2013 International Conference on Advances in ICT for Emerging Regions (ICTer): 172 - 176

Churn Prediction Methodologies in the Telecommunications Sector: A Survey

W.M.C. Bandara[#1], A.S. Perera[#2], D. Alahakoon[*3]
[#]Department of Computer Science and Engineering
University of Moratuwa, Sri Lanka
[*]School of Information and Business Analytics
Deakin University, Australia

[1]cbsoftware@gmail.com [2]shehan@uom.lk [3]d.alahakoon@deakin.edu.au

Abstract— *Predicting customer churn is a critical requirement of many if not all companies dependent on customer subscription services. The telecommunication sector is especially impacted due to the rival competition being very high and since tariff rates are maintained at a lower level. This paper describes the efforts made by researchers to build successful churn prediction models highlighting their special characteristics.*

Keywords— *Telecommunication, churn prediction, customer retention*

I. INTRODUCTION

A. What is customer churn

With the saturation of telecommunication services in the modern day world, undoubtedly the telecommunication service companies are facing rival competition day by day. Companies face this competition in two main ways, customer acquisition and customer retention. Literature reveals that, the cost of customer acquisition is up to 5 times more than the customer retention cost[1]. Therefore, organizations generally focus mainly on maintaining a long term relationship with the existing customers than acquiring new ones. In this paradigm, the word customer churn mainly refers to switching a customer from a service provider to another for some reason[2][3][4].

B. What is churn prediction

Churn prediction is the term used to determine the churning customers from a given service provider. What important is to determine whether a given customer is planning to churn and also identify reasons for it, not necessarily the target service provider.

C. Why predict churn

A company's main intention is to make profit and reducing loss making situations helps in increasing profit. Literature reveals that, average churn in telecommunication companies a month is 2.2% [4]. If a customer is going to be churned, a company sees it as a loss of money. If a company can predict whether a given customer is going to churn or not in future, that company has the opportunity to take action to try and provide the customer with a better service or address any unsatisfactory situations, if the planned churn is due to such a reason. Not only individual customers, having an idea of customer groups also beneficial to companies when making their long/short term policies, tariffs and special packages. For a telecommunication service provider company, customer churn is very common due to the switching cost being very low in this sector.

D. Different directions used in churn prediction

For more than a decade, people have researched on how to predict churn in the telecommunication sector. Decision trees (DT)[5], Artificial neural networks (ANN)[5], Genetic algorithms (GA)[6], Support vector machines (SVM)[5][7], Word of mouth[8], Bayesian Belief Networks (BBN)[3] and many other techniques have been used.

In the churn prediction paradigm, not only the churning customers but also the reason for their churn is very valuable due to the simple reason that it is difficult to address the problem without knowing the reasons behind the situation. Although techniques such as ANNs alone may not be sufficient to capture the reasons for churning they will give a high accuracy on results, a technique like BBN may identify the reasons for churn. A combination of these techniques can be used to obtain the benefits from each as well as avoid the limitations in using a single technique.

E. What will be discussed in the paper and the order of presenting

This paper is arranged as follows. Section II consists of the main subject matter of the paper, literature survey highlighting the main techniques and issues. Section III describes datasets and the evaluation methodologies being used. Section IV emphasizes on challenges and finally, Section V describes the future work, followed by the references.

II. LITERATURE SURVEY

In Aug-2002, Chih-Ping Wei and I-Tang Chiu[4] used Decision Trees to build their model for churn prediction. They have purposefully got rid of artificial neural networks (ANN) mainly due to the reasons given below, although ANNs are capable of providing decent predictions,

- For large amount of training data, ANNs take considerable time to train.
- ANNs are not easily interpretable.

978-1-4799-1276-6/13/$31.00 ©2013 IEEE

One important factor of their study was to more focus on the contractual level rather going in to the subscriber level. This means, rather than focus on the customer demographics they have used what are relevant to the contracts. One reason for this decision was that, the company which invested to their research did not have much of customer information at the period of study. On their study, they have divided the time into several periods and they experimented with how the results get changed with the sampling period of the call information. (E.g. how many days of call information should be aggregated?)

