang-Hsiang Cheng, Ching-Ting Yang. "Nearest-neighbor-based approach to time-series classification." *Decision Support Systems* , 2012: 207–217.

[31] Yiqing Huang, Fangzhou Zhu, Mingxuan Yuan, Ke Deng, Yanhua Li, Bing Ni, Wenyuan Dai, Qiang Yang, Jia Zeng. "Telco Churn Prediction with Big Data." *SIGMOD* 15 (May-June 2015): 607-618.

[32] 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.

