A Dynamic Classification Approach to Churn Prediction in Banking Industry

Hoiyin Christina Leung
University of Central Florida, cleung456@gmail.com

Wingyan Chung
Western Carolina University and University of Hong Kong, wchung9@gmail.com

A Dynamic Classification Approach to Churn Prediction in Banking Industry

Emergent Research Forum (ERF)

Hoiyin Christina Leung
University of Central Florida
cleung456@gmail.com

Wingyan Chung
College of Business, Western Carolina University
Dept. of Computer Science, University of Hong Kong
wingyan@acm.org

Abstract

Churn prediction is the process of using transaction data to identify customers who are likely to cease their relationship with a company. To date, most work in churn prediction focuses on sampling strategies and supervised modeling over a short period of time. Few have explored the area of mining customer behavior pattern in longitudinal data. This research developed a dynamic approach to optimizing model specifications by using time-series predictors, multiple time periods, and rare event detection to enable accurate churn prediction. The study used a unique three-year dataset consisting of 32,000 transaction records of a retail bank in Florida, USA. It uses trend modeling to capture the change of customer behavior over time. Results show that data from multiple time periods helped to improve model precision and recall. This dynamic churn prediction approach can be generalized to other fields for which mining long term customer data is necessary.

Keywords

Banking, churn prediction, decision support, customer retention, feature engineering, business analytics.

Introduction

Churn prediction is the process of using transaction data to identify customers who are likely to cancel their subscriptions. Popular among many service industries like telecommunication, finance and e-commerce, churn prediction can help organizations to prevent loss of future revenue and to increase customer loyalty (NGData 2013, Fiworks 2019). To date, most work in churn prediction focuses on sampling strategies, feature engineering and supervised modeling over a fixed period of time. Predictors used in existing work, such as user demographics and credit status, are static in nature and are not adequate to provide accurate prediction (Ali and Ariturk 2014). Few studies have explored the area of using dynamic predictors to mine customer behavior over longitudinal data.

This research aims to develop and validate a dynamic classification approach to accurately capture customer behavior for churn prediction. The approach considers dynamic time-series predictors, multiple time periods, and rare event detection and modeling to enable accurate and long-term dynamic prediction. The approach was applied to churn prediction on a unique three-year, fine-grained dataset consisting of 32,000 transaction records provided by a retail bank located in Florida, USA. The banking industry needs churn prediction because each lost customer costs on average $750 while adding new customers is expensive (acquiring a new customer costs $200 on average) in today's saturated market (Fiworks 2019). Using the approach, the study finds that its trend modeling helped to capture the change of customer behavior over multiple periods of time. The approach's rare event detection and its use of data from dynamic time periods to aggregate the number of churn cases helped to improve model precision and recall. The empirical results demonstrate a strong potential to improve profitability of businesses that need accurate and scalable churn prediction. The research provides a useful approach to optimizing churn prediction modeling and a unique case for the banking industry and for business analytics education (Chung 2015).

Literature Review

Churn prediction allows banks to detect potential churners early (NGData 2013), enable them to take preemptive actions to retain customers. Previous work used rare event detection, feature engineering, and data mining models to facilitate churn prediction. Their strengths and weaknesses are reviewed below.

Rare Event Prediction

One challenge in churn prediction is imbalanced sample, in which negative instances out-number positive instances significantly, causing poor learning performance. In banking industry, churn is a rare event, its annual rate is less than 10%. Various sampling techniques like under-sampling, over-sampling, both under- and over-sampling, and SMOTE sampling have been studied in the context of churn prediction (Kumar and Ravi 2008). SMOTE over-samples minority instances by creating synthetic churn examples based on the characteristics of the real churn cases. The result showed SMOTE performs best in terms of accuracy, precision and recall. Another way of tackling the rare event problem is to provide the model with more examples. Multiple training periods contain an observation for every time period within the analysis time frame (Ali and Ariturk 2014) when the accounts are active. Although Ali and Ariturk (2014)'s work showed that multiple training periods outperformed standard one observation per customer framework, their data was limited to 1 year. Customer churn cases in different time periods were sampled repeatedly to provide multiple training periods' data. However, this sampling failed to provide realistic dynamic modeling of churn behavior especially in today's fast-changing banking industry.