Hee-Su Kim, and Choong-Han Yoon[9] focused on the loyalty of the customers of a mobile service provider in terms of churning as well as recommending the service to others. The factors such as the "level of satisfaction with alternative-specific service attributes including call quality, tariff level, handsets, brand image, as well as income, and subscription duration" were affected for the customer loyalty of retaining and "the factors such as call quality, handset type, and brand image affect customer loyalty as measured by the intention/non-intention to recommend the service provider to other people".

Yu, W. Sobey et al., [10] emphasized the weaknesses in existing churn prediction methodologies in terms of determining the reasons behind the customer churn.

Shin-Yuan Hung, David C. Yen and Hsiu-Yu Wang[11] used Customer demography (age, tenure, gender), bill and payment analysis (monthly fee, billing amount, count of overdue payment), call detail records analysis (within network call duration, call type) and customer care/service analysis as features in order to build the models. In their paper they have mentioned specifically to avoid special festival seasons when taking the records as they have experienced some abnormalities in call pattern in special seasons (e.g. Chinese New Year). They have used the decision trees and neural networks in building their models.[11]

Jae-Hyeon Ahna, Sang-Pil Hana, Yung-Seop Lee [12] more focused on customers' status with regard to the company. In detail, they have defined customer statuses as "active use, non-use, suspended (by the service provider), churned" and considering the deflections as partial deflections or full deflections. Turning from active use to either non-use or suspended are being considered as a partial deflection and swapping the provider was considered as a full deflection.

Cao Kang and Shao Pei-ji[7] introduced the Support vector machine-recursive feature elimination attribute selection algorithm. In their words, "It could identify key attributes of customer churn, rule out the related and redundant attributes, and reduce the dimensions of data". In addition to that, it is mentioned two types of feature elimination algorithms, namely "Filters" and "Wrappers". The Filter type algorithms use the characteristics in the dataset to select the appropriate features without involving any learning algorithm. The good side of this method is that, it does not inherit any bias from the learning algorithm and also the computation is less expensive. The other side of this method is that, it does not guarantee an optimal feature subset. Also, it may contain a significant amount of noise features as well. On the other hand, the idea of the Wrapper type algorithms is to use the learning algorithm which uses to learn the model to evaluate the optimality of a given feature set to either keep or discard a feature.

Parag C. Pendharkar[6] proposed two Genetic algorithms based neural network models to predict churn in wireless networks. Their first GA based NN model uses a cross entropy[13] based criterion to predict customer churn, and their second GA based NN model attempts to directly maximize the prediction accuracy of customer churn. Their results show that, "both GA based NN models outperform the statistical z-score model on all performance criteria". Further, they have observed that "medium sized NNs perform best and the cross entropy based criterion may be more resistant to over fitting outliers in training dataset". They have used both GANN[6] and MLGANN[6] and also z-score to predict churning. Clearly the fastest training was obtained by z-score but its accuracy was around 60%. NNs took few hours to train and accuracy was around 98%. Also they have tested the models trained with low/medium/high hidden nodes and for the three inputs, and observed NNs with medium no of nodes were optimal. (6 in GANN and 9 in MLGANN)

Nicolas Glady, Bart Baesens, Christophe Croux[14] had a little different strategy for churn prediction. Their first contribution was to "redefine the notion of customer loyalty by considering it from a customer-centric viewpoint instead of a product-centric one." They then "used the customer lifetime value (CLV) defined as the discounted value of future marginal earnings, based on the customer's activity. Hence, a churner is defined as someone whose CLV, thus the related marginal profit, is decreasing".

Rong Liu, Yuanquan Li, Jiayin Qi[15] mainly focused on who to attend more after finding the churn probabilities of the customers rather than giving urgent attention to the customer with highest churn rate, on their research.

