Proposal for ISyE6416 Project

Profit-based classification in customer churn prediction: a case study in banking industry

Ashkan Zakaryazad

Industrial & Systems Engineering
Georgia Institute of Technology
Atlanta, USA
ashkan.zakaryazad@gatech.edu

Taewoon Kong

Industrial & Systems Engineering
Georgia Institute of Technology
Atlanta, USA
twkong@gatech.edu

1 Problem Statement

Churn prediction is one of the common applications of the classification in the business settings. The word "churn" means to stop consuming products of a specific company and use fungible product of another company because of its better quality or service or less price. There are lots of studies such as A. D. Athanassopoulos. (2000); C. B. Bhattacharya (1998); M. Colgate. et al. (1996) which show that acquiring a new customer for a company is five or six times more expensive than retaining an existing one. Accordingly, nowadays most of the financial institutions are concerned with customer retention studies to prevent losing their market share and maximize their gained profit from existing customers. The primary objective of customer retention is to maximize the potential profit which can come from existing customers. In most of the churn prediction studies, the objective of classification is to minimize the prediction error and accordingly maximize the accuracy of the prediction. This approach is definitely an optimal approach when the objective is to correctly classify the customers as much as possible, however, it may reach suboptimal solution when the objective is to maximize the profit of churn prediction for the company. In our case, the bank has information about customers' lifetime value for the next period (one year) which can be used as a profit metric to show the importance of each of the customers. In this study, we have two objective:

1. Developing a profit-based classification algorithm which classifies churners and non-churners such that it maximizes the total potential profit of the bank by giving more weight to detection of profitable churner customers.

2. Finding appropriate individual incentive offer value for each of the churner customers instead of giving fixed offers to all of them to ensure that more profitable customers are getting more valuable offers than other churners and accordingly minimize their corresponding churn (leaving) probability.

2 Data Source and Description

In this study, we gathered the data set from a well-known Turkish bank. There are totally 20000 samples (customer), each of which has 24 attributes (features) where one of them is response variable (dependent) and 23 are predictors (independent) variables. There is a restriction about introducing all of the features; however, there are three types of independent variables: biographical variables, account information and profit-based variable. Biographical variables include customer's age, gender, region, and so on. Account information includes number of accounts, number of transaction, transactions' amount, and so on. The profit-based variable here is the customer's lifetime value (CLV) which has been calculated for the next one year. The response is represented with a binary variable where the value of one indicates that the corresponding customer is churner and the value of zero represents non-churner customer. Since the response variable is binary, churn prediction problem can be studied using binary classification methods.

3 Methodology

3.1 Data Pre-processing

Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis. Data pre-processing includes cleaning, normalization, transformation, dimension reduction, etc. For our study we will concentrate on some of them and the product of data pre-processing will be our final training set.

3.2 Classification by Ensemble

One of the objectives of this study is to maximize the total profit of classification instead of minimizing the total prediction error. For the purpose we will use profit-based objective function (loss function) and the Adaboost algorithm, a commonly used ensemble method, in the model building step. The contribution of this study is to use the classifiers whose objective function is modified to be the total net profit of classification instead of the total prediction error.

Ensemble methods use multiple learning algorithms to obtain better predictive performance than one could be obtained from any of the constituent learning algorithms.Opitz, D. et al. (1999); Polikar, R. (2006); Rokach, L. (2010). Unlike a statistical ensemble in statistical mechanics, which is usually infinite, a machine learning ensemble refers only to a concrete finite set of alternative models, but typically allows for much more flexible structure to exist among those alternatives.

Among many methods, Boosting is one of the most commonly used ensemble methods L, Breiman. (1996); Z. Zhi-Hua. (2012). While boosting is not algorithmically constrained, most of the Boosting algorithms consist of iteratively learning weak classifiers with respect to a distribution and adding them to a final strong classifier. When they are added, they are typically weighted in some way that is usually related to the weak learners' accuracy. After a weak learner is added, the data is re-weighted: examples that are misclassified gain weight and examples that are classified correctly lose weight. Thus, future weak learners focus more on the examples that previous weak learners misclassified, which makes the boosting method involve incrementally building an ensemble by training each new model instance to emphasize the training instances that previous models mis-classified. In this study, although there are many boosting algorithms, we will use AdaBoost which is adaptive so that take full advantage of the weak learners. For the weak learners, we will consider a variety of classification algorithms such as Logistic Regression, Artificial Neural Network (ANN), Support Vector Machines (SVM), Decision Tree, Naive Bayes, Discriminant Analysis (LDA or QDA), and k-NN and so on.

3.3 Profit Maximization by Incentive Offer Selection

In the post-processing step of the study we will maximize the total profit of customer retention with solving an optimization problem to find individual incentive offer for each of the customers and make sure that the offer selection system works fairly considering customers' profit for the company. We will compare two policies about retention promotions (offers). First of them is to giving some fixed offers to all of the target customers and second, the variable incentive offer for each of the customers. The objective is to maximizing the profit of the company and minimizing the churn probability of customers. Different offers may effect differently on customers churn probability which will be studied in this project.

4 Expected Results

In our study we expect to reach a profit-based model which maximizes the total profit of classification compared to the error-based statistical classification models even if the error of the proposed methods are not better than the statistical methods. Since the primary objective of the bank administrators is to maximize the total profit, the proposed methods are more acceptable to them in terms of practice. Also, we are able to expect a better result with ensemble method than only with each individual weak learner according to its origin. In the promotion offer selection step, we will maximize the total profit of the bank by assigning appropriate promotion offers to target customers considering its effect on their decision. This optimization problem may reach an optimum policy in assigning offers to the target customers.

5 Responsibility

1. Ashkan Zakaryazad: Problem Statement, Data Source and Description, and Methodology (Profit Maximization by Incentive Offer Selection), and Expected results).

2. Taewoon Kong: Methodology (Data Pre-processing, Classification by Ensemble), and Expected Results.

References

A. D. Athanassopoulos. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *J. Bus. Res.* **47(3)**, 191-207.

Dumas, M., Van der Aalst, W.M.P., and Ter Hofstede, A.H. (2005). When Customers Are Members: Customer Retention in Paid Membership Contexts. *J. Acad. Mark. Sci.*, **26(1)**, 31-44.

M. Colgate, K. Stewart, and R. Kinsella. (1996). Customer defection: a study of the student market in Ireland. *Int. J. Bank Mark.* **14(3)**, 23-29.

Y. Freund and R. E. Schapire. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences.* **55(1)**, 119-139.

Opitz, D. and Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research.* **11**, 169-198.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine.* **6(3)**, 21-45.

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review.* **33(1-2)**, 1-39.

L, Breiman. (1996). Bias, Variance, and arcing classifiers. *Technical Report.*

Z. Zhi-Hua. (2012). Ensemble Methods: Foundations and Algorithms. *Chapman and Hall.* 23.