Churn-Related Features

Another challenge in churn prediction is how to capture the pattern of customer behavior. Dynamic predictors like trend factors are used to unearth the underlying trend of customer behavior. Trend factor is commonly used in financial institution to determine the volatility of an asset with respect to underlying market. It was used in churn prediction in the telecom industry (Vyas et al. 2018). Weights were assigned to each month, with increasing weight in the later months, mimicking the recency effect. In E-retailer research, behavioral features like clickstream/web logs data were also highly related to churn prediction (Subramanya and Somani 2017). Besides capturing the pattern of customer behavior, other features had been added to improve model performance. Social network features could improve prediction accuracy significantly in telecom industry (Ahmad et al. 2019) and in online social behavior on Reddit (Chung et al. 2019). Macroeconomic indicators like consumer confidence index, future oil prices could also improve model performance in predicting churn in trust accounts in financial industry (Dash and Das 2017). Churn-related features are industry-specific. Traditional bank data is mainly composed of customer demographics and transaction history. Past research on churn prediction in this area mainly focused on modeling approaches, few discussed which features are the most relevant. There is a need to identify the most important features in churn prediction to enable timely and accurate alert to managers.

Modeling Approaches

Regarding modeling approaches, the performance of seven statistical modeling approaches were compared in churn prediction (Umayaparvathi and Iyakutti 2016): logistic regression, k-nearest neighbor, random forest, support vector machine, ridge classifier, decision tree and gradient boosting. Gradient boosting and random forest outperformed other approaches in terms of accuracy, precision and recall. Similar results were shown in analyzing participants in a churn prediction analytics tournament, logistic regression and tree methods performed better than other methods like discriminant analysis and clustering (Neslin et al. 2006). Modified models like using lasso regression to extract features that were highly correlated with churn, then fed those features into radial basis function neural network (RBF). This method resulted in better performance than using RBF, Log-R or boosting alone (Xiong et al. 2019). Another modified model was improved balanced random forest (IBRF) which combined cost-sensitive learning with random forest to alter class distribution and penalized more heavily on misclassification of the minority class. This technique showed improvement over other methods like SVM, random forest and decision tree (Ying et al. 2008). However, these two models are computationally intensive, were applied to small datasets without considering the temporal dimension, and may not be suitable for large, dynamic transaction data used by banks. Furthermore, these processes are complicated with low interpretability. The banking industry is highly regulated and transparency is essential. There is a need to provide a clear and efficient way to process large volume of longitudinal data.

In summary, the literature shows that previous research used sampling techniques, new predictors and modeling approaches in churn prediction. Data used in these studies are static and cover only a short period of time. Research on modeling the dynamic nature of churn data (e.g., data sampling strategies, duration of data used, prediction window, number of multiple time periods) is rare. These specifications, like sampling techniques and modeling approaches are highly related to model performance. Therefore, this study tries to fill this research gap by developing a novel approach and by applying it to a 3-year dataset in a retail bank to optimize the above model specifications.

A Dynamic Classification Approach to Churn Prediction

The objective of this study is to develop and validate a dynamic classification approach to optimizing model specifications for predicting customer churn decisions over time. The approach consists of using trend factors to capture dynamic customer behavior, combining under-sampling and SMOTE to balance the sample, and using multiple training time periods to optimize dynamic modeling and prediction.

Experiments and Preliminary Results

The dataset includes over 32,000 customer records from August 2016 to July 2019 from a retail bank in Florida, USA. It is a well-established bank offering savings and loan products to consumers. Selected features in the dataset are listed in Table 1.

Static Predictors	Account Activities
Age, No. of Savings Products, Tenure, No. of Loan Products, Indirect Loan Indicators Credit/Debit Card Indicators, Dates of Open/Close Account	No. and Sum of Direct Deposits, No. of Bill Payments, No. and Sum of Loan Payment, No. of In-branch Visits, No. of Debit Card Transactions, No. of Online Transactions

Table 1: Selected features indicating customer characteristics and account activities

The number of active accounts throughout the year remains stable but churn cases are few in a narrow prediction window. For example, the proportion of churn cases is less than 2% in a 2-month prediction window. This study combines under-sampling and the SMOTE technique to balance the sample before modeling. Five nearest neighbors in the minority class are used to generate new cases, setting the proportion of minority class to 50%. Even with SMOTE sampling, there are still not sufficient positive examples for accurate churn prediction. This study makes use of data from multiple time periods to aggregate the number of churn cases in order to provide a more diversified profile of churners. This study tries to provide a generalized churn prediction framework by optimizing the following specifications:

- Training data time span (4 months vs 6 months)
- Prediction window (2 months vs 3 months)
- Number of multiple time periods (MTPs: 1 to 7)

A series of trials were carried out to search for the optimal specifications. Blocks of data from the specific timeframe were extracted and analyzed. The three-year-long dataset was segmented into multiple training datasets to facilitate dynamic learning of customer churn behavior.