Jiayin Qi, Yuanquan Li [16] focused on the ways of selecting the effective input variables in a telecommunication customer churn model. They have proposed a procedure to select the input parameters in a step by step manner and applied their method to a real life dataset. Their variable selection methodology consisted as below.

First, they have used the AUC index of each variable in the initial variable set. The variables with the AUC value higher than 0.5 are selected for the next round of variable selection. In their initial data set of 216, this procedure reduced it to 155.

As the second step, they have measured the mutual information for each two variables in the 155 variables. The variables with mutual information less than 0.4 were selected as the final selected input variables for the churn prediction model which left 60 variables.

Ghorbani, A. ,Taghiyareh, F[17] proposed a framework to improve the management of customer churn.

B.Q. Huang et al., [2] have changed the time period which used for the prediction and accompanied with new topologies for the prediction model combining predictors with window techniques.

Bingquan Huang, B. Buckley, T.M. Kechadi[18] conducted a research for improved churn prediction by having multiple objectives in their goal for a land line telecommunication company. The following three objectives were to be maximized.

- The overall accuracy (OA) - proportion of the total number of predictions that were correct.
- The accuracy of the true churn (TC) - proportion of churn cases that were correctly identified.
- The accuracy of true nonchurn (TN) - the proportion of nonchurn cases that were classified correctly.

In order to do that, they have taken a method NSGA-II[19] and made some improvements to the algorithm. Their methodology describes in their paper abstract as "select local feature subsets of various sizes, and then to use the method of searching non-dominated solutions to select the global non-dominated feature subsets. Finally, the method FBSM[18] which yields the fitness thresholds is proposed to choose the global solutions with the lowest ranks as the final solutions."

Marcin Owczarczuk [20] tested the usefulness of various data mining techniques for churn prediction in the telecommunications area. Their main finding and recommendation was "linear models, especially logistic regression, are a very good choice when modelling churn of the prepaid clients. Decision trees are unstable in high percentiles of the lift curve, and we do not recommend their usage."

Huang Y, Kechadi M [21] were looking after new feature sets for churn prediction in their research. In their paper they have mentioned, "The main idea of this approach is to calculate the dependency between each input feature and the class. Finally, the comparative experiments were carried out, and the results show that the new proposed feature selection approach is very effective for the churn prediction."

Pınar Kisioglu, Y. IlkerTopcu[3] applied Baysian Belief Networks in order to see what features are directly/ indirectly relevant to the churn.

TorstenDierkes, Martin Bichler, Ramayya Krishnan [8] predicted the impact of one customer's behaviour to another customer's decision making on whether to churn and cross buying decisions.

Hai-fei Qin, Jian-hua Hu[22] worked on OLAP analysis to find out the customer churn.

Yongbin Zhang et al., [23] conducted a research based on only customer service usage information using a clustering algorithm. They have mentioned the advantage of not having "missing or non-reliable data and the correlation among inputs" in their paper abstract, using this way.

Bingquan Huang, Mohand Tahar Kechadi, Brian Buckley [24] proposed a set of features and predicted the churn using many predictors (Logict Regression, Naive Bayes, Linear classifiers, Decision Tree C4.5 (C4.5), Multilayer perceptrons artificial neural networks, Support Vector Machines and the Evolutionary Data Mining Algorithms.).

Wouter Verbeke et al., [25] had two parts in their research. In the first part, they have developed a novel, profit centric performance measure. They have calculated the maximum profit that can be generated by incorporating an optimal fraction of customers with the highest predicted probabilities to attire in a retention campaign. In the second part, they have conducted an extensive benchmark experiment to evaluate classification methodologies applied in eleven real-life data sets around the world. Based on profit centric and statistics, they claim that small number of variables is enough to predict customer churn with high accuracy. They also claim that the technique of oversampling (used in the churn model training stage) generally does not improve the results significantly.

Adnan Idris, Muhammad Rizwan, Asifullah Khan[26] investigated the significance of a Particle Swarm Optimization (PSO)[27] based under sampling method to handle the imbalance data distribution in collaboration with different feature reduction techniques such as Principle Component Analysis (PCA)[28], Fisher's ratio[29], Fscore[30] and Minimum Redundancy and Maximum Relevance (mRMR)[31].

