Improving Churn Prediction with Voice of the Customer

Yiou Wang Koji Satake Takeshi Onishi Hiroshi Masuichi
Fuji Xerox Co., Ltd. Japan
{Yiou.Wang, Koji.Satake, Takeshi.Onishi,Hiroshi.Masuichi}@fujixerox.co.jp

1 Introduction

Customers, who switch to competitors or move out from service providers, become critical concerns for companies to retain customer loyalty. Churn prevention through churn prediction, which is a task to detect costumers who are about to quit, is one of the methods to ensure customer loyalty with service providers. To detect and analyze early churn is a proactive step to ensure that existing customers do not move out or switch to the product of competitors[1].

On the other hand, customers voice opinions and advices about some brands, companies, products or services. Today voice of the customer(VOC), capturing a view of customer's behaviors, needs, and feedbacks, can be obtained through center calls, emails, questionnaire, web reviews or SNS[2]. Therefore, it is attractive to consider exploiting VOC analysis in churn prediction task. However few research has been conducted in this direction. One exception is an investigative study showing that integrating the information of call center emails resulted in an increase in predictive performance[3]. They used a weighted term-by-email matrix to represent a collection of emails and used Latent Semantic Indexing (LSI) via Singular Value Decomposition (SVD) to reduce the matrix to k dimension in order to overcome disadvantages of large and sparse matrix. However, it is not possible to know what value of k will lead to an optimal solution in different business situations. They made great effort to determine the critical k for their task. Their techniques required specialized pre-processing and dimension reduction steps.

In this paper, we present a simple and easy-to-implement approach to improve the performance of churn prediction by incorporating VOC analysis into a conventional churn prediction model. We first identify the sentimental polarity of VOC, classify opinion types of VOC, calculate VOC prediction scores and then generate new features by these three kinds of information and train new models. We evaluate the usefulness of our approach in a series of experiments and demonstrate that VOC analysis provides substantial performance gains in churn prediction task.

2 The Baseline Churn Prediction Model

The churn models that exploit traditional predictors, such as demographic information, contractual details, usage facts or other service-related information, have been extensively studied[1]. The same as other conventional churn models, we introduced the following baseline features:

- Demographic information: age and gender.
- Usage facts: monthly, quarterly, half-yearly and yearly usage facts; customer internal comparison of quarterly, half-yearly and yearly usage facts
- Other service-related information: customer satisfaction ratings.
- Loyalty card details: types of members, accumulation points.

In other words, our baseline features are based on the structured information. We train our baseline prediction model on random forest method, which is introduced by Breiman[4]. It is demonstrated to be one of the most effective classification algorithms[5] and is shown to perform very well compared to many other classifiers in handling imbalanced data classifications and churn prediction tasks [6][7].

3 Proposed Method

In this section, we describe our approach of effectively integrating useful information from VOC (unstructured information) into the above baseline models through features. As we know VOC is different from profile information (demographic information),

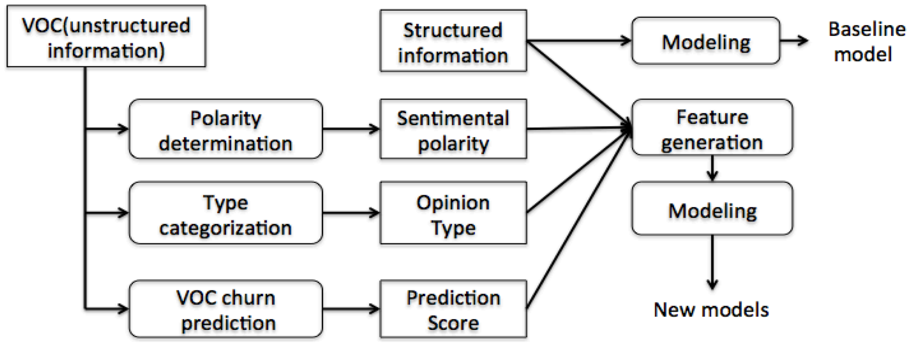

Figure 1: Proposed method

which is structured information. The core problem is to extract useful structured information from the unstructured VOC. We first preprocess VOC by the following three steps: (i) determine the sentimental polarity of VOC; (ii) categorize opinion types of VOC; (iii) make the model of VOC churn prediction. We then get the sentimental polarity of VOC, opinion types of VOC and VOC prediction scores from the above preprocessing. Finally, we incorporate new features by these three kinds of information into the above random forest algorithm based baseline churn prediction method and train new models. Figure 1 shows an overview of our approach. The rest of this section describes our features in detail.

3.1 Features of sentimental polarity

In this subsection we explore the utilization of sentiment analysis in churn prediction. The idea is motivated by the observation that the increase of customer satisfaction will reduce churn. A customer is likely to quit if he is unsatisfied with a service or a product. The sentimental polarity of VOC, which shows if VOC is a praise or a complaint of a service or a product, may indicate the customer satisfaction in some degree.

