Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 31 (2014) 423 – 430

Procedia Computer Science

2nd International Conference on Information Technology and Quantitative Management, ITQM 2014

Prediction of customer attrition of commercial banks based on SVM model

Benlan He[a,b]*,Yong Shi[c],Qian Wan[d],Xi Zhao[c]

[a] *Post-Doctoral Research Station of University of Chinese Academy of Science,Zhongguancun East Road 80, Haidian District,Beijing,100190,China*

[b] *Post-Doctoral Research Center of Industrial and Commercial Bank of China, No.55 Fuxingmen Street,Xicheng District,Beijing,100140,China*

[c] *Research Center on Fictitious Economy & Data Science,University of Chinese Academy of Science, Zhongguancun East Road 80, Haidian District,Beijing,100190,China*

[d] *Industrial and Commercial Bank of China, No.55 Fuxingmen Street,Xicheng District,Beijing,100140,China*

Abstract

Currently, Chinese commercial banks are facing triple tremendous pressure, including financial disintermediation, interest rate marketization and Internet finance. Meanwhile, increasing financial consumption demand of customers further intensifies the competition among commercial banks. To increase their profits for continuing operations and enhance the core competitiveness, commercial banks must avoid the loss of customers while acquiring new customers. This paper discusses commercial bank customer churn prediction based on SVM model, and uses random sampling method to improve SVM model, considering the imbalance characteristics of customer data sets. The results show that this method can effectively enhance the prediction accuracy of the selected model.

Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.

Keywords: SVM Customer Attrition Prediction Random Sampling

* Corresponding author. Benlan He. Tel.: 86-15210804429; .
E-mail address: hebenlan99@163.com.

Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.
doi:10.1016/j.procs.2014.05.286

1. Introduction

At present, domestic commercial banks are undergoing huge and complex changes, and face many challenges. Information technologies represented by cloud computing, mobile Internet are surging, financial regulation specialized in capital regulation is increasingly strengthened, and the process of financial disintermediation and interest rate marketization is gradually accelerating, resulting in the sharply narrowed interest margin for banks. Besides, the e-Business enterprises represented by the third-party payment, which use both the Internet technology and big data technology, are advancing to the field of traditional business of commercial banks, which subvert the advantage of commercial banks channels. Meanwhile, financial consumption demand of consumers is increasing gradually. Customers pay more attention to the experience, personalized service, diversity and agility, which further intensifies the competition among commercial banks.

One of the main competitions among commercial banks is for customers, especially for high-grade customers. As customers are directly related to profits, commercial banks must avoid the loss of customers while acquiring new customers. Harvard Business Review believes that by reducing the customer defection rate by 5%, companies can increase profits by 25% to 85%, while Business Week thought the profits will increase by 140%. Bhattacharya (1998) showed that the cost of developing a new customer is 5 times to 6 times than retaining an old customer. As can be seen, reducing customer attrition has a significant impact not only on increasing profits for commercial banks, but also on enhancing their core competitiveness . Therefore, it is urgent for commercial banks to improve the capabilities to predict customer attrition, thereby taking timely measures to retain customers and preventing other clients from churning.

2. Overview of the Domain Problem

In order to predict customer attrition for commercial banks, many scholars carried out research by using data mining methods. Some scholars used classification methods to predict customer churn. Chandar et al (2006) used CART, TreeNet and C5.0 classification method for predicting the loss of customers of commercial banks, the result showed that CART algorithm was the best one to predict the possible loss of customers. Popovic and Basic (2009) proposed Fuzzy C-Means clustering algorithm for retail banks churn prediction model. Prasad and Madhavi (2012) applied both CART and C5.0 classification techniques to a commercial bank in India. It is found that the prediction accuracy of CART algorithm is better than that of C5.0 classification.

Some scholars used the random forest method. For example, Burez and Poel (2009) used three ways, including sampling techniques, gradient Boosting and weighted random forests for prediction, and the result showed that the weighted random forest method is the best one for predicting. Xie et al (2009) proposed improved balanced random forests integrating sampling techniques and cost-sensitive learning techniques and compared with ANN, decision trees and CWC-SVM methods were compared, the results displayed IBRF work most effectively.