Jianping Peng, Jing Quan, Shaoling[32] investigated the existing customer retention strategies of a company by examining the effect of such strategies on extending the customer's agreement with company.

Ying Huang, Tahar Kechadi[33] proposes a hybrid learning system for telecom churn predictions. They use weighted k-means clustering and a rule induction method to do the prediction. They claim that the system outperforms many other churn prediction models and provide superior results.

III. Datasets And Evaluation

Chih-Ping Wei and I-Tang Chiu[4] used a customer base of 114,000 with 4500 churners in their decision tree based model. They also had around 9,100,000 call records for their evaluation.

They have done analysis of their model for the following criteria.

- Effects of desired class ratios
- Effects of number of sub-periods
- Effects of length of retention period
- Analysis of temporal sensitivity
- Comparison to previous studies

Shin-Yuan Hung, David C. Yen and Hsiu-Yu Wang[11] have used a data set of 160,000 with 14,000 churners. They have consulted experts in the domain to identify possible attributes related to the churning of the customers and possibility of using a statistical measure like z-score is mentioned.

Their performance measured under,

- Overall performance trend
- Test modeling technique differences
- Sample size impact
- Robustness of the models
- Performance comparison with prior studies

Cao Kang and Shao Pei-ji[7] used the telecommunication churn dataset from Center for Customer Relationship Management at Duke University. After random sampling, finally they obtained two data sets, 1458 samples (713 churners, 745 non churners) for training and 1256 samples (656 churners, 600 non churners) for testing. They have compared the results of their churn prediction against PCA, Information gain and Chi-squat test and claims SVM-REF[34] gets to the top by having 0.6933 of hit rate and 0.6677 of accuracy.

Parag C. Pendharkar[6] obtained the customer information from the Teradata Center for Customer Relationship Management at Duke University and contained the information of 195,956 customers from a wireless company.

Rong Liu, Yuanquan Li, Jiayin Qi[15] used the local call details of SCDMA[35] customers. In their data set, they have captured total of 2,161,248 call events made by 10242 SCDMA customers, and 78.89% were local calls.

Jiayin Qi, Yuanquan Li [16] used a data set of 2000 customers (246 churned and 1754 non-churned). Since their research based on variable selection, they have compared the ROC[36] curves with both initial variable set as well as final variable set in their paper.

B.Q. Huang et al., [2] selected 47,391 picked customers from telecommunications company in Ireland in which 9999 churners and 18196 non churners in training data set and testing data set with 1,000 churners and 18,196 non churners. The experiment was carried out with existing variables as well as proposed variables with four window techniques and four modelling techniques (MLPs[37], Decision Tree C4.5 and standard SVMs and one-Class SVMs).

Bingquan Huang, B. Buckley, T.M. Kechadi[18] randomly picked 18,600 customers from Ireland base telecommunications company (Eircom, 2008) to have a training set of 15,000 customers and a testing set of 3,600 customers (600 churners and 3,000 non churners).

Marcin Owczarczuk[20] collected three samples of data, train, calibration and test. The train sample consisted of 85,274 observations and the calibration sample with 36,824 observations and test sample with 45,497 observations. The train sample and the calibration sample data were collected at the same time and then split randomly into the train and validation part. The test sample is then collected after six months than the train and calibration sample.

Pınar Kisioglu, Y. Ilker Topcu[3] used a data set from a Turkey based telecommunications provider. It consisted of 2000 total subscribers included with 534 churners. The initial number of variables of 23 has reduced to 9 after their data preparation. As per their results, it is identified that only Average minutes of usage, Tenure and Trend in billing amount are directly related to the customer churning.

Yongbin Zhang et al., [23]used a call record generator [38] to generate customer call records.

Bingquan Huang, Mohand Tahar Kechadi, Brian Buckley [24] used a training and testing datasets of 13,562 churners and 400,000 non churners. Each customer was represented by 738 features.