Trained annotators judged whether a VOC is a positive or negative evaluation of a service or a product and tagged each VOC with sentimental polarity tag. There are three kinds of tags: a positive tag, a negative tag and a neutral tag. In this paper, we used the manual tags to generate sentimental polarity features. The process can also be automatically realized. We built a classification model with the annotated data, made a 10-fold cross validation test and achieved 88% in F result.

The final feature setting is summarized as $\{N+, N-, N0, R+, R-, R0\}$. Here, $N+/N-/N0$ is the numbers of VOC with positive/negative/neutral sentimental polarity tags. $R+/R-/R0$ shows if latest VOCs contain positive/negative/neutral ones. Note that one customer may have multiple VOC data.

3.2 Features of opinion type

In this subsection we introduce opinion type features in churn prediction. The idea is motivated by the observation that the importance of opinion types is different in different types of industries. For example, although a customer in a restaurant complained about the connection of wifi, if he was very satisfied with the food, to view the situation as a whole, he may be satisfied with the restaurant.

There are various kinds of VOC such as praises or complaints about staffs, foods, network services and so on. The opinion types of VOC are divided into 24 categories according to the targets of opinions. Due to the confidentiality of the data, we can not provide details of them. The same as sentimental polarity annotation, the trained annotators classified the VOC data into 24 opinion types. In this paper, we used the manual annotated opinion types to generate new features. This process can also be realized automatically. The classification model built on the annotated corpus provided 86% in accuracy. The final feature setting is the VOC number of each opinion type (24 types), the VOC number of each opinion type in positive tag, and the VOC number of each opinion types in negative tag.

3.3 Features of prediction score

In this subsection, we introduce VOC churn prediction models and encode the prediction score of each VOC as features in customer churn prediction.

Table 1: Feature templates for VOC churn prediction

type	Feature	Description
unigram	$w_1, b_1, ...w_i, b_i, ...w_n, b_n,$	For the words in VOC, word surface form and word base form are added as unigram features
bigram	$w_1\&w_2$, $b_1\&b_2$,... $w_i\&w_{i+1}$, $b_i\&b_{i+1}$,...$w_{n-1}\&w_n$, $b_{n-1}\&b_n$	For the words in VOC, word surface form bigram and word base form bigram features are added

The VOC from a churn customer is supposed to be a churn VOC and the VOC from a loyal customer is supposed to be a loyal VOC. In this way, we can obtain a corpus for VOC churn prediction. We predicted the churn score of VOC using polynomial kernel support vector machine (SVM). We used the features shown in Table 1 for VOC churn prediction. Here w_i, b_i and n denote the word surface form, word base form (a form of word stem) of the i-th word, and the numbers of words in VOC, respectively.

We then divided the corpus into ten equal-sized sets, as in the data preparation for 10-fold cross-validation. For each set, we used the remaining nine sets to train a VOC churn prediction model and used this model to generate VOC churn prediction scores for VOC from this set. In this way, for each VOC, a churn prediction score was provided with the VOC prediction model in this 10-fold cross validation technique. One customer may have multiple VOC data. We added the following new features to encode the prediction score information for a customer:

- the maximum prediction score of all VOC data
- the minimum prediction score of all VOC data
- the mean of prediction scores of all VOC data
- the median of prediction scores of all VOC data
- the mean of prediction scores of the latest VOC data

4 Experiments

4.1 Data Set

In this paper, we chose the customers who answered the web questionnaire as our prediction targets. VOC are collected from web questionnaire. We use the comparison of usage information of FY2014 and FY2013 to judge whether a customer is churn or loyal (See Figure 2). A customer whose utilization frequencies dramatically drops is prone to churn. In particular, we define $t = 2/3$. Table 2 provides the statistics of the customers and VOC numbers.

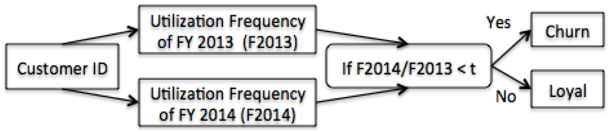

Figure 2: Customer churn definition

Table 2: Statistics of customers and VOC data

	number of customers	number of VOC
churn	1,770	4,878
loyal	3,519	10,644
all	5,289	15,522

4.2 Experimental Results

We evaluated the effectiveness of new features in a series of experiments. We used recall (R), precision (P) and F as evaluation metrics. Table 3 shows the final results for all experiments of our random forest models for customer churn prediction with the threshold of $probability = 0.5$. Precision-recall curves of all the experiments are shown in Figure 3.

