

A Survey on Customer Churn Prediction using Machine Learning Techniques

Saran Kumar A.

M.E. Scholar

Kumaraguru College of Technology
Coimbatore, India

Chandrakala D., PhD

Professor

Kumaraguru College of Technology
Coimbatore, India

ABSTRACT

The fast expansion 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 fast set up, service providers have realized the importance of retaining the on-hand customers. It is therefore essential for the service providers to prevent churn- a phenomenon which states that customer wishes to quit the service of the company. This paper reviews the most popular machine learning algorithms used by researchers for churn predicting, not only in banking sector but also other sectors which highly depends on customer participation.

Keywords

Customer retention, neural networks, Ensemble classifier, Boosting, Genetic Algorithm

1. INTRODUCTION

The volume of data has been growing at a fast rate over the last two decades due to advancements in technology. Many new methods and techniques have been introduced to process the data and gather valuable information which is hidden in the raw data. Data mining is defined as the process of extracting valuable information from the data. Many data mining methods have been successfully applied in various domains.

Customers are the most important assets in any industry since they are considered as the main profit source. Nowadays, companies have become observant that they should put much effort not only to convince the customers, but also to retain their existing customers. Churners are persons who move to other company for various reasons. To reduce customer churn, the company should be able to predict the behaviour of customer correctly and establish connections between customer attrition and keep factors under their control. Churn prediction is a binary classification task, which differentiates churners from non-churners.

Customer churn is defined as the movement of people from one bank to other bank. The main reasons for churn are dissatisfaction with the customer service, high costs, unattractive plans, bad support. It is an expensive problem in many sectors since acquiring new customer costs five to six times more than retaining existing ones [1]. The ability to predict that a specific customer is at a high risk of churning represents a huge additional potential revenue source for every business. In addition to the direct loss of revenue that results from a customer leaving the business, the costs of initially acquiring that customer may not have been covered by the customer's spending to date. The aim of customer churn prediction is to detect customers with high tendency to

leave a company. In order to retain the existing customers, the banking industry need to know the reasons of churn, which can be found through the knowledge extracted from gathered data.

Burez and Van den Poel [2] showed that there are two types of approaches for managing customer churn: reactive and proactive. When the company follows a reactive approach, it waits until the customer requests the company to cancel their service relationship. In this scenario, the company will offer the customer an incentive to stay. When the company adopts a proactive approach, it tries to find customers who are likely to churn before they do so. Then, the company provides special incentives for these customers to keep them from churning.

Machine learning is method of data analysis which automates analytical model building. Using algorithms that iteratively learn from data, machine learning allows systems to explore hidden patterns without being explicitly programmed where to look.

There are three types of machine learning approaches: unsupervised, semi-supervised, and supervised. Supervised learning is the machine learning task of finding the hidden patterns from labelled datasets. Unsupervised learning is the machine learning task of finding the hidden patterns from unlabeled data. Semi-supervised learning is a class of supervised learning tasks and techniques which also make use of unlabeled data for training – typically a small set of labelled data with a large set of unlabeled data. Semi-supervised learning falls between unsupervised learning and supervised learning [3].

2. LITERATURE SURVEY

M.A.H. Farquad [4] proposed a hybrid approach to overcome the drawbacks of general SVM model which generates a black box model (i.e., it does not reveal the knowledge gained during training in human understandable form). The hybrid approach contains three phases: In the first phase, SVM-RFE (SVM-recursive feature elimination) is employed to reduce the feature set. In the second phase, dataset with reduced features is then used to obtain SVM model and support vectors are extracted. In the final phase, rules are then generated using Naive Bayes Tree (NBTree which is combination of Decision tree with naive Bayesian Classifier). The dataset used here is bank credit card customer dataset (Business Intelligence Cup 2004) which is highly unbalanced with 93.24% loyal and 6.76% churned customers. The experimental showed that the model does not scalable to large datasets.

Chih-Fong Tsai [5] introduced the hybrid neural networks techniques to predict the customer churning in a CRM dataset provided by American telecom companies. Here, they built two hybrid models by combining two different neural network

techniques like back-propagation artificial neural networks (ANN) and self organizing maps (SOM) for churn prediction. The hybrid models are ANN combined with ANN (ANN + ANN) and SOM combined with ANN (SOM + ANN). In particular, the first method of the two hybrid models performs the data reduction task by filtering out unrepresentative training data. Then, the outputs of the previous step are used to create the prediction model based on the second technique. For performance evaluation of these models, three different kinds of testing sets like one general testing set and two fuzzy testing sets based on the filtered out data by the first technique of the two hybrid models, i.e. ANN and SOM respectively are used. The results indicate that the hybrid models outperform the single neural network baseline model in term of prediction accuracy and in particular, the ANN + ANN hybrid model performs better than the SOM + ANN hybrid model.