Some other scholars employed an algorithm based on sequence patterns, such as Chiang et al (2003) . And Mutanen (2006) established Logistic regression model and believed that this method was better for predicting the loss of customers, while Zhao and Dang (2008) suggested that SVM model was better.

Although the previous studies have used numerous data mining methods to predict customer attrition, there is not yet a unified conclusion on the application effect of these models, the accuracy of these models is not ideal, and the capability of processing large data sets are to be improved. Therefore, considering the characteristics of banking customer attrition, this paper uses SVM model combined with random sampling method to improve the performance of customer churn prediction.

3. Methodology

SVM is a machine learning method raised by Vapnik in the early 1990’s, which arises from optimal linearly separable SVM classification surface. Optimal classification surface requires the separating line can not only separate two dimensions correctly (the training error rate is 0), but can also maximize the margin between the two classes. SVM aims to find a hyperplane which can meet classification requirements and make the trained points far from the classification surface, namely looking for a classification face to make the margin on its both sides maximum.

3.1. SVM Model

Given a linearly separable sample set:

$$T=\{(x_1,y_1),\cdots,(x_l,y_l)\}, i=1,2,\cdots,l,$$
$$x_i \in R^n, \quad y_i \in \{+1,-1\}$$

(1)

where x_i is called an input and y_i is called the corresponding output, namely class label.

The general form of the equation of the separating line is given as

$$f(x)=(\mathbf{w}\cdot\mathbf{x})+b \tag{2}$$

where $(\mathbf{w}\cdot\mathbf{x})$ represents the inner product of the vector $\mathbf{w}$ and the vector $\mathbf{x}$ .The optimal classification face corresponding to the linear discriminator function is given as:

$$(\mathbf{w}\cdot\mathbf{x})+b=0 \tag{3}$$

If the linear discriminator function is normalized so that all samples meet $|f(x)|\geq 1$,then the margin between the classification face $(\mathbf{w}\cdot\mathbf{x})+b=1$ and $(\mathbf{w}\cdot\mathbf{x})+b=-1$ is $2/\|\mathbf{w}\|$ (namely the classification interval).

Minimizing the distance $2/\|\mathbf{w}\|$, it is equivalent to maximizing $\frac{1}{2}\|\mathbf{w}\|^2$, and then we can get the optimal classification face. Thus, the problem of seeking the optimal classification face is transformed into the following optimization problem:

$$\begin{cases} \min\limits_{w,b} \frac{1}{2}\|\mathbf{w}\|^2 \\ s.t. y_i\left((\mathbf{w}\cdot\mathbf{x})+b\right)\geq 1, i=1,\cdots,l \end{cases}$$

(4)

Solving the above equation and we get the optimal classification face. It can completely divide the sample set into two categories. But the actual situation is not so. We cannot set the sample apart entirely. Therefore, the slack variable ξ_i ($i=1,\cdots,l$) is introduced in the formula, allowing to weaken the constraint to a certain degree. The degree of constraint is represented by penalty parameter C . Thus, the constrained optimization problem of solving optimal classification face is transformed into:

$$\begin{cases} \min\limits_{w,b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \\ s.t. y_i\left((\mathbf{w}\cdot\mathbf{x})+b\right)+\xi_i\geq 1, i=1,\cdots,l \\ s.t. \xi_i\geq 0, i=1,\cdots,l \end{cases}$$

(5)

The larger C is, the greater penalty it will give to the wrongly classified sample. Using Lagrangian method for solving the optimum classification face is converted into the dual problem:

$$\begin{cases} \min\limits_{w,b} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^{l} \alpha_i \\ s.t. \sum_{i=1}^{l} y_i \alpha_i = 0, \\ 0 \le \alpha_i \le C, i = 1, \cdots, l \end{cases}$$

(6)

Solving the formulation and we obtain the optimal classification function under the linear condition:

$$f(\mathbf{x}) = \mathrm{sgn}(\sum_{i=1}^{l} \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}) + b^*)$$

(7)

where $b^* = y_i - \sum_{i=1}^{l} y_i \alpha_i^* (\mathbf{x}_i \cdot \mathbf{x}_j)$。