Table 3: The results of churn prediction model

Methods	Precision	Recall	F
baseline	0.670	0.447	0.536
+ (a) sentimental polarity	0.675	0.451	0.541
+ (b) opinion type	0.677	0.449	0.540
+ (c) prediction score	**0.686**	**0.490**	**0.571**
+ (a) +(b) +(c)	0.686	0.479	0.565

The results of Table 3 show that sentimental polarity features and opinion type features achieved slight improvement in both recall and precision. Prediction score features were very effective. The combination of all features can not provide further improvement. This may be because the former two kinds of features are not effective. The details of feature effect will be discussed in Section 5. Finally we only added the prediction score features to baseline features as our final feature set. The final results show that adding VOC analysis into a conventional churn prediction model results in a significant increase in predictive performance.

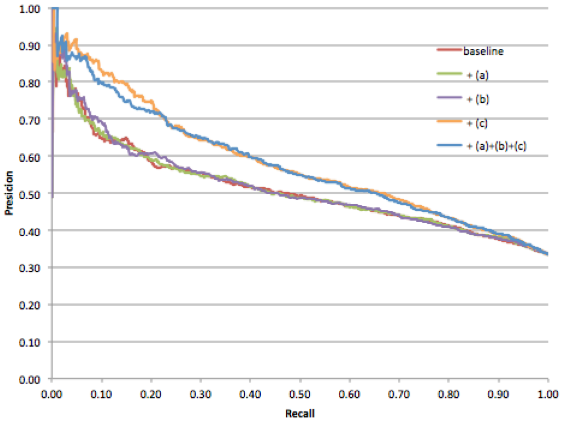

Figure 3: Precision-recall curves

5 Discussion

In this section, we discuss the effect of new features.

1. Why sentimental polarity features and opinion type features are not very effective?

We analyzed the VOC data with positive and negative opinions. Positive opinions account for 41% and negative opinions account for 59% of VOC data from loyal customer, while positive opinions account for 55% and negative opinions account for 45% of VOC data from churn customer. Such statistics contradict our expectation. This may because Japanese customers are prone to make evaluations in a polite way, even they are not satisfied with the services or products.

We can not encode the real satisfaction information by only using the limited classification information of sentimental polarity and opinion type. Degrees and strengths of praises and complaints are very important for churn prediction. For example, "*the food is really amazing*" indicates a very strong positive opinion. "*the food is good*" indicates a weak positive attitude. The customer expressed VOC like the former example is likely to be a loyal customer.

2.Why prediction score features are very effective?

Prediction scores provide more information than binary classification and are in some degree able to represent the strength of sentiment. By comparing differences of prediction results of our baseline model and the model with prediction score features, the effect of VOC can be summarized in the following two directions:

(i) The churn predictions of the customers, who voiced the opinions with strong attitude or in an extreme way, achieved an improvement in performance. Such as the following VOC examples:

(e1) The whole experience there was extremely good and we really appreciate it.

(e2) The staff made a fatal mistake and did not apologize to us.

(ii) For the customers who have multiple VOC data, the sequential evaluation of services or products was encoded by the VOC features and provided an improvement in prediction performance, such as the following VOC examples from the same loyal customer.

(e3) The food was too cold and tough. (Jan 2013)

(e4) The dessert was very good and we ordered many for takeout . (Aug 2013)

(e5) The food was hot and very delicious. (Dec 2013)

6 Conclusion

In this paper, we presented a simple and easy-to-implement approach to improve the performance of churn prediction by incorporating VOC analysis into a conventional churn prediction model. We explored the utility of VOC sentimental analysis, VOC opinion type classification and VOC churn prediction and demonstrated that sentimental polarity and opinion type features achieved slight improvement and VOC prediction score features provided substantial improvement over the baseline method through a series of experiments.

Reference

[1] Palupi D. Kusuma, Dejan Radosavljevik and Frank W. Takes, "Combining Customer Attribute and Social Network Mining for Prepaid Mobile Churn Prediction, " in Proceedings of BENELEARN, 2013.

[2] L. Venkata Subramaniam, Tanveer A. Faruquie, Shajith Ikbal, Shantanu Godbole and Mukesh K. Mohania, "Business Intelligence from Voice of Customer, " iIEEE International Conference on Data Engineering, 2009.

[3] Kristof Coussement and Dirk Van den Poel, "Integrating the voice of customers through call center emails into a decision support system for churn prediction, " Information & Management, Volume 45, Issue 3, Pages 164-174, 2008.

[4] L.Breiman, "Random forests," Machine Learning, volume 45, pages 5-32, 2001

[5] Manuel Fernandez-Delgado, Eva Cernadas, Senen Barro,and Dinani Amorim, "Do we Need Hundreds of Classifiers to Solve Real World Classification Problems," Journal of Machine Learning Research 15, pages 3133-3181, 2014

[6] Yaya Xie, Xiu Li, E. W. T. Ngai, and Weiyun Ying, "Customer churn prediction using improved balanced random forests," Expert Systems with Applications 36, pages 5445-5449, 2009

[7] Gideon Dror, Dan Pelleg, Oleg Rokhlenko, and Idan Szpektor, "Churn prediction in new users of Yahoo! answers," Proceedings of the 21st international conference companion on World Wide Web, 2012