Adnan Idris, Muhammad Rizwan, Asifullah Khan [26]used a dataset from a French telecom company named Orange. It had 50,000 instances with 260 features with 190 numerical and 70 nominal features.

JianpingPeng, Jing Quan, Shaoling[32] used a data set of a Chinese telecommunication sector company with 414,733 customers within three different calling plans.

IV. CHALLENGES

Large datasets with large number of features makes it difficult to train and evaluate the prediction models and hence a feature selection mechanism should be accompanied with the mining process [21]. Researchers were focused on contractual data to avoid the problem of missing values [4]. The 2% usual churning for a month makes the dataset highly imbalanced and may make the prediction process difficult [39].

Having rival future package details in advance is valuable information but not feasible to obtain as in a competitive market, companies tend not to release information about such new products and services much in advance of the launch.

V. FUTURE WORKS

In the Sri Lankan context, duration of calls within and outside of the network can have a drastic impact on churn. Since usually the call tariffs are relatively high for outside networks, people tend to switch to the outside network if the customer has to make considerably high number of calls to that network. Such issues can be addressed in future research.

[1] N. Kamalraj and A. Malathi, "A Survey on Churn Prediction Techniques in Communication Sector," *International Journal of Computer Applications* , vol. 64, 2013.

[2] B.Q. Huang et al., "A new feature set with new window techniques for customer churn prediction in land-line telecommunications," in *Expert Systems with Applications.*, May 2010, vol. 37, pp. 3657-3665.

[3] Pınar Kisioglu and Y. Ilker Topcu, "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey," in *Expert Systems with Applications.*, June 2011, pp. 7151-7157.

[4] Chih-Ping Wei and I-Tang Chiu, "Turning telecommunications call details to churn prediction: a data mining approach," in *Expert Systems with Applications.*, August 2002, vol. 23, pp. 103-112.

[5] Koen W. De Bock and Dirk Van den Poel, "An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12293–12301, September 2011.

[6] Parag C. Pendharkar, "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services," in *Expert Systems with Applications.*, April 2009, vol. 36, pp. 6714-6720.

[7] Cao Kang and Shao Pei-ji, "Customer Churn Prediction Based on SVM-RFE," in *International Seminar on Business and Information Management*, 2008.

[8] Torsten Dierkes, Martin Bichler, and Ramayya Krishnan, "Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks," *Decision Support Systems*, vol. 51, no. 3, pp. 361–371, June 2011.

[9] Hee-Su Kim and Choong-Han Yoon, "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market," *Telecommunications Policy*, vol. 28, no. 9-10, pp. 751–765, October-November 2004.

[10] Wei Yu, Dawn N. Jutla, and Shyamala C. Sivakumar, "A Churn-Strategy Alignment Model for Managers in Mobile Telecom," in *Communication Networks and Services Research Conference, 2005. Proceedings of the 3rd Annual*, May 2005, pp. 48 - 53.

[11] Shin-Yuan Hung, David C. Yen, and Hsiu-Yu Wang, "Applying data mining to telecom churn management," in *Expert Systems with Applications.*, October 2006, vol. 31, pp. 515-524.

[12] Jae-Hyeon Ahn, Sang-Pil Han, and Yung-Seop Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," *Telecommunications Policy*, vol. 30, no. 10-11, pp. 552–568, November-December 2006.

[13] Reuven Y. Rubinstein and Dirk P. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning.*, 2004.

[14] Nicolas Glady, Bart Baesens, and Christophe Croux, "Modeling churn using customer lifetime value," *European Journal of Operational Research*, vol. 197, no. 1, pp. 402–411, August 2009.

[15] Rong Liu, Yuanquan Li, and Jiayin Qi, "Making Customer Intention Tactics with Network Value and Churn Rate," in *Wireless Communications, Networking and Mobile Computing, 2009. WiCom '09. 5th International Conference*, Beijing, pp. 1 - 4.