3.2. Nonlinear SVM Model

In the case of nonlinear classification, introduce the nonlinear mapping Φ to project low-dimensional sample into a higher dimensional feature space and use kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ to transform nonlinear classification into linear classification in this space. Then the optimal classification face function is:

$$f(\mathbf{x}) = \mathrm{sgn}(\sum_{i=1}^{l} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^*) \tag{8}$$

Commonly kernel functions are as following types:

（1）Linear kernel function： $K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x} \cdot \mathbf{x}_i)$

（2）Polynomial kernel function： $K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x} \cdot \mathbf{x}_i)^d$， $K(\mathbf{x}, \mathbf{x}_i) = ((\mathbf{x} \cdot \mathbf{x}_i)^d + 1)$

（3）RBF kernel function：： $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2 / \sigma^2)$

（4）Sigmoid kernel function： $K(\mathbf{x}, \mathbf{x}_i) = \tanh[k(\mathbf{x} \cdot \mathbf{x}_i) + v]$， where $k > 0$， $v < 0$ 。

3.3. Random sampling method

Due to the low proportion of churners of commercial banks, the distribution of the data set is much imbalanced, random sampling method can be used to change the distribution of data in order to reduce the imbalance of the dataset. Through the way of sampling, for one thing, the training sample size is reduced and the model training speed is improved; for another thing, the capability of identifying churners of the training model is much better due to the reduced gap between the number of customers and that of non-customers. There are two types of random sampling methods: under-sampling and over-sampling. Under-sampling does not change the number of minority-class samples but eliminates the majority-class samples to improve the prediction accuracy of minority-class, while over-sampling method does not change the number of majority-class of samples but increase the number of the minority-class samples by duplicating the minority-class samples in its simplest form. But this approach will lead to over-fitting (Drummond, Holte (2003)), and the predictive precision of the minority-class of under-sampling method is superior to that of over-sampling method (Ling and Li (1998), Chen et al (2004)).

Through increasing the proportion of minority-class, under-sampling makes the machine learning method no longer lay particular stress on predicting the majority-class sampling and ignoring the minority-class. Therefore, in this study we will use the random sampling method.

3.4. Evaluation criteria

In the case of the imbalanced data set, the classifiers are often more inclined to predict the majority class correctly, while the prediction accuracy of the minority class is much poor. At this time the general classification evolution rules like prediction accuracy can no longer effectively measure the predictive power of the model. Thus, many scholars introduced several measure evaluation criteria such as F-measure, sensitivity, ROC curve, etc. Among them, F-measure is calculated from the confused matrix based on the classification and it is harmonic mean of recall and precision, which is defined as follows:

$$F = \frac{2}{\frac{1}{\mathrm{Re}\,call} + \frac{1}{\mathrm{Pr}\,ecision}} \tag{9}$$

We will take F-measure to evaluate predictive ability of the model.

4. Empirical study

4.1. Data preprocess

In this study, we will choose customer dataset by a Chinese commercial bank from its data warehouse. Specifically, the dataset includes 50000 customer records, from January 2011 to June 2012. We take the data from January 2011 to December 2011 as the training data, and January 2012 to June 2012 as the test data. Through preprocessing the dataset, such as filling in missing values and removing outliers, finally there are 46,406 valid data records. And if the customer cancels his account during the observation period we define him as a churner. Finally we get 421 churners accounting for 0.91% and 45,985non-churners accounting for 99.09%. The ratio of churners to non-churners is 109.23. We can see that the customer dataset is serious imbalanced.

We consider the basic attribute indicators and business indicators of customers as input variables. Among them, the customer basic attribute indicators include age, sex, education, income, occupation, service stars, the asset-to-liabilities ratio, etc. The business indicators include deposit accounts, deposit balance, the number of deposits, and the amount of consumption and so on.