Wouter Verbeke [6] proposed the application of Ant-Miner+ and ALBA algorithms on a publicly available churn prediction dataset in order to build accurate as well as comprehensible classification rule-sets churn prediction models. Ant-Miner+ is a high performing data mining method based on the principles of Ant Colony Optimization which allows to include domain knowledge by imposing monotonicity constraints on the final rule-set. The advantages of Ant-Miner+ are high accuracy, comprehensibility of the generated models and the possibility to demand intuitive predictive models. Active Learning Based Approach (ALBA) for SVM rule extraction is a rule extraction algorithm, which combines the high predictive accuracy of a non-linear support vector machine model with the comprehensibility of the rule-set format.

The results which are benchmarked to C4.5, RIPPER, SVM and logistic regression showed that ALBA, combined with RIPPER, results in the highest accuracy, while sensitivity is the highest for C4.5 and RIPPER applied on an oversampled dataset. Ant-Miner+ results in less sensitive rule-sets, but allows to include domain knowledge, and results in comprehensible rule-sets that are much smaller than the rule-sets induced with C4.5. RIPPER also results in small and comprehensible rule-sets, but lead to unintuitive models that violate domain knowledge.

Ning Lu [7] proposed the use of boosting algorithms to enhance a customer churn prediction model in which customers are separated into two clusters based on the weight assigned by the boosting algorithm. As a result, a high risky customer cluster has been found. Logistic regression is used as a basis learner, and a churn prediction model is built on each cluster, respectively. The experimental results showed that boosting algorithm provides a good separation of churn data when compared with a single logistic regression model.

Benlan He [8] suggested a customer churn prediction methodology based on SVM model, and used random sampling method to improve SVM model by considering the imbalance characteristics of customer data sets. A support vector machine constructs a hyper-plane in a high- or infinite-dimensional space, which can be used for classification. Random sampling method can be used to change the distribution of data in order to reduce the imbalance of the dataset. Imbalance in dataset is caused due to the low proportion of churners.

Ssu-Han Chen [9] used a novel mechanism based on the gamma Cumulative SUM (CUSUM) chart in which the gamma CUSUM chart monitors individual customer's Inter

Arrival Time (IAT) by introducing a finite mixture model to design the reference value and decision interval of the chart and used a hierarchical Bayesian model to capture the heterogeneity of customers. Recency, another time interval variable which is complementary to IAT, is combined into the model and tracks the recent status of the login behavior. In addition, benefits from the basic nature of control charts, the graphical interface for each customer is an additional advantage of the proposed method.

The results showed that the accuracy rate (ACC) for gamma CUSUM chart is 5.2% higher than exponential CUSUM and the Average Time to Signal (ATS) is about two days longer than required for exponential CUSUM.

Koen W. De Bock [10] proposed two rotation-based ensemble classifiers namely Rotation Forest and Rotboost as modeling techniques for customer churn prediction. An ensemble classifier is a combination of several member classifier models into one aggregated model, including the fusion rule to combine member classifiers outputs. In Rotation Forests, feature extraction is applied to feature subsets in order to turn the input data for training base classifiers, while RotBoost combines Rotation Forest with AdaBoost. Four data sets from real-life customer churn prediction projects are used here. The results showed that Rotation Forests outperform RotBoost in terms of area under the curve (AUC) and top-decile lift, while RotBoost demonstrates higher accuracy than Rotation Forests. They also compared three alternative feature extraction algorithms namely: Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Sparse Random Projections (SRP) on classification performance of both RotBoost and Rotation Forest.

In general, the performance of rotation-based ensemble classifier depends upon: (i) the performance criteria used to measure classification performance and (ii) the implemented feature extraction algorithm.

Lee et al. [11] focused on building an accurate and succinct predictive model with the purpose of churn prediction by using a Partial Least Squares (PLS) based method on highly correlated data sets among variables. They not only present a prediction model to accurately predict customers churning behaviour, but also a simple but implementable churn marketing program was employed. The proposed methodology allows the marketing managers to maintain an optimal (atleast a near optimal) level of churners effectively and efficiently through the marketing programs. Here, PLS is employed as the prediction modelling method.

Koen W. De Bock [12] suggested GAMensPlus, an ensemble classifier based upon generalized additive models (GAMs), in which both performance and interpretability are reconciled and evaluated for churn prediction modeling. The proposed GAMens, which is based on Bagging, the Random Subspace Method and semi-parametric GAMs as constituent classifiers, is extended to include two instruments for model interpretability namely, the generalized feature importance scores, and bootstrap confidence bands for smoothing splines. The experimental comparison of classification performance over data sets from six real-life churn prediction projects showed that GAMensPlus provides a strong classification performance that is performing at least as good as the two individual classifiers (logistic regression, and GAM).

Ning et al. [13] performed an experimental investigation of customer churn prediction in telecom industry and proposed

the use of boosting to improve the customer churn prediction model. Unlike other boosting methods that improve the accuracy of a given basis learner, the author suggested to separate customers into two clusters based on the weight assigned by the boosting algorithm. The proposed model provides an opportunity to an “Implementation Zone” where customers with the highest churn propensity can be addressed for retention actions.

Ver-braken et al. [14] proposed a new performance measure called the expected maximum profit criterion, which is aligned with the main objectives of the end users. The proposed framework not only assists the companies with selecting the classifier that maximizes the profit, but also provides information about the fraction of the customer base to be included in the retention campaign.