[16] Jiayin Qi and Yuanquan Li, "A novel and convenient variable selection method for choosing effective input variables for telecommunication customer churn prediction model," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference*, San Antonio, TX, 2009, pp. 3217 - 3222.

[17] A Ghorbani and F Taghiyareh, "CMF: A framework to improve the management of customer churn;" in *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific*, Singapore, 2009, pp. 457 - 462.

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

[19] Yusliza Yusoff, Mohd Salihin Ngadiman, and Azlan Mohd Zain, "Overview of NSGA-II for Optimizing Machining Process Parameters," *Procedia Engineering*, vol. 15, pp. 3978-3983, 2011.

[20] Marcin Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4710–4712, June 2010.

[21] Y. Huang, B. Q. Huang, and M. T. Kechadi, "A New Filter Feature Selection Approach for Customer Churn Prediction in Telecommunications,".

[22] Hai-fei Qin and Jian-hua Hu, "OLAP analysis of churn in the telecom's role," in *Computer Science and Service System (CSSS), 2011 International Conference*, Nanjing , pp. 2347 - 2350.

[23] Yongbin Zhang Ronghua, Yanying Zheng, and Michael Berry, "Behavior-Based Telecommunication Churn Prediction with Neural Network Approach," in *International Symposium on Computer Science and Society*, 2011.

[24] Bingquan Huang, Mohand Tahar Kechadi, and Brian Buckley, "Customer churn prediction in telecommunications," in *Expert Systems with Applications.*, January 2012, vol. 39, pp. 1414-1425.

[25] Wouter Verbeke, Karel Dejaeger, David Martens, Joon Hur, and Bart Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *European Journal of Operational Research*, vol. 218, no. 1, pp. 211–229, April 2012.

[26] Adnan Idris, Muhammad Rizwan, and Asifullah Khan, "Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies," *Computers & Electrical Engineering*, vol. 38, no. 6, pp. 1808–1819, November 2012.

[27] Particle Swarm Optimization. [Online]. http://www.swarmintelligence.org/

[28] Ordination Methods for Ecologists. [Online]. http://ordination.okstate.edu/PCA.htm

[29] VIAS - Visual Institute of Applied Science. [Online]. http://www.vias.org/tmdatanaleng/cc_fisher_ratio.html

[30] Eyewire. [Online]. http://blog.eyewire.org/f-scores/

[31] H Peng, F Long, and C Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226-1238, 2005.

[32] Jianping Peng, Jing Quan, and Shaoling Zhang. Mobile phone customer retention strategies and Chinese e-commerce. [Online]. http://www.sciencedirect.com/science/article/pii/S1567422313000409

[33] Ying Huang and Tahar Kechadi, "An effective hybrid learning system for telecommunication churn prediction," *Expert Systems with Applications*, vol. 40, no. 14, pp. 5635–5647, October 2013.

[34] Google. [Online]. http://www.google.com/patents/US8095483

[35] GSMArena. [Online]. http://www.gsmarena.com/glossary.php3?term=td-scdma

[36] AHA Journals. [Online]. http://circ.ahajournals.org/content/115/5/654.full

[37] Neuroph. [Online]. http://neuroph.sourceforge.net/tutorials/MultiLayerPerceptron.html

[38] Call Detail Record Generator. [Online]. http://www.gedis-studio.com/online-calldetail-records-cdr-generator

[39] Yaya Xiea, Xiu Lia, E.W.T. Ngaib, and Weiyun Yingc, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, April 2009.

[40] Teuvo Kohonen, "Self-Organized Formation of Topologically Correct Feature Maps," in *Neurocomputing: foundations of research*. MA, USA: MIT Press Cambridge, 1988, pp. 509-521.

[41] Chih-Fong Tsai and Yu-Hsin Lu, "Customer churn prediction by hybrid neural networks," *Expert Systems with Applications*, vol. 36, no. 10, pp. 12547-12553 , December 2009.

[42] Mu sigma. [Online]. http://www.mu-sigma.com/analytics/thought_leadership/cafe-cerebral-chaid.html