4.2. Results

Our algorithm code is written in MATLAB2010.In order to compare the model results; we establish SVM models in the case of non-sampling and sampling respectively, and select Logistic regression model for comparison. These three models are compared to test the sampling effectiveness. We select two types of SVM models: linear SVM and SVM with radial basis kernel function. In the case of sampling, we use under-sampling method and select five kinds of ratio of churners to non-churners: 2:1, 1:1, 1:2, 1:5, 1:10 and employ 10-fold cross-validation to estimate the accuracy. Final training model results are as follows:

Table 1: Prediction Results of the Customer Churn classification models

	Ratio of churners to non-churners	Accuracy（%）	Customer Churn Recall Rate（%）	Customer Churn Precision Rate（%）
Logistic	Non-sampling	99.09	0	——
	2:1	64.74	94.13	2.37

	1:1	77.78	86.13	3.42
	1:2	86.84	73.41	4.90
	1:5	96.36	49.62	12.39
	1:10	98.32	38.82	23.81
Linear SVM	Non-sampling	99.09	0	——
	2:1	57.90	94.69	2.00
	1:1	77.92	84.85	3.39
	1:2	90.11	70.29	6.21
	1:5	95.16	58.86	10.67
	1:10	98.24	32.43	20.47
RBF SVM	Non-sampling	98.95	26.84	39.10
	2:1	68.07	89.99	2.50
	1:1	80.84	87.39	4.00
	1:2	91.76	72.81	7.63
	1:5	97.63	57.71	20.86
	1:10	98.39	56.09	29.50

From the table above, both the Logistic regression model and linear SVM model predict all of the churners to be non-churners and the churners recall rate is 0.Although the model prediction accuracy rate is 99.09%, but it is still unacceptable. The churners recall rate of RBF SVM model is 26.84%, the precision rate is 39.10% and the model prediction accuracy was 98.95%.

In the case of sampling, when the ratio of churners to non-churners is 2:1, the number of churners is higher than the one of non-churners. The churners recall rates of all the three models are higher than those in the case of all the other ratio values, while the model prediction accuracy rate and precision rate are the lowest.

As the ratio of churners to non-churners decreases, the number of churners increases and then the model accuracy rate is rising. When the ratio is 1:10, the prediction accuracy rate and the precision rate reach the highest point, but the recall rate is the lowest compared with other sampling ratio.

Compare with these three models, we can find that in the case of the same ratio the prediction accuracy, churner recall rate and precision rate of the RBF SVM model are the highest, while those of linear SVM model are the lowest. When the ratio is 1:10, the recall rate, the precision rate and prediction accuracy of RBF SVM are 56.09%, 29.50% and 98.39% respectively, while those of the linear SVM models are 32.43%, 20.47% and 98.24%.

From the table above F-measure is calculated as can be seen as table 2:

Table 2: F-measures under Different Sampling Ratios of the Three Models

Model / Churner to non-churner	Logistic	Linear SVM	RBF SVM
Non-sampling	——	——	0.32
2:1	0.05	0.04	0.05
1:1	0.07	0.07	0.08
1:2	0.09	0.11	0.14
1:5	0.20	0.18	0.31
1:10	0.30	0.25	0.39

According to table 2, F-measure of the three models all reaches its maximum point when the ratio of churners to non-churners is 1:10. Among the F-measure values, the one of RBF SVM is the highest 0.39 and the one of linear SVM is the lowest 0.25. The F value is plotted in Fig.1:

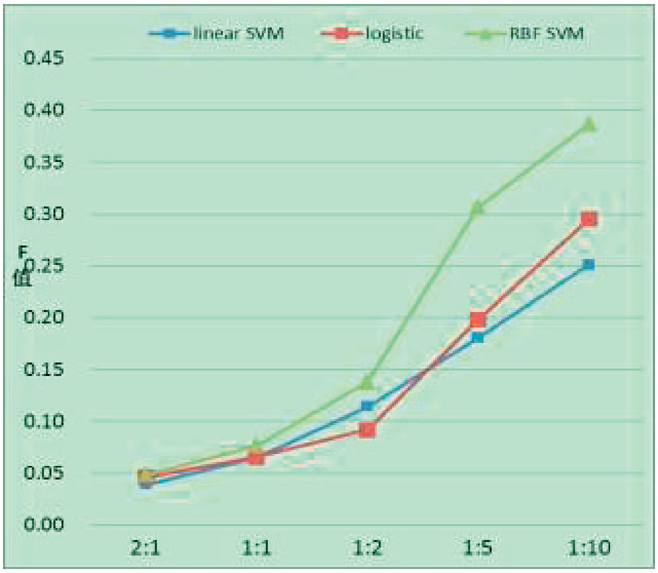

Fig.1 F-measure of different ratios and models

From the results, by under-sampling method the prediction effect of the three classification models have significantly improved and different sampling ratios have different impact on the prediction performance. However, no matter what the sampling ratio is, the prediction effect of RBF SVM is optimal. As can be seen, SVM model can predict churners effectively and has a relatively high accuracy. However, as the customer churn data of commercial banks often displays imbalanced features, we should combine the random sampling method with SVM model and select appropriate kernel functions (This choice is RBF kernel) to enhance customer churn prediction effectively .

5. Conclusion

As the external situation has undergone profound changes in operations, customer competition among commercial banks is increasingly fierce. Customer attrition analysis has become an significant research topic for commercial banks. This study aims to establish SVM model to predict customer attrition of commercial banks. Due to the imbalanced characteristics of the actual commercial bank customer churn dataset, SVM model cannot predict the churners effectively and only general evaluation criteria cannot measure the predictive power of the model.

By changing the sample distribution, Random sampling method has a higher degree of recognition. Therefore, we use random sampling method to improve SVM method, and select F-measure to evaluate the predictive power. To compare the prediction effect, we also establish Logistic regression model. The results show that the combination of random sampling method and the SVM model can significantly improve the predictive power and help commercial banks predict churners more accurately.

Acknowledgements

We would like to thank Daobin Chen who provided the suggestions and the environment for data laboratory. We would also like to thank Lei Qin, Ruiming Li and Xin Li for their help with data preprocessing.

References

1. Bhattacharya, C. B. .*When customers are members: Customer retention in paid member ship contexts* [J] .Journal of the Academy of Marketing Science, 1998, 26(1) : 31- 44..
2. Chandar,M., Laha,A., and Krishna,P.. Modeling churn behavior of bank customers using predictive data mining techniques.[J]. National Conference on Soft Computing Techniques for Engineering Applications(SCT-2006). 2006,(3):24-26.
3. Popovic,D. and Basic,B.D..Churn Prediction Model in Retail Banking Using Fuzzy C-Means Algorithm[J].Informatica:2009(33):243-247.
4. Prasad,D. and Madhavi,S.. Prediction of Churn Behavior of Bank Customer Customers Using Data Mining Tools[J].Business Intelligence Journal.2012,5(1):96-101.
5. Burez,J. and Poel,D.V..Handling Class Imbalance in Customer Churn Prediction[J]. Expert System with Applications.2009(36):4626-4636.
6. Xie,Y.Y.,Li,X.,Ngai,E.W.T. and Ying Weiyun. Customer Churn Prediction Using Improved Balanced Random Forests[J].Expert Systems with Applications. 2009(36):5445-5449.
7. Chiang,D.,Wang,Y., Lee,S., and Lin,C.. Goal-oriented sequential pattern for Network Banking Churn Analysis. Expert Systems with Applications, 2003,25(3):293-302.
8. Mutanen,T.,Ahola,J. and Nousiainen,S..Customer Churn Prediction-A Case Study in Retail Banking[C]. ECML/PKDD2006 Workshop.2006:13-18.
9. Zhao,J. and Dang X.H..Bank Customer Churn Prediction Based on Support Vector Machine: Taking a Commercial Bank's VIP Customer Churn as the Example[J].IEEE.2008:1-4.
10. Vapnik,V.N..The Nature of Statistical Learning Theory[M]. New York: Springer.1995
11. Drummond,C., Holte,R.C..C4.5, Class Imbalance, and cost sensitivity: Why under-sampling beats over-sampling. Workshop on Learning from Imbalanced data sets Ⅱ,International Conference on Machine Learning.
12. Ling,C.,Li,C..Data Mining for Direct Marketing problems and solutions. Proceedings of the fourth international conference on knowledge discovery and data mining. New York, NY:AAAI Press.
13. Chen,C.,Liaw,A. And Breiman,L.. Using Random Forests to Learn Imbalanced Data[J]. Technical Report666, Statistics Department, University of California at Berkeley, 2004.