P.C.Pendharkar [15] suggested two Genetic Algorithm(GA) based neural network (NN) models to predict the customer churn. The first GA-based NN model used a cross entropy based criteria to predict customer churn, and the second GA based NN model made some efforts to directly increase the prediction accuracy of customer churn. Using real-world customer dataset and three various sizes of NNs, they compared the two GA-based NN models with a statistical z-score model using model evaluation criterion like prediction accuracy, top 10% docile lift and area under Receiver Operating Characteristics (ROC) curve. The results of experiments indicated that both GA-based NN models outperform the statistical z-score model on all performance criteria.

Y.Xie et al., [16] used an improved balance random forest (IBFR) model which is a combination of balanced random forests and weighted random forests in order to overcome the data distribution problem. The nature of IBRF is that the best features are iteratively learned by altering the class distribution and by putting higher penalties on misclassification of the minority class. The experiments are carried out with Chinese bank dataset which showed that IBRF is better than artificial neural network, decision tree and support vector machines in terms of accuracy.

3. CONCLUSION

This paper provides a detailed study on the methods used for the process of customer churn prediction. Each of the above churn prediction models has low accuracy and prediction. Hence a good prediction model is required in order to avoid the customer churn problem. This can be achieved by combining SVM with boosting algorithms for higher accuracy and performance which can be considered as a future work for Churn prediction. Good prediction models have to be constantly developed and a combination of the proposed methods has to be used.

4. REFERENCES

- [1] T.Vafeiadis, K.I. Diamantaras, G.Sarigiannidis, K.Chatzisavvas “Customer churn prediction in telecommunications”, *Simulation Modelling: Practice and Theory* 55 (2015) 1-9.
- [2] Burez J., & Van den Poel, D “Crm at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services”, *Expert Systems with Applications* 32, 277–288.
- [3] S.Parvathavardhini and Dr. S.Manju “Analysis on Machine Learning Techniques” *International Journal of Computer Sciences and Engineering (IJCSE)*, Vol-4(8), pp 59-77 Aug 2016, E-ISSN: 2347-2693.
- [4] M.A.H. Farquad, Vadlamani Ravi, S. Bapi Raju “Churn prediction using comprehensible support vector machine: An analytical CRM application”, *Applied Soft Computing* 19 (2014) 31–40.
- [5] Chih-Fong Tsai, Yu-Hsin Lu “Customer churn prediction by hybrid neural networks”, *Expert Systems with Applications* 36 (2009) 12547–12553.
- [6] 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.
- [7] Ning Lu, Hua Lin, Jie Lu, Guangquan Zhang “A Customer Churn Prediction Model in Telecom Industry Using Boosting”, *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, may 2014.
- [8] Benlan He, Yong Shi, Qian Wan, Xi Zhao “Prediction of customer attrition of commercial banks based on SVM model”, *Proceedings of 2nd International Conference on Information Technology and Quantitative Management (ITQM)*, Procedia Computer Science 31 (2014) 423 – 430.
- [9] Ssu-Han Chen, “The gamma CUSUM chart method for online customer churn prediction”, *Electronic Commerce Research and Applications*, 17 (2016) 99–111.
- [10] Koen W. De Bock, Dirk Van den Poel, “An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction”, *Expert Systems with Applications* 38 (2011) 12293–12301.
- [11] H. Lee, Y. Lee, H. Cho, K. Im, Y.S. Kim, “Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model”, *Decision Support System* 52 (2011) 207–216.
- [12] Koen W. De Bock, Dirk Van den Poel, “Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models”, *Expert Systems with Applications* 39 (2012) 6816–6826.
- [13] L. Ning, L. Hua, L. Jie, Z. Guangquan, “A customer churn prediction model in telecom industry using boosting”, *IEEE Trans. Ind. Inform.* 10 (2014) 1659–1665.
- [14] T. Verbraken, W. Verbeke, B. Baesens, “A novel profit maximizing metric for measuring classification performance of customer churn prediction models”, *IEEE Transaction on Knowledge and Data Engineering* 25 (2013) 961–973.
- [15] P.C. Pendharkar, “Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services”, *Expert System Application* 36 (2009) 6714–6720.
- [16] 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.

5. AUTHOR PROFILE

Mr. A. Saran kumar received his B.Tech. Degree in Information Technology from Coimbatore Institute of Technology, Coimbatore, India. He is currently pursuing M.E. Degree in Computer Science and Engineering in Kumaraguru College of Technology, Coimbatore, India. His areas of interest are Data mining, Big Data and Web Technology.

Dr. D. Chandrakala is a Professor in the Department of Computer Science and Engineering, Kumaraguru College of

Technology, Coimbatore, India. She received her Ph.D. degree in the area of Data Mining from Anna University in 2013. She has published several papers in National / International Journals and Conferences. Her current research interest includes Soft Computing, Artificial Intelligence, Evolutionary Algorithms, Database management systems, Data Mining and Image Processing. She is a life member of Indian Society for Technical Education.