Feature Engineering

In order to incorporate variations in the economic and competitive environment, account activities are analyzed in a time-specific manner. Since the training data covers 3 years, data in specific blocks of time (e.g., 4 or 6 months) are extracted at a time. Account activities during that period are aggregated. Trend factor is calculated to capture the trend of account activities during this period (Vyas et al. 2018). It magnifies the influence of recent activities by putting a heavier weight on recent months. Its value ranges from 0 to 2. Activity can be the frequency of monthly transactions or the amount of monthly transactions. For example, if an account has a monthly payment of $100 for 4 months, a weight from 1 to 4 is assigned to each month with the later the month, the heavier the weight. Plug in the numbers into the formula below, TF1 equals to 1. Trend factors of all account activities are calculated and served as inputs in supervised models.

$$Trend\ Factor\ (TF) = \frac{\sum(Weight\ \times\ Acticity) \times \frac{No.of\ Months}{\sum Weight}}{\sum Activity} = \frac{1000\ \times \frac{4}{10}}{400} = 1$$

TF < 1 → decrease
TF = 1 → no change
TF > 1 → increase

Predictive Modeling

This study uses data in 2019 as test set, whereas all previous data is used as train set. In order to mimic the real-life situation, a filtering process is carried out to remove all inactive accounts during that period (e.g. accounts not yet opened or closed before that time). Confusion matrix, accuracy, recall and precision are used to measure model performance. The following formulae show how the performance measures are calculated. TP is true positive, TN is true negative, FN is false negative and FP is false positive. In an imbalanced dataset like this, accuracy does not truly reflect model performance because the majority class is over 97%. Recall and precision are essential to measure how well the model in predicting the target, i.e. churn in this case.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \qquad Precision = \frac{TP}{TP+FP} \qquad Recall = \frac{TP}{TP+FN}$$

Training Time Span

In the banking industry, 4 or 6 months of data are commonly used in churn prediction (Dash and Das 2017). The following test was conducted to see whether 4 months or 6 months of data make better prediction. Data from two 6-month and two 4-month periods are extracted in 2018 as train set. The validation period is a sliding window - three months immediately after each block of training data. The prediction period is from May to July in 2019. The prediction data is extracted 4 or 6 months prior to the prediction period. The data in each time period is compiled and goes through all three supervised learning models: logistic regression, random forest and gradient boosting (GBM). The result in Table 3 indicates 6 months of data outperforms 4 months of data in terms of prediction accuracy and recall. Random forest and GBM give higher accuracy and recall than logistic regression. Since churn is a long-term decision, six month training data is able to capture customer behavioral pattern better, thus producing better result.

<table>
<tr><th>Training Data</th><th>No. Of Months</th><th>Validation Period</th><th>Testing Data</th><th>Prediction Period</th><th>Average Accuracy</th><th>Average Recall</th><th>Average Precision</th></tr>
<tr><td>2018_1to6</td><td>6</td><td>2018_7to9</td><td>2018_11to2019_4</td><td>2019_5to7</td><td rowspan="2">94.78%</td><td rowspan="2">31.14%</td><td rowspan="2">24.8%</td></tr>
<tr><td>2018_7to12</td><td>6</td><td>2019_1to3</td><td>2018_11to2019_4</td><td>2019_5to7</td></tr>
<tr><td>2018_1to4</td><td>4</td><td>2018_5to7</td><td>2019_1to4</td><td>2019_5to7</td><td rowspan="2">90.81%</td><td rowspan="2">13.89%</td><td rowspan="2">30.3%</td></tr>
<tr><td>2018_7to10</td><td>4</td><td>2018_11to2019_1</td><td>2019_1to4</td><td>2019_5to7</td></tr>
</table>

Table 2: Comparison of 6 months vs 4 months of data

Prediction Window

A number of factors like cost, accuracy, number of positive cases affect the decision of the length of prediction window. This study compares 2-month and 3-month prediction windows, using the same 2 blocks of 2018 6-month data described in the previous section. The result shows 2-month prediction window has higher accuracy (96.1% vs 95%) and recall (37.19% vs 31.14%) than 3-month window. The result indicates accuracy decays fast as prediction window extends.

Number of Multiple Time Periods (MTP)

As the number of churn cases increases over time, the number of active accounts also increases but the overall churn proportion remains at 2%. Combining data can improve accuracy, however, it increases computational cost inevitably. This part tries to find out the optimal number of multiple time periods in the train set. Multiple blocks of 6-month data are extracted from 2016 to 2018. Data from 1 to 7 MTPs are aggregated and fed into supervised models. It is expected that prediction accuracy will increase with MTPs. The pattern of this increase and whether marginal gain in accuracy will decrease are ongoing analyses of this study. Results should inform decisions on optimal size of training data.

Feature Importance

Our preliminary results indicate that "number of savings products" and "total sum of fees" are the most influential churn predictors. Customers with more savings products tend to stay, those who are charged with high fees tend to leave. This information can help banks to build an alert system to monitor these two features, alarming them which customers are likely to leave soon.

Conclusion

Churn prediction is a powerful tool for businesses to maintain long-term relationship with their customers. However, limited work has been done in capturing customer behavior on longitudinal data. This research developed a dynamic classification approach to optimizing model specifications for predicting customer churn decisions over time. Using a 3-year transaction dataset of a U.S. bank, the approach was used to identify duration of training data, length of prediction window and the number of multiple time periods to support accurate prediction. The result shows that a longer period of training data (6 months) captures customer behavioral pattern better than a shorter period (4 months). A shorter prediction window (2 months) causes less decay in accuracy than a longer window (3 months). Overall, the research contributes a new approach to dynamic churn prediction and a new case to business analytics education (Chung 2015). The research was limited by a potential lack of independence due to multiple observations of the same customers. Since the training data is drawn from multiple periods, the data of the same customers is used multiple times. Although customers may change their behavior in different time periods, static predictors remain the same, thus endangering the independence assumption. For future directions, it will be interesting to use multiple time periods to test whether the marginal gain in accuracy will decline after a certain quantity of data is added. Another possibility is to compare the use of autoregressive model and trend factor in capturing transaction pattern. The dynamic research framework can be generalized to other areas for which mining long term customer data is necessary.

Acknowledgments

The authors thank the data provider for their generous support and the conference editors and reviewers for their valuable comments.

REFERENCES

Ahmad, A. K., Jafar, A., and Aljoumaa, K. 2019. "Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform," *Journal of Big Data* (6:28).

Ali, O. G., and Ariturk, U. 2014. "Dynamic Churn Prediction Framework with More Effective Use of Rare Event Data: The Case of Private Banking," *Expert Systems with Application* (41), pp. 7889-7903.

Chung, W. "Business Analytics in Context: Evaluating New Curricular Modules in an Undergrad. Stat. Course," *Proc. NSF Workshop on Curricular Dev. for Computing in Context*, ACM Press, 2015.

Chung, W., Toraman, C., Huang, Y., Vora, M. and Liu, J. "A Deep Learning Approach to Modeling Temporal Social Networks on Reddit," *Proc. IEEE Int'l Conf. Intelligence & Security Informatics*, Jul. 2019.

Dash, M., and Das, K. 2017. "Customer Attrition Analytics in Banking," *International Journal of Business Analytics and Intelligence* (5:2), pp. 7-14.

Fiworks. 2019. "Statistics: Industry Stats Fueling the Need for Targeted Sales and Marketing Execution." from https://www.fiworks.com/resources/statistics.

Kumar, D., and Ravi, V. 2008. "Predicting Credit Card Customer Churn in Banks Using Data Mining," *International Journal of Data Analysis Techniques and Strategies* (1:1), pp. 4-28.

Neslin, S., Gupta, S., Kamakura, W., Lu, J., and Mason, C. 2006. "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research* (XLIII May 2006), pp. 204-211.

NGData. 2013. "Predicting & Preventing Banking Customer Churn by Unlocking Big Data." from https://www.ngdata.com/wp-content/uploads/BankingChurnWP-1mbg.pdf

Subramanya, K. B., and Somani, A. 2017. "Enhanced Feature Mining and Classifier Model to Predict Customer Churn for E-Retailer," in: *7th Int'l Conf. Cloud Comp., Data S. & E.* Noida, India: IEEE.

Umayaparvathi, V., and Iyakutti, K. 2016. "Attribute Selection and Customer Churn Prediction in Telecom Industry," in: I*nt'l Conf. on Data Mining & Adv. Comp. (SAPIENCE)*. Ernakulam, India: IEEE.

Vyas, R., Prasad, B. G. M., Nanda, S. K., and Vamshidhar, H. K. 2018. "Predicting Inactiveness in Telecom (Prepaid) Sector: A Complex Big Data Application," *Int'l Conf. on Info. Tech*: IEEE, pp. 39-43.

Xiong, A., You, Y., and Long, L. 2019. "L-Rbf : A Customer Churn Prediction Model Based on Lasso+Rbf," in: *2019 Int'l Conf. Internet of Things (iThings / GreenCom / CPSCom / SmartData)*. Atlanta, USA: IEEE.

Ying, W., Li, X., Xie, Y., and Johnson, E. 2008. "Preventing Customer Churn by Using Random Forests Modeling," in: *2008 IEEE International Conference on Information Reuse and Integration*. Las Vegas, NV, USA: IEEE.